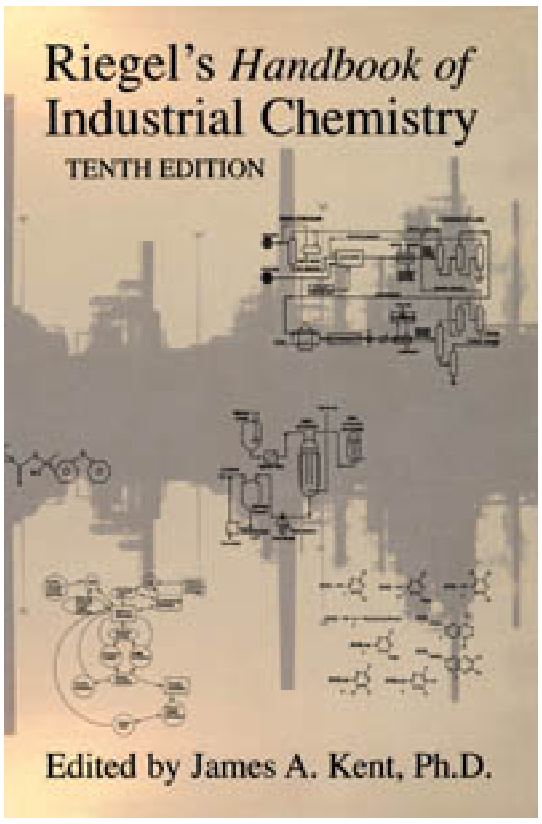
Cover



---

title : Riegel's Handbook of Industrial Chemistry
author : Riegel, Emil Raymond.; Kent, James Albert
publisher : Springer Science & Business Media
isbn10 | asin : 0306474115
print isbn13 : 9780306474118
ebook isbn13 : 9780306485411
language : English
subject Chemistry, Technical.
publication date : 2003
lcc : TP145.R53 2003eb
ddc : 660
subject : Chemistry, Technical.

1.1 Definition of the Chemical Industry

Early in the twentieth century, the chemical industry was considered to have two parts: the discovery, ammonia synthesis synthesis, and manufacture of inorganic and organic chemicals. Later, and until about 1997, the Standard Industrial Classification (SIC) of the U.S. Bureau of the Census defined "Chemical and Allied Products" as comprising three general classes of products: (1) basic inorganic chemicals such as acids, alkalis, and salts and basic organic chemicals; (2) chemicals to be used in further manufacture such as synthetic by type fibers, plastic materials, dry colors, pigments; and (3) finished chemical products to be used for ultimate consumer consumption as architectural paints, cosmetics, drugs, and soaps or to be used as materials or supplies in other industries such as industrial paints and coatings, adhesives, fertilizers, and explosives.1 The SIC system was a series of four-digit number codes that attempted to classify all business by product and service type for the purpose of collection, tabulation, and analyses of data. It used a mixture of market-based and production-based categories.

In 1997, the SIC classification was replaced by the "North American Industry Classification System" (NAICS).2 The system is a major revision based on six-digit numerical codes, and it allows for new or relatively new industries to be included in what is termed "Chemical Manufacturing." It also reorganizes all categories on a production/process-oriented basis. Further, NAICS establishes a common numerical code between Canada, Mexico, and the United States that is compatible with the two-digit level of the United Nations' "International Standard Industrial Classification of All Economic Activities" (ISIC).

The NAICS code for "Chemical Manufacturing" is "325" and there are 49 subclassifications with four- to six-digit codes. The four-digit codes, which are a description of the manufacturing segments included in chemical manufacturing, the value of shipments, and the number of employees in the manufacturing segment are listed in Table 1.1.2 Each of these four-digit segments may have five-digit subclasses associated with them, and the five-digit subclasses in turn may have six-digit subclasses associated with them. This hierarchy is exemplified for Manufacturing Segment 3251, which is titled "Basic Chemical Manufacturing," and one of its sub components, Code 32519, in Table 1.2.

TABLE 1.1 Chemical Manufacturing, NAICS Code 325, and Its Four-Digit Area Components. Shipment Value and Employees are from 1997 U.S. Economic Census2

NAICS Codea	Description of Area	Shipments Value ($1000)	Percentage of Total	Employees
325	Chemical manufacturing	419,617,444	100	884,321
3251	Basic chemical manufacturing	115,134,992	27.44	202,486
3252	Resin, Rubber synthetic rubber, artificial and synthetic fibers, and filament manufacturing	63,639,476	15.17	114,792
3253	Pesticide, fertilizer, and other agricultural chemical manufacturing	24,266,513	5.78	37,206
3254	Pharmaceutical and medicine manufacturing	93,298,847	22.23	203,026
3255	Paint, coating, and adhesive manufacturing	26,594,550	6.34	75,100
3256	Soap, cleaning compound, and toilet preparation manufacturing	57,507,318	13.7	126,895
3259	Other chemical product manufacturing	39,175,748	9.34	124,816

aCodes 3257 and 3258 were not used.

TABLE 1.2 Basic Chemical Manufacturing, NAICS Code 3251, and its Five-Digit Components and Other Basic Organic Chemical Manufacturing, Code 32519, and Its Six-Digit Components. Shipment Value and Employees are from 1997 U.S. Economic Census2

NAICS Codea	Description of Area	Shipments Value ($1000)	Percentage of Total	Employees
	NAICS Code 3251 and Its Components			
3251	Basic chemical manufacturing	115,134,992	100.00	202,486
32511	Petrochemical manufacturing	20,534,750	17.84	10,943
32512	Industrial gas manufacturing	5,231,468	4.54	12,492
32513	Dye and pigment manufacturing	6,427,357	5.58	17,289
32518	Other basic inorganic chemical manufacturing	20,716,361	17.99	60,056
32519	Other basic organic chemical manufacturing	62,225,056	54.05	101,706
	NAICS Code 32519 and Its Components			
32519	Other basic organic chemical manufacturing	62,225,056	100.00	101,706
325191	Gum and wood chemical manufacturing	815,201	1.31	2,267
325192	Cyclic crude and intermediate manufacturing	6,571,093	10.56	8,183
325193	Ethyl alcohol manufacturing	1,287,273	2.07	1,890
325199	All other basic organic chemical manufacturing	53,551,489	86.06	89,366

aCodes 325194 through 325198 were not used.

While it may seem that Code 325199, "All Other Basic Organic Chemical Manufacturing," is too general in nature for its size, one needs to consider that by delving into the makeup of this component, about 150 individual compounds or groups of compounds are found. These contain a diverse group of chemicals including manufacturing of acetic acid and anhydride, calcium citrate, cream of tartar, ethylene glycol ethers, ethylene oxide, solid organic fuel propellants, hexyl and isopropyl alcohols, perfume materials, peroxides, silicone, sodium alginate, sugar substitutes, tear gas, synthetic vanillin, vinyl acetate, and so on. All these compounds have the code number 325199. Compounds with the code number 325211, "Plastic Material and Resin Manufacturing" are about 80 in number and may be exemplified by acrylic and methacrylic polymers; cellulose derivatives such as acetates, nitrates, xanthates, and the like; phenolics, polyesters, polyolefins, polystyrene, poly(vinyl halide)s, polyurethanes, and, again, and so on.

The new NAICS has broadened the definition of the chemical industry, and it now is more encompassing than in the past. The broadening is reasonable, and it improves on the goals of collecting and tabulating data so that it is available for study and analysis. One might say the data could be timelier, but collecting, amassing, and breaking down the information so it is understandable is a difficult, time consuming task that is dependent on many people. The Internet is a major factor in making the data available to the general public almost as rapidly as it is compiled. The NAICS system is recognized and accepted by the North American countries, and the system appears to be global in nature by being at least partially in line with the United Nations' classifications. Those interested in markets and market areas and in their size including their relation to other markets will find the U.S. Census Bureau's web site pages well worth visiting.

1.2 The Place of the Chemical Industry in the Economy

Because the chemical industry is a major sector of any advanced national economy, a forecast of trends in the chemical industry must fall within certain general guidelines that are established by the national economy. A forecast for the chemical industry in the United States must be within the general boundaries set for the overall societal, financial, environmental, governmental, and economic forecasts for the country. However, such forecasts should be carefully considered for they may or may not accurately predict the future.

It had been said that it was clear for many years that certain demographic and societal issues would have a dominant effect on the U.S. economy of the 1990s. In the previous edition of this Handbook, it was pointed out that there was an expectation that from the late 1980s through the year 2000 there would be a decline in the growth of the work force in the United States. This was predicated on the number of women within the usual childbearing age group of 1835 and by family-size decisions that were made in the 1960s. Shortages of chemists, chemical engineers, and other scientists were predicted for the 1990s. Supposedly, such predictions can be made from census data that was obtained in the prior two or three decades. There is a direct relation between the growth of the workforce and the growth of Gross Domestic Product (GDP). Although this was the prediction, it is not what happened.

During the 1990s and through the start of the twenty-first century, due to events put in motion during the late 1980s, the United States and many world economies experienced unprecedented growth. During this time unemployment decreased and reached very low percentages on an absolute and a historical level. This factor was coupled with significant productivity increases throughout the economy. The productivity increases resulted from a better-trained workforce, from new tools such as computers and allied software, and from just plain harder, more conscientious working during regular hour and overtime hours. An important factor during this period of growth was that the productivity increases were obtained without inflation raising its ugly head. Company mergers and the spinning off or the selling off of business segments to stockholders or to allied businesses played an important role through these years. These actions resulted in new stand-alone businesses that were operated by new owners and managers when the units were spun off. The mergers or unit-sales resulted in a restructuring or downsizing as duplicated efforts were eliminated. Productivity increased because of these actions, and many workers were displaced. But, the man power hungry economy quickly absorbed for the most part these displaced workers. The hunger for manpower was partially, but importantly, related to the electronic, computer, telecommunications, and related industries that provided many jobs in previously non-existent sectors. Chemicals were used in various ways in these new growth areasas, for example, wire coatings, solder masks, conformal coatings, optical fiber coatings and marking materials, magnetic tape coatings, and so on.

Mergers certainly played a role in shaping today's chemical manufacturing industry. Included among the notable mergers are the Pfizer Inc. merger with Warner Lambert Company, The Dow Chemical Company combination with Union Carbide Corporation, Exxon merging with Mobil, and many others in the United States. In Great Britain, Glaxxo Holdings PLC first merged with Wellcome to form Glaxo Wellcome, this combination then merged with Smith Kline to form GlaxoSmithKline PLC. Larger, improved efficiency companies resulted. One result of such mergers is a shrinking of Chemical and Engineering News' top 100 companies to the top 75 companies.4

The electronics/computer industry grew rapidly during the past decade or so, and new company names appeared during this time period. Its growth was spurred by the productivity increase even as it was a participant in causing the increased productivity. Computers began to be used to control processes and training personnel with the skills to run such computer-operated processes was high on many companies' lists of important projects. A decade or so ago, computers were available in companies on a limited basis. Today, there is a computer on essentially every desk and portable computers to carry out work during trips, and the like. The "dot com" companies started their appearance through the Internet, and they grew rapidly. Later, when business turned down, many of these companies disappearedthey merged with or were purchased by other companies. However, overall prosperity reigned during this time period, and, as it did, the chemical manufacturing industry, which was allied with a broad variety of these industries, also prospered.

At the turn of the twenty-first century, it was becoming apparent that the economy was at a high point and could be expanding too rapidly. Inflation was still low, and there was even talk that deflation might come into play. They latter did not happen. Price-to-earnings ratios were very high for many companies, and it did not appear that future growth would expand sufficiently to accommodate such high price-to-earnings ratios and large additions to the work force. The national economy, which certainly includes chemical manufacturing, entered a recession in March 2000. However, the economy grew in the first three months of 2001 indicating that the economic recovery could be beginning. The improvement was led by new automobile purchases and increased government spending. The fourth quarter of 2001 was small and considered by some as flat, but it built on the preceding quarter and in early 2002 there was a belief that the economic recovery has begun.

The events of September 11, 2001, changed many aspects of our lives with chemical manufacturing included. The terrorist attacks rocked many markets on a short-term basis, but before long the markets stabilized, but did not really grow in the recent past. Overall, in the first half of 2002 the world economy remained in a recession. However, because of the constant threat of terrorism, national corporate spending will increase as military, security, and other government expenses increase and transportation costs and its allied security measures come into play. Chemical manufacturing of basic chemical, polymer, and pharmaceuticals are expected to increase. Yet, there are no expected productivity increases as was seen in the 1990s associated with the increased spending.5 In early 2002, the Chairman of the Federal Reserve predicted that the recovery was apparent, but would be a mild recovery.6

As can be seen from the previous discussion, economic forecasts are subject to all of the uncertainties and unpredictabilities of national, international, and societal events. With this in mind, at the present time the forecast for the ensuing part of this decade is for improved growth in national chemical manufacturing with growth and the profit picture beginning an upturn in the second or third quarter of 2002. (Note that in the 1980s, it was not predicted that the 1990s would show strong growth, yet strong growth did take place.) At present the concerns with global terrorism, low interest rates, high but constant productivity, oil and gas prices, and other factors are components of a mixture that will dictate the future. None of these factors will remain constant. Rather, they will change individually at times and with some factors in concert at other times. These variations along with the size of the workforce and its attitude will dictate the future for chemical manufacturing and the global economy.

Against this brief discussion of the general demographic, societal, and economic factors that govern forecasting economic prospects, a general picture of the economy of the United States can be given by the GDP and chemical and allied products portion of GDP as described in Table 1.3.

TABLE 1.3 U.S. Economy and Chemical Manufacturing7

Year	U.S. Gross Domestic Product (GDP) Current Dollars, Billions	Chemicals and Allied Products Portion of GDP, Current Dollars, Billions
1987	4,742.5	83.8
1990	5,803.2	109.9
1995	7,400.5	150.8
1996	7,813.2	153.6
1997	8,318.4	164.8
1998	8,781.5	164.8
1999	9,268.6	175.1
2000	9,872.9	191.1

This reasoning is a way to highlight the sensitivity and place chemical manufacturing has in the national economy, which is becoming more and more entangled with the countries of the North American Free Trade Agreement (NAFTA) and with the global economy. Thus the chemical manufacturing industry is worldwide and interconnected in many ways. These factors play important roles in the importance of imported raw materials such as Petroleum refining, Crude oil petroleum products and the cost of labor. Businesses or parts of businesses can be transported across the southern U.S. border to take advantage of more favorable labor costs. Through this, successful partnerships have been forged and welded together between border countries. As mentioned earlier, other partnerships are developing through purchase of assets in other countries by the United States and by other countries in the United States. Today, the United States is entrenched in the age of a global economy and all it ramifications.

The United States imports and exports a wide variety of raw materials and chemical products. Major U.S.-based chemical companies have manufacturing and sales facilities abroad and a large number or foreign-based companies have similar facilities in the United States. The U.S. economy is dependent on the balance of trade, that is, on the difference between the dollar value of exports and imports. A negative trade balance means that dollars spent abroad to import goods and services exceed the value of goods and services exported. In effect such an imbalance increases the cost of goods and services purchased in the United States and results in a net inflationary effect. To a large extent during the 1980s, this potential inflationary effect was offset by foreign investment in the United States. In the 1990s and through the early years of the twenty-first century, foreign investment in the United States has increased, productivity has increased without major wage increases, and interest rates were managed with the net result that inflation remained low.

In foreign trade, the chemical industry of the United States has consistently performed in an outstanding manner. While the overall balance of trade has been negative, the chemical industry has been one of the truly strong sectors in the economy of the United States, Table 1.4. Year after year, the trade balance of chemicals has been positive and thus has had a positive impact on the national economy. When the total world export market for chemicals is considered, that is the sum of all the chemicals exported by all the world's national economies, the U.S. chemical manufacturing industry has held a significant market share, about 15 percent, for the past three decades.

TABLE 1.4 U.S. Balance of Trade[8]

	Total Trade Balance[a] (Billions of Dollars)			Chemical Trade[b] (Billions of Dollars)		
Year	Export	Import	Balance	Export	Import	Balance
1987	250.2	409.8	-159.6	26[c]	16	+10
1990	389.3	498.3	-109.0	39[c]	22	+17
1995	575.8	749.6	-173.8	32.18	20.59	11.59
1996	612.1	803.3	-191.2	31.4	21.81	9.59
1997	678.4	876.5	-198.1	34.6	23.5	11.10
1998	670.4	917.1	-246.7	33.32	23.38	9.94
1999	684.6	1,030.0	-345.4	34.09	23.82	10.27
2000	772.2	1,224.4	-452.2	38.42	27.12	11.30

[a] International Trade Accounts (ITA); [b] Chemicals-Fertilizer, -Organic, -Inorganic, and -Other; [c] Amounts for 1987 and 1990 are taken from Bailey and Koleske.[9]

The export and import values for chemical segments described in Table 1.4, chemical-fertilizer, chemical-organic, chemical-inorganic, and chemical-other, are detailed in Table 1.5. The magnitude of the individual items varies from year to year, but overall, the balance is favorable and these four segments of chemical manufacturing are usually positive values. It should be pointed out that various items (plastic materials, pharmaceuticals, etc.) that make up chemical manufacturing have been excluded, but this was done without bias. The four items used in Tables 1.4 and 1.5 are directly related to what has been traditionally known as the "chemical industry." The less favorable Total Trade Balance of the United States is principally due to imports of manufactured good and petroleum products.

TABLE 1.5 Chemical Export/Import Segments for Year 2000

Chemical Segment	Exports (Billions of Dollars)	Imports (Billions of Dollars)	Balance (Billions of Dollars)
Fertilizer	4.098	3.388	+0.710
Inorganic	4.180	4.414	-0.234
Organic	16.505	13.779	+2.726
Other	13.636	5.525	+8.111

To support the U.S. chemical manufacturing economy (Code 325) in 1997 (see Table 1.1), there was a workforce of more than 884,000 of which about one-fourth were employed in basic chemical manufacturing (Code 3251) and about one fourth were employed in pharmaceutical and medicine manufacturing (Code 3254). The next largest area of employment, about 14 percent of the workforce, was the soap, cleaning compound, and toilet preparation manufacturing component (Code 3256), which was closely followed by the polymer manufacturing area (Code 3252) at 13 percent. In such comparisons, one might argue that the paint, coating, and adhesive manufacturing component (Code 3255), with about 8.5 percent of the employment figure, should be included with polymer manufacturing. The remainder of the workforce is employed in the agricultural chemical and other chemical manufacturing components. The value of the chemical manufacturing business produced by this workforce was $419,617,444,000 in 1997. To maintain market share and grow this huge business, the companies in chemical manufacturing invest to various degrees in research and development efforts, which are carried out by scientists within the organizations. The percentage of sales varies with the particular component, and the pharmaceutical firms will spend much more than say a fertilizer manufacturer. The average for many chemical companies varies from about 3 to 5 percent of sales. In the 199997 period, about seven billion dollars were spend annually on research and development by the chemical industry.

< previous page

next page >

1.3 Characteristics of the Chemical Industry

Investment Trends

The U.S. chemical industry is the world's largest and it accounts for about one fourth of global chemical production. The industry, which is a part of the non-durable goods manufacturing industry, is a high capital investment business. Capital spending by the chemical and allied products industry in the United States has been a sizable percentage of that spent for all manufacturing. In 1999, non-durable goods manufacturers spent about $80 billion on capital goods. This was a decrease of about 7 percent from that of 1998, which was approximately $85 billion. Most of this decline can be attributed to decreased spending by the basic chemical industry. In 1996 and 1997, capital expenditures in the chemical industry were about $15.5 and $16.4, respectively. During the late 1980s and 1990s, a significant portion of these capital expenditures was made for pollution control and other environmentally related efforts.

Much of the capital investment in the chemical industry is spent for facilities to produce major chemicals in enormous quantities. The huge volume of the chemicals produced and consumed is reflected in the size of plants being built to achieve the required economies of scale, which in turn allow for competitive pricing. The fact that such economies are achieved is seen in the relatively modest increases in chemical producers' price indices relative to the inflation levels in the general economy. The competitive nature of the chemical business also plays a role in this matter of price. (Economy of scale refers to the relative cost of building a larger plant; a rule of thumb is that the relative cost of building a smaller or larger plant is the ratio of the productivities of the two plants being considered raised to the 0.6 power. In other words, the unit cost of producing a chemical markedly decreases as the size of the plant producing it is increased, providing the plant can be operated near capacity.)

Along with these very large plants and the associated enormous investments, most of the chemical industry is characterized by high investment versus low labor components in the cost of manufacture. The National Industrial Conference Board statistics list the chemical industry as having one of the highest capital investments per production worker. The investment per worker in a base petrochemicals olefins plant may be in the neighborhood a half-million dollars. A profitable chemical specialties manufacturer may have capital investments as low as 10 percent of such values per employee. Of course, sales per employee are also important and large. From Table 1.1, it can be seen that annual sales per employee in the overall chemical manufacturing area (Code 325) are about $475,000. Such ratios vary with the market segment and depend on the labor intensity needed within the segment. For example, the number is about $569,000 for basic chemical manufacturing (Code 3251), $652,000 for agriculture chemicals (Code 3253), $459,541 pharmaceuticals (Code 3254), $354,000 coating chemicals (Code 3255), and so on. Note that number for pharmaceutical sales per employee is quite close to the overall chemical manufacturing sales per employee. The average hourly pay of production workers in the chemical industry was $18.15 in 2000.

Commercial Development and Competition Factors

During the earlier period of the chemical industry's development, chemical companies were generally production oriented, wherein they would exploit a process to produce a chemical and then sell it into rapidly expanding markets. The investments and plant sizes required for participation were small fraction of that required to participate today. Raw materials were often purchased to produce chemical intermediates for sale. Small-sized units operating in small manufacturing facilities do not present the obvious problems of environmental pollution, a factor about which everyone has become more aware in the past two or three decades. A new investment in chemical production facilities today must include a sizable proportion of the total outlay for pollution abatement and control of environmental intrusion. The chemical industry spends about $5 billion annually on pollution abatement.

As the chemical industry has grown, there has been a strong tendency toward both forward and backward integration. Petroleum producers have found opportunities based on their raw materialsnatural gas, condensates, and oilto move into chemical refining. Chemical companies, on the other hand, have moved to assure their access to low-cost raw materials through contract purchases and hedging contracts. Similarly, producers of basic plastic materials have forward integrated to produce compounder materials and fabricated products such as consumer items, fibers, and films. At the same time, fabricators have installed equipment to handle and formulate or compound the basic plastic materials and thus provide a ready, constant supply at the lowest possible cost. With the global economy in place and relative ease of moving around the world, large investments are now made in far-off countries such as Malaysia and Saudi Arabia, for example, to be near raw material supplies and to meet large market needs.

With ever-larger investment costs and increasing cross-industry competition, markedly greater sophistication has been required of marketing analysis coupled with cost analysis when selections of investment opportunities are made. The enormity of investment capital required in today's marketplace to successfully participate does not permit multiple approaches for the private investor. Consequently, a high degree of market orientation tends to predominate in the chemical industry along with increasingly targeted and pinpointed research and development programs. In 2000, the industry spent about $31 billion on such research and development efforts.10

A major trend in industrial chemistry has been an emphasis on improved processes for the production of major chemicals such as ethylene, propylene, vinyl chloride, styrene, alkylene oxides, methanol, terephthalates, and so on. The necessity for higher efficiency, lower cost processes has been accentuated by the relatively slow growth rates of major industrial chemicals over the past two decades or so. The fertilizer portion of the agricultural chemicals market as described in Table 1.6 is an example of the slow growth.

TABLE 1.6 Annual Production of Inorganic Chemicals Used in the Fertilizer Industry (Note Break in Years between 1996 and 1993)8

	Production Amount (Billions of Pounds)		
Chemical	1997	1996	1993
Ammonia	34.68	35.85	34.39
Ammonium Nitrate	17.21	17.00	16.56
Ammonium Sulfate	5.42	5.32	4.87
Nitric Acid	18.87	18.41	16.51
Phosphoric Acid	26.32	26.42	23.03
Sulfuric Acid	96.04	95.54	79.68
Super Phosphates and others	20.86	21.09	17.6
Urea	15.33	17.10	16.66

How well has the chemical industry developed? At the beginning of the twenty-first century the United States accounted for 27 percent of the world's chemical production, making it the world's largest chemical producer.10 In 2000, chemical shipments reached $460 billion, and, at this level, it provides about 1.2 percent of the national GDP and almost 12 percent of the manufacturing GDP. As such, it is the largest factor in the manufacturing segment of the economy. The chemical industry continues to grow, and it attained an all-time high in profits by netting $44 billion. Globally, chemicals are almost a $1.5 trillion dollar business.

With its large size, the chemical industry is a large user of energy, and it consumed about 7 percent of all domestic energy and about 25 percent of all energy used in manufacturing. In 1985, the industry used 3,567 trillion Btu, in 1991 usage increased to 5,051 trillion Btu, and in 1994 it had increased still further to 5,328 trillion Btu.11 The energy is used to supply heat and power for plant operations and as a raw material for petrochemicals, plastics, and fibers production. Feedstocks represent a little less than half of the total usage, a number that varies from year to year.

Thus, the chemical industry is a key component in the U.S. economy. It converts raw materials such as gas, oil, condensates, water, metals, and minerals into more than 70,000 products that are used in a variety of ways. In some fashion, this industry impacts the daily lives of everyone. Industrial customers for chemicals are many, but some of the major ones are apparel, plastic and Polymers rubber products, Crude Oil, Petroleum petroleum refining, textiles, pulp and paper, primary metal, and the like.

Information technology and E-commerce have become increasingly important assets to the chemical industry. Spending on information technology reached $10.2 billion in 2000 and this represented a 75 percent increase over such expenditures of 1990.10 Selling via the Internet or E-commerce resulted in sales of $7.2 billion in 2000. Projections indicate that this type of business will grow rapidly and are expected to reach $150 billion by 2006. This means about one third of shipments will be via E-commerce transactions in 5 years.

Technological Orientation

The chemical industry is a high technology industry, albeit now is more marketing oriented and competitive than in its earlier period of development. Chemists and materials scientists held about 92,000 jobs in 2000. Over half of these are employed in manufacturing companies and most of these companies are in the chemical manufacturing industry, that is in firms that produce synthetic materials, plastics, drugs, soaps and cleaners, paints, industrial organic and inorganic chemicals, and other chemicals.12(a) Other chemists and chemical engineers are found in various government Departments and Agencies, in teaching, and in research, development, and testing firms. The latter firms are becoming more and more a growth area in the chemical industry. A bachelor's degree in chemistry or a related discipline is the minimum education requirement for these technical positions. To work and grow in research positions, a Ph. D. is required. There will be strong demand for those people who have a masters or Ph.D. in the future with job growth concentrated in the pharmaceutical companies and in research, development, and testing services firms.

The contemporary scientist or engineer engaged in research and development in the chemical industry is a highly trained individual who is a part of a high-investment occupation. Since about the mid 1950s, much of chemistry has become increasingly an instrumental science, and the instruments routinely used by investigators are highly sophisticated, reliable, and costly. In the laboratory, a scientist has available mass, infrared, visible, and ultraviolet spectrometers; various chromatographs; physical and chemical property determination devices such as those used for molecular structure, size, and conformation determinations; and others used for reaction kinetic studies. Pilot plants and many production facilities are highly instrumented and automated. The basic scientist doing research, laboratory workers, pilot plant and process development chemical engineers, and plant production workers require at a minimum access to excellent computer facilities. All engineers and scientists require computers to analyze the massive amounts of data that are generated and to aid in the design of manufacturing processes and equipment.

Employment of chemists and chemical engineers is expected to grow at about a 1020 percent rate between 2000 and 2010.12(a) Predictions indicate that job growth will be concentrated in drug manufacturing and in research, development, and testing services companies. Demands will be for new and better pharmaceuticals, personal care products, and specialty chemicals designed to solve specific problems or applications. Demand will be high for personnel who have a Ph.D. degree and the opportunities will be in biotechnology and pharmaceutical firms. An aging, better informed population will want products that treat aging skin, that are milder on the body, new and innovative drugs, reliable medical devices, and so on. The population in general will be interested in chemical processes that are more benign in nature to produce all types of products and thus in an industry that is more friendly to the environment.

In the year 2000, the median salary of chemists was $50,080.12(a) The lowest 10 percent earned less than $29,620, and the highest 10 percent earned more than $88,030. The middle 50 percent earned between $37,480 and $68,240. It is interesting to point out that the median annual salary of chemists employed in the Federal Government was $65,950 or about 30 percent higher than the overall median. In 2001, chemists in non-supervisory, supervisory, and managerial positions in the Federal Government averaged $70,435.12(a) As is the usual case, chemical engineer salaries were higher than those of chemists by about 1025 percent. Median experienced and starting salaries for the various degrees can be found in Table 1.7.

TABLE 1.7 Chemist12(a) and Chemical Engineer12(b) Salaries in 2000

	Year 2000	
Degree	Overall Median Salary	Inexperienced Median Starting Salary
Chemists		
Bachelor	$55,000	$33,500
Master	$65,000	$44,100
Ph.D.	$82,200	$64,500
Chemical Engineers		
Bachelor	a	$51,073
Master	a	$57,221
Ph.D.	a	$75,521

aThe median annual salary for all chemical engineers was $65,960. The salary of the middle 50% ranged between $53,440 and $80,840. The salary of the lowest 10% was less than $45,200 and of the highest 10% was greater than $93,430.

Historical

How did the chemical manufacturing industry get its beginning? To get this answer, we need to go back to the latter part of the eighteenth century.13 The availability of alkali or soda ash (soda ash sodium carbonate) for the growing manufacture of glass, soap and textiles in France was becoming a major concern. At that time, the chemical was obtained from plant materials, principally from wood ashes that were leached with hot water to obtain potash and from marine plants such as *barilla*, which grows mainly along the Spanish Mediterranean coast and in the Canary Islands. Other plant sources existed. The main exporter of soda ash was Great Britain with whom France was at odds and there was concern about the chemical's availability. In 1783, the French Academy of Sciences was offered a handsome prize by Louis XVI to develop a simple process for "decomposing" sea salt on a large scale and securing alkali from it. Eight years later Nicholas Leblanc, a 49-year old French physician, devised a scheme to commercially obtain soda ash from sea salt. The process became known as the Leblanc process, and this process is considered the basis for development of the first chemical industry. For almost a century, this process was the most important method known for producing chemicals. Basically, Leblanc's process involved reacting sodium chloride with sulfuric acid to produce sodium sulfate. The product was then reacted with calcium carbonate and carbon to form a "black ash" that contained sodium carbonate and other compounds. The "black ash" was extracted with water followed by an evaporation process to obtain soda ash.

Leblanc's process had many disadvantages; it was complicated, was dirty and polluting, and was materials and fuel inefficient. This set other scientists to working on development of a new process. In about 1872, Ernst Solvay developed what became known as the Solvay process, and this resulted in establishment of the French firm Solvay & Cie. By 1890 the Solvay process dominated the world's alkali production.14 Leblanc's process was obsolete.

Obsolescence and Dependence on Research

The high technology level that characterizes the chemical industry, and which is reflected in heavy research and development investments, generally concerns discovery and development of new products as well as improvements in the manufacture of known products. New product discovery and development may be typified by a new pharmaceutical product for a specific disease, by a stealth aircraft and all its special polymer and composite needs, by development of a new non-polluting technology for a known process, a uniform molecular weight polymer designed and made by nanotechnology, and so on. Improvements in the manufacture of known products might be typified by producing a modified form of a pharmaceutical that is easier to dissolve, by a new or modified higher efficiency catalyst for a known process, by toughening a brittle plastic material, by improving the strength of a composite, and so on. The development of a new, lower cost process for a commercial product can permit development of a profitable opportunity or it can spell disaster for a company with existing investment in a plant made obsolete by the competitor's new process as in the preceding soda ash example.

Major reductions in manufacturing cost can be achieved, for example, by reducing the number of reaction steps required in a process, by changing to a lower cost or more available raw material, or by eliminating by-products and co-products, costly separation, and environmental intrusions. The ability of a process scheme to contain or avoid a pollutant can be a deciding factor in continuance of a manufacturing operation. At times new regulations, such as the Clean Air Act (CAA), or shortages can spawn new ideas and technologies if the people involved are astute and react positively to the new, developing environment. The brief discussion of Leblanc's process being replaced by Solvay's process for soda ash is an example of how economic consequences can change if a competitor finds a process better than the one being practiced. The following detailed examples will make the matter even more clear.

Acetic Acid

Acetic acid production in the United States has increased by large numbers in the last half century, since the monomer has many uses such as to make polymers for chewing gum, to use as a comonomer in industrial and trade coatings and paint, and so on. In the 1930s, a three-step synthesis process from ethylene through acid hydrolysis to ethanol followed by catalytic dehydrogenation of acetaldehyde and then a direct liquid-phase oxidation to acetic acid and acetic anhydride as co-products was used to produce acetic acid

$$CH_2=CH_2 \xrightarrow{H_2SO_4/H_2O} C_2H_5OH$$

$$C_2H_5OH \xrightarrow{Cu/Cr} CH_3CHO$$

$$CH_3CHO \xrightarrow{[O], Co} CH_3COOH + (CH_3CO)_2=O$$

Then, in the 1940s, a major process change was introduced. In this new process, butane was directly oxidized to acetic acid and co-products such as methylethylketone.

$$C_4H_{10} \xrightarrow{[O]} CH_3COOH + CH_3COC_2H_5 + \text{others}$$

The novel synthesis required fewer process steps, and this resulted in lower costs and investment. In 1969, another advance was announcedthe synthesis of acetic acid from methanol and carbon monoxide with essentially no by-products or co-products.15,16

$$CH_3OH + CO \xrightarrow{I/Rh} CH_3COOH$$

The use of readily available raw materials and absence of co-products reduces production costs and investment needed for distillation and other separation systems. Such simplification results in a very attractive process in an industry where the principally accepted measure of business quality is return-on-investment.

Acetic Anhydride

Acetic anhydride is required as a process intermediate in acetylations. To obtain acetic anhydride from acetic acid, acetic acid is first pyrolyzed to ketene, which then reacts with recovered acetic acid to yield the anhydride.

$$CH_3COOH \xrightarrow{\text{Heat}} CH_2=C=O$$

$$CH_2=C=O + CH_3COOH \longrightarrow (CH_3CO)_2{-}O$$

In 1980, the Tennessee Eastman unit of Eastman Kodak announced that it would begin construction of a facility to make acetic anhydride from coal, which was readily available at reasonable cost.17,18 This decision reflected a changing of the raw materials base of much of the chemical industry due to such factors as the rising cost of natural gas and petroleum and the large coal reserves of the United States.

In the new Eastman process, synthesis gas (carbon monoxide and hydrogen) is made from coal. Then, from the generated synthesis gas, methanol was prepared. (Prior to this time, methanol had been made from methane, i.e., natural gas.)

$$CO + 2H_2 \longrightarrow CH_3OH$$

Methanol was next reacted with acetic acid to form methyl acetate.

$$CH_3OH + CH_3COOH \longrightarrow CH_3COOCH_3 + H_2O$$

Acetic anhydride was then obtained by the catalytic carbonylation of methyl acetate with carbon monoxide.16

$$CH_3COOCH_3 + CO \longrightarrow CH_3CO{-}O{-}OCCH_3$$

There are two major points that make this process attractive. First, the raw material base of synthesis gas is coal. The second point is avoidance of the energy-consuming manufacture of ketene by pyrolyzing acetic acid.

Vinyl Chloride

The increase in the production of vinyl chloride, which is the principal monomer for poly(vinyl chloride) plastics and various vinyl copolymers that are used in vinyl flooring, shower curtains, car-seat upholstery, house siding, pipe, beverage can coatings, and so on, is an even more spectacular example. This polymer is used in multibillion pound quantities. It is an interesting sidelight to point out that the polymer has poor thermal stability, and its huge penetration into the marketplace is attributable to the development of highly efficient thermal stabilizers.

During the early monomer development in the 1930s, vinyl chloride was produced by means of a catalytic addition of hydrogen chloride to acetylene.19

$$CH{\equiv}CH + HCl \xrightarrow{HgCl} CH=CHCl$$

Later, what was called a "balanced" process was introduced. In this process, chlorine was added to ethylene and ethylene dichloride was produced.

$$CH_2=CH_2 + Cl_2 \xrightarrow{[O]/Cu} CH_2ClCH_2Cl$$

The ethylene dichloride was then cracked to vinyl chloride and hydrochloric acid with the hydrochloric acid recycled to produce vinyl chloride from ethylene as shown above.

$$CH_2ClCH_2Cl \xrightarrow{\text{Heat}} CH_2=CHCl + HCl$$

At this point in time, vinyl chloride was being produced from chlorine, acetylene, and ethylene. After these processes, a catalytic oxychlorination has been developed in which vinyl chloride is produced from ethylene and hydrogen chloride in the presence of oxygen.20,21

$$CH_2=CH_2 + HCl \xrightarrow{[O]/Cu} CH_2=CHCl + H_2O$$

If desired, the hydrochloric acid can be obtained via cracking of ethylene dichloride. The oxychlorination process freed vinyl chloride production from the economics of a more costly raw material, acetylene. Deliberate acetylene manufacture is energy intensive and relatively expensive. By-product acetylene from gas cracking is less expensive, but it has not been available in sufficient supply for the large, approximately billion-pound-per-year plus vinyl chloride production units.

During the long development and commercialization of poly(vinyl chloride) into one of the major plastic materials, several basic processes of making the polymer evolved. In all of these processes, vinyl chloride was handled as a liquid under pressure. Other than the relative ease with which the monomer could be free radically polymerized, vinyl chloride was regarded as an innocuous, relatively inert chemical. During the 1960s, the monomer sold for five or six cents a pound. Because of the low cost, it was uneconomical to recover and compress the monomer for recycle during stripping and drying operations at the end of the process. The monomer was often vented into the atmosphere.

Then, in the 1970s, a number of poly(vinyl chloride) producers were completely surprised when it was found that long-term (20-year) exposure to vinyl chloride could cause rare forms of tumors.22 After the discovery that vinyl chloride was a carcinogen, venting was not permissible. Containment and recovery of the monomer was mandatory. As a result, some older processes and manufacturing facilities could not be economically modified to incorporate containment, and as a result such operations were discontinued. This case is but one example of the impact that necessary and regulated environmental controls can have on manufacturing processes and operations.

Coatings Technology

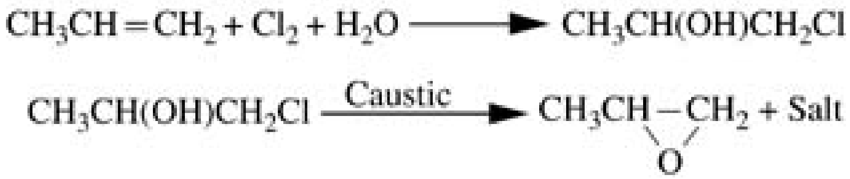
Environmental regulation also had a major impact on the coatings industry. Before, around 1970, almost all (between 90 and 95%) industrial coatings were applied at low-solids (about 1020%) contents from solvents. Many trade or house paints were also solvent based, since aqueous latex technology did not yet have its dominant position. Solvents were inexpensive and they did an excellent job of dissolving the high molecular weight polymers needed to obtain good performance characteristics. The high molecular weights used necessitated the large quantities of solvent to be usedabout 49 lb of solvent were venting to the atmosphere for each pound of final coating film.

Then in the early 1970s, two factors affected the coatings industry. One of these was cartel oil pricingboth unexpected and quickly imposed. This was a factor that increased solvent cost and potentially its availability in needed quantities. At almost the same time, Government regulations requiring less solvent usage (the CAA) were imposed on the industry. Solvent cost was not a major problemjust raise prices, but many coatings are not inelastic commodities and there still was the threat of non-availability. The availability of oil to manufacture solvents was a totally different matter, and anyone who suffered through the gasoline shortages of this time knows well what effect oil availability can have on an economy. Also, government regulations were not just a temporary measure. Many of the regulations were difficult to meet and could not be met with the technology in hand.

Here a large industry (NAICS Code 3255) that represented about 6 percent of all chemical manufacturing was being asked to change the way they had been doing business for many, many years. Some companies responded well, but others thought solvents were too important to literally be taken out of such a large industry. Many companies innovated and came up with radiationcurable coatings, powder coatings, high-solids coatings, two-package coatings, and others. The new technologies did not take over the marketplace overnight, but with time, each found a niche and in so doing took away a portion of the original market. Today one does not find low-solids, solvent-borne coatings in the market to any great extent. One does find companies that have changed the nature of their business favorably by innovating and changing, that have lost market share by taking a "wait and see" attitude, and that are new on the scene and growing. Thus, increased raw material pricing, shortagesreal or created, and regulation can have a positive effect on the overall chemical economic picture.

Propylene Oxide

Propylene oxide is another basic chemical used in manufacturing intermediates for urethane foams (cushioning and insulation), coatings, brake fluids, hydraulic fluids, quenchants, and many other end uses.23 The classic industrial synthesis of this chemical has been the reaction of chlorine with propylene to produce the chloro-hydrin followed by dehydrochlorination with caustic to produce the alkylene oxide, propylene oxide, plus salt.

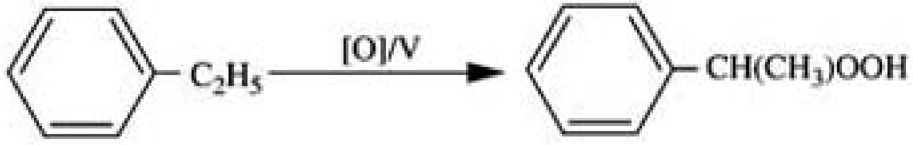


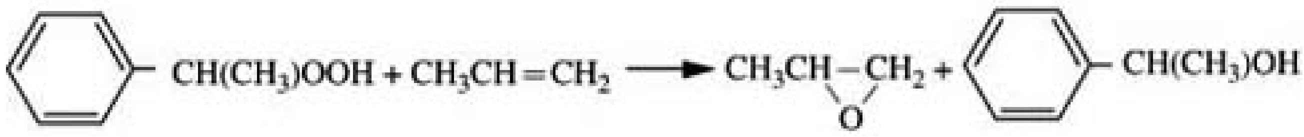
$$CH_3CH=CH_2 + Cl_2 + H_2O \longrightarrow CH_3CH(OH)CH_2Cl$$

$$CH_3CH(OH)CH_2Cl \xrightarrow{\text{Caustic}} CH_3CH\overset{O}{-}CH_2 + \text{Salt}$$

In this reaction sequence, both the chlorine and the caustic used to effect the synthesis are discarded as a valueless salt by-product.

A more economic process has been commercialized. In one version, the hydroperoxide is produced by catalytic air-oxidation of a hydrocarbon such as ethylbenzene. Reaction of this hydroperoxide with propylene yields propylene oxide as a co-product.





This direct peroxidation scheme can be carried out with other agents to give different co-products such as *t*-butanol or benzoic acid.[24,25]

When the economics of the direct peroxidation system are balanced, a significant cost reduction in the preparation of propylene oxide is achieved by eliminating the co-product, salt, which is of low to nil value and thus it presents a disposal problem coupled with all of the related environmental ramifications. Note that in the previous reaction scheme the weight of salt is almost the same as the weight of the epoxide produced and thus great quantities of salt would be produced. In addition, the process can be designed to produce a co-product that can be used or sold as a chemical intermediate. For example, in the case of using isobutene as the starting hydrocarbon, the by-product is *t*-butanol, which can then be converted to methyl *t*-butyl ether, which is a gasoline additive.

Ramifications

If a company is in the business of making and selling products such as those used in the above examples, that is, acetic acid, vinyl chloride, and propylene oxide, as well as other chemicals and if it has plans to stay in business and to expand its facilities and workforce thereby serving growing markets, it must have at least economically competitive processes. Today, this means being competitive with not only any new process developed in the United States, but also with any new process technology developed in the global economy. Another need for market maintenance and growth is a management system that is forward-looking, keenly aware of potential competitive threats, and ready to change directions when needed. The same management team must maintain a highly skilled research, development, and engineering staff to provide the new products and processes as well as product and process innovations and thereby create and maintain the pipeline to future sales, growth, and profitability as existing products or processes become obsolete. Anticipation of market needs must be recognized by both the scientific and management components of a successful firm. In the year 2000 the top ten global companies in sales are listed in Table 1.8.[26]

TABLE 1.8 Sales for the Top 10 Global Chemical Companies in 2000[26]

Company	2000 Chemical Sales (Billions of Dollars)	Rank in 1999
BASF	30.8	1
DuPont	28.4	2
Dow Chemical[a]	23.0	4
Exxon-Mobil	21.5	5
Bayer	19.3	3
Total Fina Elf	19.2	11
Degussa	15.6	9
Shell	15.2	7
ICI	11.8	6
British Petroleum	11.3	10

[a] Sales numbers were before the merger with Union Carbide Corp. Merger is expected to raise Dow Chemical to the No. 1 position.

The companies spending the most for research and development are given in chronological order in Table 1.9.[27]

TABLE 1.9 Research and Development Spending by Chemical and Pharmaceutical Companies[27]

	Spending in Millions of Dollars					Spending as % of Sales
	2000	1999	1998	1997	1996	2000
Chemical Companies						
Dupont	1,776	1,617	1,308	1,116	1,032	6.3
Dow Chemical	892	845	807	785	761	3.9
Rohm and Haas	259	236	207	200	187	3.8
Union Carbide	152	154	143	157	159	2.3
Eastman Chemical	149	187	185	191	184	2.8
Air Products and Chemicals	124	123	112	114	114	2.3
International Flavors	123	104	98	94	94	7.7
Pharmaceutical Companies						
Pfizer	4,435	4,036	3,305	2,536	2,166	15.0
Pharmacia	2,753	2,815	2,176	2,144	1.936	15.2
Merck	2,344	2,068	1,821	1,684	1,487	5.8
Ely Lilly	2,019	1,784	1,739	1,382	1,190	18.6
Bristol-Myers Squibb	1,939	1,843	1,577	1,385	1,276	10.6
American Home Product's	1,688	1,740	1,655	1,558	1,429	12.7

Not noted above is the profound effect that environmental concerns have on new products and processes. It was mentioned earlier that the chemical manufacturing industry spends about five billion dollars annually on pollution and environmental control. Thus, the entire staff of an organization must be aware of these costs and have concerns for the environment when new products and processes are created.

1.4 The Future

What will this huge manufacturing giant called the chemical manufacturing industry look like five or ten years from now? That changes will be made is certain, and an ability to predict and anticipate those changes and to guide the industry or segments to certain changes will certainly bode well for the economic health of any particular company. It may be recalled that earlier in this chapter, it was mentioned that more and more sales of chemicals would be done via E-commerce on the Internet. If the projections are correct, there will be a compounded annual growth rate of 66 percent in chemical E-commercefrom $7.2 billion in 2000 to $150 billion in 2006. Such sales will be carried out through buyers and sellers of chemicals as they develop agreements on purchases; cost estimates including the carrier methodland, water, truck, train; custom matters such as documentation, regulatory fees and taxes; insurance; warehousing matters; and so on. This will have an economic effect on the industry.

The coating and paint industry has undergone huge changes in the past two decades, but there probably will be further change as the new technological areas are sorted out. In paint or trade sales, it would appear that aqueous latexes will dominate the industry for the foreseeable future. It would take a major breakthrough to dislodge them from their place. The industrial-coating industry and their customers must continue to sort through a number of new technologiesconventional solvent-based, high-solids, powder, radiation, water-borne, two-package, and othersin the future. In the past some technologies have been given emphasis at coating companies because they fit existing technology, but they are not necessarily the final winners. Powder coating has captured more than 10 percent of the industrial coating market, and it will probably grow in the future but perhaps not as rapidly as it did in the past decade. Radiation curing appears to be poised for rapid growth. It is on an upsweeping growth curve, and if new products, new end uses, and interest through meeting attendance is any measure, this technology will have a significant portion of the market fiveten years from now.

In the pharmaceutical area, there are many new compounds that will reach the marketplace in the near future. One in particular is insulin for diabetics that can be inhaled or taken orally rather than by injection. The former is in final testing stages and could be on the market in the very near future. The oral type many take longer. Either of these will be "blockbuster" drugs and will mean major sales, growth, and profit for the winning company or companies. The revolution that has taken place in automated drug synthesis and screening will continue and improve. Workers entering the industry or relocating will be expected to have knowledge of parallel and recombinational synthesis methods,28 or they will be rapidly trained within the company in the area. Such screening methods and facilities can screen 200,000 drug candidates in a single day.29 This technology is so useful that it will also creep into other areas of chemical manufacturing that can benefit by screening large numbers of candidate materials as, for example, the paint, coating, and adhesive area.

Interrelated with the pharmaceutical industry is the biotechnology area. Biotechnology products or products derived from biotech-nology processes, are expected to account for 30 percent of the total chemicals market by 2010. Large national and international chemical companies are getting started and have a position in the biotechnology area. Examples of such companies are The Dow Chemical Company, DuPont Company, and Monsanto Company in the United States and Bayer AG, BASF AG, Alusuisse Lonza Group AG, and Degussa AG in Europe. Biotechnology offers both cost effective and environmentally friendly technology and products. The technology will produce proteins and vitamins for animal feed; genetically modified vegetative plants that will resist drought, insects, and cold; new enzyme controlled processes for production of specific chemicals, new fibers for textiles that are derived from renewable raw materials and are biodegradable. Such agro-growth aspects will result in significant losses in other agrochemicals areas such as herbicides and insecticides as resistant, modified plants are developed.

Research, development, and testing will be carried out more and more at independent facilities according to Federal Government reports, and this function will be a significant growth area. This is related to the high cost of specialized investigative tools, which can be shared by a number of companies. Also important is the ability to temporarily hire highly skilled personnel to carry out testing and developmental efforts when they are needed rather than have them as permanent members of a firm's staff. The independent agencies are the ones who will be able to set up the combinational programs for some of the chemical manufacturing segments so set-up costs, in effect, are shared by a number of companies. These programs are expensive to develop and maintain. Many smaller firms would not be able to afford the technology unless some centralized, independent source was available to allow them to share the cost rather than the whole cost burden.

Composites have not been previously mentioned, but they form an important area that is sizable and that will grow in the future. These graphite- or glass-reinforced materials are useful in many markets that need strong, shaped articles including the aircraft/aerospace, automotive, recreation, general industrial, and similar markets. In addition to strength, composites often offer weight savings plus the ability to rapidly produce complex-shaped, small to large articles.

Nanotechnology has great promise for the chemical industry. This is an emerging technology whose aim is to place atoms and molecules in particular arrays, a technique termed "positional assembly," and to have this done repetitively through "self-replication."30 This sounds more like science fiction than chemistry. But, nanotechnology already produces significant sales, and the sales are predicted to grow from a base of $200 million in 2002 to $25 billion by 2012.31 This is an astounding 62 percent compounded growth rate for 10 years.

What is nanotechnology? Imagine having a machine that can go forward, backward, right, and left as well as up and down at various angles. Now imagine that this machine is very, very smallin fact, so small that it is approaching atomic dimensions and it is measured in terms of nanometers. A nanometer is 1 × 10-9 meters, and we certainly cannot see things this small. However, regardless of these difficulties, we want to build and control this machinethat is, have it do whatever we want it to do, which is to place atoms and molecules in particular arrays so we end up with a desired product. Still not satisfied, we want the machine to do this over and over and over again.

One might ask why we want to do this. We all know that everything is made up of atoms. Chemicals are everywhere and everything. The difference in the carbon in coal and in diamonds is the way the atoms of carbon are arranged. Arrange the atoms properly and a worthless pile of carbon becomes a precious diamond. If we were able to arrange the atoms in air, dirt, and water into a desired configuration, it would be possible to make, for example, carrots, beets, potatoes, and so on. At present, the transformations are made by nature, using a gene system to combine the ingredients in the proper way. Properly align the atoms of a material that is to be used as a filler or reinforcing material, and one can envision super strong composite materials resulting. If one were to rearrange the atoms of sand and in so doing add a few trace elements, the end result could be a computer chip.32 The goals of nanotechnology are to:

1. Arrange every or almost every atom in a desired structure in its proper place,
2. Make effectively any structure that can be atomically specified and that does not violate laws of chemistry and physics, and
3. Have manufacturing costs that are basically energy and raw material costs.

The other concepts associated with nanotechnology are those described above, positional assembly and self-replication. Nanotechnology is currently being used for light-emitting polymer films, in computer applications, electrically conductive adhesives, and other areas.

These are but only a few of the potential areas for chemical manufacturing during the first decade of the twenty-first century. Most probably, many of the products and processes that will be in place in 2012 or so cannot even be imagined today.

1.5 References

Chemical Economics Handbook, Stanford Research Institute, Menlo Park, CA.
"Frequently Asked Question on NAICS, SIC, and Business to Business Market," Internet, http://www.naics.com/faq.htm (Jan. 11, 2002).
Anon, "Top 100 Shrinks to 75," Chem. Eng. News, 77, 18 (May 3, 1999).
Anon, "The Urge to Merge," Chem. Eng. News, 77, 36 (Sept. 6, 1999).
Smith, Greg A., "A Kinder, Gentler, Less Productive Capitalism," in Fearless Forecast, p. 5, Prudential Financial, Jan. 2002.
"Greenspan Sees Increasing Signs of U.S. Recovery," Internet, http://www.foxnews.com/0,3566,46664,00.html (Feb. 27, 2002).
"Gross Domestic Product by Industry in Current Dollars," U.S. Census Bureau, 2001.
U.S. Department of Commerce, Bureau of Economic Analysis, "U.S. International Trade in Goods and Services," BEA News Releases for various years.
Bailey, Jr., F. E., and Koleske, J. V., Chapter 1, "Economic Aspects of the Chemical Industry," in Riegel's Handbook of Industrial Chemistry (9th ed.), James A. Kent (Ed.), Van Nostrand Reinhold, 1992.
"Industrial Profile," U.S. Office of Industrial Technology, http://www.oit.doe.gov/chemicals/profile.shtml.
"Chemical-Energy Use," U.S. Office of Industrial Technology, http://www.eia.doe.gov/emeu/mecs/iba/chemicals/page2.html.
"Chemists and Material Scientists," http://stats.bis.gov/oco/ocos049.htm.
"Chemical Engineers, Earnings," U.S. Department of Labor, http://stats.bls.gov/oco/ocos029.htm.
Kiefer, David M., "It Was All About Alkali," Today's Chemist, 11(1), 45 (Jan. 2002).
Kiefer, David M., "Soda Ash, Solvay Style," Today's Chemist, 11(2), 87 (Feb. 2002).
Belgian Patent 713,296 issued to Monsanto Co.
Parshall, G. W., "Organometallic Chemistry in Homogeneous Catalysis," Science, 208, 1221 (1980).
Anon, Chem. Week, 126(3), 40 (1980).
Anon, Chem. Eng. News, 58, 6 (1980).
Koleske, J. V., and Wartman, L. H., Poly(vinyl chloride), Gordon and Breach, NY, 112 pp, 1969.
British Patent 1,2027,277 to PPG Industries and British Patent 1,016094 to Toyo Soda.
Fedor, W. S., Chem. Eng. News, 43, 80 (Sept. 12, 1965).
Anon, Chem. Week, 125(6), 22 (1979).
Bailey, Jr., F. E., and Koleske, J. V., Alkylene Oxides and Their Polymers, 261 pp. Marcel Dekker, Inc., NY, 1991.
French Patent 1,460,575 to Halcon Corp.
Anon, Chem. Week, 127, 58, No. 1, 6 (1980).
"2001 Chemical Industry Review," Chem. Eng. News, 79(52), 13 (Dec. 24, 2001).
Anon, "Facts and Figures for Chemical R&D," Chem. Eng. News (Oct. 29, 2001), Internet, http://Pubs.ACS.org/CEN.
Czarnik, A. W., and DeWitt, S. H., "A Practical Guide to Combinational Chemistry," Am. Chem. Soc., 360 (1997).
Drews, J., Science, 287, 1960 (2000).
Drexler, K. Eric, Nanosystems: Molecular Machinery, Manufacturing, and Computation, Wiley, 1992.
Kramer, Karen L., "Nanotechnology Set for Exponential Growth," Adhs. Seal. Ind., 9(2), 6 (Mar. 2002).
Merkle, Ralph C., Internet Website http://www.zyvex.com/nano/.

2.1 Introduction

This chapter is intended to familiarize the reader with the concepts of pollution prevention and waste minimization, and to help understand how these practices are preferable to the more traditional practices of pollution control and waste management.

1. Every year, U.S. industry emits literally billions of pounds of pollutants into the environment. For example, in 1998 facilities reported releases of 7.3 billion pounds of chemicals tracked in the Toxics Release Inventory to air, water, or land.1
2. These releases have environmental impacts, leading to Federal and State regulations. Pollution control costs exceed $100 billion annually.2
3. Over 20,000 generators produced 40.7 million tons of hazardous waste in 1997.3 Improper management of hazardous and other industrial wastes has required remediation and restoration of over 1,400 "Superfund" sites, and parties have agreed to over $15 billion in cleanup settlement costs.4 Improper management of hazardous waste can also result in large fines. One firm in the northwest United States agreed to pay over $170 million to settle charges that it violated hazardous waste laws.5

The idea behind pollution prevention is simply that it is easier and cheaper, and better for the environment, to design a process to prevent pollution in the first place, rather than to clean it up later. The U.S. Congress set forth a clear policy with the passage of the Pollution Prevention Act (PPA) of 1990:

(b) POLICY: The Congress hereby declares it is to be the national policy of the United States that pollution should be prevented or reduced at the source whenever feasible; pollution that cannot be prevented should be recycled in an environmentally safe manner, whenever feasible; pollution that cannot be prevented or recycled should be treated in an environmentally safe manner whenever feasible; and disposal or other release into the environment should be employed only as a last resort and should be conducted in an environmentally safe manner. [42 U.S.C. 13101.]

The Congress also made a number of relevant findings. See Section 6602(a). In short, the Congress found that industry and government were relying too much on the end of pipe controls and missing opportunities for source reduction.

The Clean Air Act (CAA), Clean Water Act (CWA), and the Solid Waste Disposal Act (SWDA) also contain provisions encouraging pollution prevention (P2). See, for example, Section 1003(b) of the SWDA, which states:

NATIONAL POLICY. The Congress hereby declares it to be the national policy of the United States that, wherever feasible, the generation of hazardous waste is to be reduced or eliminated as expeditiously as possible. Waste that is nevertheless generated should be treated, stored, or disposed of so as to minimize the present and future threat to human health and the environment. [42 U.S.C. 6902.]

The terms "pollution prevention" and "waste minimization" are often used interchangeably. Pollution prevention has a broader connotation (e.g., eliminating air and water emissions are P2), but in this chapter, since we are talking only about waste reduction, the meanings are essentially the same. Waste minimization includes both activities to reduce waste at the source, and environmentally sound recycling.

The Agency charged with formulating and leading the U.S. effort to reduce or eliminate hazardous waste generation is the U.S. Environmental Protection Agency (EPA). EPA administers to the two preceding statutes (the SWDA and PPA). EPA conducts activities to support waste minimization out of both its headquarters office in Washington, DC, and in its 10 regional offices. Also, virtually every state now has an agency charged with helping to facilitate waste minimization. Often, these agencies are the same as the environmental control agency, although several states have set up separate agencies for P2. EPA has determined that two important roles it can play to facilitate P2 is to set goals for the nation, and measure progress towards those goals.

In 1994, EPA issued the Waste Minimization National Plan.6 The Plan was the result of a series of stakeholder meetings. Although the Plan is not a consensus document, EPA found there was strong public support for certain key elements of a national waste minimization program. EPA identified three major goals for industry and government:

1. To reduce, as a nation, the presence of the most persistent, bioaccumulative, and toxic (PBT) constituents in hazardous waste by 50 percent by the year 2005 (from a 1991 baseline).
2. To avoid transferring these constituents across environmental media.
3. To ensure these constituents are reduced at their source whenever possible, or, when not possible, that they are recycled in an environmentally sound manner.7

EPA noted that not every generator of hazardous waste can achieve the 50 percent goal at their particular facility. The goal is meant as an aggregate benchmark that industry, states, and EPA can use to measure progress in the United States.

2.2 Incentives

First and foremost, pollution is inefficiency. No process is 100 percent efficient, but pollutants, whether air or water discharges, energy, or solid waste, represent material that cannot be sold for revenue but instead are lost. Furthermore, in the United States and in most industrialized countries, pollutant discharges are tightly regulated, and investment in pollution control is required. Solid wastes, and especially hazardous waste, require special and costly management. Also, under the U.S. law, disposal of waste in accordance with current regulations is not necessarily a shield against future liability.

The release of toxic chemicals and production of hazardous waste, even in compliance with all applicable laws, has a cost in lost public goodwill. Every year, EPA publishes the Toxic Release Inventory (TRI). The TRI is carried out by EPA under authority of Section 313 of the Superfund Amendments and Reauthorization Act of 1986. TRI covers most manufacturing sectors, and was expanded in 1998 to cover the mining and utility industries, among others. This report documents releases of toxic chemicals to air, water, and land from most of the major manufacturers in the United States. Community and environmental groups pay close attention to these statistics, and it is not unusual for the local newspaper to publish ranked lists of the "worst polluters" in the county or state.

Smart companies have found that they can turn environmental improvements into good publicity. A number of State agencies now give special awards for P2 achievements. For example, the state of Minnesota awards an annual Governor's Award for Excellence in Waste and Pollution Prevention. The National Pollution Prevention Roundtable (NPPR) also has provided recognition for P2 achievements, including the Most Valuable Pollution Prevention (MVP2) Awards, and the PBT Cup.8 The MVP2 Award honors the most innovative and successful P2 programs in the country. In 2000, the first place winner was the Michigan Department of Environmental Quality and the Michigan Pulp and Paper Environmental Council for their cooperative work to reduce air, water, and waste emissions from the pulp and paper industry. The PBT Cup (given in partnership with EPA) was awarded to Morton Powder Coatings for their complete elimination of cadmium from their industrial coating operation. Companies that won the PBT Cup awards in 1999 included:

1. American Video Glass: reduced lead in their products
2. Stanley Works: optimized metal finishing and plating to recover zinc

Additionally, both the U.S. Department of Defense (DOD) and the Department of Energy have set up P2 award programs for their installations and contractors. For example, the DOD grants an annual PPA to recognize P2 efforts from any level of the military departments, defense agencies, military or civil installation worldwide.9

Numerous other companies have achieved benefits from P2 projects. Table 2.1 identifies just some of the success stories.

TABLE 2.1 Examples of Successful Waste Minimization and P2 Projects

Company/Operation	Changes Made	Estimated Annual Savings	Environmental Benefits
GM Hamtramck/auto painting plant paint use	Changed timing of paint jets to reduce	$85,000	Reduced VOC air emissions
Ford Yipsilanti/auto starters	Changed cleaning solution from TCE to water	$25,000	Reduced TCE air emissions and waste generation
FMC Naval Systems/part cleaning	Eliminated use of 1,1,1-tricholoroethane	$100,000	Reduced incineration of hazardous waste
Harris Broadcast Division/radio & TV manufacturer	Several changes to reduce paint, solvents	$125,000	Reduced air emissions
HADCO/printed wiring boards	Eliminated chlorinated solvents	$600,000	Reduced air and water emissions, eliminated hazardous waste generation
Charles H. Lilly/batch pesticide formulator	Reuse of in-plant wastewater	$30,000	90% reduction in hazardous waste generation
PPG/auto coatings	Recycle plant wastewater	$205,000	80% reduction in hazardous waste

Sources: U.S. EPA Reports, EPA-530-F-97-09, 2023, 2526, all August of 1997.

2.3 Methods of Eliminating Waste

By law, a generator of hazardous waste must have a program in place to reduce the volume or toxicity of their waste "…to the degree determined to be economically practicable." See Section 3002(b)(1) and 3005(h) of the SWDA. If the generator is also required to obtain a hazardous waste permit, the generator must maintain at the facility a written certification saying that he has a program in place to reduce the volume or toxicity of the waste. See 40 CFR 264.73(b)(9) in the Code of Federal Regulations (CFR). The generator determines what is practicable. EPA provided guidance10 on what a program would include:

1. top management support, such as making waste minimization part of organization policy, and designating a waste minimization coordinator
2. characterization of waste generation and waste management costs
3. periodic waste minimization assessments
4. a cost allocation system, that is, charge waste management costs to the operation that generates the waste, not to "overhead"
5. encourage technology transfer within the company and industry, and take advantage of government and university technical assistance
6. use program implementation and evaluation to see what worked and what did not

Some states require more extensive measures, such as a detailed plan showing how the generator will reduce their waste. For example, Massachusetts and California require formal plans to reduce waste. As described previously, EPA also recommends that waste minimization efforts focus on certain key chemicals, by source reduction or environmentally sound recycling. Table 2.2 shows one way of systematically identifying P2 opportunities.

TABLE 2.2 Example Methodology to Identify P2 Opportunities

1. Understand the types and amounts of chemicals and wastes emitted by your plant	Identify emissions and waste generation information from sources such as: 1. Toxics Release Inventory (TRI) 2. Biennial Hazardous Waste Report (BRS) 3. Clean Air and Clean Water Act permits 4. Hazardous and solid waste permits
2. Develop a short list of candidate chemicals for reduction or elimination	Evaluate and prioritize emissions and waste generation by considering factors such as: 1. Potential long-term liability 2. Concerns of regulatory agencies, workers, and community groups 3. Costs of emission control and waste management
3. Consider P2 options for candidate chemicals	Evaluate technical and economic feasibility of changes such as: 1. Substitution of raw materials 2. Revisions to process flows, schedules 3. Reuse of chemical in the same process 4. Reuse of chemical in another process in the plant 5. Reuse of chemical in another process elsewhere in the company 6. Recycling of the chemical by commercial vendor
4. Design projects to reduce or eliminate top priority candidate chemicals	Develop detailed project plans including: 1. Resource requirements (in house and contractor) 2. Schedules and milestones 3. Expected benefits in costs savings, reduced liability, improved regulatory compliance 4. Time frame for expected benefits (payback)
5. Secure support and approvals for P2 projects	Obtain support and approval from: 1. Company management 2. Affected employees 3. Regulatory agencies (if required) 4. Community groups (as appropriate)
6. Implement changes	Carry out project plans: 1. Purchase equipment and materials 2. Retrain employees as needed 3. Obtain permit modifications as needed
7. Evaluate impact of changes	Determine if benefits were achieved: 1. Cost savings 2. Improved worker and community relations 3. Improved regulatory compliance
8. Repeat Steps 1 and 2	Reassess emissions and waste generation to determine if another round of P2 changes are warranted

The types of waste minimization changes generally can be grouped under source reduction and recycling (defined in Appendix I). Source reduction is preferred because waste is never produced in the first place. In practice, the line between source reduction and recycling is not that clear, for example, a by-product can be sold as a feedstock to another company, who finds it an improved input to their own manufacturing process. As long as toxic constituents are not released or transferred to another media, either form of waste minimization yields environmental benefits.

2.4 Technical Assistance

A leading organization in the P2 field is the NPPR. NPPR is a non-profit organization that serves as a national forum "… promoting the development, implementation, and evaluation of efforts to avoid, eliminate, or reduce waste generated to air, land, and water."11 NPPR sponsors activities in the areas of information exchange, public policy at federal and state levels, general education on P2, and external partnerships with organizations that have compatible missions. NPPR maintains several services that can provide P2 technical assistance, including a searchable on-line data base of P2 experts, yellow pages of P2 service providers, and a P2 reference collection.

The U.S. EPA, and a number of state governments, also provide technical assistance for P2. For example, EPA Region III in Philadelphia, Pa., maintains a Business Assistance Center in part to provide P2 help to small and medium sized businesses in the mid-Atlantic states. The Waste Reduction Resource Center provides support to all of the mid-Atlantic and Southeastern states through technical experts and a clearinghouse of 14,000 documents. The Waste Reduction Resource Center is located in Raleigh, North Carolina. The center is supported financially by the EPA and the Tennessee Valley Authority. In the Midwest, the Waste Management and Research Center (associated with the University of Illinois) provides a range of services, including P2 assessments, compliance assistance, and technology evaluation testing.

2.5 Measuring Progress

As discussed previously, an individual facility can measure its progress in achieving P2 by examining releases under its TRI submission, its Biennial Hazardous Waste Report, and Clean Air and Water Act discharge reports. A facility may chose a baseline of emissions prior to introduction of P2-related changes to measure actual reductions in emissions that are achieved. A company may chose to measure its overall progress across all of its facilities, to compare versus a corporate goal. Some companies have set goals for the eventual elimination of all waste, and other ambitious achievements in reduced effluents on a corporate basis. For example, a major manufacturer of aluminum, set the following goals as a corporation: reduce SO2 emissions by 60 percent, VOC emissions by 50 percent, NOX emissions by 30 percent, mercury emissions by 80 percent, landfilled waste by 50 percent, and water usage by 60 percent, greenhouse gas emissions by 25 percent, all generally over 710 years with a baseline of 2000.12

EPA reports that total releases of toxic chemicals to air, land, water, and underground injection have decreased by 1.5 billion pounds, or 45 percent, from 1988 to 1998.13

2.6 International Perspectives

There is tremendous interest in P2 across the globe. The United Nations Environment Programme (UNEP) launched The International Declaration on Cleaner Production in 1998.14 This is a voluntary but public statement of commitment to the strategy and practice of cleaner industrial production. By November 2000, over 200 governments and non-government entities had signed the Declaration. UNEP has also sponsored an international treaty to reduce or eliminate emissions and discharges of 12 specified persistent organic pollutants (POPs).

The Organization of Economic and Community Development (OECD) through its Working Group on Waste Management Policy, has conducted extensive work to promote proper waste management as well as P2. The Working Group is cooperating with the UN Secretariat of the Basel Convention to identify areas of possible harmonization of OECD and UN programs. The Working Group sponsored workshops on developing standards for environmentally sound management of wastes destined for recovery operations in 1999 and 2000. The Working Group is presently developing a Waste Minimization work program focusing on:

1. Government Policy Options and Self-assessment Guide on Strategic Waste Prevention
2. International waste prevention performance evaluation tools, including quantitative indicators.15

2.7 Evolving Trends and Challenges

Moving into the twenty-first century, a number of trends have begun to take shape in industries across the developed nations, which are sometimes grouped under the umbrella of "industrial ecology."16 Under this evolving philosophy, products are viewed as long-term assets, and not as materials that are consumed. Recycling, re-manufacture (disassembling worn out parts to make a rebuilt product), and redesign of products with energy and environmental implications in mind, are all part of a rethinking of material and energy use.

A related concept is "product stewardship," where a manufacturing industry retains some level of responsibility for products it produces, even when the general public is the ultimate user of the product. In the United States, there are examples today of industries voluntarily setting up "take back" systems to prevent their products from entering the municipal solid waste stream. For example, the Rechargeable Battery Recycling Corporation collects rechargeable nickelcadmium batteries and the Thermostat Recycling Corporation collects used mercury switch thermostats. Both of these organizations are voluntary, industry-supported corporations to encourage collection and proper management of these potentially hazardous consumer products.

The goal behind these efforts, and P2 in general, is to develop an economic and ecosystem that is virtually self-sustaining. Economic improvement can be achieved through increased efficiency, rather than solely based on exploitation of large quantities of materials and energy. Industries may eventually be clustered physically to allow one industry to use the byproducts of another as feedstocks in making new products in what have been termed "eco parks." Eventually, the word "waste" may become antiquated, as materials are seen as useful sources of energy and materials, and as products are designed "from cradle to re-birth."17

A related concept is chemical management service (CMS). Under a CMS arrangement, a customer (such as a manufacturing firm) no longer simply buys chemicals from a chemical supplier. Rather, the customer enters into an arrangement with a chemical service provider, under which the latter manages the entire chemical use system, from purchasing, to inventory, to waste.18 The service provider is paid for delivering and managing chemicals. Financial incentives are built in to help customers reduce inventory and waste.

Government is facing a number of challenges ranging from environmental regulation to tax and land use policy, even to antitrust legislation, to facilitate this evolution to national and regional self-sufficiency. Businesses of all sizes are faced with the challenge of creating value for shareholders as raw materials become more scarce and more expensive and efficiency is a main aspect of competition.

2.8 Appendix I: Glossary of P2 Terms

By-product means a material that is not one of the primary products of a production process and is not solely or separately produced by the production process. Examples are process residues such as slags or distillation column bottoms. 40 CFR 261.1(c)(3).

Hazardous waste means a waste identified by EPA as hazardous according to 40 CFR Part 261, or by the appropriate state environmental agency.

Pollutant as used in P2 and waste minimization terminology is an unwanted toxic chemical or substance that is released to the environment, or becomes part of a waste.

Pollution prevention (P2) is defined by EPA as source reduction (defined below) and other practices that reduce or eliminate the creation of pollutants through increased efficiency in the use of raw materials, energy, water, or other resources, or protection of natural resources by conservation.

Reclamation is a type of recycling. A material is reclaimed if it is processed to recover a useable product, or if it is regenerated. Examples are recovery of lead values from spent batteries and regeneration of spent solvents. 40 CFR 261.1(c)(4).

Recycling includes material use, reuse, and reclamation. 40 CFR 261.1(c)(7).

Source reduction means any practice which reduces the amount of any hazardous substance, pollutant, or contaminant entering a waste stream or otherwise released to the environment (including fugitive emissions) prior to recycling, treatment, or disposal; and reduces the hazards to public health and the environment associated with the release of such substances, pollutants, or contaminants. The term includes technology or equipment modifications, process or procedure modifications, reformulation or redesign of products, substitution of raw materials, and improvements in housekeeping, maintenance, training, or inventory control. The term does not include any practice which alters the physical, chemical, or biological character or the volume of a hazardous substance, pollutant, or contaminant through a process or activity which itself is not integral to and necessary for the production of a product or the providing of a service.19

Use or *reuse* (generally used interchangeably) means a material is (1) employed as an ingredient (including use as an intermediate) in an industrial process to make a product (for example, distillation bottoms from one process used as a feedstock in another process), or (2) employed in a particular function or application as an effective substitute for a commercial product. 40 CFR 261.1(c)(5).

Waste is an unwanted material or material with no economic value. The SWDA (Section 1004(27)) defines "solid waste" as "…any garbage, refuse, sludge from a waste treatment plant, water supply treatment plant, or air pollution control facility and other discarded material, including solid, liquid, semisolid, or contained gaseous material resulting from industrial, commercial, mining, and agricultural operations, and from community activities…" EPA has issued a regulatory definition of solid waste for hazardous waste regulatory purposes that includes activities such as land disposal, incineration, and includes certain kinds of recycling activities (e.g. burning for energy recovery) but excludes others (direct use in a manufacturing process).20

Waste minimization is defined by EPA as source reduction or environmentally sound recycling.21 EPA believes that recycling activities closely resembling conventional waste management activities do not constitute waste minimization, for example, generally, burning a waste for energy value is not waste minimization.

2.9 Notes

See the Summary of 1998 Toxics Release Inventory Data, U.S. EPA, May 2000.

The major studies on the cost of pollution control include: Environmental Investments: The Cost of a Clean Environment. Report of the Administrator, EPA, Washington, DC, Island Press, 1991; and The Department of Commerce study, Pollution Abatement and Control Expenditures, 19721992. Survey of Current Business, May 1994.

National AnalysisThe National Biennial RCRA Hazardous Waste Report (Based on 1997 data). *EPA530-R-99-036c*, Sept. 1999.

See the EPA report Superfund Reforms Annual Report, FY 1998, EPA-540-R-99-002, Mar. 1999.

See the press release from EPA Region 10, FMC to Pay Largest RCRA Settlement in Environmental Enforcement History, Oct. 1998.

See the EPA Waste Minimization National Plan, *E*PA/530-R-94-045, Nov. 1994.

EPA Waste Minimization National Plan, p. 3.

Information on these awards can be obtained by contacting the National Pollution Prevention Roundtable and the U.S. EPA.

Memorandum of Sherri Goodman, Deputy Under Secretary of Defense (Environmental Security), FY 1998, Secretary of Defense Environmental Security Awards Program.

The EPA guidance was published in the Federal Register of May 28, 1993. 58 FR 31.1.1.4-20.

The National Pollution Prevention Roundtable Mission Statement, Goals and Objectives are available on the NPPR web site (www.p2.org).

Alcoa press release dated Apr. 25, 2001.

Environmental Protection Agency Fiscal Year 2000 Annual Report, EPA-190-R-01-001, pp. 1176, Mar. 2001.

See the UNEP Declaration on Cleaner Production Signatory Summary, Nov. 16 2000, Information about the Declaration may be obtained from the UNEP web site (see www.uneptie.org).

See the OECD web site.

See the report, Materials: A Report of the Interagency Working Group on Industrial Ecology, Material and Energy Flows. Several agencies including the EPA, the Department of the Interior, and the Department of Energy (under the auspices of the White House Council on Environmental Quality and Office of Science and Technology Policy) along with non-government participants issued a document in August 1998 as an overview to industry ecology and materials flow analysis.

Materials: A Report of the Interagency Working Group on Industrial Ecology, Material and Energy Flows, p. 18, Aug. 1998.

Kaufman-Johnson, Jill, Chemical Management Services: Good for Business, Good for the Environment, California Manufacturing and Technology Association On-Line Journal, Jan. 2001.

Section 6603(5) of the Pollution Prevention Act of 1990.

See 40 CFR 261.2 and 261.4 in the U.S. Code of Federal Regulations.

See 58 FR 31115, published in the Federal Register, May 29, 1993, and EPA's Waste Minimization National Plan, EPA530-R-94-045, Nov. 1994.

3.1 Introduction

As described in the previous chapter, profitability in today's global marketplace demands that companies minimize wastes, both in the form of discharges to surface water, groundwater, and air, and generation of solid wastes sent to landfills. In addition to the valuable raw materials lost, the cost of disposing of these wastes in compliance with the protective state and federal regulatory programs in place for the new millennium is high. Furthermore, the cost of remediation of releases from wastes inadequately treated and disposed prior to these regulations has convinced many companies that they must build consideration of waste management costs into the design of their products and manufacturing processes.

Despite this strong incentive, elimination or recycling of some of the wastes cannot yet be achieved due to prohibitive costs or technology challenges yet to be solved. When this was last estimated in 1991,** it was projected that by 1995 environmental regulations would be costing the United States $50 billion annually for water pollution control, $30 billion annually for air pollution control, and $35 billion annually for land and groundwater pollution control. This is not surprising since environmental problems have spawned 50 categorical industrial wastewater standards, and 186 categorical industrial air emission standards, some of which are still being completed.† EPA also has issued standards regulating virtually all of the 40 million tons of hazardous waste disposed each year.

**National Academy of Public Administration, *Setting Priorities, Getting Results A New Direction for the Environmental Protection Agency*, Washington DC, May 1995.

†Data only exist in EPA Website@www.epa.gov.ttn/atw/socatlst/socatpg.html & www.epa.gov.waterscience/guide

3.2 Federal Laws and Regulations

There are currently 17 key federal environmental protection laws related to the proper management of industrial wastewater, solid/hazardous waste and potentially hazardous materials. Table 3.1 provides a list of these laws, their objectives, and the primary federal agency charged with implementing each law. For the convenience of the reader all major statutes are shown, not just those related to the primary topics of this chapter. In addition, there are many state laws and regulations, which can add different and more stringent requirements than the current federal requirements.

TABLE 3.1 Summary of Key Federal Environmental Laws and Lead Agencies Related to Management of Industrial Wastewater, Solid/Hazardous Waste, and Potentially Hazardous Materials

Objective	Federal Law	Lead Agency
Water pollution control	Clean Water Act (CWA)	EPA
	Safe Drinking Water Act (SDWA)	EPA
	Marine Protection Research and Sanctuaries Act (MPRSA)	EPA/USACOE
Hazardous/solid waste control	Hazardous and Solid Waste Amendments of 1984 (HSWA)	EPA
	Resource Conservation and Recovery Act (RCRA)	EPA
	Comprehensive Environmental Response Compensation and Liability Act (CERCLA)	EPA
	Hazardous Materials Transportation Act	DOT
Air pollution control	Clean Air Act (CAA)	EPA
Promoting pollution prevention	Pollution Prevention Act (PPA)	EPA
Emergency planning/activities disclosure	Superfund Amendments and Reauthorization Act (SARA)	EPA
Management of risk from toxic chemicals	Toxic Substances and Control Act (TSCA)	
	Federal Insecticide, Fugicide, and Rodenticide Act	EPA
	Federal Environmental Pesticide Act	EPA
Impact of federal government on the environment	National Environmental Policy Act (NEPA)	CEQ/EPA

EPA: U.S. Environmental Protection Agency (www.epa.gov).
DOT: U.S. Department of Transportation (http://hazmat.dot.gov/).
CEQ: Council on Environmental Quality (http://ceq.eh.doe.gov/nepa/nepanet.htm).
USACOE: U.S. Army Corps of Engineers (no headquarters public Internet site available at this time).

Through the use of the Internet sites of these agencies (provided in the table), an expansive collection of materials is available to understand the specific requirements and programs associated with these laws.

3.3 Industrial Wastewater Treatment

Treatment of industrial wastewater often will involve multiple unit operations. The number of operations needed will be determined by the chemical composition of the wastewater. Various combinations of up to 18 basic types of operations maybe required. Some direct assistance describing both the pollutants present and the technologies applicable to wastewater from specific industrial sectors is available from the EPA Internet site. A list of categories that EPA has studied is provided in Table 3.2. (See http://www.epa.gov/OST/guide/ for additional information.)

TABLE 3.2 Industry-Specific EPA Industrial Wastewater Technical Documents

Documents Related to Regulations

1. Aquaculture
2. Aluminum Forming
3. Asbestos Manufacturing
4. Battery Manufacturing
5. Builders' Paper and Board Mills
6. Carbon Black Manufacturing
7. Cement Manufacturing
8. Centralized Waste Treatment
9. Coal Mining
10. Coil Coating
11. Commercial Hazardous Waste Combustors
12. Construction and Development
13. Concentrated Animal Feeding Operations
14. Copper Forming
15. Dairy Products Processing
16. Electrical and Electronic Components
17. Electroplating
18. Explosives Manufacturing
19. Feedlots
20. Ferroalloy Manufacturing
21. Fertilizer Manufacturing
22. Fruits and Vegetables Processing
23. Glass Manufacturing
24. Grain Mills
25. Gum and Wood Chemicals Manufacturing
26. Hospitals
27. Iron and Steel Manufacturing
28. Industrial Containers and Drum Cleaning
29. Industrial Laundries
30. Ink Formulating
31. Inorganic Chemicals
32. Landfills
33. Leather Tanning and Finishing
34. Meat and Poultry Products
35. Metal Finishing
36. Metal Molding and Casting (Foundries)
37. Metal Products and Machinery
38. Mineral Mining and Processing
39. Nonferrous Metals Forming and Metals Powders
40. Nonferrous Metals Manufacturing
41. Oil and Gas Extraction
42. Ore Mining aud Dressing
43. Organic Chemicals, Plastics and Synthetic Fibers
44. Paint Formulating
45. Paving and Roofing Materials (Tars and Asphalt)
46. Pesticide Formulating, Packaging, and Repackaging
47. Petroleum Refining
48. Pharmaceutical Manufacturing
49. Phosphate Manufacturing
50. Photographic Processing
51. Plastics Molding and Forming
52. Porcelain Enameling
53. Pulp, Paper, and Paperboard
54. Rubber Manufacturing
55. Seafood Processing
56. Soap and Detergent Manufacturing
57. Steam Electric Power Generating
58. Sugar Processing
59. Textile Mills
60. Timber Products Processing
61. Transportation Equipment Cleaning

Miscellaneous Documents

1. Concrete Products Industry
2. Drum Reconditioning
3. Ethanol-for-Fuel Industry
4. Fish Hatcheries
5. Foods and Beverages (Miscellaneous)
6. Livestock Market Industry
7. Low BTU Gasification
8. Polychlorinated Biphenyls
9. Printing and Publishing
10. Shipbuilding Industry
11. Solvent Recycling Industry
12. Used Oil Reclamation and Re-Refining Industry
13. Water Supply

Primary Treatment

This initial process captures and removes large and heavy solids (grit) and debris (such as rags or tree limbs) that may interfere with or damage subsequent treatment equipment. Devices such as screens, comminutors (to grind up this material), and settling tanks with short residence times are used for this purpose.

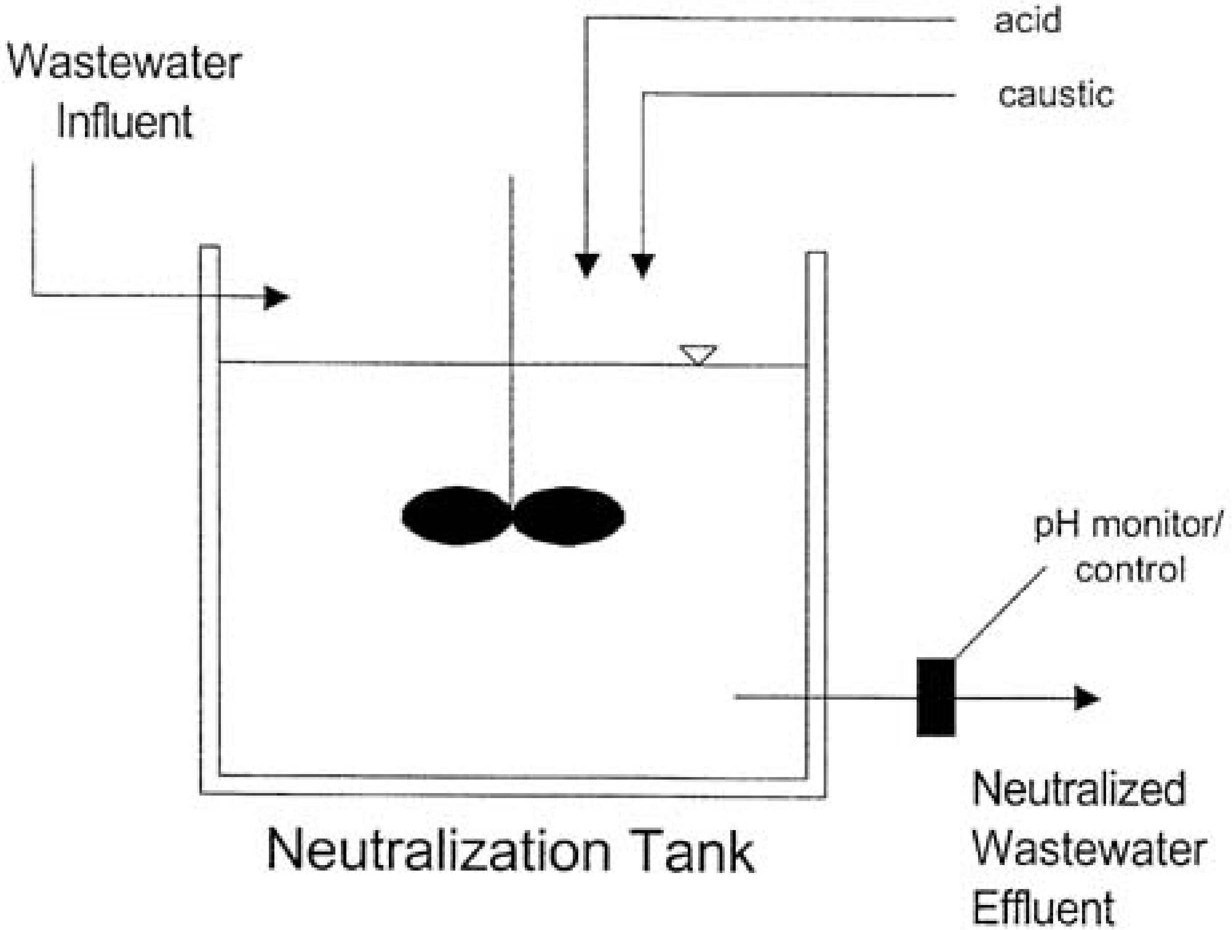
Equalization

Unless the flow of wastewater is steady and of uniform composition, an equalization tank often is advisable. Equalization allows consolidation of smaller or intermittent flows from process operations, dampening the effect of peak and minimum flows from continuous operations, and the mixing of wastewater to provide a more uniform chemical composition for treatment operations that cannot be optimized if that composition is widely variable. It also allows batch treatment when that is the most effective or economical choice. Equalization tanks frequently include agitators or aerators to prevent solids from settling in the tank. It is important to make sure that an equalization tank does not have a negative impact on treatment effectiveness. If excessive dilution of pollutants occurs during equalization, subsequent treatment may become more difficult and expensive. Intentional dilution to avoid regulatory requirements based on concentrations often is prohibited by regulations and permits, since the pollutants will then pass into the environment untreated.

Neutralization

Control of pH is critical to the effectiveness of several wastewater treatment operations, including biological digestion of organics and chemical precipitation to remove metals. Acids such as sulfuric or hydrochloric acid or alkalies such as sodium hydroxide are used to adjust the pH to the desired range for treatment. A last step of pH control also may be needed following all other unit operations to assure that the discharge does not damage the receiving water or municipal treatment works that receives the discharge. Most facilities are subject to regulatory requirements that discharges met the pH range of 69. Figure 3.1 shows a diagram of a neutralization tank.

Fig. 3.1 Neutralization system diagram. (*Courtesy of EPA*)



Coagulation

Coagulation is the enhancement of settling of solids (either biological solids or chemical precipitates) by increasing the mass of the particles by causing them to "stick together" through the use of chemical compounds. Coagulation usually occurs in a flocculation tank in which coagulants are added in a highly mixed environment then followed by a less aggressive mixing to encourage joining and bridging of the solids into larger, heavier particles. This treatment increases the rate of gravity settling in settling tanks (clarifiers, as described later) and are more easily removed by filters, which can be used as a polishing step after the clarifier.

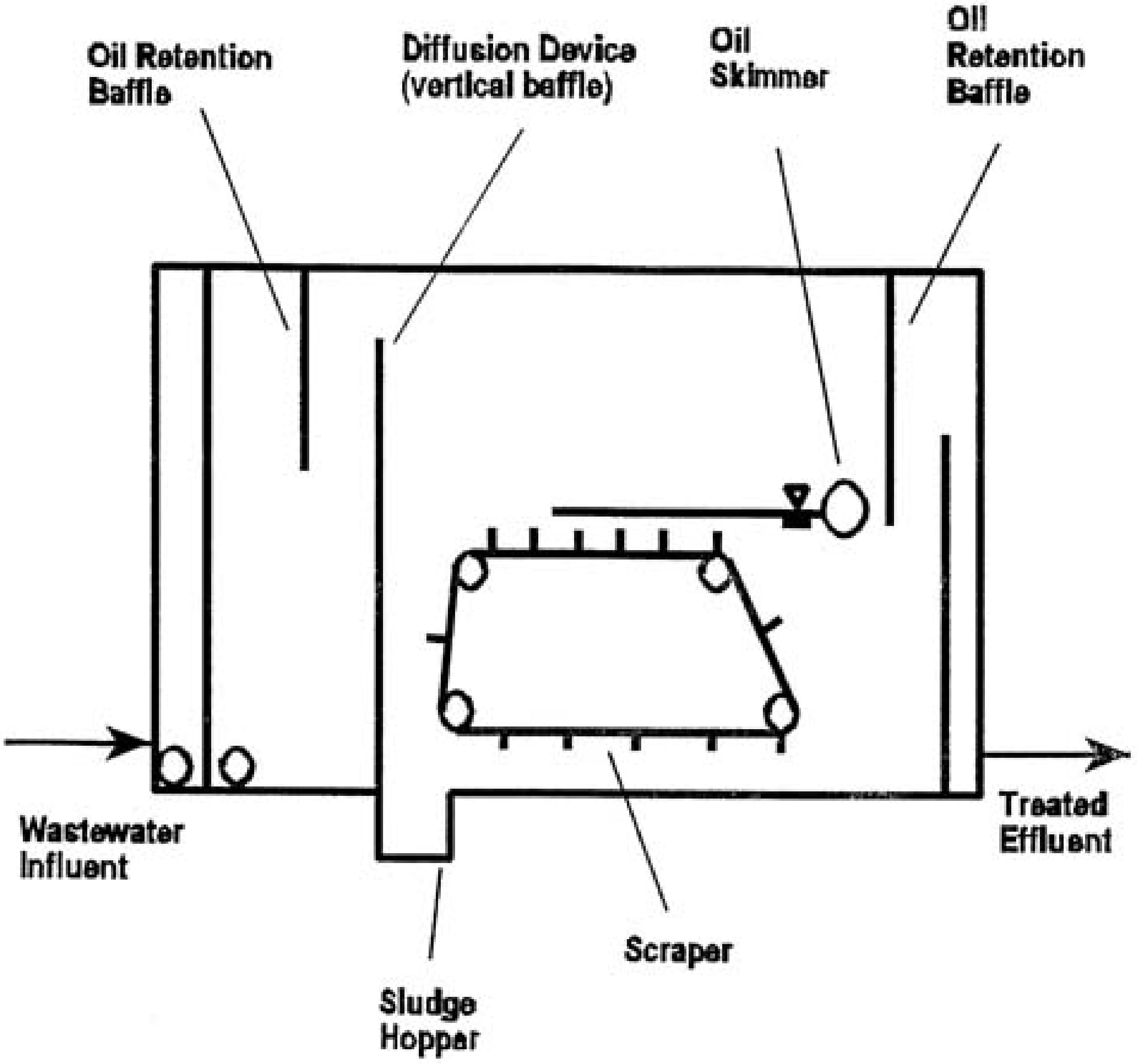
Emulsion Breaking

This process is used to counteract the effect of chemicals (emulsifiers) that allow the dispersion of oil into water. These stable emulsions, including metal working coolants, lubricants, and antioxidents, can interfere with the effectiveness of subsequent treatment steps or may allow some pollutants to pass through treatment with no removal. Emulsion breaking can be achieved through acid cracking at a pH of 12, or the addition of surfactants or coagulants. Once the emulsion has been broken, the oils usually can be removed through gravity separation processes (see below) that allow the oil and water to separate. In a separate step the oil is then skimmed or filtered for removal. Separation can be enhanced by heating the mixture (usually to 100150°F) or the use of dissolved air flotation (described later).

Gravity Oil/Water Separation

Without the presence of emulsifiers, treatment of wastewater containing non-soluble oils can rely on the lower specific gravity of the oils. This usually occurs in tanks designed to provide quiescent separation of the oil to a surface layer that can be removed using skimmers, baffles, plates, slotted pipes, or dip tubes. The most common device use for this purpose is the API separator. Figure 3.2 shows a gravity separation system for oily wastewater.

Fig. 3.2 Gravity separation system diagram. (*Courtesy of EPA*)



Clarification

Clarification provides the opportunity for the gravity separation of solids (including coagulated/flocculated materials) and chemical precipitates of metals. This operation usually will occur in a tank designed to provide quiescent flow over a weir with sufficient tank volume to allow solids to settle to an area where they can be mechanically removed for subsequent disposal. These solids normally are dewatered prior to disposal with technologies that are described in a later section. Figure 3.3 shows a clarification unit, and Fig. 3.4 shows how clarification can be combined with coagulation and flocculation.

Fig. 3.3 Clarification system diagram. (*Courtesy of EPA*)

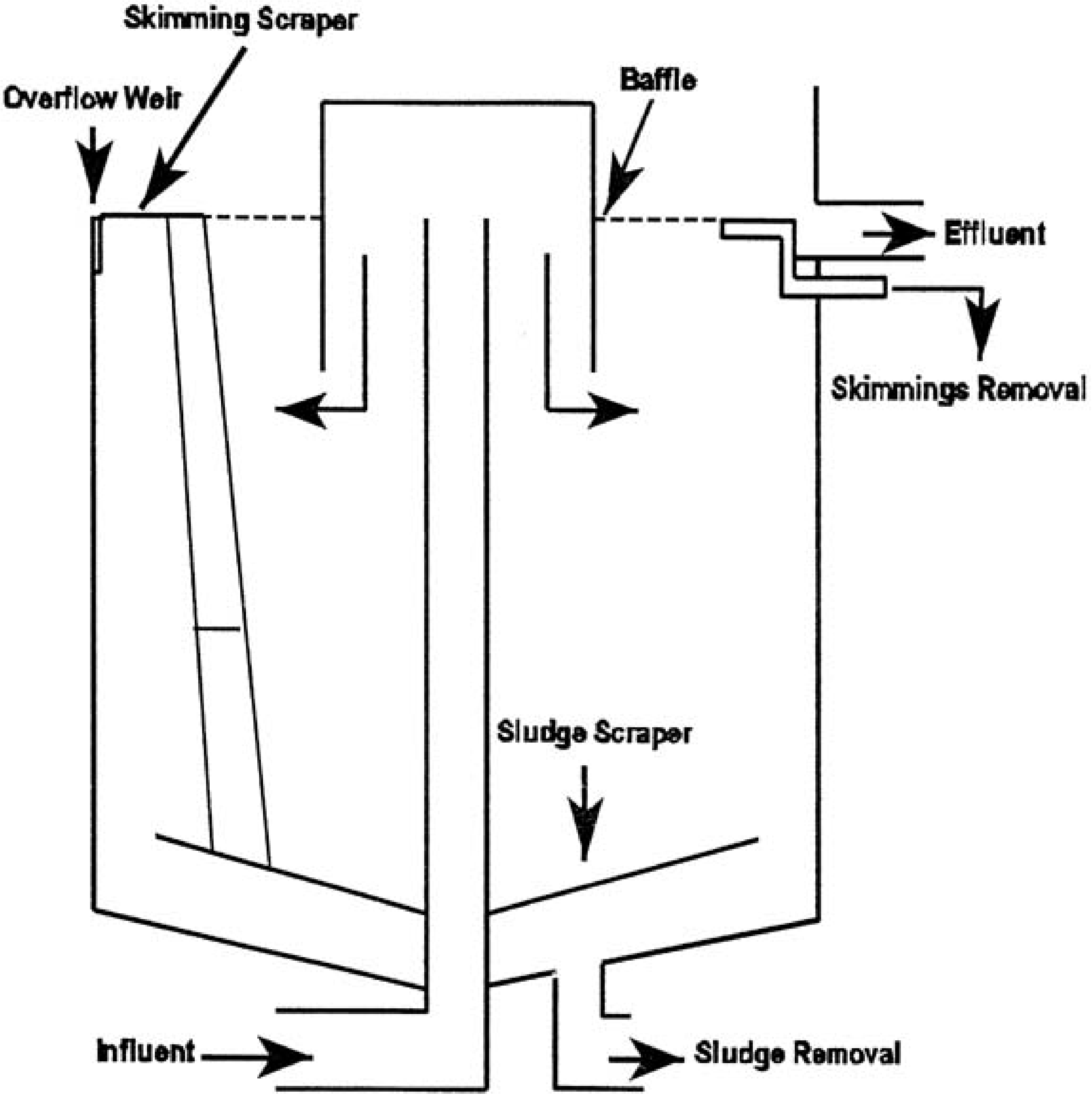
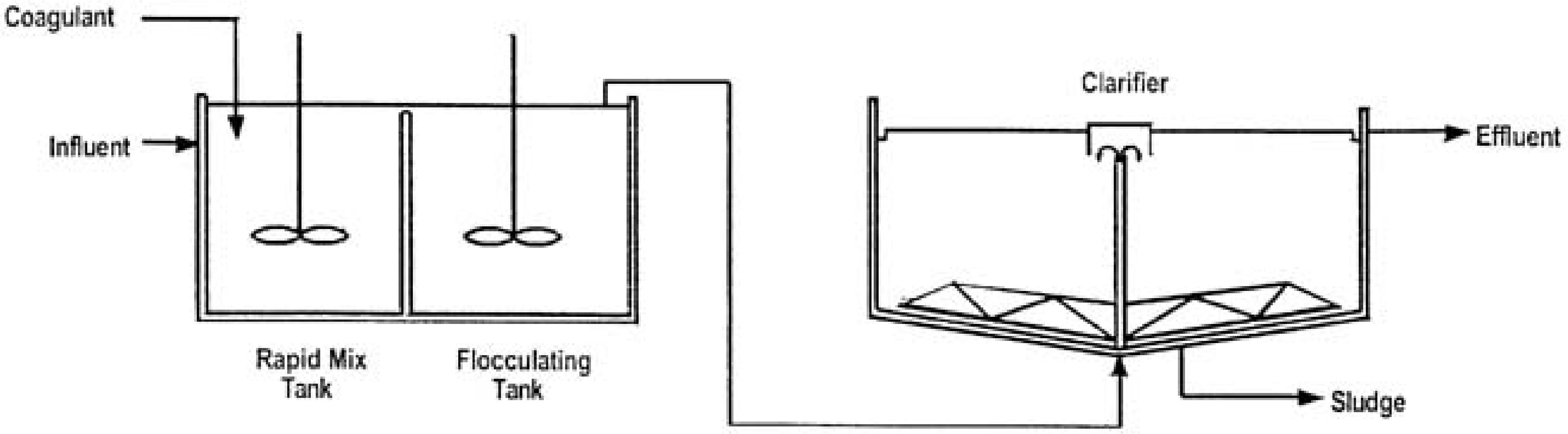


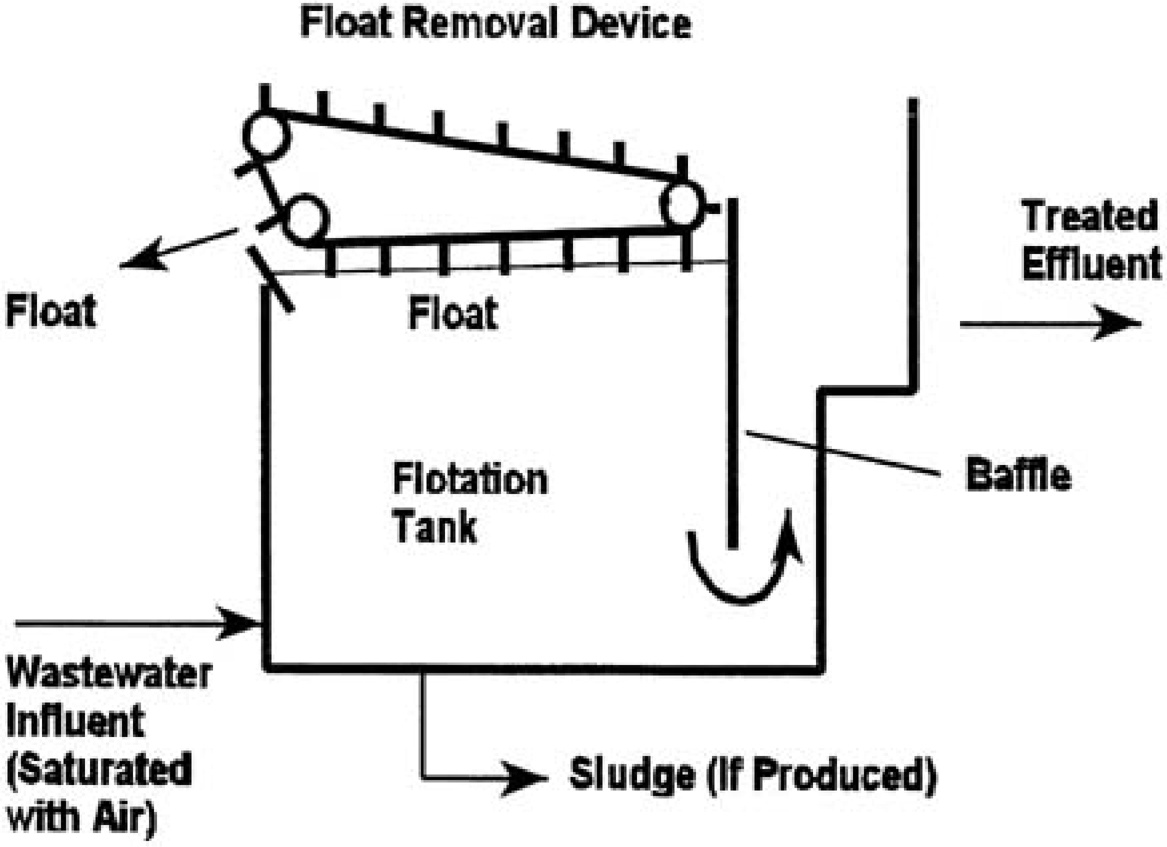
Fig. 3.4 Clarification system incorporating coagulation and flocculation. (*Courtesy of EPA*)



Dissolved Air Flotation

Rather than using a clarification system, or as an enhancement step to conventional clarification, a dissolved air flotation system can be used to remove solids. Dissolved air flotation generates a curtain of small air bubbles to force solids to the surface of a tank where the solids can be skimmed for removal. This process is especially useful in removing oily solids, solids with relatively low density, or when the space available for conventional settling is not available. Figure 3.5 provides a diagram of dissolved air flotation.

Fig. 3.5 Dissolved air flotation system diagram. (*Courtesy of EPA*)

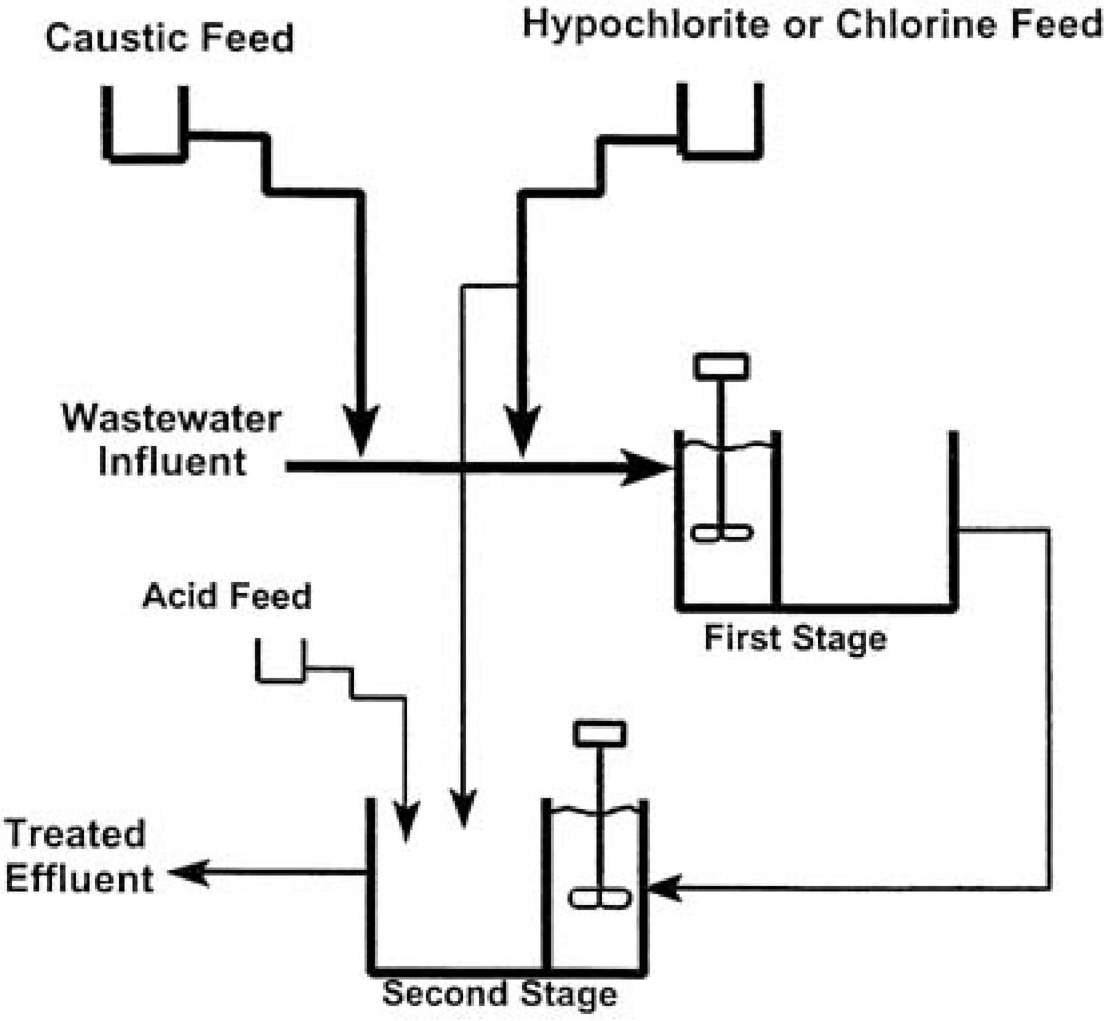


Chemical Oxidation

Chemical oxidation can be used to destroy organic pollutants, especially those pollutants resistant or inhibiting to other forms of treatment such as biological treatment. It also is used to destroy the organic molecule in organo-metal complexes, allowing treatment of the metal component. This process is the most common approach for cyanide treatment, and also can be used to convert sulfides to sulfates. Sometimes it is used to alter the valence states of metals such as arsenic to forms that are less soluble and, therefore, more amenable to precipitation and settling for removal.

Oxidation of cyanide normally is accomplished with alkaline chlorination. This is a two stage process in which cyanide is first converted to cyanate at a pH above 10, then converted to nitrogen and bicarbonate at a lower pH. However, this process is relatively ineffective for cyanide complexes containing metals such as iron, copper, or nickel. Figure 3.6 displays a cyanide destruction system.

Fig. 3.6 Cyanide destruction by alkaline chlorination. (*Courtesy of EPA*)



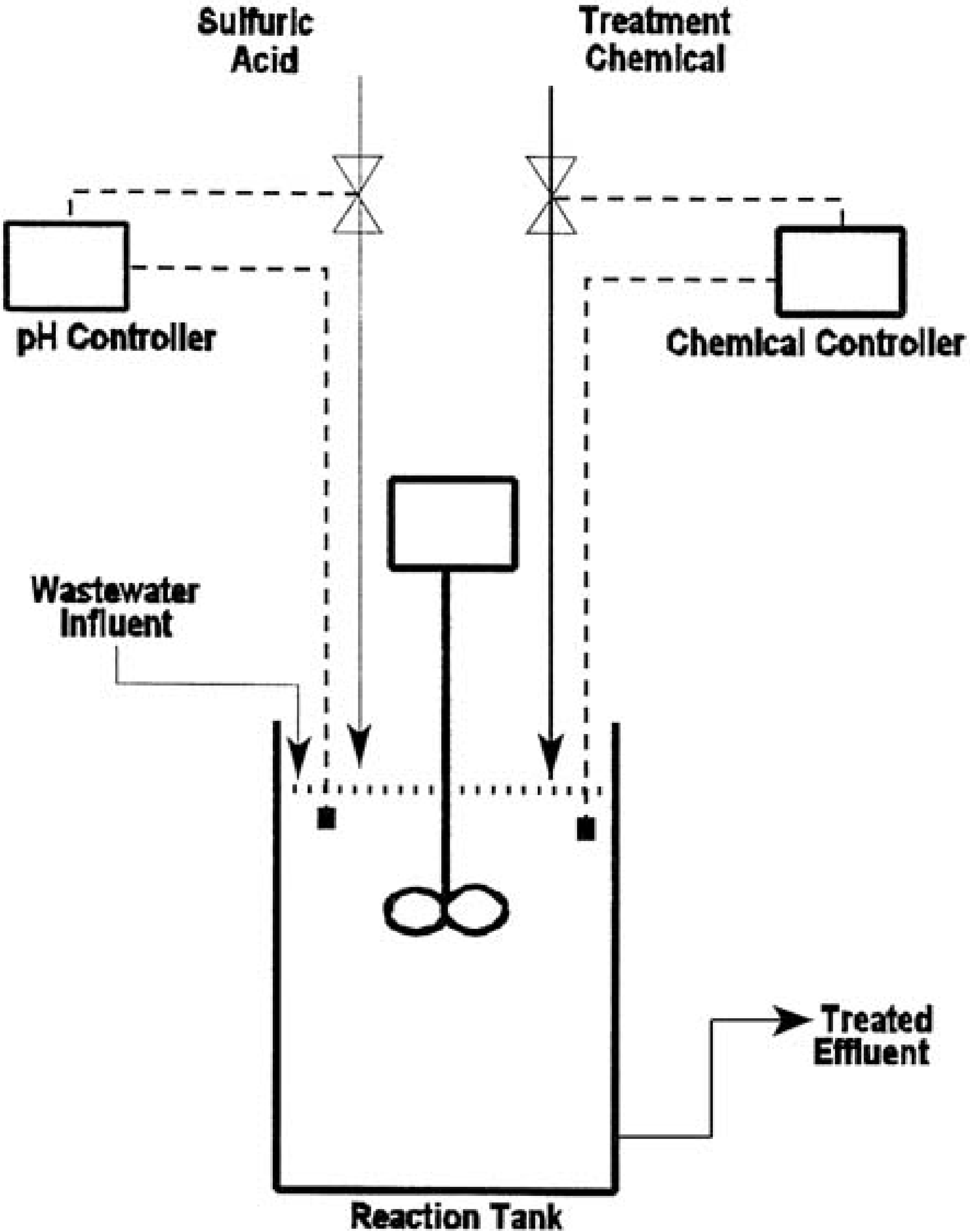
Electrolytic Recovery

If metals are present in relatively high concentrations, it may be feasible to recover those metals with an electrolytic process. While it can be very effective for copper, zinc, silver, cadmium and gold, it is not useful for nickel. The process relies on inserting electrodes into the wastewater and allowing the electric potential that is created to cause the migration of the metal ions to the positive electrode (cathode). The metal film that forms on the cathode is periodically stripped and sent for recovery.

Chromium Reduction

In order to facilitate the removal of chromium with other metals in the chemical precipitation operation described in a later section, it is essential that all hexavalent chromium be pretreated with a process that converts it to trivalent chromium. This process also significantly reduces the toxicity of any chromium not removed by subsequent treatment. Chromium reduction is achieved by the addition of chemicals such as ferrous sulfate, sodium dioxide, or sodium bisulfate/metabisulfate into a tank where pH is controlled to low levels (usually pH 23). Figure 3.7 shows a chromium reduction system.

Fig. 3.7 Chromium reduction system diagram. (*Courtesy of EPA*)



Chemical Precipitation

Chemical precipitation uses the relatively low solubility of metal hydroxides, carbonates, or, less frequently, metal sulfides to power the removal of toxic metals from industrial wastewater. Effective precipitation can be critical to lowering the toxicity of the wastewater, and may be essential in removing metals that may inhibit biological treatment of organic pollutants in later processing steps. Once the metal ions (or other soluble species) are converted to low solubility species through the addition of chemicals such as lime, sodium hydroxide, sodium carbonate, sodium sulfide or ferrous sulfate, a flocculation and clarification step is used to remove the metals.

The effectiveness of chemical precipitation based on hydroxide precipitation (by far, the most common technology) is very dependent on pH control. As show in Fig. 3.8, most of the metals removed in this step have minimum solubilities between 8 and 12. The actual optimum pH for any wastewater will be dependent on the specific metals present and the concentrations of metals requiring treatment. Sometimes, multiple step precipitations are needed in order assure effective treatment. This is especially true when removal of arsenic or antimony is needed. These two metals require removal at much lower pH than other metals. Figure 3.9 shows a common configuration for chemical precipitation.

Fig. 3.8 Calculated solubilities of metal hydroxides. (*Courtesy of EPA*)

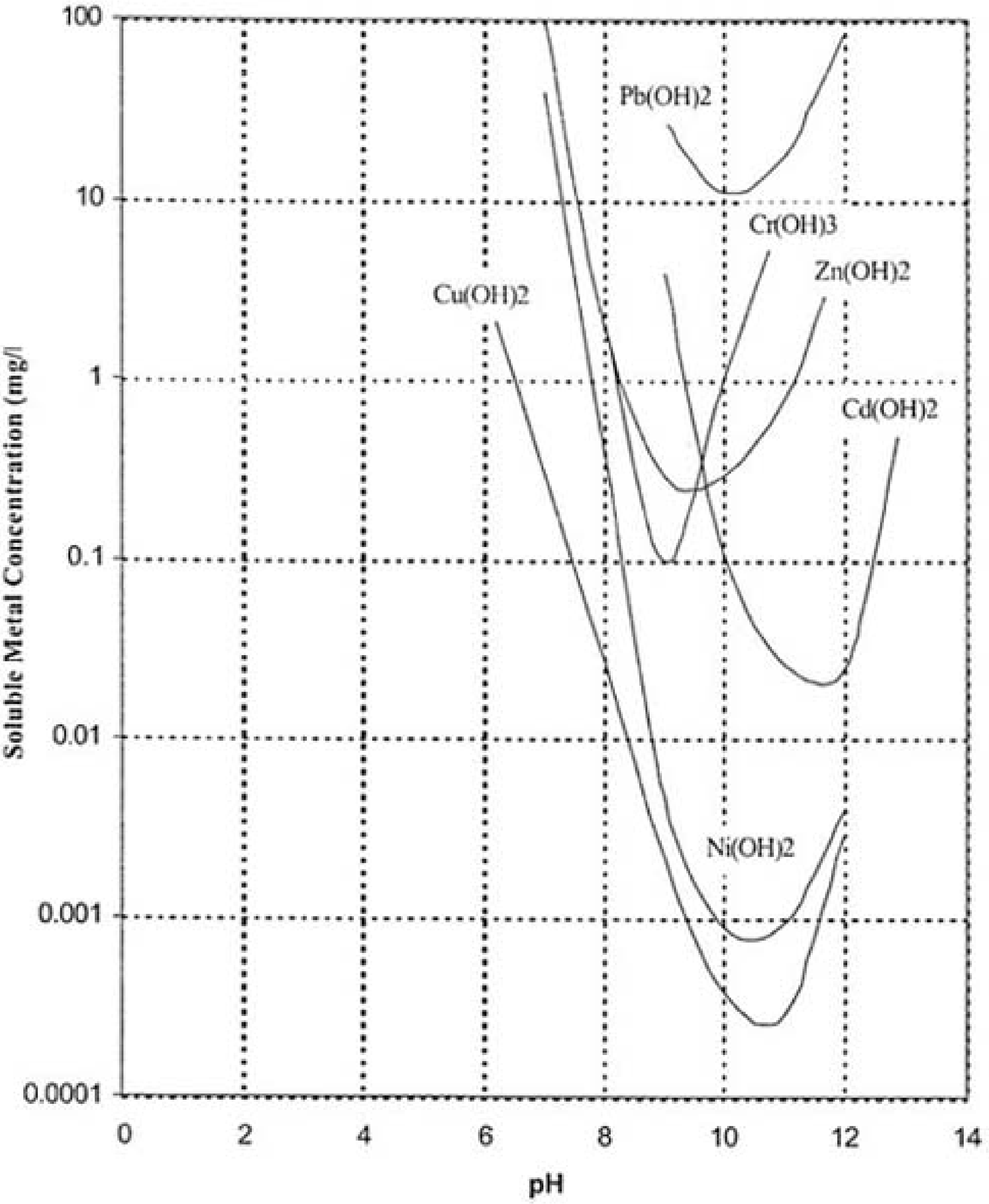
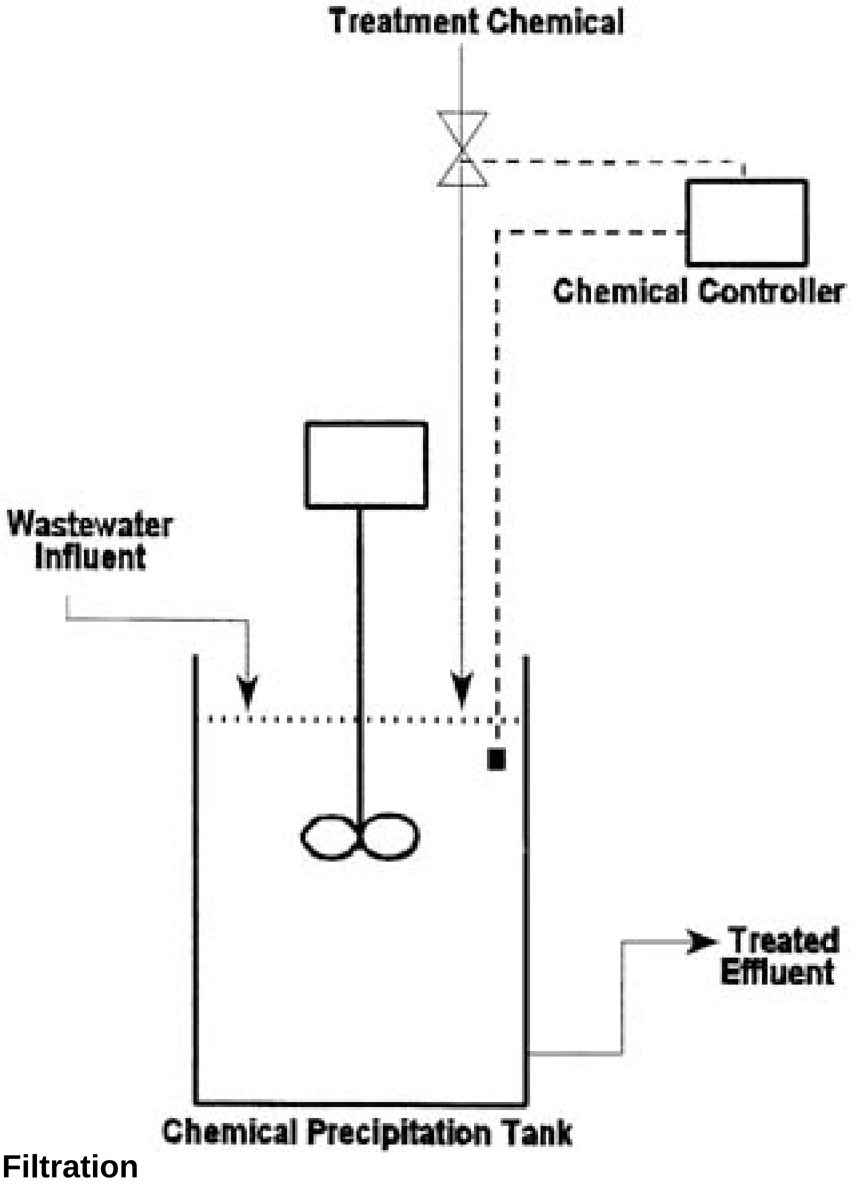


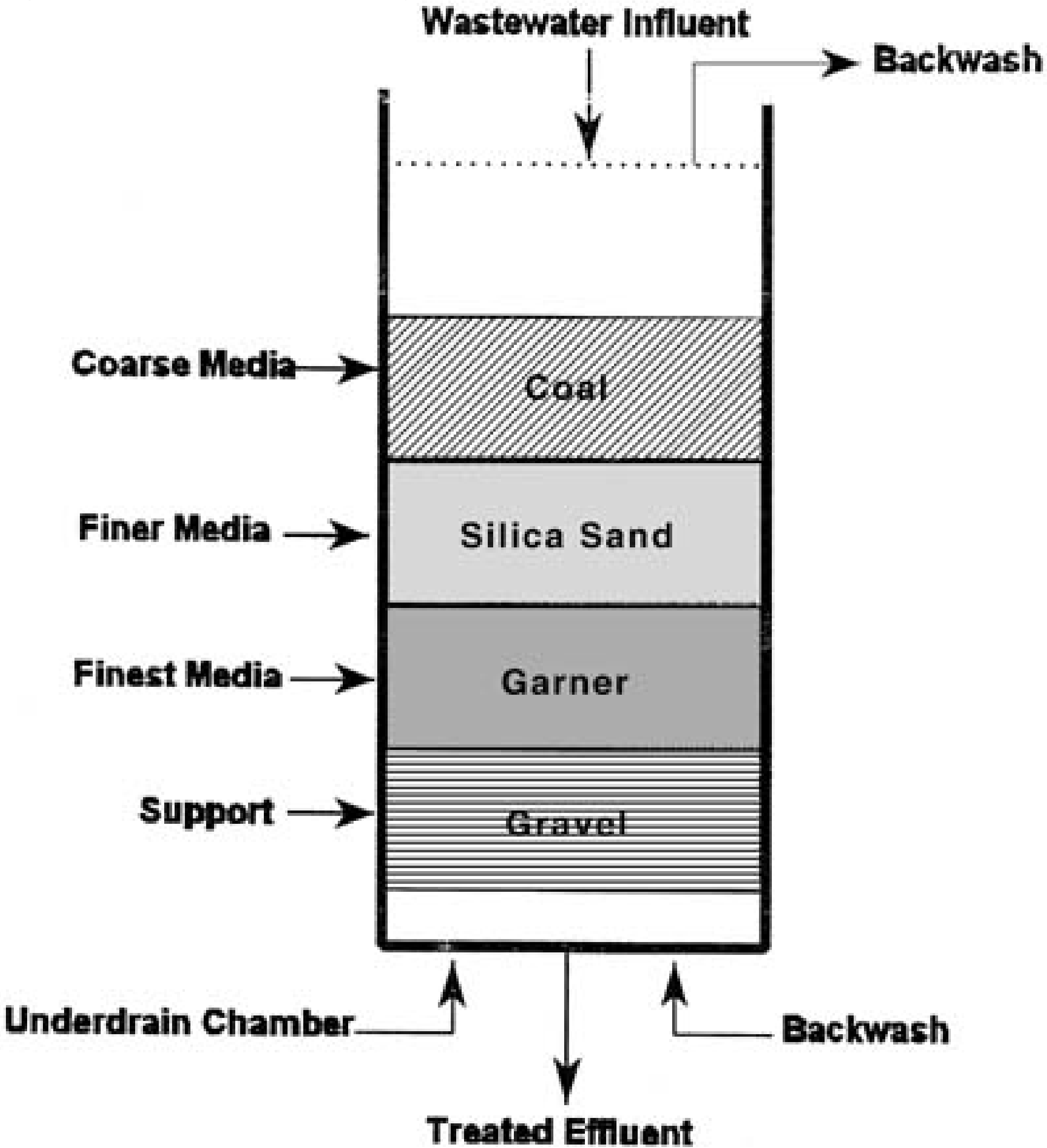
Fig. 3.9 Chemical precipitation system diagram. (*Courtesy of EPA*)



Filtration

Additional removal of solids can be achieved by passing wastewater through a filter. The amount of solids removed is controlled by the pore size of the filter and the rate of filtration can be enhanced by use of pressure, vacuum, or even centrifugal force. The most simple form of filtration is sand filtration. Multimedia filtration, using anthracite, sand, and finely packed garnet on top of a gravel base can provide more efficient removal, and may even add some adsorption of organics or metals. Periodically, the filter will require a backwash to remove the trapped solids, with the concentrated solids returned for treatment elsewhere in the wastewater treatment system. A diagram of a multimedia filtration unit is provided in Fig. 3.10.

Fig. 3.10 Multimedia filtration system design. (*Courtesy of EPA*)



Membrane Filtration

When the situation demands more removal of suspended solids, oil and grease, or some dissolved pollutants (e.g. metals) than can be provided by the more conventional means described in the previous sections, membrane filtration is another level of treatment that is available. The two most common forms of this technology are ultrafiltration and reverse osmosis. Both rely on using pressure to force wastewater through a semipermeable membrane, concentrating pollutants in an exit stream, which is recycled to previous waste treatment operation. Ultrafiltration (shown in Fig. 3.11) is used when the pollutant's molecular size is larger than the rest of the wastewater stream. Reverse osmosis (Fig. 3.12) relies on a more complex membrane that can selectively sort molecules of similar size. It is used when the goal is to remove dissolved pollutants, usually dissolved metals, salts, or solids.

Fig. 3.11 Ultrafiltration system diagram. (*Courtesy of EPA*)

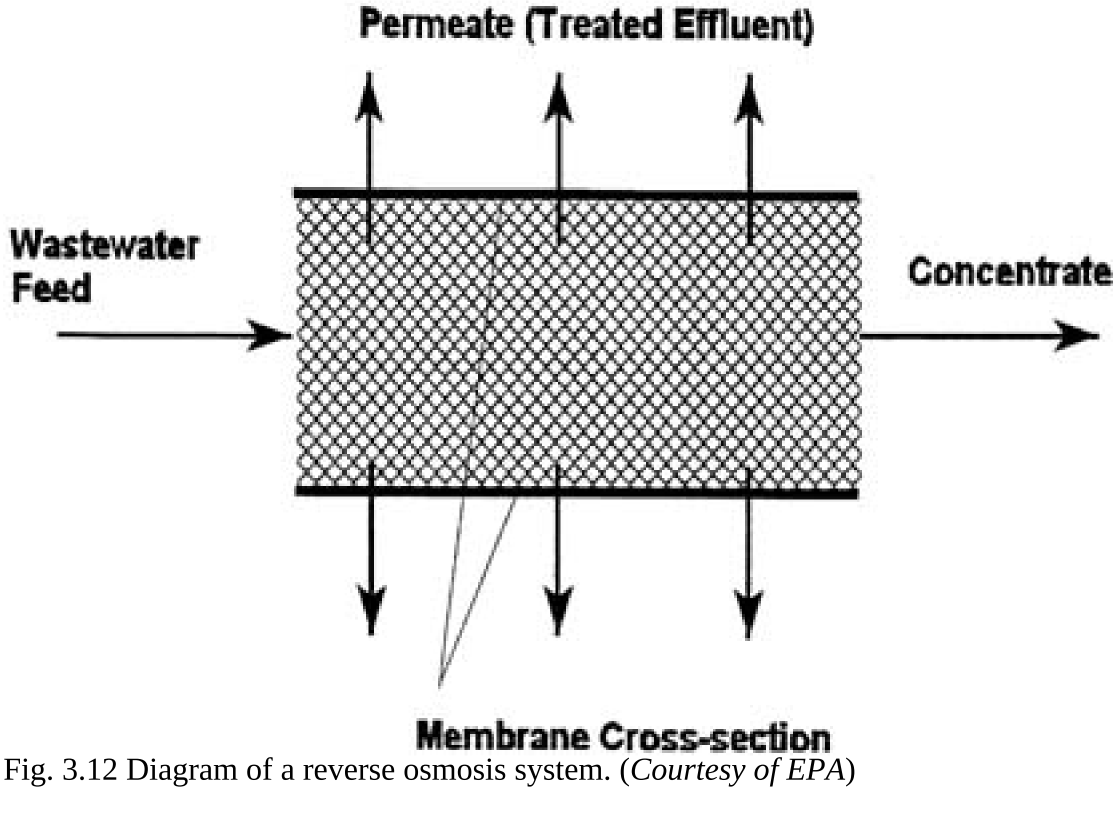
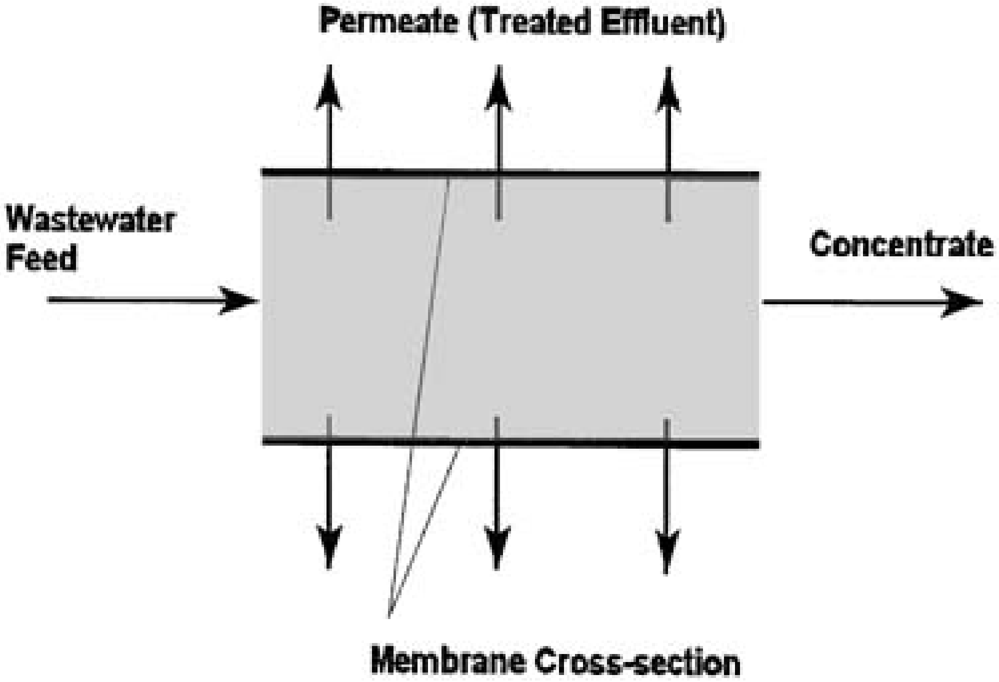


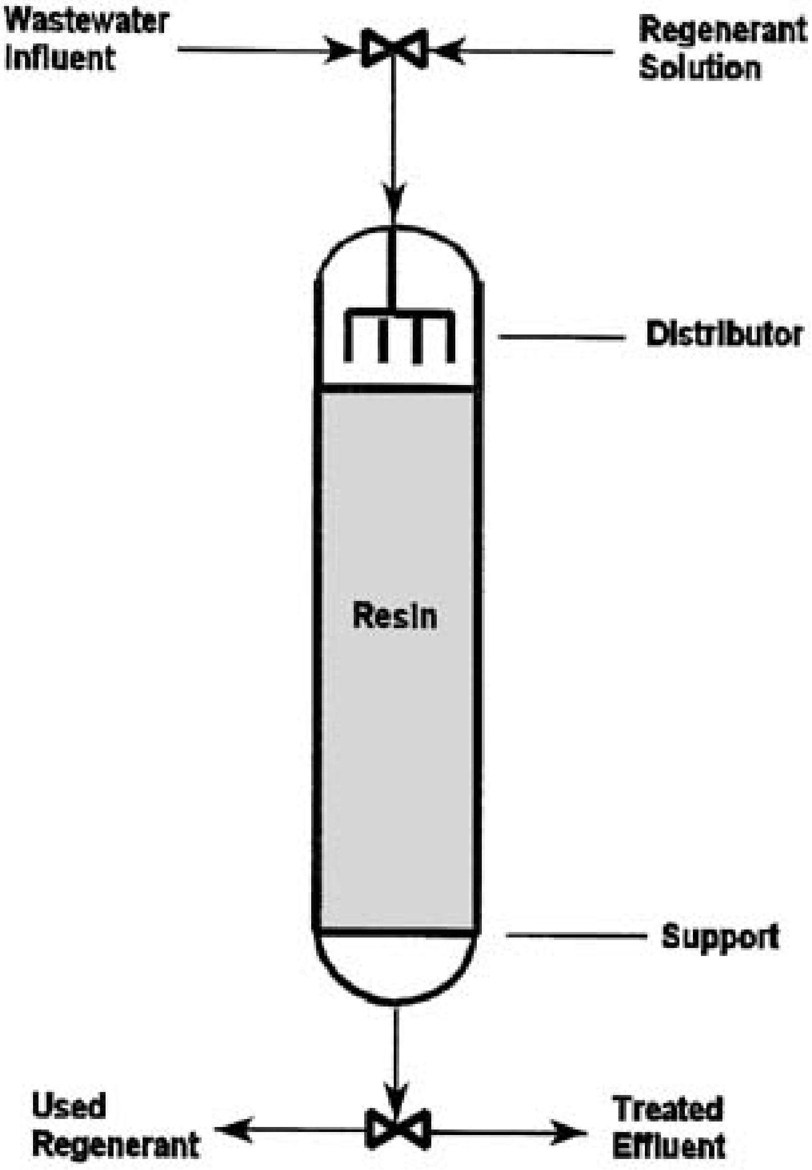
Fig. 3.12 Diagram of a reverse osmosis system. (*Courtesy of EPA*)



Ion Exchange

When metals are present in relatively low concentrations, but still require treatment because of the potential toxicity of the wastewater, ion exchange can be used. Ion exchange is based on passing the wastewater over a bed of chemically engineered resin that will selectively adsorb the metal. Periodically the resin bed will require backwash and regeneration. The advantage of ion exchange is that, in some cases, it can be cost-effective to recover the metals captured in the resin that are released in the regeneration process. However, ion exchange resins can be fouled or rendered ineffective by other pollutants that are present in the wastewater, especially organics. Figure 3.13 provides a diagram of an ion exchange system.

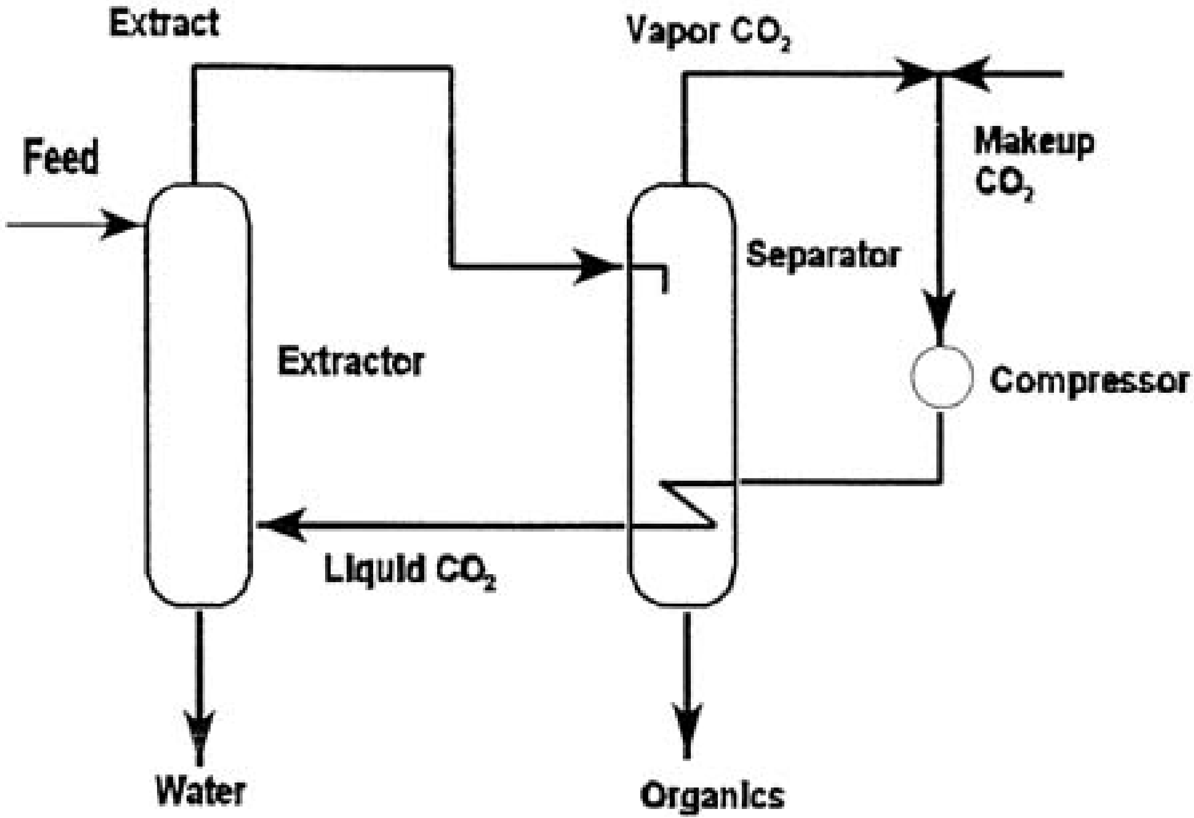
Fig. 3.13 Ion exchange system diagram. (*Courtesy of EPA*)



Stripping of Volatile Organics

If the wastewater contains significant quantities of dissolved volatile organic pollutants, it is wise to consider stripping technologies. This allows the organics to be efficiently removed and captured before passing the wastewater through additional treatment steps where the organics might be released directly to the atmosphere unless emission controls are present at each of those treatment operations. While most stripping is done with air, it also is possible to strip with steam or solvents such as liquified carbon dioxide (Fig. 3.14). Organics from air stripping must be captured in a control device such as a carbon column or incinerator, but steam stripping allows the organics to be condensed and captured. Solvent stripping allows for the solvent to be regenerated and reused leaving a concentrated organic waste stream for beneficial recycling or disposal.

Fig. 3.14 Liquid CO2 extraction system diagram.



Biological Treatment

Once the metals that can inhibit biological growth have been removed, the treatment of organic pollutants can be initiated, most commonly with biological treatment. Many, but not all, organic pollutants can be readily consumed by populations of microbes that convert the pollutants to a biomass that can be more readily removed from the wastewater. The microbial population must be acclimated to the food source, and the oxygen concentrations, pH, and specific nutrients (especially phosphorus and nitrogen) can be critical to supporting biological treatment. Most industrial biological treatment systems are aerobic systems, where an excess of oxygen is provided. However, there are options in which anerobic (little or no oxygen) or facultative (cycling between aerobic or anerobic conditions) designs are used. In virtually all situations it is critical to provide equalization of wastes prior to the biological treatment, to avoid significant variation in the feed concentrations or pH. In some cases, this equalization is provided by the biological treatment reactor itself.

Biological treatment can be provided by a wide array of equipment designs, including activated sludge digestion (Fig. 3.15), trickling filters (Fig. 3.16), biotowers (Fig. 3.17), and sequencing batch reactors (Fig. 3.18). All of these systems rely on active growth of microorganism populations to consume the organic pollutants. In the natural cycle of these populations there will be a need to address removal of a fraction of the biomass containing dead organisms or any microorganisms that may be generated in excess of what is needed to digest the available food (pollutant) supply. This waste biomass (biosludge) is usually removed through gravity settling in clarifiers, similar to the suspended solids removal described previously. Efficiency of the clarification process may be enhanced with a flocculent. Filtration also can be provided as an enhancement or substitute for clarification.

Fig. 3.15 Activated sludge system. (*Courtesy of EPA*)

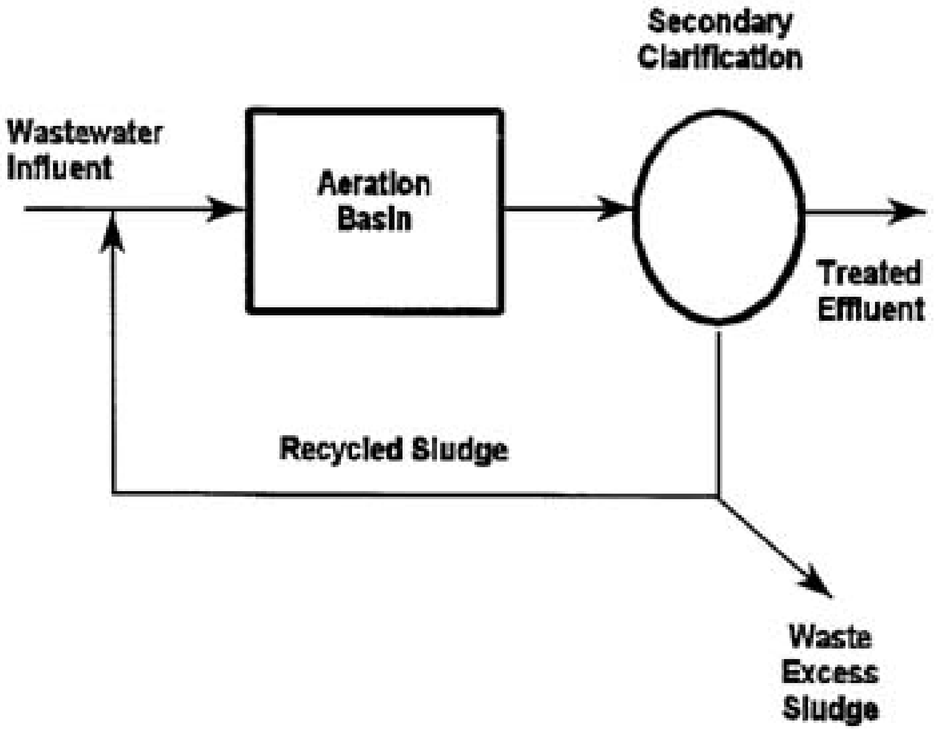


Fig. 3.16 Trickling filter system. (*Courtesy of EPA*)

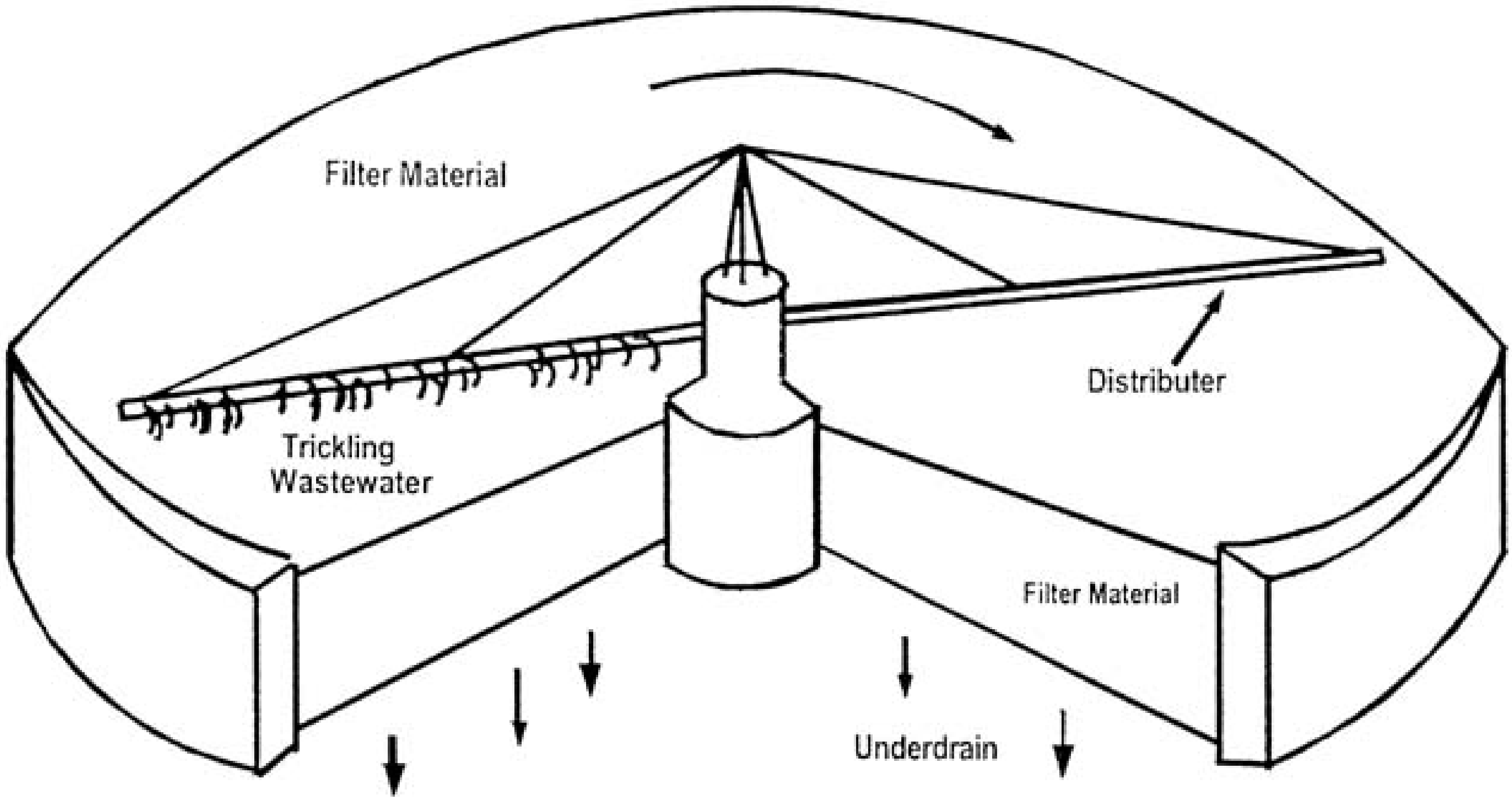


Fig. 3.17 Biotower system scheme. (*Courtesy of EPA*)

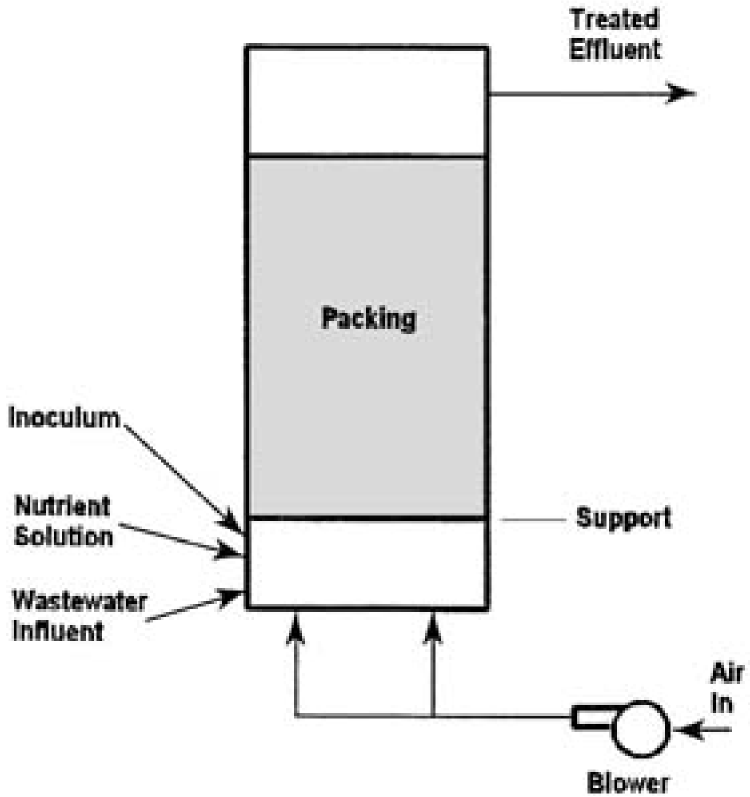
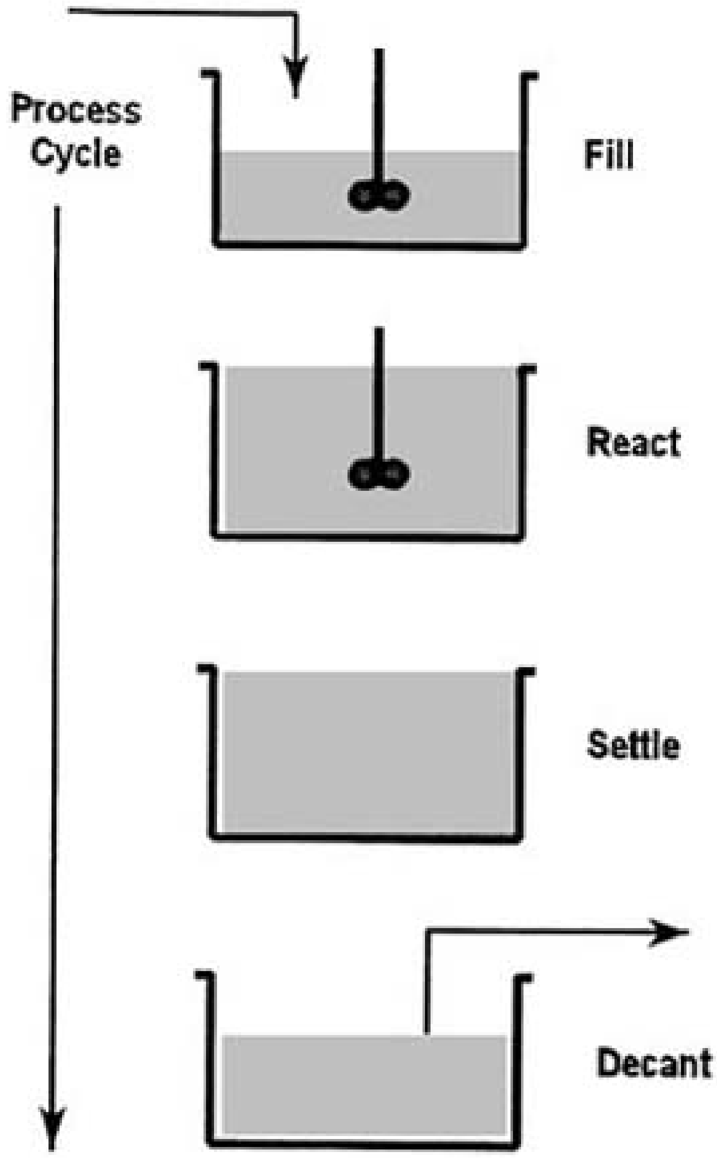


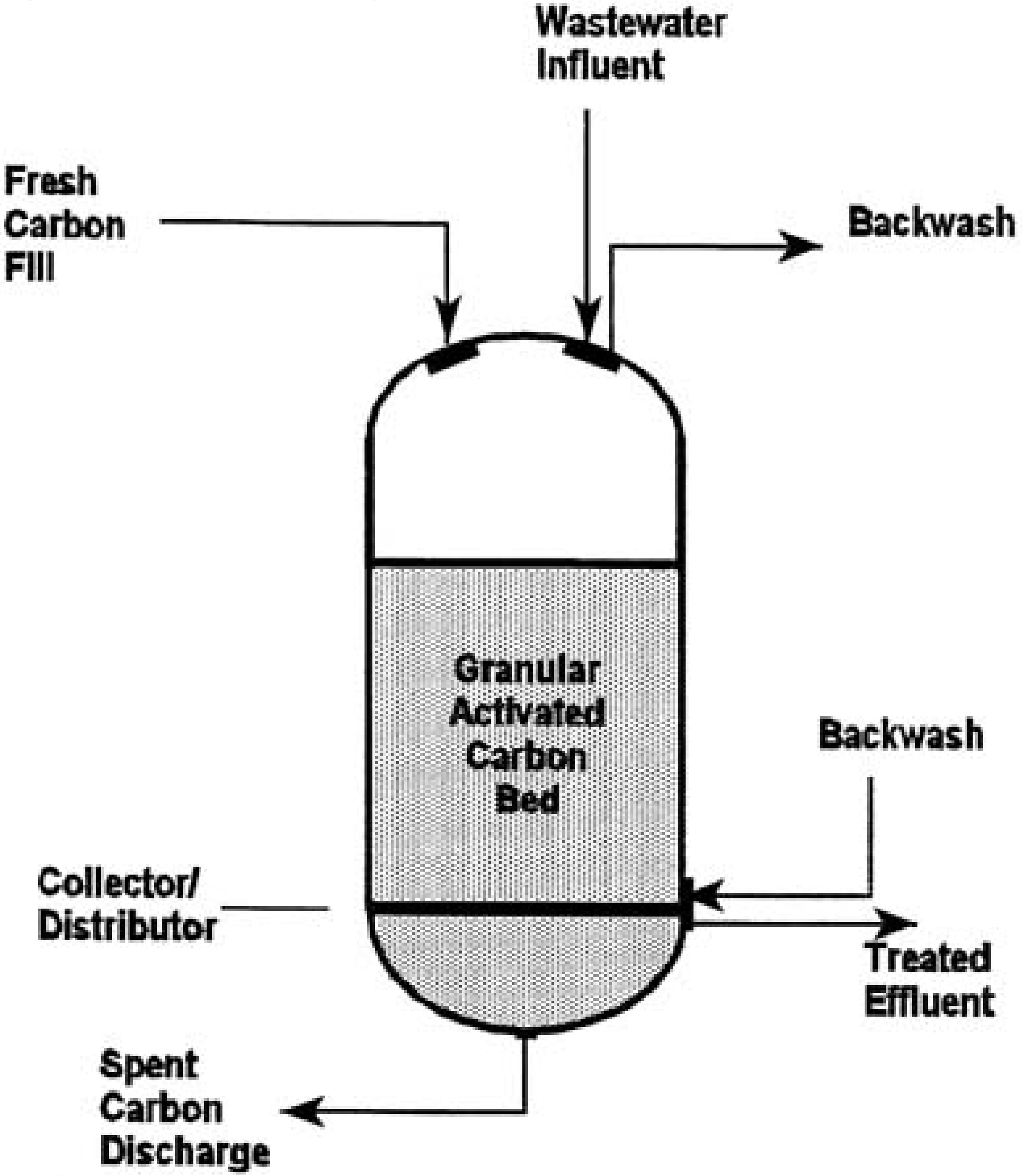
Fig. 3.18 Diagram of sequencing batch reactor system. (*Courtesy of EPA*)



Carbon Adsorption

Dissolved organic pollutants in relatively low concentrations, but significant because the residual toxicity is presented, often can be removed with activated carbon adsorption. Activated carbon is produced from coal, coke wood, or other carbon sources by exposing the material to an oxidizing gas to create a product with significant amounts of pore space within it. This creates a large internal surface area that provides many locations for organic molecules to chemically adsorb to the carbon surface. The organic molecules concentrate within the carbon until there are few remaining sites. At this point the carbon is regenerated thermally or by passing a solvent through the carbon that collects the organic pollutants for waste treatment and disposal. Alternatively, the carbon can be disposed. Carbon adsorption is usually applied through the use of fixed carbon beds, or addition of powdered activated carbon to wastewater tanks and subsequent settling and removal of the carbon for reuse or disposal. Figure 3.19 provides a diagram of a carbon adsorption column.

Fig. 3.19 Carbon adsorption system. (*Courtesy of EPA*)



Sludge Dewatering

Prior to treating wastewater treatment sludges (e.g. biological or chemical precipitation sludges from clarifiers) either on-site or offsite, it is usually less expensive to remove any entrained wastewater and return it back to the wastewater treatment system. This can be done with a variety of devices that use pressure (Fig. 3.20), gravity, vacuum (Fig. 3.21) or centrifugal force to increase the solids content to 1550 percent solids. This alters the sludge from a fluid to a dry cake, easing handling and disposal.

Fig. 3.20 Belt pressure filtration system. (*Courtesy of EPA*)

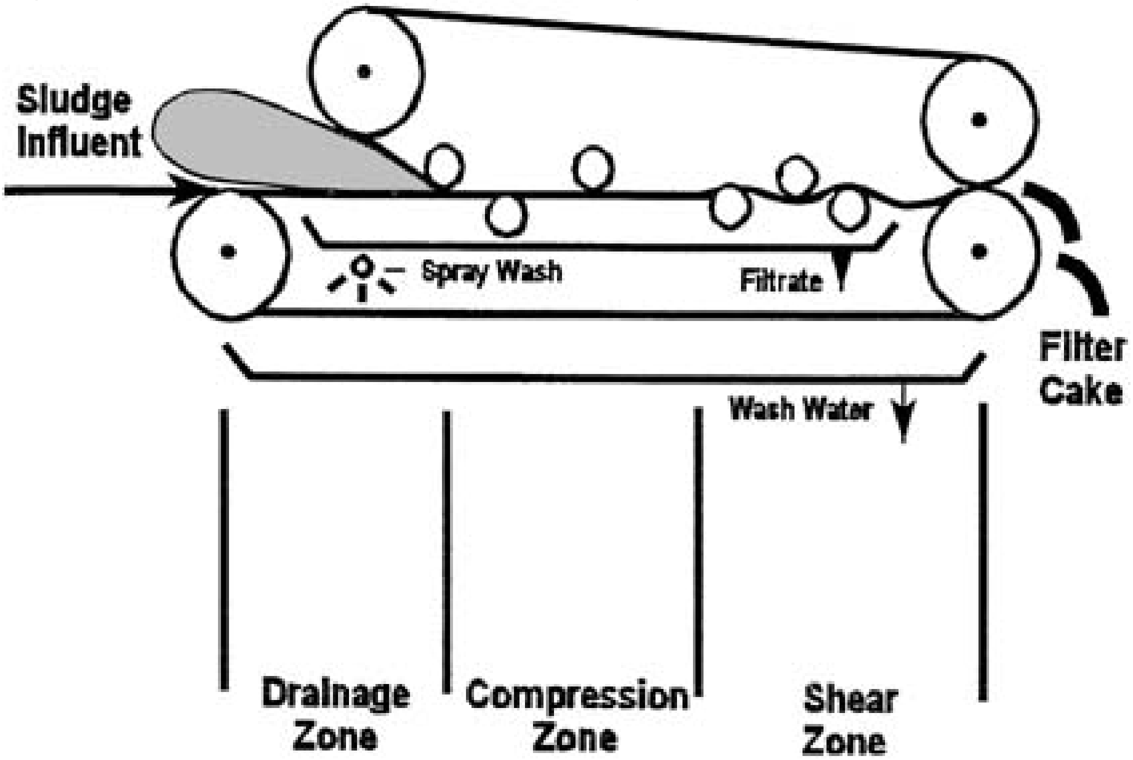
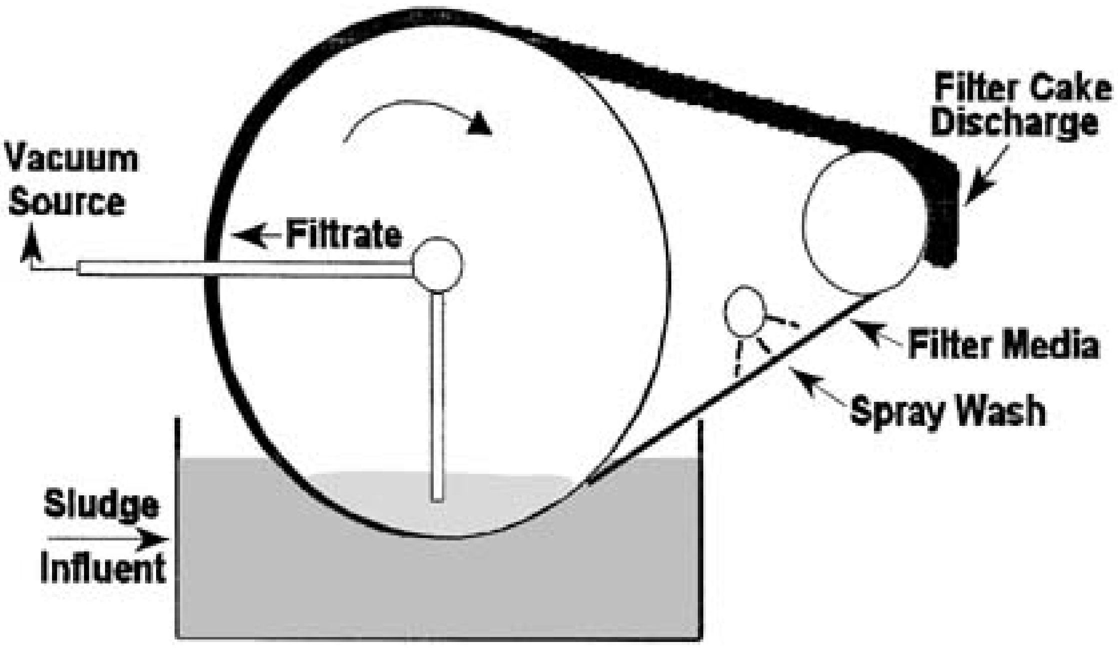


Fig. 3.21 Schematic of a vacuum filtration system. (*Courtesy of EPA*)


Filter Cake Discharge
Vacuum Source
Filtrate
Filter Media
Spray Wash
Sludge Influent

3.4 Hazardous and Industrial Solid Waste Treatment

Hazardous and solid waste management can be subject to very comprehensive regulatory oversight, especially if the waste meets the U.S. Environmental Protection Agency's (EPA) definition of a hazardous waste. States also are allowed to place additional, more protective requirements on hazardous waste. Industrial nonhazardous waste generally is not subject to federal regulations, but is subject to state government oversight. However, EPA has issued draft nonbinding guidance to states describing the elements of effective nonhazardous waste management (see www.epa.gov/epaoswer/non-hw/industd/guide/index.htm).

Therefore, it is important to be aware of the requirements that apply where the waste will be generated, treated, or disposed. It also is essential to understand that solid waste under these regulations need not be solid, the regulations can apply to liquids and contained individual gases gases as well as solid materials. EPA's website (www.epa.gov/epaoswer/osw/index.htm) provides a wealth of information to help understand the requirements of its regulations. It also provides a toll-free Hotline (800) 424-9346 to answer questions by the public. States often provide similar services.

As part of its overall program, EPA has set minimum requirements with respect to hazardous waste treatment. These requirements are expressed as performance requirements for most wastes, but EPA has mandated that specific technologies be used for some wastes where performance monitoring is not feasible. For the wastes with performance requirements, any technology may be used as long as it does not achieve the performance standards through impermissible dilution. 40 CFR 288.48 (available at the website www.epa.gov/epaoswer/hazwaste/ldr/rules01.htm) provides the uniform treatment standards that are the basis for the performance requirements for most wastes.

Treating hazardous waste to achieve EPA's standards usually requires one or more treatment processes. For hazardous wastes that are wastewater, the processes described for the similar industrial wastewater are applicable equally to the hazardous waste. A summary of the other treatment processes most commonly used for nonaqueous hazardous wastes is provided below.

Distillation

In this process the waste is heated and the volatile organics captured through a condensation step that follows. Distillation can occur as batch distillation or fractional distillation. Batch distillation utilizes a heated vessel (pot) in which the waste is heated. The condensate is produced in a single cooling step. This process is favored when only crude separation is needed, for example, there are high concentrations of organics to be removed from the waste for destruction or disposal, or there is only a single organic present.

When relatively high concentrations (e.g. above 7%) of valuable organics such as solvents are present in a waste, it may be economical to apply fractional distillation to recover those organics for recycle. Fractional distillation uses a column with trays or packing to allow separation of the solvents and differential recovery.

In either case, the operator must be concerned about emission controls for organics that escape the condenser, and for treatment and disposal of the distillation bottoms which can contain significant concentrations of low volatility organics and metals.

Solvent Extraction

Organics or metal theoretically can be removed and concentrated from wastes using various solvent extraction processes. However, in practice this operation is usually performed for organics treatment. The key to providing effective solvent extraction is to identify a solvent that has a strong chemical affinity for the organic pollutants but will not entrain the remainder of the waste. Common solvents used for this purpose are aliphatic amines, and critical fluids including carbon dioxide and propane as described previously in the discussion of solvent stripping of wastewaters. This treatment occurs in closed vessels or columns that allow high efficiency contact with the solvent in well mixed tanks, then provide separation through gravity or reduction of pressure on critical fluids (allowing the fluids to return to a gas phase.) Solvent extraction can be operated in either batch or continuous modes. Solvents are recycled until impurities extracted from the waste gradually build up to concentrations that impact the removal efficiency of the process. At this point, the solvent must either be regenerated or sent for further treatment and disposal. However, the amount of waste requiring treatment is substantially reduced compared to the amount treated by the process.

Air Stripping

Volatile organics may be removed from wastes just as for wastewater and concentrated for further treatment by passing air through the waste. In some cases the organics removed may not be a significant contributor to air emissions, but it is usually a good idea to provide for subsequent removal of the organics. Carbon adsorption is frequently the technology used to provide that removal.

Combustion of Organic Pollutants

Combustion of organics can occur in a variety of devices. Incinerators, cement kilns, lightweight aggregate kilns, and industrial boilers can provide the conditions needed for destruction of organics: high temperature, adequate residence time at temperature, excess available oxygen from adequate air feeds, effective air emission controls, and monitoring of residuals to assure effective destruction. Combustors, depending on design, can accept liquids, solids or both. Regardless of physical states of waste that are accepted, wastes must be blended or equalized to provide a relatively uniform feed. Hazardous waste combustors are precisely tuned to the characteristics of the wastes fed to the unit. Rapid changes in feed are not consistent with effective destruction of organics.

Cement and aggregate kilns do not produce ash, because any residual becomes part of their product. However, incinerators can produce large quantities of ash that may require additional treatment. Many newer slagging incinerators reduce the potential need for treatment due to leaching of metals from the ash by operating at higher temperatures that convert the ash to a vitreous (glass-like) solid.

Metals Stabilization

Many of the metals that create the toxicity of hazardous can be rendered less leachable and, therefore, safer for disposal following stabilization with binders such as Portland cement or lime. Stabilization is based on formation of a lattice structure or chemical bonds that bind the metals to a solid lattice structure that reduces their availability to leach when exposed to water. Stabilization requires more than physical entrapment, because this type of binding is more susceptible to the action of freeze/thaw and wet/dry cycles that can break down physical barriers. High concentrations of organics, including oil and grease, sulfides or chlorides can inhibit metals stabilization. Many companies have developed proprietary additives such as clays and synthetic binders to reduce the impact of interferences and enhance chemical and physical binding.

This process is usually the treatment of choice for wastes containing lead, trivalent chromium, cadmium, zinc, and nickel. The toxic metals often not treated effectively by stabilization include mercury, arsenic, antimony, and hexavalent chromium. Pretreatment to remove organics, oil and grease, or change the valence state of chromium may be necessary. Because the process can be sensitive to interferences and the specific metal compound present, bench-scale optimization is usually a sound practice.

Stabilization involves weighing or measuring the waste, binders and water (if needed), mixing the materials together, and providing an adequate cure time to allow a solid product to form. This process may be accomplished with sophisticated mixing and blending equipment, but also is achieved with low technology approaches such as front-end loader on concrete pads.

3.5 Air Pollution Control

Air emissions from industrial processes are subject to both, federal and state regulations and permits. The most recent requirements applicable are derived from the Clean Air Act (CAA) Amendments of 1990. The federal program for control of toxic air pollutants, the Maximum Achievable Control Technology (MACT) program is now resulting in requirements to remove 188 hazardous air pollutants (see http://www.epa.gov/oar/toxicair/index.html). This program relies on the average performance associated with the technologies used by the best 12 percent of facilities in the nation for that category of industrial activity. Table 3.3. lists the categories for which standards have been issued by late 2001. States also exercise control requirements under state laws that can be more stringent than federal requirements. There is a wide array of pollutants and control options. This section will try to summarize only some of the technology options for the most commonly controlled pollutants.

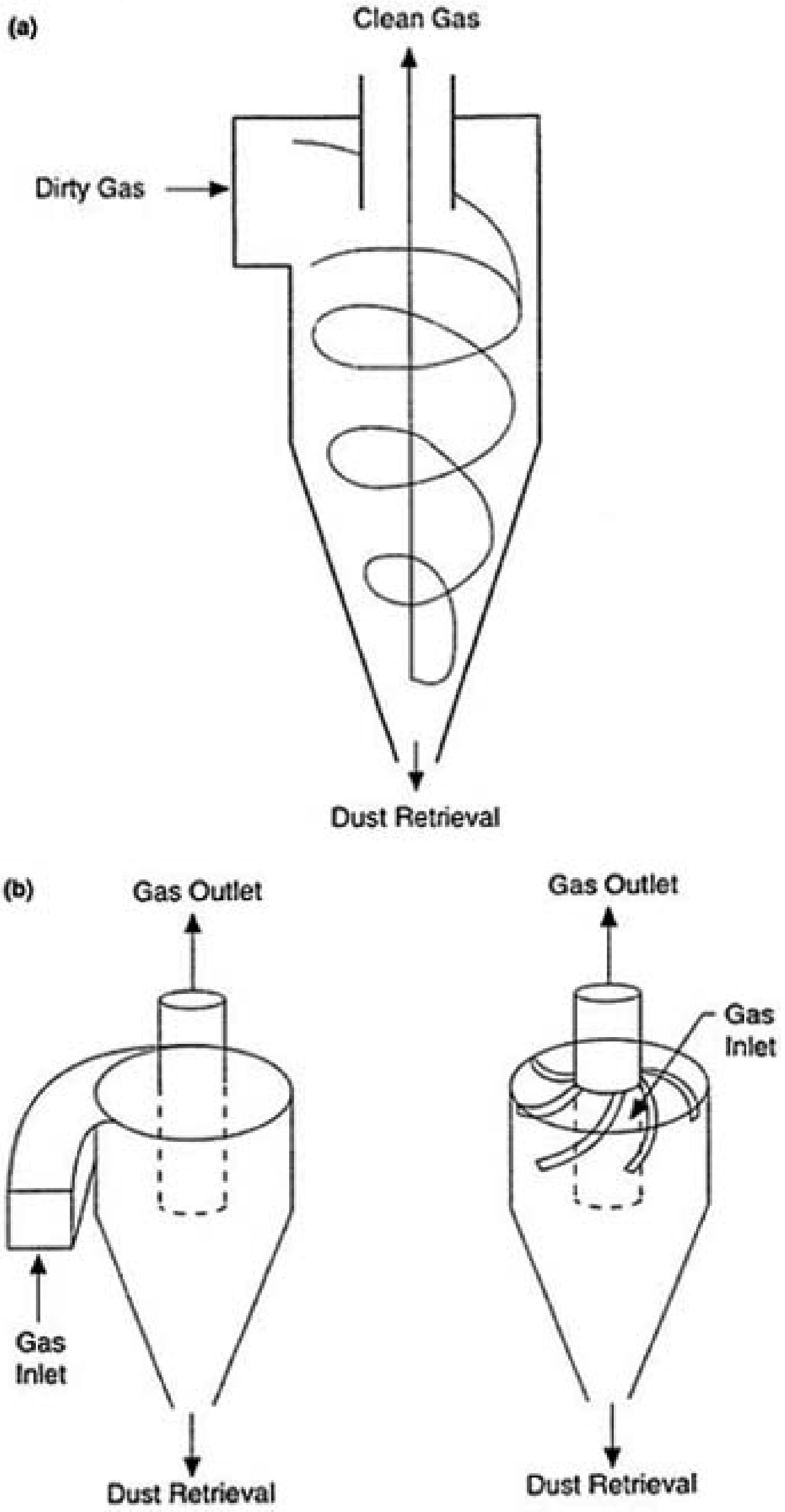
TABLE 3.3 Industrial Categories Subject to MACT Standards (Fall 2001)

Dry Cleaners
Coke Oven Batteries at Steel Plants
Organic Chemical Production Plants
Industrial Process Cooling Towers
Halogenated Solvent Cleaning Machines
Commercial Sterilization and Fumigation Operations
Gasoline Distribution Facilities
Magnetic Tape Manufacturing
Chromium Electroplating and Anodizing Operations
Basic Liquid Epoxy Resins and Non-Nylon Polyamide Resins Manufacture
Secondary Lead Smelter Industry
Petroleum Refining Industry
Aerospace Manufacturing and Rework Industry
Marine Tank Vessel Loading Operations
Wood Furniture Manufacturing
Shipbuilding and Ship Repair Industry
Printing and Publishing
Off-Site Waste Operations
Elastomer Production
Polyethylene Terephthalate Polymer and Styrene-Based Thermoplastic Polymers Production
Primary Aluminum Reduction Industry
Pulp and Paper Mills
Pharmaceutical Production
Flexible Polyurethane Foam Production
Ferroalloys Production: Ferromanganese and Swilicomanganese
Polyether Polyols Production
Mineral Wool Production
Primary Lead Smelters
Phosphoric Acid Manufacturing and Phosphate Fertilizer Production (two rules)
Wool Fiberglass Manufacturing
Portland Cement Manufacturing
Oil and Natural Gas Production and Natural Gas Transmission and Storage (two rules)
Steel Pickling
Pesticide Active Ingredient Production
Generic Rule (acetal resin production, acrylic and modacrylic fiber production, hydrogen fluoride production, and polycarbonate production)
Publicly Owned Treatment Works
Amino Resins and Phenolic Resins Production
Secondary Aluminum Production

Particulate Matter/Metals

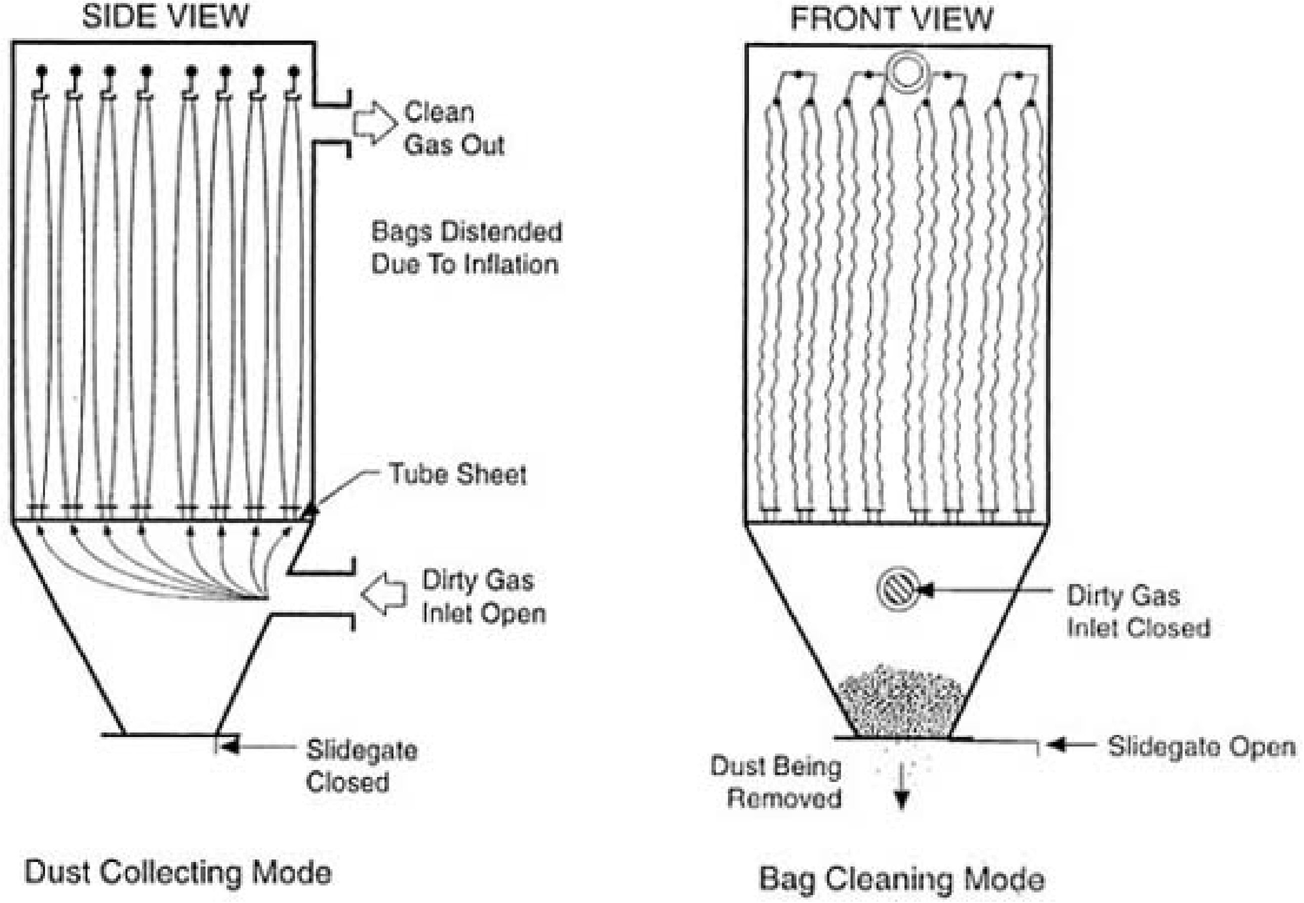
Particulate matter is controlled to prevent the direct impact on human health, such as respiratory diseases (e.g. asthma) and indirect toxicity associated with the content of the particles from metals that may adsorb to the particles. Environmental control must focus on two sizes of particles, coarse and fine. Removal of coarse particles, often with a cyclone device (Fig. 3.22) can be used as pretreatment step prior to some types of fine particle treatment, or when relatively low efficiency is required and when little fine particulate is present.

Fig. 3.22 Cyclone particle collector. (a) Particle collection procedure; (b) flue gas inlet options.



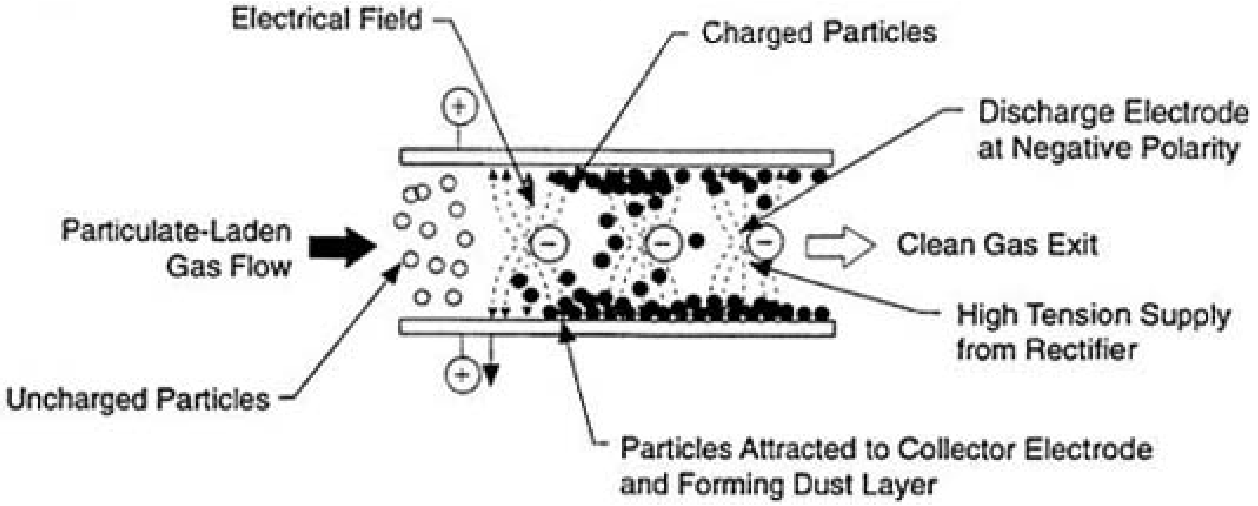
Higher efficiency particulate removal is achieved with a variety of devices. Fabric filters (baghouses) are used when gas temperatures are below 500°F, either as generated or after being quenched with water. The effectiveness of the bags can be reduced below the dew point of acids or water, since the condensate can plug and corrode some of the cloth bags used. Bags are periodically cleaned and the captured particles fall into a hopper for disposal as solid waste as shown in Fig. 3.23.

Fig. 3.23 Shaker cleaning baghouse. (*Courtesy of EPA*)



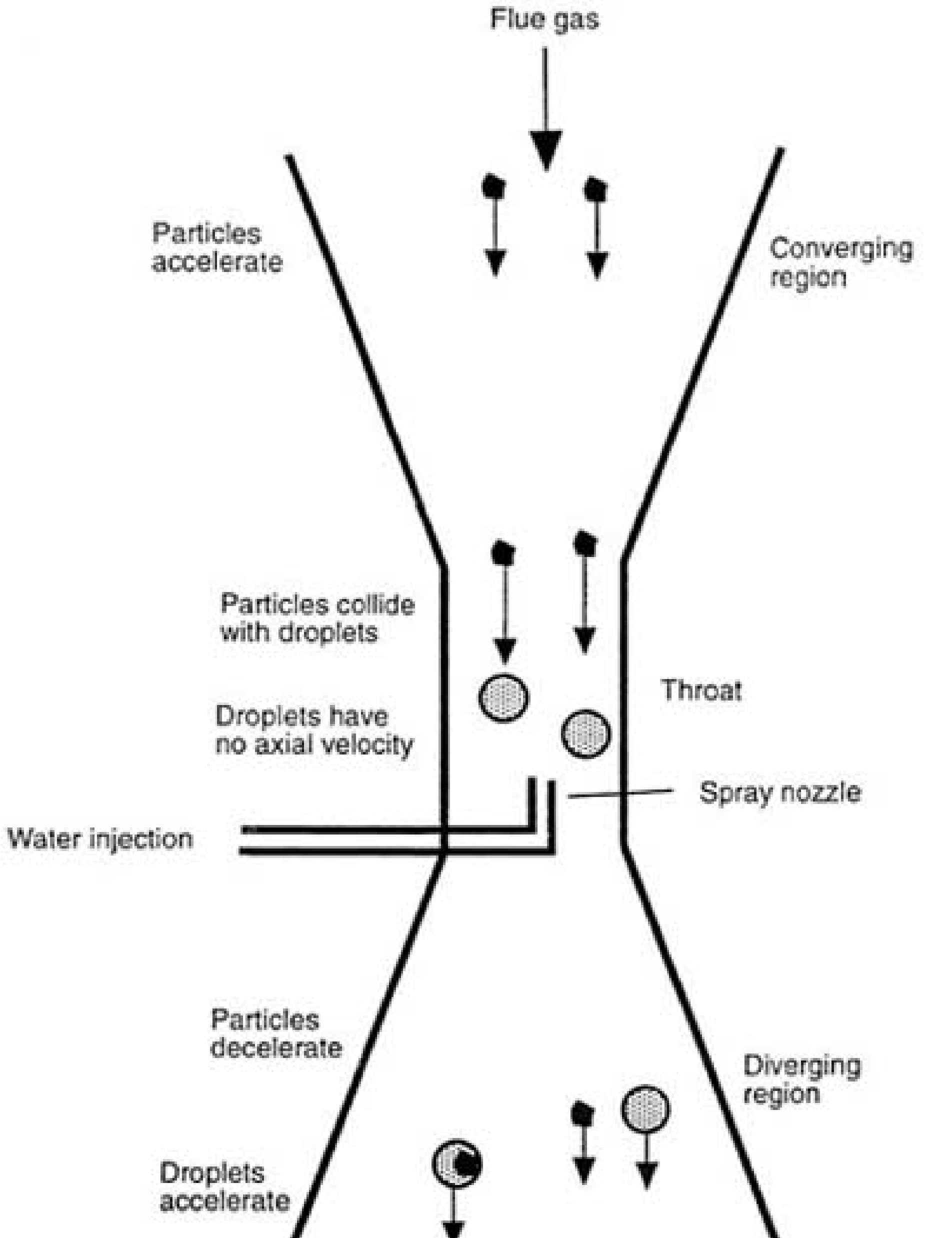
Electrostatic precipitators are also used for high efficiency removal. Although their performance may sometimes be slightly less effective than the best baghouses, these devices can be designed to operate at temperatures above those feasible for a baghouse. As shown in Fig. 3.24, these units, depend on creating an electrical charge on the particles and then attracting them to a charged collection plate. Either a wet or dry cleaning mode is used to remove the solids from the plate, depending on the gas stream treatment requirements.

Fig. 3.24 ESP particle collection procedure. (*Courtesy of EPA*)



If a gas stream already contains significant quantities of moisture, has pollutants that have good adhesion properties, or contains acid gases that require neutralization, a venturi scrubber may be used for particulate control. It relies on contacting the gas stream with an alkaline liquid while at high speed. The fine particulates are agglomerated by the liquid droplets, allowing removal of the solids in a cyclone, demister, or other separator device when the gas speed is reduced. Fig. 3.25 shows the design of a venturi scrubber.

Fig. 3.25 Principles of a Venturi scrubber. (*Courtesy of EPA*)



Mercury

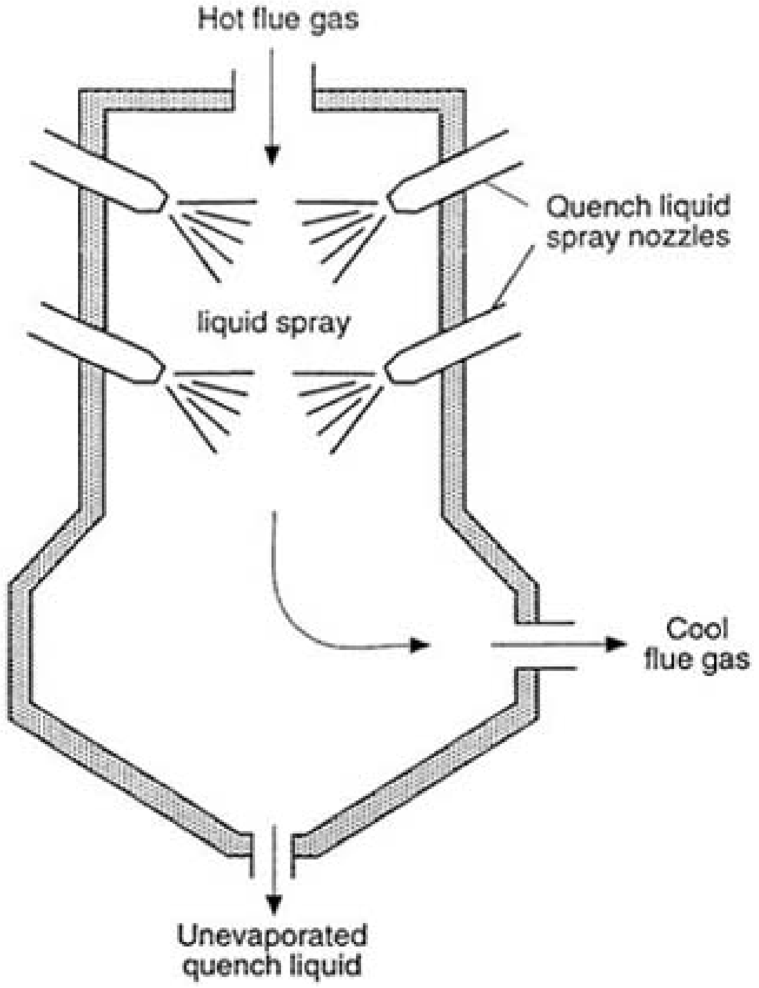
Mercury is a high priority air pollutant target for control by both federal and state government regulations. It is persistant, bioaccumualtive and toxic, and is creating water quality problems in streams and lakes in many parts of the country. However, mercury is not easily removed once it is part of the gas emissions from a facility. It is highly volatile and not amenable to adequate reductions through wet scrubbing, particulate matter removal or other conventional technologies. Removals of mercury using these technologies seldom exceed 40 percent and can be as low as zero. Instead, technologies that specifically remove mercury are now becoming more prevalent. These technologies include selenium coated filters, sodium sulfide injection, mercury scrubbers, and activated carbon adsorption. As described in the next section, activated carbon adsorption can provide additional control of dioxins and furans, leading to the choice of this technology when both types of pollutants are present and require control. Removals of mercury using optimized carbon adsorption have been shown to reach 9099 percent.

Dioxins/Furans

Dioxins and furans are extremely persistent, bioaccummulative, and toxic pollutants. The most dangerous isomers can be of concern when present in emissions at levels below one part per trillion. Therefore, in recent years it has been a priority for all levels of government to minimize the presence of dioxins and furans in emissions from all sources. Although these pollutants can be present as a contaminant from some manufacturing processes, it is far more common for the pollutants to be created or reformed in emissions that have the necessary conditions (especially temperature) and ingredients (chlorine and organic precursors) to allow formation. In fact, some pollution control devices may provide additional reaction times for dioxin/furan formation unless the system has been designed to minimize the conditions that may lead to formation. Other devices, such as heat recovery boilers also can increase the reaction times and the potential for dioxin formation.

While the conditions leading to dioxin and furan formation must be evaluated for each site, temperature frequently has been found to be a critical parameter. The amount of gas residence time in the range between 250 and 700°F is thought to be essential in formation of dioxin. Temperatures between 400 and 700° have been observed to be most significant for some hazardous waste combustion facilities. As a result, the most prevalent control technology for dioxins and furans is rapid quench cooling as the first step of air pollution control. As shown in Fig. 3.26, this technology often uses a dedicated vessel that injects a finely atomized spray of water to quickly bring the gas temperature through the range where dioxins and furans are formed. The system must be controlled to match spray rates to temperature variations in order to assure a relatively constant exit temperature for the gas. If the temperature is too high, dioxin and furans may be formed. If the temperature drops too low, it can cause operational problems in downstream operations, including condensation that may cause corrosion of devices. This device also can be upgraded to provide acid gas control as described in the next section.

Fig. 3.26 Water quench cooling procedure. (*Courtesy of EPA*)



In some cases, temperature control may prove ineffective or impractical. This has been the case when a waste heat boiler is used to recover energy from the gas stream. In these situations, carbon adsoption is a viable option. It can provide effective removal of dioxins/furans (below 5 ng per dry standard cubic meter) and provides removal of mercury and other metals as well. Carbon adsorption usually is achieved in one of three types of devices, fixed beds, fluidized beds, or duct injection. Duct injection, which frequently occurs upstream of the particulate control device can offer the least carbon usage and operational complexity. However, it is not recommended if it would be integrated with a electrostatic precipitator (safety concerns associated with carbon combustibility), or a wet scrubber (effective performance has not yet been shown). Spent carbon from these systems must be evaluated for further treatment and possible disposal as hazardous waste.

Acid Gases

The removal of chlorine, fluorine, and sulfur-based compounds, especially sulfur dioxide and chlorine/chloride is important in preventing the formation of sulfuric and hydrochloric acid in emissions. These emissions can have respiratory impacts on humans and wildlife, and ecological impacts on plants and aquatic organisms. The most common approach to controlling acid gas emissions is to use wet or dry scrubbing. The purpose of the scrubber is to contact the gas stream with lime, limestone, or sodium carbonate to produce a solid salt such as calcium sulfate or calcium chloride that can be removed through conventional solids removal technologies.

Wet scrubbers include spray towers, tray towers, packed towers and ionized wet scrubbers. All of these devices then use mist eliminators to assure effective removal of the scrubbing solution from the gas stream. Wet scrubbers often use a slurry of lime to capture and convert the acid emission precursors, then use wastewater treatment technologies to remove and concentrate the solids that are produced.

Dry scrubbers involve injection of the sorbent (e.g. lime) into the flue gas. A variation of this technology injects a wet slurry that is designed to allow the water to completely evaporate into dry solid identical to the product of a dry scrubber. Dry scrubbers, when paired with a fabric filter for particulate removal often provide a less complex solution to acid gas control. These systems do not create a wastewater stream requiring treatment and disposal, but instead only add relatively low toxicity inert solids to the particulate removal required. Mist elimination also is not needed. Although dry scrubbers do not achieve the 9599 percent removal of wet scrubbers, well-designed dry scrubbers can achieve 8090 percent acid gas removal.

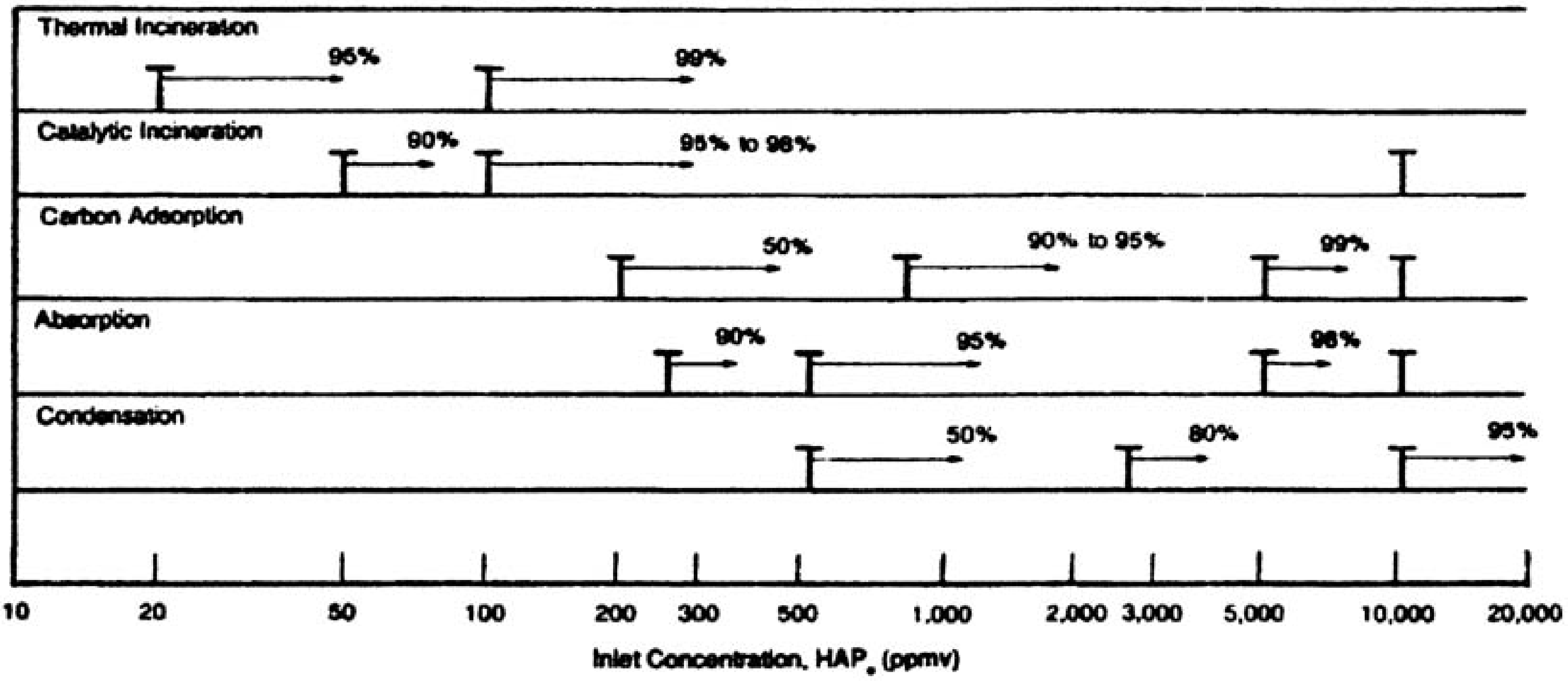
Volatile Organic Compounds

Volatile Organic Compounds (VOC) can be captured or destroyed if present in emissions in environmentally significant amounts. The most common forms of control devices are flares, incinerators (thermal or catalytic) or carbon adsorption units. However, if the VOC concentrations are high enough, recovery of the raw material or heat value is possible through the use of gasifiers or industrial boilers.

Incinerators (or thermal oxidizers as they have been renamed of late) or flares often are the control technology of choice. This technology has the advantage of high destruction efficiencies, can handle virtually any source of VOC, and produce little in the way of residuals requiring additional, management. However, the devices must be designed and well operated to avoid risky products of incomplete combustion that present significant risk in the emissions. Flares are used most often as emergency controls or when VOC loading will be highly variable or unpredictable. Flares also have the ability to handle high volume emissions that might overwhelm an incinerator.

Carbon adsorption fixed bed units are the other commonly used control device. It is best suited for low to moderate concentrations of organics (10,000ppm or less) and compounds with molecular weights between 45 and 130. It is the technology of choice when the source of the VOC is a high value material. Operating parameters for the carbon can be critical. For example, a solids removal technology may be necessary to prevent clogging of the carbon columns. It is also important to select the correct carbon and appropriate method of carbon regeneration. The estimated performance range of control technologies for organic emissions is shown in Fig. 3.27.

Fig. 3.27 Approximate percent reduction ranges for various categories of add-on equipment. (*Courtesy of EPA*)



< previous page

next page >

3.6 References

Freeman, Harry M., Standard Handbook of Hazardous Waste Treatment and Disposal, McGraw Hill Publishing, New York, NY, 1998.

Heinsold, Robert Jennings and Kabel, Robert Lynn, Sources and Control of Air Pollution, Prentice Hall Inc., Saddle River, NJ, 1999.

U.S. Environmental Protection Agency, Technical Support Document for Hazardous waste Combuster MACT Standards, Office of Solid Waste, Available on line only (Sept. 1999).

U.S. Environmental Protection Agency, Handbook: Control Technologies For Hazardous Air Pollutants, Office of Research and Development, EPA/625/6-91/014, June 1991.

U.S. Environmental Protection Agency, National Emission Standards for Source Categories: Paper and Other Web-Coating OperationsBackground Information for Proposed Standards, Office of Air Quality Planning and Standards, EPA-453/R-00-002, Apr. 2000.

U.S. Environmental Protection Agency, Treatment Technology Background Document, Office of Solid Waste, PB91-160-556, June 1991.

U.S. Environmental Protection Agency, Guide for Industrial Waste Management, Office of Solid Waste, EPA-R-99-001, June 1999.

U.S. Environmental Protection Agency, Background Document for the Centralized Wastewater Treatment Point Source Category, Office of Water, EPA-821-R-00-020, Dec. 2000.

4.1 Introduction

The discipline of statistics is the study of effective methods of data collection, data summarization, and (data based, quantitative) inference making in a framework that explicitly recognizes the reality of nonnegligible variation in many real-world processes and measurements.

The ultimate goal of the field is to provide tools for extracting the maximum amount of useful information about a noisy physical process from a given investment of data collection and analysis resources. It is clear that such a goal is relevant to the practice of industrial chemistry. The primary purposes of this chapter are to indicate in concrete terms the nature of some existing methods of applied statistics that are particularly appropriate to industrial chemistry, and to provide an entry into the statistical literature for those readers who find in the discussion here reasons to believe that statistical tools can help them be effective in their work.

This chapter will begin with some simple ideas of modern descriptive statistics, including numerical and graphical data summarization tools, and the notions of fitting equations to data and using theoretical distributions. Next, some tools for routine industrial process monitoring and capability assessment, concentrating primarily on the notion of control charting, will be presented. This will be followed by a more extensive discussion of common statistical data collection strategies and data analysis methods for multifactor experimental situations met in both laboratory and production environments. This section will touch on ideas of partitioning observed variation in a system response to various sources thought to influence the response, factorial and fractional factorial experimental designs, sequential experimental strategy, screening experiments, and response surface fitting and representation. Next come brief discussions of two types of special statistical tools associated specifically with chemical applications, namely, mixture techniques and nonlinear mechanistic model building. A short exposition of chemical industry implications of relationships between modern business process improvement programs and the discipline of statistics follows. The chapter concludes with a reference section listing sources for further reading.

4.2 Simple Tools of Descriptive Statistics

There are a variety of data summarization or description methods whose purpose is to make evident the main features of a data set. (Their use, of course, may be independent of whether or not the data collection process actually employed was in any sense a "good" one.) To illustrate some of the simplest of these methods, consider the data listed in Table 4.1. These numbers represent aluminum impurity contents (in ppm) of 26 bihourly samples of recycled PET plastic recovered at a Rutgers University recycling pilot plant.

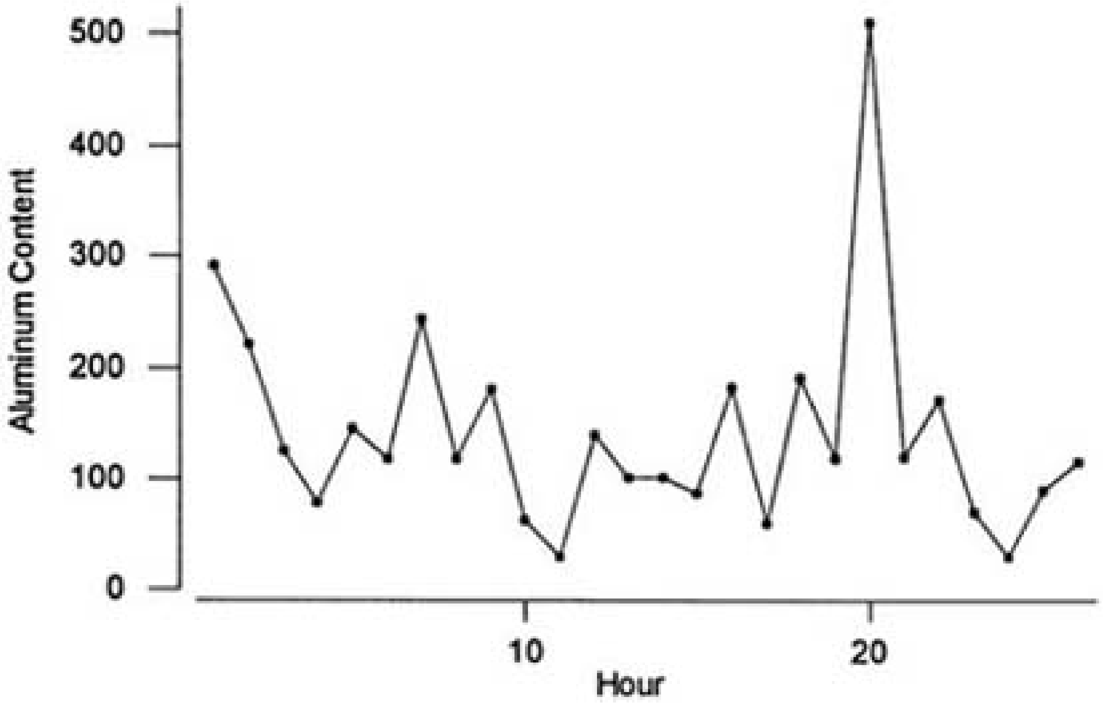
TABLE 4.1 Twenty-Six Consecutive Aluminum Contents (ppm)a

291,	222,	125,	79,	145,	119,	244,	118,	182,	63,	30,	140,	101
102,	87,	183,	60,	191,	119,	511,	120,	172,	70,	30,	90,	115

aBased on data in Albin.1

A simple plot of aluminum content against time order, often called a *run chart*, is a natural place to begin looking for any story carried by a data set. Figure 4.1 shows such a plot for the data of Table 4.1, and in this case reveals only one potentially interesting feature of the data. That is, there is perhaps a weak hint of a downward trend in the aluminum contents that might well have been of interest to the original researchers. (If indeed the possible slight decline in aluminum contents is more than "random scatter," knowledge of its physical origin, whether in actual composition of recycled material or in the measurement process, presumably would have been helpful to the effective running of the recycling facility. We will save a discussion of tools for rationally deciding whether there is more than random scatter in a plot like Fig. 4.1 until the next section.)

Fig. 4.1 A run chart for 26 consecutive aluminum contents.



The run chart is a simple, explicitly dynamic tool of descriptive statistics. In those cases where one decides that there is in fact little information in the time order corresponding to a data set, there are a variety of simple, essentially static, statistical tools that can be used in describing the pattern of variation in a data set. Figures 4.24.5 show graphical representations of the data of Table 4.1 in, respectively, *histogram*, *stem and leaf plot*, *dot plot*, and *box plot* forms.

Fig. 4.2 A histogram for 26 aluminum contents.

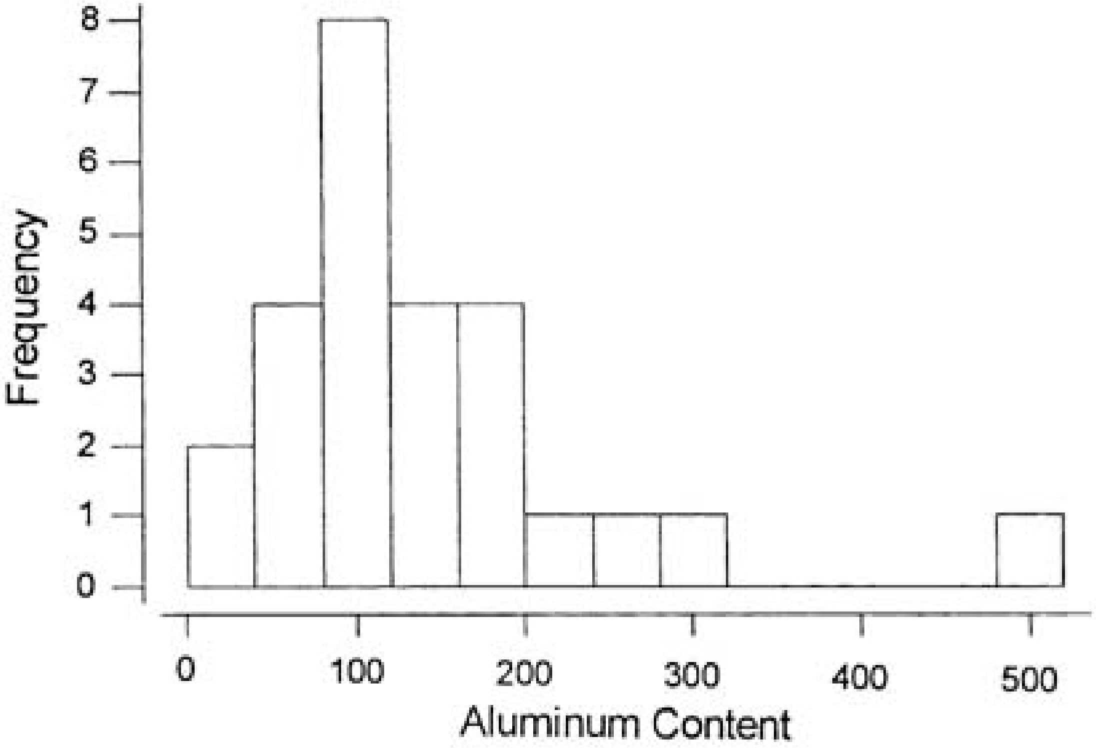


Fig. 4.3 A stem and leaf plot for 26 aluminum contents.

5 | 11
4 |
4 |
3 |
3 |
2 | 91
2 | 22,44
1 | 72,82,83,91
1 | 01,02,15,18,19,19,20,25,40,45
0 | 60,63,70,79,87,90
0 | 30,30

Fig. 4.4 A dot plot for 26 aluminum contents.

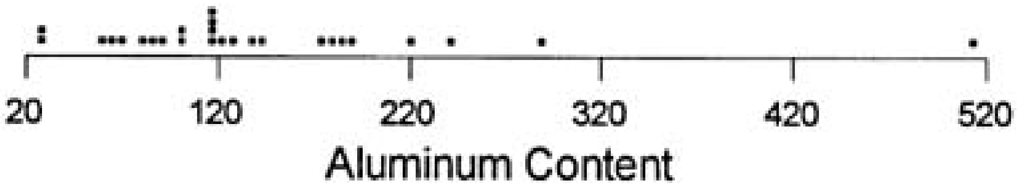
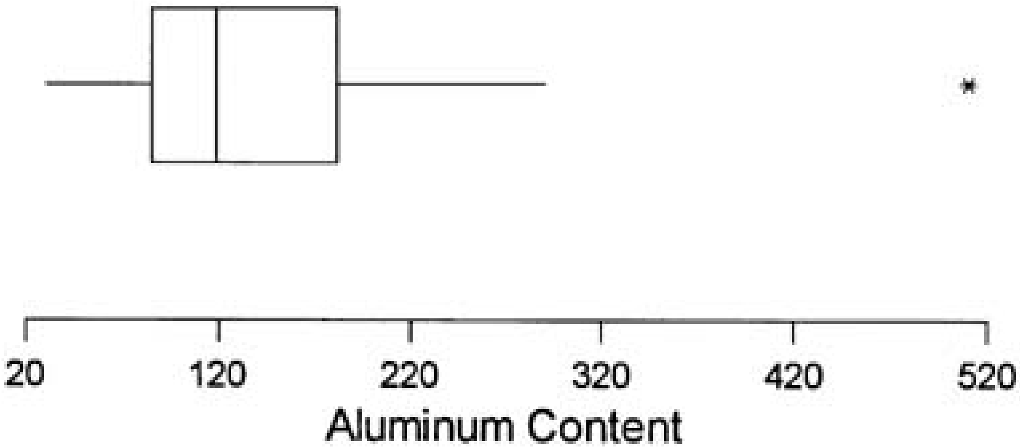


Fig. 4.5 A box plot for aluminum contents.



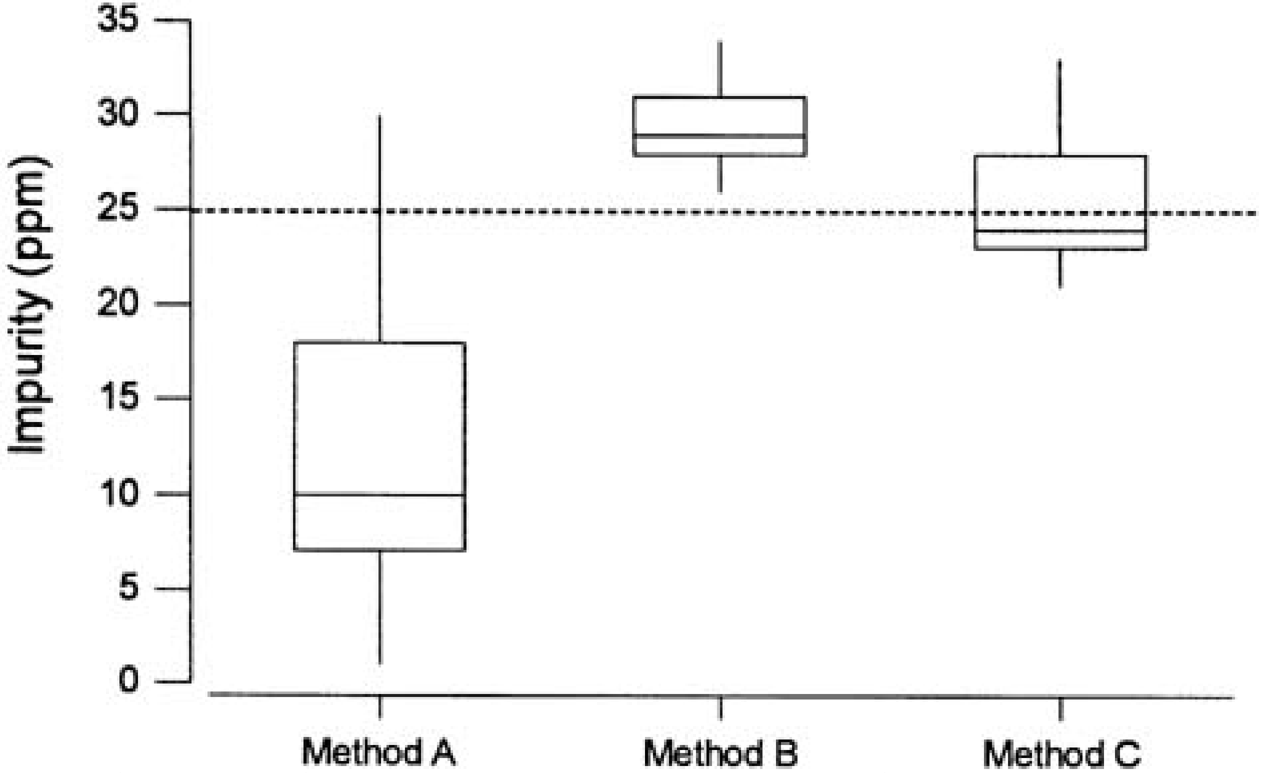
The histogram/bar chart idea of Fig. 4.2 is likely familiar to most readers, being readily available, for example, through the use of commercial spreadsheet software. It shows how data are spread out or distributed across the range of values represented, tall bars indicating high frequency or density of data in the interval covered by the base of the bar. Figure 4.2 shows the measured aluminum contents to be somewhat asymmetrically distributed (statistical jargon is that the distribution is "skewed right"), with a "central" value perhaps somewhere in the vicinity of 120 ppm.

Histograms are commonly and effectively used for final data presentation, but as working data analysis tools they suffer from several limitations. In one direction, their appearance is fairly sensitive to the data grouping done to make them, and it is usually not possible to recover from a histogram the exact data values used to produce it, should one wish to try other groupings. In another direction, histograms are somewhat unwieldy, for example, not being particularly suitable to the comparison of, say, 10 or 12 data sets on a single page. The graphical devices of Figs 4.34.5 are less common than the histogram, but address some of these shortcomings.

The stem and leaf diagram of Fig. 4.3 and the dot plot of Fig. 4.4 carry shape information about the distribution of aluminum contents in a manner very similar to the histogram of Fig. 4.2. But the stem and leaf and dot diagrams do so without losing the exact identities of the individual data points. The box plot of Fig. 4.5 represents the "middle half" of the data with a box divided at the 50th percentile (or in statistical jargon, the median) of the data, and then uses so-called whiskers to indicate how far the most extreme data points are from the middle half of the data.

Box plots preserve much of the shape information available from the other displays (e.g. portraying lack of symmetry through differing sizes of box "halves" and/or whisker lengths), but do so in a way that is conducive to simultaneous representation and comparison of many data sets on a single graphic, through the placement of box plots side by side. Figure 4.6 illustrates this point with a graphical comparison of three laboratory test methods to a standard.

Fig. 4.6 Side-by-side box plots for three laboratory test methods.



A total of 90 samples of a stock solution known to contain 25 ppm of an impurity were analyzed by a single lab team using three different test methods (30 of the samples being allocated to each of the three methods), and the box plots in Fig. 4.6 portray the measured impurity levels for the different methods. The figure shows quite effectively that Method A is neither precise nor accurate, Method B is quite precise but not accurate, and Method C is somewhat less precise than B but is accurate. This kind of knowledge can form the basis of an informed choice of method.

Figures 4.24.6 give only a hint of the spectrum of tools of statistical graphics that are potentially helpful in data analysis for industrial chemistry. For more details and much additional reading on the subject of modern statistical graphics, the reader is referred to the book by Chambers et al.2 listed in the references section.

Complementary to graphical data summaries are *numerical summarizations*. For the simple case of data collected under a single set of conditions, the most commonly used measures deal with the location/center of the data set and the variability/spread of the data. The (*arithmetic*) *mean* and the *median* are the most popular measures of location, and the *variance* and its square root, the *standard deviation*, are the most widely used measures of internal variability in a data set.

For n data values $y1, y2, \ldots, yn$ the median is

$$\tilde{y} = \text{the ``middle'' or } \frac{n+1}{2} \text{th ordered data value} \qquad (4\text{-}1)$$

and the mean is

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad (4\text{-}2)$$

The reader is invited to check that upon ordering the $n = 26$ values in Table 4.1, the 13th smallest value is 119 and the 14th smallest value is also 119, so that the only sensible interpretation of (4-1) for the aluminum content data is that

$$\begin{aligned} \tilde{y} &= \text{the 13.5th ordered data value} \\ &= \frac{119+119}{2} = 119\,\text{ppm} \end{aligned}$$

On the other hand, from (4-2) the mean of the aluminum contents is

$$\begin{aligned}\bar{y} &= \frac{1}{26}(291+222+125+\cdots+30+90+115)\\ &\approx 142.7\,\mathrm{ppm}\end{aligned}$$

The media and mean are clearly different measures of location/center. The former is in the middle of the data in the sense that about half of the data are larger and about half are smaller. The latter is a kind of "center of mass," and for asymmetrical data sets like that of Table 4.1 is usually pulled from the median in the direction of any "skew" present, that is, is pulled in the direction of "extreme" values.

The variance of *n* data values *y*1, *y*2,…, *yn* is essentially a mean squared deviation of the data points from their mean. In precise terms, the variance is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2 \qquad (4\text{-}3)$$

and the so-called standard deviation is

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2} \qquad (4\text{-}4)$$

For the example of the aluminum contents, it is elementary to verify that

$$\begin{aligned}s^2 &\approx \frac{1}{26-1}\left[(291-142.7)^2+(222-142.7)^2\right.\\ &\qquad \left.+\cdots+(115-142.7)^2\right]\\ &\approx 9,644\,(\mathrm{ppm})^2\end{aligned}$$

so that

$$s=\sqrt{s^2}\approx 98.2\,\mathrm{ppm}$$

An appropriate interpretation of *s* is not completely obvious at this point, but it does turn out to measure the spread of a data set, and to be extremely useful in drawing quantitative inferences from data. (In most, but not all, circumstances met in practice, the *range* or largest value in a data set minus the smallest value is on the order of four to six times *s*.) The variance and standard deviation are time-honored and fundamental quantifications of the variation present in a single group of measurements and, by implication, the data-generating process that produced them.

When data are collected under several different sets of conditions, and those conditions can be expressed in quantitative terms, effective data summarization often takes the form of *fitting an approximate equation* to the data. As the basis of a simple example of this, consider the data in Table 4.2. The variable *x*, hydrocarbon liquid hourly space velocity, specifies the conditions under which information on the response variable *y*, a measure of isobutylene conversion, was obtained in a study involving the direct hydration of olefins.

TABLE 4.2 Seven Liquid Hourly Space Velocity/Mole % Conversion Data Pairsa

Liquid Hourly Space Velocity, x	Mole % Isobutylene Conversion, y
1	23.0, 24.5
2	28.0
4	30.9, 32.0, 33.6
6	20.0

aBased on a graph in Odioso et al.3

For purposes of economy of expression, and perhaps some cautious interpolation between values of *x* not included in the original data set, one might well like to fit a simple equation involving some parameters $\underline{b}$, say,

$$y \approx f(x|\,\underline{b}) \qquad (4\text{-}5)$$

to the data of Table 4.2. The simplest possible form for the function $f(x|\,\underline{b})$ that accords with the "up then back down again" nature of the conversion values *y* in Table 4.2 is the quadratic form

$$f(x|\,\underline{b}) = b_0 + b_1x + b_2x^2 \qquad (4\text{-}6)$$

and a convenient method of fitting such an equation (that is linear in the parameters *b*) is *the method of least squares*. That is, to fit a parabola through a plot of the seven (*x*,*y*) pairs specified in Table 4.2, it is convenient to choose *b*0, *b*1, and *b*2 to minimize the sum of squared differences between the observed conversion values *y* and the corresponding fitted values *y* on the parabola. In symbols, the least squares fitting of the approximate relationship specified by (4-5) and (4-6) to the data of Table 4.2 proceeds by minimization of

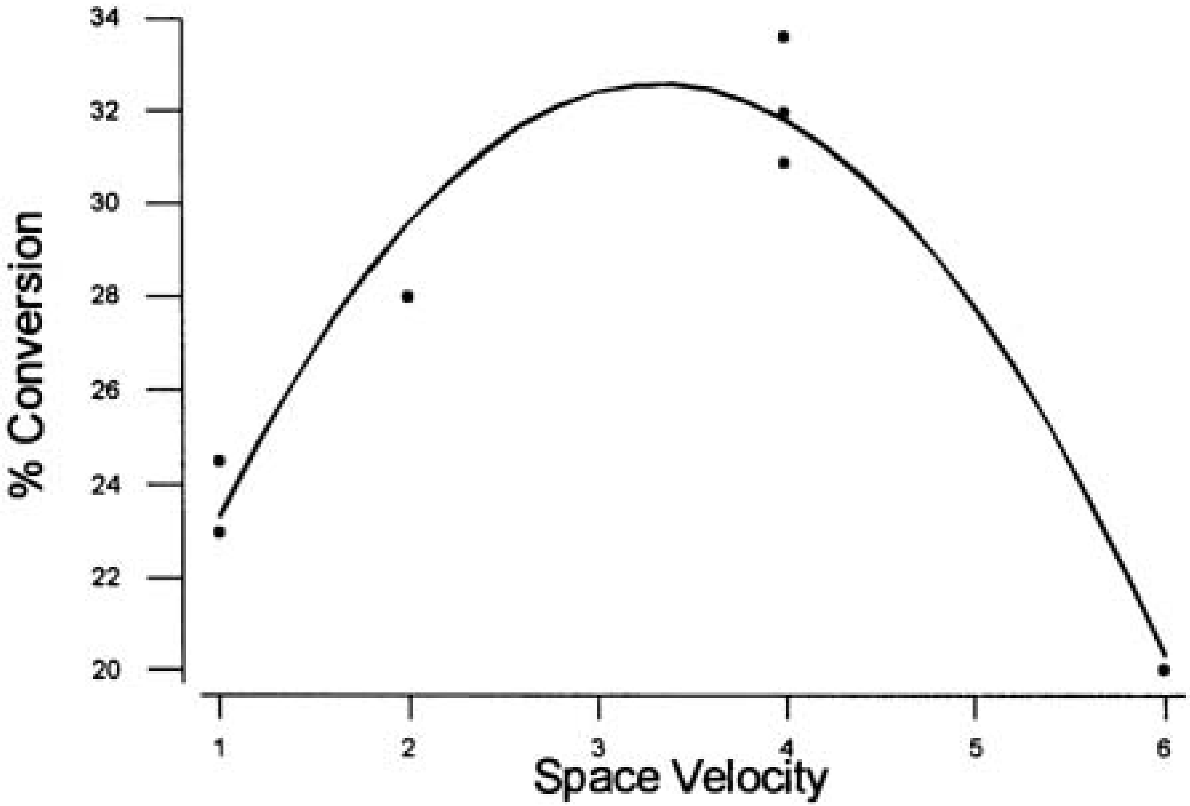
$$\sum_{i=1}^{7}\left[y_i-\left(b_0+b_1x_i+b_2x_i^2\right)\right]^2$$

over choices of the coefficients *b*. As it turns out, use of standard statistical "regression analysis" software shows that the fitting process for this example produces the approximate relationship

$$y \approx 13.64 + 11.41x - 1.72x^2$$

and Fig. 4.7 shows the fitted (summarizing) parabola sketched on the same set of axes used to plot the seven data points of Table 4.2.

Fig. 4.7 A scatter plot of seven space velocity/mole % conversion data pairs and a fitted parabola.

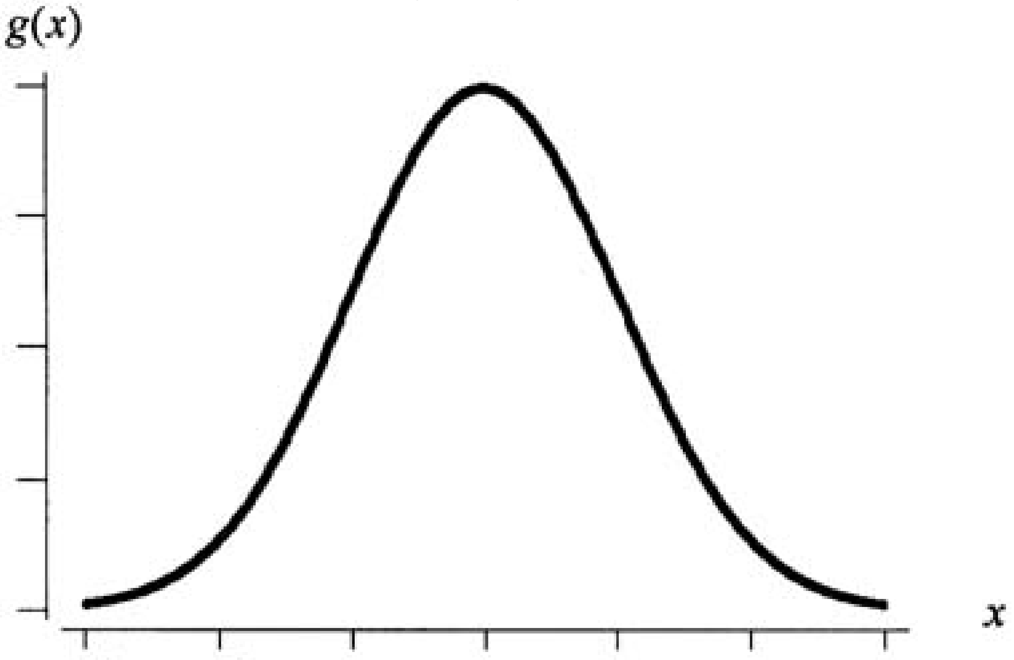


The least squares fitting of approximate functional relationships to data with even multidimensional explanatory variable *x* typically goes under the (unfortunately obscure) name of *multiple regression* analysis, and is given an introductory treatment in most engineering statistics textbooks, including, for example, the ones by Devore,4 Vardeman and Jobe,5 and Vardeman6 listed in the references. A lucid and rather complete treatment of the subject can also be found in the book by Neter et al.7

A final notion that we wish to treat in this section on descriptive statistics is that of representing a distribution of responses and/or the mechanism that produced them (under a single set of physical conditions) by a *theoretical distribution*. That is, there are a number of convenient theoretical distributional shapes, and it is often possible to achieve great economy of expression and thought by seeing in a graphical representation such as Figs 4.24.5 the possibility of henceforth describing the phenomenon portrayed via some one of those theoretical distributions. Here we will concentrate on only the most commonly used theoretical distribution, the so-called *Gaussian* or *normal* distribution.

Figure 4.8 is a graph of the function of *x*

Fig. 4.8 The Gaussian probability density with mean μ and standard deviation σ.



$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2} \right) \qquad (4\text{-}7)$$

where $g(x)$ specifies the archetypical "bell-shaped curve" centered at the number μ, with spread controlled by the number σ (and is in fact usually called the Gaussian probability density with mean μ and standard deviation σ).

Figure 4.8 can be thought of as a kind of idealized histogram. Just as fractional areas enclosed by particular bars of a histogram correspond to fractions of a data set with values in the intervals represented by those bars, areas under the curve specified in (4-7) above particular intervals might be thought of as corresponding to fractions of potential data points having values in those intervals. (It is possible to show that the total area under the curve represented in Fig. 4.8, namely, $\int_{-\infty}^{\infty} g(x)dx$, is 1.) Simple tabular methods presented in every elementary statistics book avoid the need to regularly use numerical integration in evaluating such areas. These methods can, for example, be used to show that roughly 68 percent of a Gaussian distribution lies between $\mu - \sigma$ and $\mu + \sigma$, roughly 95 percent lies between $\mu - 2\sigma$ and $\mu + 2\sigma$, and roughly 99.7 percent lies between $\mu - 3\sigma$ and $\mu + 3\sigma$. Part of the convenience provided when one can treat a data-generating process as approximately Gaussian is that, given only a theoretical mean μ and theoretical standard deviation σ, predictions of fractions of future data values likely to fall in intervals of interest are thus easy to obtain.

At this point let us return to the aluminum content data of Table 4.1. The skewed shape that is evident in all of Figs 4.24.5 makes a Gaussian distribution inappropriate as a theoretical model for (raw) aluminum content of such PET samples. But as is often the case with right skewed data, considering the *logarithms* of the original measurement creates a scale where a normal distribution is more plausible as a representation of the phenomenon under study. Thus, Table 4.3 contains the natural logs of the values in Table 4.1, and the corresponding stem and leaf plot in Fig. 4.9 shows the transformed data to be much more symmetrically distributed than the original data. The possibility opened up by this kind of transformation idea is one of using statistical methods based on the normal distribution to reach conclusions about ln*y* and then simply exponentiating to derive conclusions about the original response *y* itself. The applicability of statistical methods developed for normal distributions is thereby significantly broadened.

TABLE 4.3 Twenty-Six Logarithms of Aluminum Contents

5.67,	5.40,	4.83,	4.37,	4.98,	4.78,	5.50,	4.77,	5.20,	4.14,	3.40,	4.94,	4.62
4.62,	4.47,	5.21,	4.09,	5.25,	4.78,	6.24,	4.79,	5.15,	4.25,	3.40,	4.50,	4.74

Fig. 4.9 A stem and leaf plot for the logarithms of 26 aluminum contents.

6 | .24
5 | .50,.67
5 | .15,.20,.21,.25,.40
4 | .50,.62,.62,.74,.77,.78,.79,.79,.83,.94,.98
4 | .09,.14,.25,.37,.47
3 |
3 | .40,.40

In addition to providing convenient conceptual summarizations of the nature of, response distributions, theoretical distributions such as the normal distribution form the mathematical underpinnings of methods of formal quantitative *statistical inference*. It is outside our purposes in this chapter to provide a complete introduction to such methods, but thorough and readable accounts are available in engineering statistics books such as those of Devore4 and Vardeman and Jobe.5 Here, we will simply say that, working with a Gaussian description of a response, it is possible to quantify in various ways how much information is carried by data sets of various sizes. For instance, if a normal distribution describes a response variable *y*, then in a certain well-defined sense, based on *n* = 26 observations producing a mean $\bar{y}$ and a standard deviation *s*, the interval with end points

$$\bar{y} - 2.060s\sqrt{1 + \tfrac{1}{26}}$$

and

$$\bar{y} + 2.060s\sqrt{1 + \tfrac{1}{26}} \qquad (4\text{-}8)$$

has a 95 percent chance of predicting the value of an additional observation. For instance, applying formula (4-8) to the log values in Table 4.3, the conclusion is that the interval from 3.45 to 6.10 ln(ppm) has a 95 percent chance of bracketing an additional log aluminum content produced (under the physical conditions of the original study) at the recycling plant. Exponentiating, the corresponding statement about raw aluminum content is that the interval from 31 to 446 ppm has a 95 percent chance of bracketing an additional aluminum content. Methods of statistical inference like that represented in (4-8) are called *prediction interval* methods. The book by Hahn and Meeker8 provides a thorough discussion of such methods, based not only on the Gaussian distribution but on other theoretical distributional shapes as well.

4.3 Tools of Routine Industrial Process Monitoring and Capability Assessment

Probably the two most basic generic industrial problems commonly approached using statistical methods are those of (1) monitoring and maintaining the stability/consistency of a process and (2) assessing the capability of a stable process. This section provides a brief introduction to the use of tools of "control" charting in these enterprises.

Working at Bell Labs during the 1920s and 1930s, Walter Shewhart developed the notion of routinely plotting data from an industrial process in a form that allows one to separate observed variability in a response into two kinds of variation. The first is that variation which appears to be inherent, unavoidable, short-term, baseline, and characteristic of the process (at least as currently configured). This variation Shewhart called *random* or *common cause variation*. The second kind of variability is that variation which appears to be avoidable, long-term, and/or due to sources outside of those seen as legitimately impacting process behavior. This variation he called *assignable* or *special cause variation*. Shewhart reasoned that by plotting summary statistics from periodically collected data sets against time order of collection, one would be able to see interpretable trends or other evidence of assignable variation on the plots, and could intervene to eliminate the physical causes of that variation. The intention was thereby to make process output stable or consistent to within the inherent limits of process precision. As a means of differentiating plotted values that should signal the need for intervention from those that carry no special message of process distress, he suggested drawing so-called control limits on the plots. (The word "control" is something of a misnomer, at least as compared to common modern engineering usage of the word in referring to the active, moment-by-moment steering or regulation of processes. The nonstandard and more passive terminology "monitoring limits" would actually be far more descriptive of the purpose of Shewhart's limits.) These limits were to separate plausible values of the plotted statistic from implausible values when in fact the process was operating optimally, subject only to, causes of variation that were part of standard conditions.

By far the most famous implementations of Shewhart's basic logic come where the plotted statistic is either the mean, the range, or, less frequently, the standard deviation. Such charts are commonly known by the names *x-bar charts*, *R charts*, and *s charts*, respectively. As a basis of discussion of Shewhart charts, consider the data given in Table 4.4. These values represent melt index measurements of specimens of extrusion grade polyethylene, taken four per shift in a plastics plant.

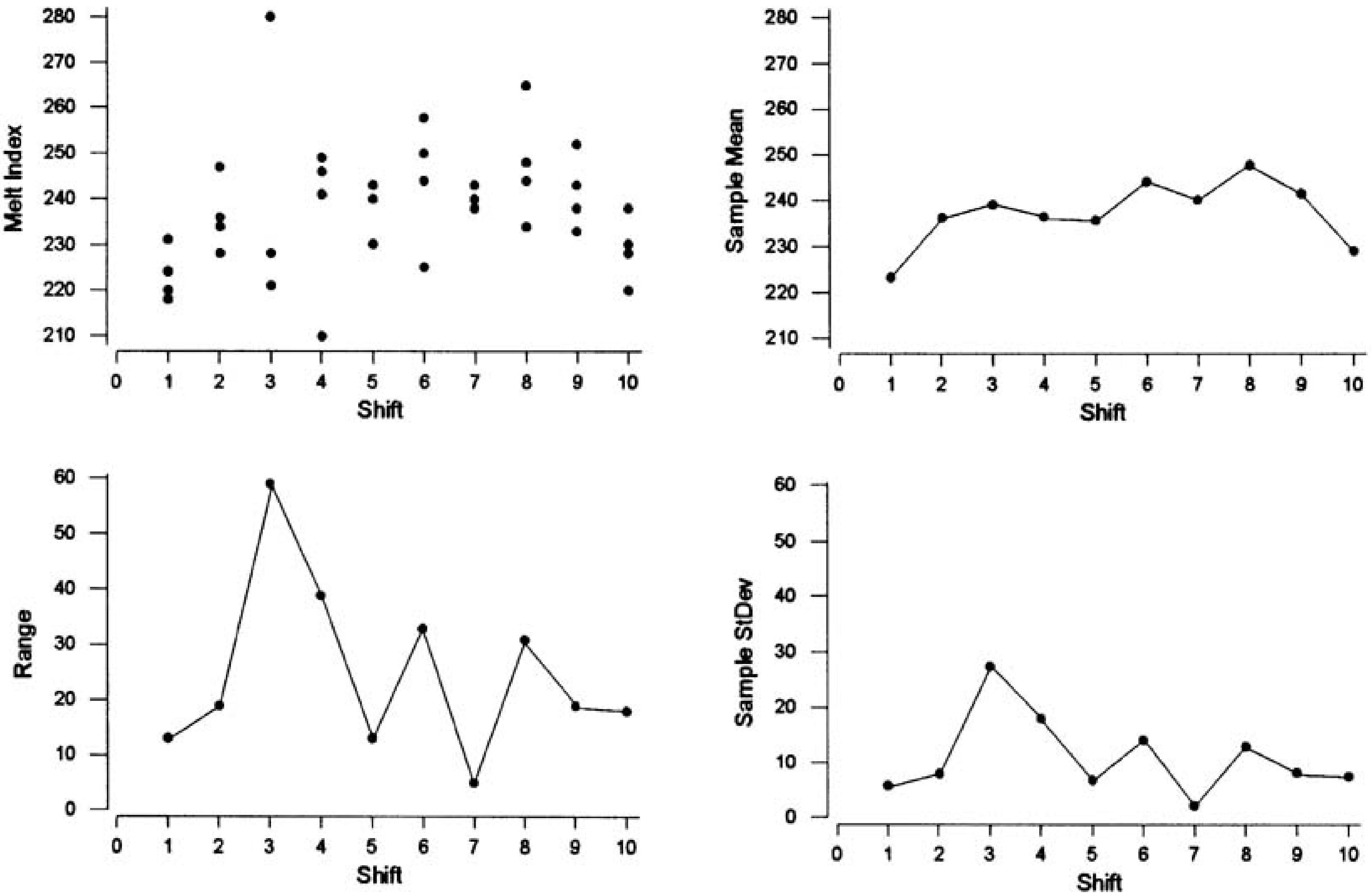
TABLE 4.4 Measured Melt Indices for Ten Groups of Four Specimens[a]

Shift	Melt Index	$\bar{y}$	R	s
1	218, 224, 220, 231	223.25	13	5.74
2	228, 236, 247, 234	236.25	19	7.93
3	280, 228, 228, 221	239.25	59	27.37
4	210, 249, 241, 246	236.50	39	17.97
5	243, 240, 230, 230	235.75	13	6.75
6	225, 250, 258, 244	244.25	33	14.06
7	240, 238, 240, 243	240.25	5	2.06
8	244, 248, 265, 234	247.75	31	12.92
9	238, 233, 252, 243	241.50	19	8.10
10	228, 238, 220, 230	229.00	18	7.39

[a] Based on data from page 207 of Wadsworth, Stephens, and Godfrey.[9]

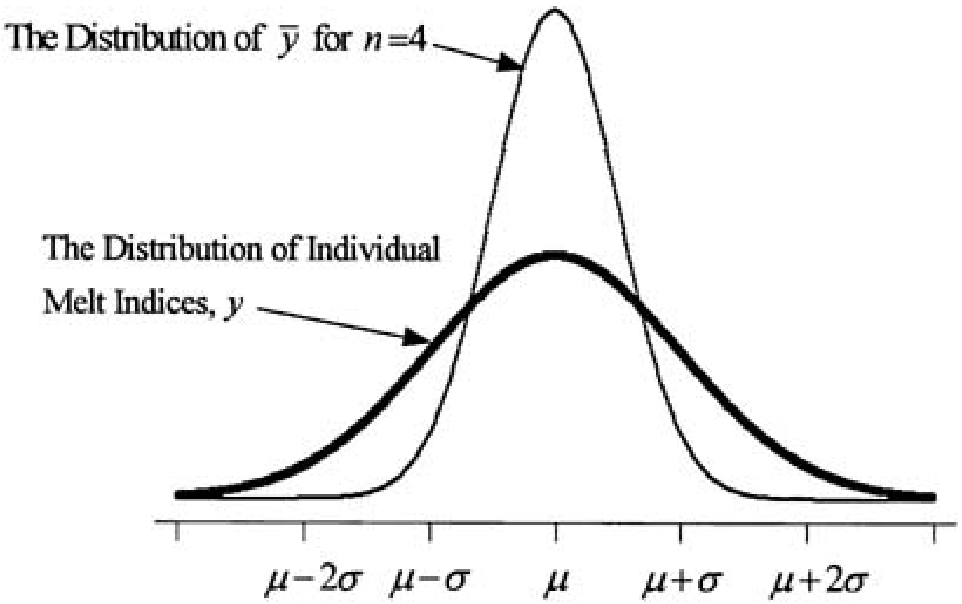
Figure 4.10 shows plots of the individual melt indices, means, ranges, and standard deviations from Table 4.4 against shift number. The last three of these are the beginnings of so-called Shewhart $\bar{x}$, *R*, and *s* control charts.

Fig. 4.10 Plots of melt index, $\bar{y}$, *R* and *s* against shift number.



What remain to be added to the plots in Fig. 4.10 are appropriate control limits. In order to indicate the kind of thinking that stands behind control limits for Shewhart charts, let us concentrate on the issue of limits for the plot of means. The fact is that mathematical theory suggests how the behavior of *means* $\bar{y}$ ought to be related to the distribution of *individual* melt indices *y*, provided the data-generating process is stable, that is, subject only to random causes. If individual responses *y* can be described as normal with some mean μ and standard deviation σ, mathematical theory suggests that averages of *n* such values will behave as if a different normal distribution were generating them, one with a mean $\mu_{\bar{y}}$ that is numerically equal to μ and with a standard deviation $\sigma_{\bar{y}}$ that is numerically equal to $\sigma/\sqrt{n}$. Figure 4.11 illustrates this theoretical relationship between the behavior of individuals and the behavior of means.

Fig. 4.11 The distribution of individuals, *y*, and sample means, $\bar{y}$.



The relevance of Fig. 4.11 to the problem of setting control chart limits on means is that if one is furnished with a description of the typical pattern of variation in *y*, sensible expectations for variation in $\bar{y}$ follow from simple normal distribution calculations. So Shewhart reasoned that since about 99.7 percent (most) of a Gaussian distribution is within three standard deviations of the center of the distribution, means found to be farther than three theoretical standard deviations (of $\bar{y}$) from the theoretical mean (of $\bar{y}$) could be safely attributed to other than chance causes. Hence, furnished with standard values for μ and σ (describing individual observations), sensible control limits for $\bar{y}$ become

$$\begin{aligned}\text{Upper Control Limit (UCL) for } \bar{y} &= \mu_{\bar{y}} + 3\sigma_{\bar{y}} \\ &= \mu + 3\frac{\sigma}{\sqrt{n}}\end{aligned}$$

and

$$\begin{aligned}\text{Lower Control Limit (LCL) for } \bar{y} &= \mu_{\bar{y}} - 3\sigma_{\bar{y}} \\ &= \mu - 3\frac{\sigma}{\sqrt{n}} \qquad (4\text{-}9)\end{aligned}$$

Returning to the context of our example represented by the data of Table 4.4, Wadsworth et al.9 state that the target value for melt index in the original application was in fact 235. So if standard process behavior is "on target" behavior, the value μ = 235 seems appropriate for use in (4-9). No parallel value for σ was provided by the authors. Common practice in such situations is to use the data in hand (the data of Table 4.4) to produce a plausible value for σ to use in (4-9). There are many possible ways to produce such a value, but to understand the general logic behind the standard ones, it is important to understand what σ is supposed to measure. The variable σ is intended as a theoretical measure of baseline, short-term, common cause variation. As such, the safest way to try to approximate it is to somehow use only measures of variation *within* the groups of four values in Table 4.4 *not* influenced by variation *between* groups. (Measures of variation derived from considering all the data simultaneously, e.g., would reflect variation between shifts as well as the shorter-term variation within shifts.) In fact, the most commonly used ways of obtaining from the data in hand a value of σ for use in (4-9) are based on the averages of the (within-group) ranges or standard deviations. For example, the 10 values of *R* given in Table 4.4 have a mean

$$\overline{R} = \frac{1}{10}(13 + 19 + 59 + \cdots + 19 + 18) = 24.9$$

and some standard mathematical theory suggests that because the basic group size here is *n* = 4, an appropriate multiple of $\overline{R}$ for use in estimating σ is

$$\frac{\overline{R}}{2.059} \approx 12.1 \qquad (4\text{-}10)$$

(The divisor above is a tabled factor commonly called *d*2, which increases with *n*.)

Finally, substituting 235 for μ and 12.1 for σ in (4-9) produces numerical control limits for $\bar{y}$:

$$\text{LCL} = 235 - 3\frac{(12.1)}{\sqrt{4}} = 216.9$$

and

$$\text{UCL} = 235 + 3\,\frac{(12.1)}{\sqrt{4}} = 253.1$$

Comparison of the $\bar{y}$ values in Table 4.4 to these limits reveals no "out of control" means, that is, no evidence in the means of assignable process variation. Figures 4.12 and 4.13 show control charts for all of $\bar{y}$, *R*, and *s*, where control limits for the last two quantities have been derived using standard calculations not shown here.

Fig. 4.12 Control charts for $\bar{y}$ and *R* based on melt indices.

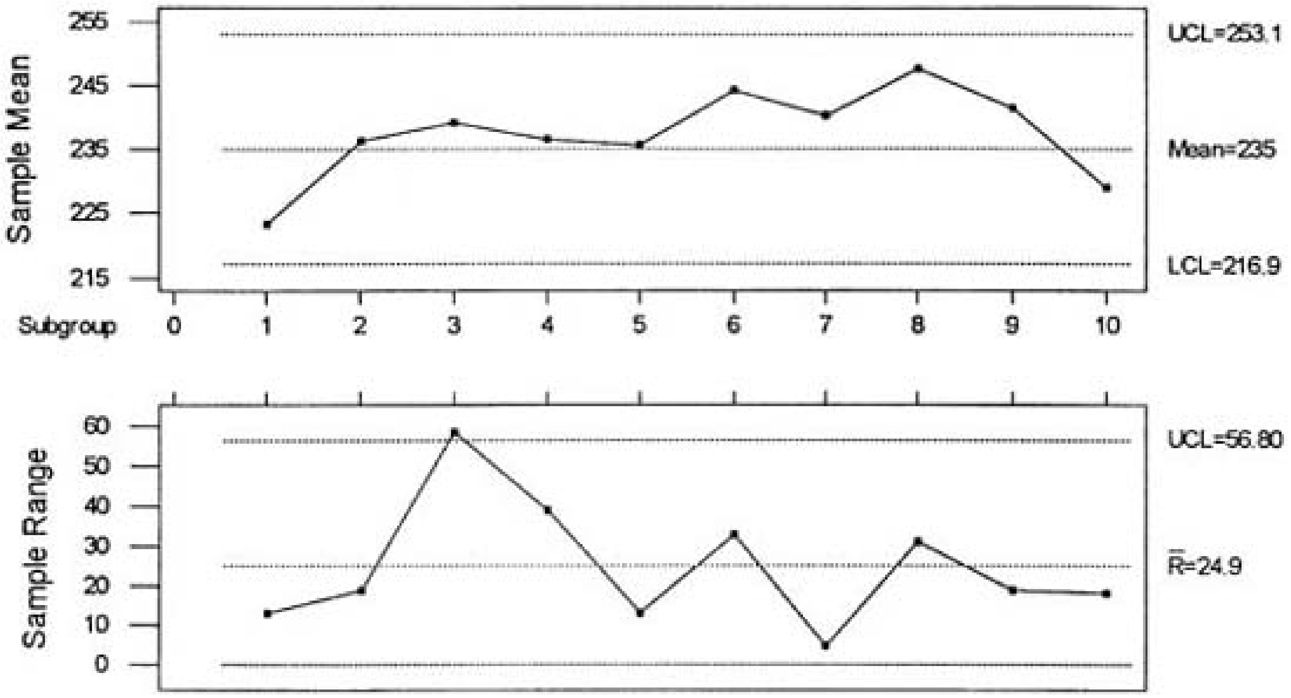
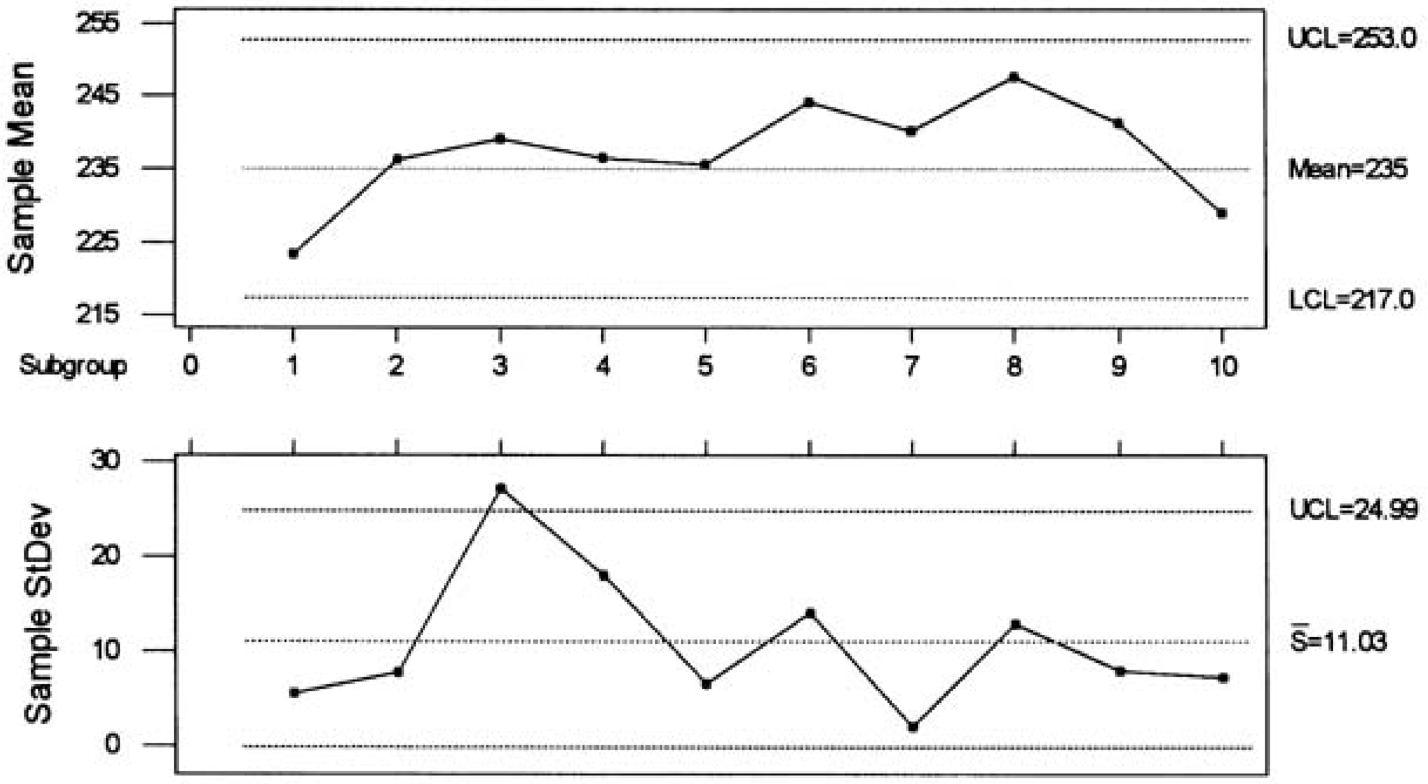


Fig. 4.13 Control charts for $\bar{y}$ and *s* based on melt indices.



The *R* and *s* charts in Figs 4.12 and 4.13 are related representations (only one is typically made in practice) of the shift-to-shift behavior of melt index *consistency*. It is seen that on both charts, the shift #3 point plots above the upper control limit. The strong suggestion thus is that melt index consistency was detectably worse on that shift than on the others, so that from this point of view the process was in fact *not stable* over the time period represented in Table 4.4. In practice, physical investigation and hopefully correction of the origin of the instability typically would follow, as well as some reconsideration of our earlier assessment of 12.1 as a plausible figure to represent the inherent short-term variability of melt index. (If shift #3 could be treated as a special case, explainable as an unfortunate but correctable situation that was not expected to reoccur, there might be reason to revise $\overline{R}$ downward by deletion of shift #3 from the calculation, and thereby to reduce one's view of the size of baseline process variability. Notice that, in general, such a downward revision of $\overline{R}$ might well also have the effect of causing one to need to rethink his or her assessment of the constancy of the melt index *mean*.)

There is a variation on the basic "$\overline{x}$ and *R* chart" idea that we wish to illustrate here next, because of its frequent application in chemical industry applications. That is the making of a so-called *x* and MR chart pair. The motivation for this modification of the ideas outlined thus far in this section is that in many chemical process monitoring contexts the natural "group size" is $n = 1$. A mean of $n = 1$ observation(s) is simply that observation itself, and the limits of (4-9) make perfectly good sense for the case of $n = 1$. That is, the analog of an $\overline{x}$ chart for $n = 1$ cases is clear, *at least if one has an externally provided value for* σ. But what, if anything, to do for an $n = 1$ counterpart of the *R* chart and how to develop an analog of (4-10) in cases where σ is not a priori known are perhaps not so obvious. Table 4.5 contains data representing moisture contents in 0.01 percent of bihourly samples of a polymer, and the question at hand is what besides simply the bihourly *y* values might be plotted in the style of a Shewhart control chart for such data.

TABLE 4.5 Moisture Contents for 20 Polymer Samples[a]

Sample	Moisture, y	Moving Range, MR
1	36	
2	20	16
3	16	4
4	21	5
5	32	11
6	34	2
7	32	2
8	34	2
9	23	11
10	25	2
11	12	13
12	31	19
13	25	6
14	31	6
15	34	3
16	38	4
17	26	12
18	29	3
19	45	16
20	27	18

[a] Based on data from page 190 of Burr.[10]

The final column of Table 4.5 gives 19 so-called moving ranges of pairs of successive moisture contents. It is often argued that although these MR values are actually affected not only by variation within a 2-hr production period but by some variation between these periods as well, they come as close to representing purely short-term variation as any measure available from $n = 1$ data. Accordingly, as a kind of $n = 1$ analog of an *R* chart, moving ranges are often charted in addition to individual values *y*. Further, the average moving range is used to estimate σ in cases where information on the inherent variability of individuals is a priori lacking, according to the formula

$$\text{estimated } \sigma = \frac{\mathrm{MR}}{1.128}$$

where MR is the mean of the moving ranges (and plays the role of $\overline{R}$ in (4-10)), and 1.128 is the $n = 2$ version of the factor $d2$ alluded to immediately below (4-10).

In the case of the data of Table 4.5,

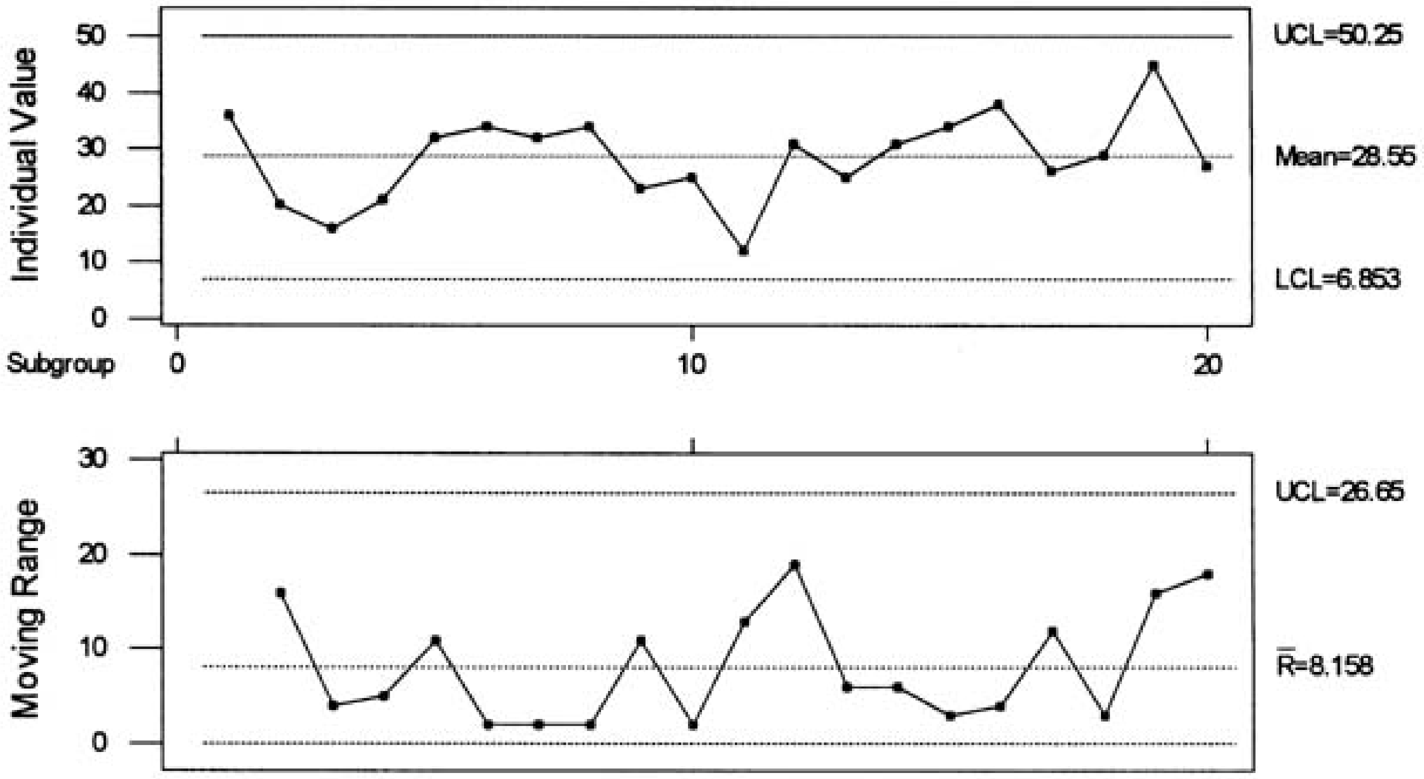
$$\mathrm{MR} = \frac{1}{19}\left(16 + 4 + 5 + \cdots + 16 + 18\right) \approx 8.2$$

so that a (possibly somewhat inflated due to between period variation) data-based estimate of within-period variability σ for use, for example in limits (4-9) is

$$\frac{8.2}{1.128} \approx 7.2$$

Figure 4.14 shows both an *x* (individuals) chart and an MR (moving range) chart based on these calculations. As no standard value of moisture content was provided in Burr's text,[10] the value $\bar{y} = 28.55$ was used as a substitute for μ in (4-9). The MR chart limits are based on standard $n = 2$ (because ranges of "groups" of two observations are being plotted) *R* chart control limit formulas. Figure 4.14 shows no evidence of assignable variation in the moisture contents.

Fig. 4.14 Control charts for *y* and MR based on moisture contents of 20 polymer samples.



Statistical research in the last decade has cast serious doubt on the wisdom of adding the MR chart to the *x* chart in *n* = 1 situations. The price paid for the addition in terms of "false alarm rate" is not really repaid with an important increase in the ability to detect process change. For a more complete discussion of this issue see Section 4.4 of Vardeman and Jobe.14

The use of Shewart control charts is admirably documented in a number of statistical quality control books including those by Vardeman and Jobe,14 Wadsworth et al.,9 Duncan,11 Burr,10 Grant and Leavenworth,12 and Ott et al.13 Our purpose here is not to provide all details necessary for their use, but only to give the reader an introduction to the overall function that they serve. It should be said, however, that in recent years other statistical process monitoring tools such as the so-called CUmulative SUM (CUSUM) schemes and Exponentially Weighted Moving Average (EWMA) schemes have been developed as competing methodologies, and can in some circumstances be practically more effective than the original Shewhart charts. Indeed, many computerized controllers for real-time chemical process monitoring and adjustment now employ some form of CUSUM or EWMA logic. For more on these topics, including their integration with model-based process controllers, the reader is referred to Sections 4.1 and 4.2 of Vardeman and Jobs14 and Vander Wiel et al.15

Shewhart's basic conceptualization of common and special cause variation not only leads to control charts as quantitative, rational tools to guide one in knowing when (and when not!) to intervene in an industrial process to correct potential ills, but it also provides a framework for considering the question of what is the best/most consistent performance one can hope for from a particular version of a process. That is, it provides a framework for discussing process capability assessment.

If $\hat{\sigma}$ is some (standard deviation type) estimate of the baseline variation inherent in an industrial process (obtained, e.g., from a calculation such as (4-10) or from data taken from the process after eliminating all physical sources of assignable variation), it essentially specifies what is possible in terms of consistency of process output. There are, however, several common ways of using such an estimate to produce related measures of process capability.

For one thing, remembering again the fact that an interval from μ - 3σ to μ + 3σ (i.e. of length 6σ) will bracket about 99.7 percent of a normal distribution, the figure 6σ is sometimes stated as "the process capability." This usage would say that in the context of the polyethylene melt index example of Table 4.4 the $\hat{\sigma} = 12 : 1$ figure from (4-10) implies a melt index process capability of approximately 6 · (12.1) ≈ 72.6. If properly monitored, the process appears capable of producing almost all individual melt indices in a 73-point range.

Where there are stated specifications for individual measurements *y*, σ is sometimes turned into a kind of index comparing it to the difference in upper and lower engineering specifications. For example, one such *process capability index* is

$$C_{\mathrm{p}} = \frac{\mathrm{USL} - \mathrm{LSL}}{6\sigma}$$

where USL - LSL is the difference in specifications. Fairly obviously, the larger the value of *C*p, the more comfortably (properly targeted) process output values will fit in an interval from LSL to USL.

Another process capability measure that is frequently used in the industrial chemistry sector is

$$C_{\mathrm{pk}} = \mathrm{minimum}\left\{ C_{\mathrm{pu}} = \frac{\mathrm{USL} - \mu}{3\sigma},\ C_{\mathrm{pl}} = \frac{\mu - \mathrm{LSL}}{3\sigma} \right\}$$

where μ is an overall process average for an in-control/stable/predictable process, and σ is as before. This measure is clearly similar to *C*p, but it takes into account the placement of the process mean in a way that is ignored by *C*p. A large value of *C*pk indicates that not only is the process short-term variation small enough for the process output values to potentially fit comfortably between LSL and USL, but that the process is currently so targeted that the potential is being realized.

4.4 Statistical Methods and Industrial Experimentation

One of the most important areas of opportunity for the new application of statistical methods in the chemical industry in the twenty-first century is that of increasing the effectiveness of industrial experimentation. That is, it is one thing to bring an existing industrial process to stability (a state of "statistical" control), but it is quite another to determine how to make fundamental changes in that process that will improve its basic behavior. This second activity almost always involves some form of experimentation, whether it be in the laboratory or in a plant. As we indicated in the introduction, efficient methods and strategies of such data collection (and corresponding analysis) are a central concern of applied statistics. In this section, we hope to give the reader some insight into the kinds of statistical tools that are available for use in chemical industry experimentation.

We will here take as our meaning of the term "experimentation" the observation of a (typically noisy) physical process under more than one condition, with the broad goal of understanding and then using knowledge of how the process reacts to the changes in conditions. In most industrial contexts, the "conditions" under which the process is observed can be specified in terms of the settings or so-called levels chosen for several potentially important process or environmental variables, the so-called factors in the experiment. In some cases, the hope is to identify those (often largely unregulated) factors and combinations of factors that seem to most influence an observed response variable, as a means of targeting them for attention intended to keep them constant or otherwise to eliminate their influence, and thereby to improve the consistency of the response. In other situations the hope is to discover patterns in how one or more critical responses depend on the levels of (often tightly controlled) factors, in order to provide a road map for the advantageous guiding of process behavior (e.g., to an increased mean reaction yield) through enlightened changing of those levels.

This section is organized into two subsections. In the first, we will illustrate the notion of variance component estimation through an example of a nested or hierarchical data collection scheme. In the second, we will discuss some general considerations in the planning of experiments to detail the pattern of influence of factors on responses, consider so-called factorial and fractional factorial experimental designs, illustrate response surface fitting and interpretation tools and the data requirements they imply, and, in the process, discuss the integration of a number of statistical tools in a sequential learning strategy.

Identifying Major Contributors to Process Variation

A statistical methodology that is particularly relevant where experimentation is meant to identify important unregulated sources of variation in a response is that of variance component estimation, based on so-called ANalysis Of VAriance (ANOVA) calculations and random effects models. As an example of what is possible, consider the data of Table 4.6 Shown here are copper content measurements for some bronze castings. Two copper content determinations were made on each of two physical specimens cut from each of 11 different castings.

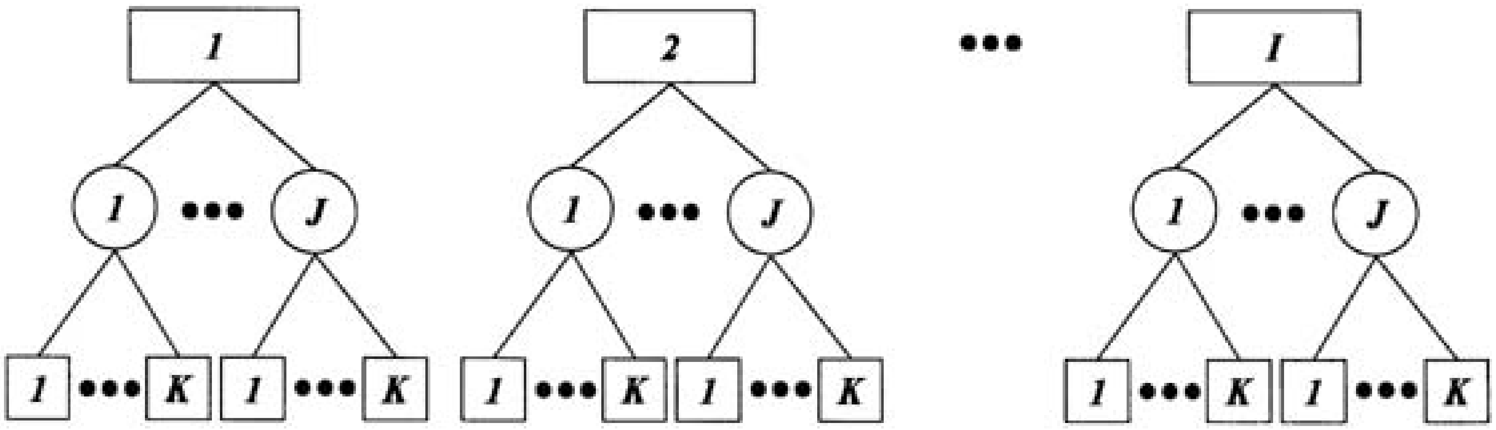
TABLE 4.6 Forty-four Copper Content Measurements from 11 Bronze Castings[a]

Casting	Specimen	Determination	Copper Content, y, (%)
1	1	1	85.54
1	1	2	85.56
1	2	1	85.51
1	2	2	85.54
2	1	1	85.54
2	1	2	85.60
2	2	1	85.25
2	2	2	85.25
3	1	1	85.72
3	1	2	85.77
3	2	1	84.94
3	2	2	84.95
4	1	1	85.48
4	1	2	85.50
4	2	1	84.98
4	2	2	85.02
5	1	1	85.54
5	1	2	85.57
5	2	1	85.84
5	2	2	85.84
6	1	1	85.72
6	1	2	85.86
6	2	1	85.81
6	2	2	85.91
7	1	1	85.72
7	1	2	85.76
7	2	1	85.81
7	2	2	85.84
8	1	1	86.12
8	1	2	86.12
8	2	1	86.12
8	2	2	86.20
9	1	1	85.47
9	1	2	85.49
9	2	1	85.75
9	2	2	85.77
10	1	1	84.98
10	1	2	85.10
10	2	1	85.90
10	2	2	85.90
11	1	1	85.12
11	1	2	85.17
11	2	1	85.18
11	2	2	85.24

[a] Based on data taken from Wernimont.

The data of Table 4.6 were *by design* collected to have a "tree type" or so-called hierarchical/nested structure. Figure 4.15 shows a diagram of a generic hierarchical structure for *balanced* cases like the present one, where there are equal numbers of branches leaving all nodes at a given level (there are equal numbers of determinations for each specimen and equal numbers of specimens for each casting).

Fig. 4.15 A balanced hierarchical data structure.



An important goal in most hierarchical studies is determining the size of the contributions to response variation provided by the different factors, that is, the different levels of the tree structure. (In the present context, the issue is how variation between castings compares to variation between specimens within a casting, and how they both compare to variation between determinations for a given specimen. If the overall variability observed were considered excessive, such analysis could then help guide efforts at variation reduction by identifying the largest contributors to observed variability.) The structure portrayed in Fig. 4.15 turns out to enable an appealing statistical analysis, providing help in that quantification.

If one lets

yijk = the copper content from the *k*th determination of the *j*th specimen from casting *i*

$\bar{y}_{ij\cdot} = \frac{1}{2}\sum_k y_{ijk}$ = the mean copper content determination from the *j*th specimen from casting *i*

$\bar{y}_{i\cdot\cdot} = \frac{1}{2}\sum_j \bar{y}_{ij\cdot}$ = the mean copper content determination from the *i*th casting

and

$\bar{y}_{\cdots} = \frac{1}{11}\sum_i y_{i\cdot\cdot}$ = the overall mean copper determination

it is possible to essentially break down the variance of all 44 copper contents (treated as a single group) into interpretable pieces, identifiable as variation between $\bar{y}_{i\cdot\cdot}$ (casting means), variation between $\bar{y}_{ijk\cdot}$ (specimen means) within castings, and variation between $\bar{y}_{ijk}$'s (individual measurements) within a specimen. That is, it is an algebraic identity that for 44 numbers *yijk* with the same structure as those in Table 4.6

$$\begin{aligned}(44-1)s^2 &= \sum_{i,j,k}(y_{ijk}-\bar{y}_{\cdots})^2 \\ &= \sum_{i,j,k}(y_{i\cdot\cdot}-\bar{y}_{\cdots})^2+\sum_{i,j,k}(\bar{y}_{ij\cdot}-\bar{y}_{\cdot\cdot})^2 \\ &\quad+\sum_{i,j,k}(y_{ijk}-\bar{y}_{ij\cdot})^2 \qquad (4\text{-}11)\end{aligned}$$

The sums indicated in (4-11) are over all data points; so, for example, the first summand on the right is obtained for the copper content data by summing each $(\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdots})^2$ a total of 2 · 2 = 4 times, one for each determination on a given casting. With the obvious meaning for the $\bar{y}$'s and the substitution of the total number of data values for 44, the identity in (4-11) applies to any balanced hierarchical data structure. It is a so-called ANOVA identity, providing an intuitively appealing partitioning of the overall observed variability in the data, an *analyzing of the (observed) variation*.

Some tedious arithmetic "by hand," or use of nearly any commercially available statistical package that includes an ANOVA program, shows that for the copper content data of Table 4.6 the numerical version of (4-11) is approximately

$$5.1385 = 3.2031 + 1.9003 + 0.0351 \qquad (4\text{-}12)$$

Although we will not provide any details here, the reader is alerted to the fact that it is common practice to present the elements of an identity such as (4-12) in a tabular form called an "ANOVA table." The use for the elements of (4-12) that we wish to illustrate here is their role in estimating casting, specimen, and determination "variance components."

That is, if one models an observed copper determination as the sum of a *random* casting-dependent *effect* whose distribution is described by a variance σsc, a *random* specimen-dependent *effect* whose distribution is described by a variance σ2s, and a *random* determination-dependent *effect* whose distribution is described by a variance σ2d, the elements of (4-12) lead to estimates of the *variance components* σ2c, σ2s, and σ2s in the model. Note that in such a random effects model of the data-generating process, copper measurements from the same casting share the same casting effect, and copper measurements from the same specimen share the both same casting and the same specimen effects. The individual σ2 values are conceptually the variances that would be seen in copper contents if only the corresponding sources of variation were present. The sum of the σ2 values is conceptually the variance that would be seen in copper contents if single determinations were made on a number of different castings.

Standard statistical methodology for estimation of the variance components (which we will not detail here but can, e.g. be found in Section 5.5 of Vardeman and Jobe14 or Chapter 11 of Hicks and Turner17) produces

$$\hat{\sigma}_{\mathrm{d}}^2 = \frac{0.0351}{11\cdot 2\cdot(2-1)} \approx 0.0016\,(\%)^2$$

as an estimate of σ2d,

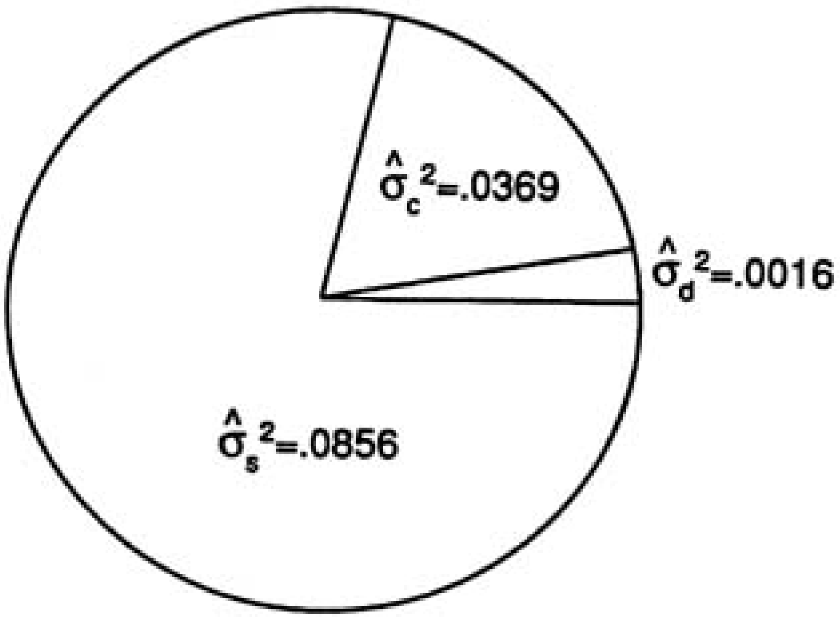
$$\hat{\sigma}_{\mathrm{s}}^2 = \frac{1}{2}\left(\frac{1.9003}{11\cdot(2-1)} - 0.0016\right) \approx 0.0856\,(\%)^2$$

as an estimate of σ2s, and

$$\begin{aligned}\hat{\sigma}_{\mathrm{c}}^2 &= \frac{1}{2\cdot 2}\left(\frac{3.2031}{(11-1)} - \frac{1.9003}{11\cdot(2-1)}\right) \\ &\approx 0.0369\,(\%)^2\end{aligned}$$

as an estimate of σ2c. Figure 4.16 is a pie chart representation of these three estimated variance components as fractions of their sum (the variance predicted if single determinations were made on single specimens from each casting), and graphically identifies inhomogeneity between specimens cut from a single casting as the biggest contributor to observed variation.

Fig. 4.16 Three estimated variance components for copper contents.



On the standard deviation scale the estimates translate to $\hat{\sigma}_{\mathrm{d}} \approx 0.04\%,\ \hat{\sigma}_{\mathrm{s}} \approx 0.29\%,\ \hat{\sigma}_{\mathrm{c}} \approx 0.19\%.$ So, for example, the data of Table 4.6 indicate that even if castings and specimens were all exactly alike, it would still be reasonable to expect measured copper contents to vary according to a standard deviation of about 0.04 percent, presumably due to unavoidable measurement error.

Variance component estimation methodology is not limited to balanced hierarchical experiments, but they do provide an important and straightforward context in which to introduce the technology. More detailed information on the case discussed here and extensions to other kinds of data structures can be found in books by Vardeman,6 Neter et al.,7 Mason, Gunst, and Hess,18 and Hicks and Turner.17

Discovering and Exploiting Patterns of Factor Influence on Responses

Having discussed statistical methodology particularly appropriate to studies whose primary purpose is simply to identify factors with the largest influence on a response, we will now consider methods aimed more directly at detailed experimental quantification of the pattern of factor influence on one or more responses. As an example, we will use a "sanitized" account of some statistical aspects of a highly successful and economically important process improvement project. (Data presented here are not the original data, but resemble them in structure. Naturally, details of the project not central to our expository purposes and those of a proprietary nature will be suppressed.) A more complete version of this case study appears as Chapter 11 of Vardeman.6

The process monitoring, capability assessment, and variance source identification ideas discussed thus far are almost logical prerequisites for industrial experimentation to detail the nature of dependence of response variables on factors of interest. When an industrial process has been made to operate in a stable manner, its intrinsic variability reduced to the extent practically possible, and that baseline performance quantified and understood, the prospects of success are greatly enhanced in subsequent efforts to understand the effects of potential fundamental process changes.

Preliminary work by various groups left a project team with a batch production process behaving in a stable but unsatisfactory fashion. Obvious sources of variation (both in the process itself and "upstream") had been identified and, to the degree practically possible, eliminated. The result was a process with an average output *purity* of 88 percent and an associated purity standard deviation of around 5 percent, and an average *yield* of 43 percent and an associated yield standard deviation of around 5 percent as well. The project team was charged with finding ways to increase the purity and yield means to, respectively, 95 percent and 59 percent while hopefully also further reducing the standard deviations. To accomplish this, the team recognized the need for an improved understanding of how various process variables under their control influenced purity (which we will call $y1$) and yield (which we will call $y2$).

Experimentation to provide this was authorized, and, in particular, attention was focused on four factors consisting of three reactant concentrations and the process run time. We will call the Reactant A mole ratio $x1$, the Reactant B mole ratio $x2$, the Reactant C mole ratio $x3$, and the run time (in hours) $x4$.

The choice of experimental factors (what to vary in data collection) is a nontrivial matter of fundamental importance that is best handled by people with firsthand process knowledge. There are a number of popular techniques and tools (such as so-called cause and effect diagrams, discussed for instance in Section 2.1 of Vardeman and Jobe14) for helping groups brainstorm and reach a consensus on such matters. Further, in cases where a priori knowledge of a process is scarce, relatively small preliminary screening experiments can help reduce a large list of potential factors to a smaller list apparently worthy of more detailed study. (The fractional factorial plans that will be illustrated shortly often are recommended for this purpose.)

Once a particular set of experimental factors has been identified, questions about exactly how they should be varied must be answered. To begin with, there is the choice of levels for the factors, the matter of how much the experimental factors should be varied. Particular experimental circumstances usually dictate how this is addressed. Widely spaced (substantially different) levels will in general lead to bigger changes in responses, and therefore clearer indications of how the responses depend upon the experimental factors, than will closely spaced (marginally different) levels. But they may do so at the expense of potentially creating unacceptable or even disastrous process conditions or output. Thus, what may be an acceptable strategy in a laboratory study might be completely unacceptable in a production environment and vice versa.

Given a set or range of levels for each of the individual experimental factors, there is still the question of exactly what combinations of levels actually will be used to produce experimental data. For example, in the process improvement study, standard process operating conditions were $x1 = 1.5$, $x2 = 1.15$, $x3 = 1.75$, and $x4 = 3.5$, and the project team decided on the ranges

$$1.0 \leqslant x_1 \leqslant 2.5,\ 1.0 \leqslant x_2 \leqslant 1.8,\ 1.0 \leqslant x_3 \leqslant 2.5,$$

$$\text{and} \qquad 2.0 \leqslant x_4 \leqslant 5.0 \tag{4-13}$$

as defining the initial limits of experimentation. But the question remained as to exactly what sets of mole ratios and corresponding run times were appropriate for data collection.

A natural, but largely discredited, strategy of data collection is the one-variable-at-a-time experimental strategy of picking some base of experimental operations (such as standard operating conditions) and varying the level of only one of the factors away from that base at a time. The problem with such a strategy is that sometimes two or more factors act on responses jointly, doing things in concert that neither will do alone. For example, in the process improvement study, it might well have been that an increase in either $x1$ or $x2$ alone would have affected yield very little, whereas a simultaneous increase in both would have caused an important increase. Modern strategies of industrial experimentation are conceived with such possibilities in mind, and attempt to spread out observations in a way that gives one some ability to identify the nature of the response structure no matter how simple or complicated it turns out to be.

There are several issues to consider when planning the combinations of levels to include in an experiment. We have already said that it is important to "vary several factors simultaneously." It also is important to provide for some replication of at least a combination or two in the experiment, as a means of getting a handle on the size of the experimental error or baseline variation that one is facing. The replication both verifies the reproducibility of values obtained in the study and identifies the limits of that reproducibility. Also, one must balance the urge to "cover the waterfront" with a wide variety of combinations of factor levels against resource constraints and a very real law of diminishing practical returns as one goes beyond what is really needed in the way of data to characterize response behavior. In addition, the fact that real-world learning is almost always of a sequential rather than a "one shot" nature suggests that it is in general wise to spend only part of an experimental budget on early study phases, leaving resources adequate to follow up directions suggested by what is learned in those stages.

It is obvious that a minimum of two different levels of an experimental factor must appear in a set of experimental combinations if any information is to be gained on the effects of that factor. So one logical place to begin thinking about a candidate design for an industrial experiment is with the set of all possible combinations of two levels of each of the experimental factors. If there are p experimental factors, statistical jargon for such an arrangement is to call it a (complete) $2 \times 2 \times 2 \times \ldots \times 2$ or $2p$ factorial plan. For example, in the process improvement situation, an experiment consisting of the running of all 16 possible combinations of

$$x_1 = 1.0 \quad \text{or} \quad x_1 = 2.5$$

$$x_2 = 1.0 \quad \text{or} \quad x_2 = 1.8$$

$$x_3 = 1.0 \quad \text{or} \quad x_3 = 2.5$$

and

$$x_4 = 2.0 \quad \text{or} \quad x_4 = 5.0$$

would be called a complete 2 × 2 × 2 × 2 or 24 factorial experiment. Notice that in geometric terms, the (*x*1,*x*2,*x*3,*x*4) points making up this 24 structure amount to the 16 "corners" in four-dimensional space of the initial experimental region defined in (4-13).

A complete factorial experimental plan is just that, in some sense "complete." It provides enough information to allow one to assess (for the particular levels used) not only individual but also joint or interaction effects of the factors on the response or responses. But when in fact (unbeknownst to the investigator) a system under study is a relatively simple one, principally driven by only a few individual or low-order joint effects of the factors, fewer data actually are needed to characterize those effects adequately. So what is often done in modern practice is initially to run only a carefully chosen part of a full 2*p* factorial, a so-called fractional factorial plan, and to decide based on the initial data whether data from the rest of the full factorial appear to be needed in order adequately to characterize and understand response behavior. We will not here discuss the details of how so-called 2*pq* fractional factorials are intelligently chosen, but there is accessible reading material on the subject in books by Box, Hunter, and Hunter,19 and by Vardeman and Jobe.5

In the process improvement study, what was actually done in the first stage of data collection was to gather information from one-half of a full 24 factorial (a 241 fractional factorial) augmented by four observations at the "center" of the experimental region (thereby providing both some coverage of the interior of the region, in addition to a view of some of its corners, and important replication as well). The data in Table 4.7 are representative of what the group obtained.

TABLE 4.7 Data from an Initial Phase of a Process Improvement Study

x1	x2	x3	x4	Purity, y1(%)	Yield, y2(%)
1.00	1.0	1.00	2.0	62.1	35.1
2.50	1.0	1.00	5.0	92.2	45.9
1.00	1.8	1.00	5.0	7.0	4.0
2.50	1.8	1.00	2.0	84.0	46.0
1.00	1.0	2.50	5.0	61.1	41.4
2.50	1.0	2.50	2.0	91.6	51.2
1.00	1.8	2.50	2.0	9.0	10.0
2.50	1.8	2.50	5.0	83.7	52.8
1.75	1.4	1.75	3.5	87.7	54.7
1.75	1.4	1.75	3.5	89.8	52.8
1.75	1.4	1.75	3.5	86.5	53.3
1.75	1.4	1.75	3.5	87.3	52.0

The order in which the data are listed is simply a convenient systematic one, not to be confused with the order in which experimental runs were actually made. The table order is far too regular for it to constitute a wise choice itself. For example, the fact that all *x*3 = 1.0 combinations precede the *x*3 = 2.5 ones might have the unfortunate effect of allowing the impact of unnoticed environmental changes over the study period to end up being confused with the impact of *x*3 changes. The order in which the 12 experimental runs were actually made was chosen in a "completely randomized" fashion. For a readable short discussion of the role of randomization in industrial experimentation, the reader is referred to Box.25

For purposes of this discussion, attention is focused on the yield response variable, *y*2. Notice first that the four *y*2 values from the center point of the experimental region have $\bar{y} = 53.2$ and *s* = 1.13 (which incidentally already appear to be an improvement over typical process behavior). As a partial indication of the logic that can be used to investigate whether the dependence of yield on the experimental factors is simple enough to be described adequately by the data of Table 4.7, one can compute some estimated "main effects" from the first eight data points. That is, considering first the impact of the variable *x*1 (alone) on yield, the quantity

$$\begin{aligned}\bar{y}_{\text{high } x_1} - \bar{y}_{\text{low } x_1} &= \frac{1}{4}(45.9 + 46.0 + 51.2 + 52.8) \\ &\quad -\frac{1}{4}(35.1 + 4.0 + 41.4 + 10.0) \\ &= 26.35\end{aligned}$$

is perhaps a sensible measure of how a change in *x*1 from 1.00 to 2.50 is reflected in yield. Similar measures for the other variables turn out to be

$$\bar{y}_{\text{high } x2} - \bar{y}_{\text{low } x2} = -15.20$$

$$\bar{y}_{\text{high } x3} - \bar{y}_{\text{low } x3} = 6.10$$

and

$$\bar{y}_{\text{high } x4} - \bar{y}_{\text{low } x4} = 0.45$$

These measures provide some crude insight into the directions and magnitudes of influence of the experimental variables on *y*2. (Clearly, by these measures *x*1 = 2.50 seems preferable to *x*1 = 1.00, and the run time variable *x*4 seems to have little impact on yield.) But they also provide strong evidence that the nature of the dependence of yield on the experimental factors is too complicated to be described by the action of the factors individually. For example, if it *were* the case that the separate actions of the experimental factors were adequate to describe system behavior, then standard statistical theory and the data indicate that the mean response for the *x*1 = 1.00, *x*2 = 1.0, *x*3 = 1.00, and *x*4 = 2.0 set of conditions would be around

$$\begin{aligned}\hat{y} &= \bar{y}_{\text{corners}} - \frac{1}{2}(-26.35) - \frac{1}{2}(-15.20) \\ &\quad -\frac{1}{2}(6.10) - \frac{1}{2}(0.45) = 27.45\end{aligned}$$

(where $\bar{y}_{\text{corners}}$ is standing for the mean of the first eight yields in Table 4.7). But the observed yield of 35.1 is clearly incompatible with such a mean and the standard deviation value (of *s* = 1.13) derived from the repeated center point. Also, other simple evidence that (at least linear and) separate action of the four factors is not enough to describe yield adequately is given by the large difference between $\bar{y}_{\text{corners}} = 35.8$ and the observed mean from the center point $\bar{y} = 53.2$. (As it turns out, calculations that we will not show here indicate the *possibility* that individual action of the factors *plus joint action* of the Reactant A and Reactant B mole ratios is sufficient to describe yield. But in any case, the point is that the data of Table 4.7 provide evidence that the pattern of dependence of yield on the experimental variables is not simple, and thus that completion of the 24 factorial is in order.)

After a complete analysis of the first round of experimental data, the project team "ran the second half fraction" of the 24 factorial, and data similar to those in Table 4.8 were obtained. (Again, no significance should be attached to the order in which the observations in Table 4.8 are listed. It is not the order in which the experimental runs were made.)

TABLE 4.8 Data from a Second Phase of a Process Improvement Study

x1	x2	x3	x4	Purity, y1(%)	Yield, y2(%)
1.00	1.0	1.00	5.0	64.0	35.3

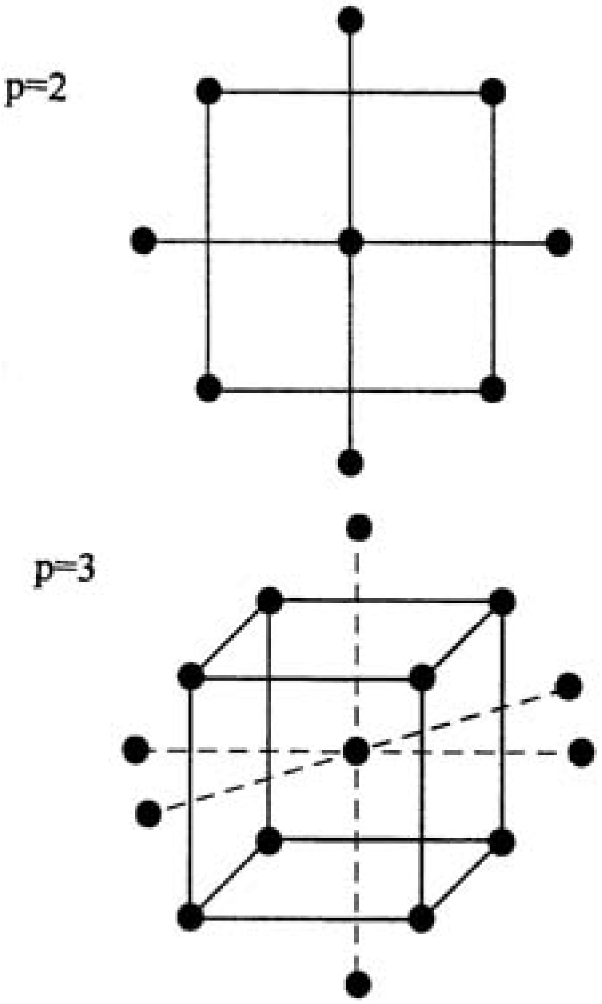
2.50	1.0	1.00	2.0	91.9	47.2
1.00	1.8	1.00	2.0	6.5	3.9
2.50	1.8	1.00	5.0	86.4	45.9
1.00	1.0	2.50	2.0	63.9	39.5
2.50	1.0	2.50	5.0	93.1	51.6
1.00	1.8	2.50	5.0	6.8	9.2
2.50	1.8	2.50	2.0	84.6	54.3

The data from the second phase of experimentation served to complete the project team's 24 factorial picture of yield and confirm the tentative understanding drawn first from the initial half fraction. It is seen that the combinations listed in Table 4.8 are in the same order as the first eight in Table 4.7 as regards levels of experimental variables *x*1, *x*2, and *x*3, and that the corresponding responses are very similar. (This, by the way, has the happy practical implication that run time seems to have little effect on final purity or yield, opening the possibility of reducing or at least not increasing the standard run time.) Thorough data analysis of a type not shown here left the project team with a clear (and quantified version of the) understanding that Reactant A and B mole ratios have important individual and joint effects on the responses, and that, acting independently of the other two reactants, Reactant C also has an important effect on the responses. However, it did *not* yet provide a solution to the team's basic problem, which was to reach a 59 percent mean yield goal.

The data of Tables 4.7 and 4.8 do hold out hope that conditions producing the desired purity and yield can be found. That is, though none of the 16 corners of the experimental region nor the center point appeared to meet the team's yield goal, the data do show that there is substantial *curvature* in the yield response. (The joint effect of *x*1 and *x*2 amounts to a kind of curvature, and the non-linearity of response indicated by a large difference between $\bar{y}_{\text{corners}} \approx 35.8$ and $\bar{y} = 53.2$ at the center of the experimental region also is a kind of curvature.) If one could "map" the nature of the curvature, there is at least the possibility of finding favorable future operating conditions in the interior of the initial experimental region defined in (4-13).

It ought to be at least plausible to the reader that 24 factorial data (even supplemented with center points) are not really sufficient to interpolate the nature of a curved response over the experimental region. More data are needed, and a standard way of augmenting a 2*p* design with center points to one sufficient to do the job is through the addition of so-called star points to produce a central composite design. Star points are points outside the original experimental region whose levels of all but one of the *p* experimental factors match those of the center point. Figure 4.17 shows graphical representations of central composite designs in *p* = 2 and *p* = 3 factors.

Fig. 4.17 *p* = 2 and *p* = 3 central composite designs.



The project team conducted a third phase of experimentation by adding eight star points to their study and obtained data similar to those in Table 4.9.

TABLE 4.9 Data from a Third Phase of a Process Improvement Study

x1	x2	x3	x4	Purity, y1(%)	Yield, y2(%)
0.6895	1.4	1.75	3.5	20.8	13.0
2.8105	1.4	1.75	3.5	95.9	54.3
1.75	0.8344	1.75	3.5	99.9	62.4
1.75	1.9656	1.75	3.5	65.9	41.2
1.75	1.4	0.6895	3.5	64.4	32.7
1.75	1.4	2.8105	3.5	64.8	40.3
1.75	1.4	1.75	1.379	88.1	52.7
1.75	1.4	1.75	5.621	88.9	50.5

The data in Tables 4.74.9 taken together turn out to provide enough information to enable one rather thoroughly to quantify the "curved" nature of the dependence of *y*2 on *x*1, *x*2, *x*3, and *x*4. A convenient and often successful method of accomplishing this quantification is through the least squares fitting of a general *quadratic response surface*. That is, central composite data are sufficient to allow one to fit an equation to a response that involves a constant term, linear terms in all the experimental variables, quadratic terms in all of the experimental variables, and cross-product terms in all pairs of the experimental variables. Appropriate use of a multiple regression program with the project data represented here produces the fitted equation

$$\begin{aligned} y_2 \approx\ & 15.4 + 37.9x_1 - 66.2x_2 + 48.8x_3 \\ & + 0.97x_4 - 16.1x_1^2 - 0.03x_2^2 \\ & - 13.6x_3^2 - 0.046x_4^2 + 26.5x_1x_2 \\ & + 0.344x_1x_3 - 0.217x_1x_4 \\ & + 1.31x_2x_3 - 0.365x_2x_4 + 0.061x_3x_4 \end{aligned}$$

This may not seem to the reader to be a particularly helpful data summary, but standard multiple regression tools can be used to deduce that an essentially equivalent, far less cluttered, and more clearly interpretable representation of the relationship is:

$$\begin{aligned} y_2 \quad \approx \quad & 13.8 + 37.8x_1 - 65.3x_2 + 51.6x_3 \\ & - 16.2x_1^2 - 13.6x_3^2 + 26.5x_1x_2 \end{aligned} \tag{4-14}$$

Equation (4-14), provides an admirable fit to the data in Tables 4.74.9, is in perfect agreement with all that has been said thus far about the pattern of dependence of yield on the experimental factors, *and* allows one to do some intelligent interpolation in the initial experimental region. Use of an equation like (4-14) ultimately allowed the project team to determine that an increase of $x1$ only would, with minimal change in the existing process, allow them to meet their yield goal. (In fact, the single change in $x1$ proved to be adequate to allow them to meet all of their yield *and* purity goals!)

Graphical representations similar to those in Figs 4.18 and 4.19 for (4-14) with $x3 = 1.75$ (the standard operating value for $x3$) were instrumental in helping the team understand the message carried by their data and how yield could be improved. Figure 4.18 is a so-called contour plot (essentially a topographic map) of the fitted equation, and Fig. 4.19 is a more three-dimensionallooking representation of the same surface. Both types of display are commonly used tools of modern statistical experiment design and analysis. The contour plot idea is particularly helpful where several responses are involved, and by overlaying several such plots one can simultaneously picture the various implications of a contemplated choice of process conditions.

Fig. 4.18 A contour plot of fitted yield when $x3 = 1.75$. (From *Statistics for Engineering Problem Solving (1st Ed.)* by S. B. Vardeman © 1994. Reprinted with permission of Brooks/Cole, an imprint of the Wadsworth Group, a Division of Thomson Learning. FAX 800-730-2215.)

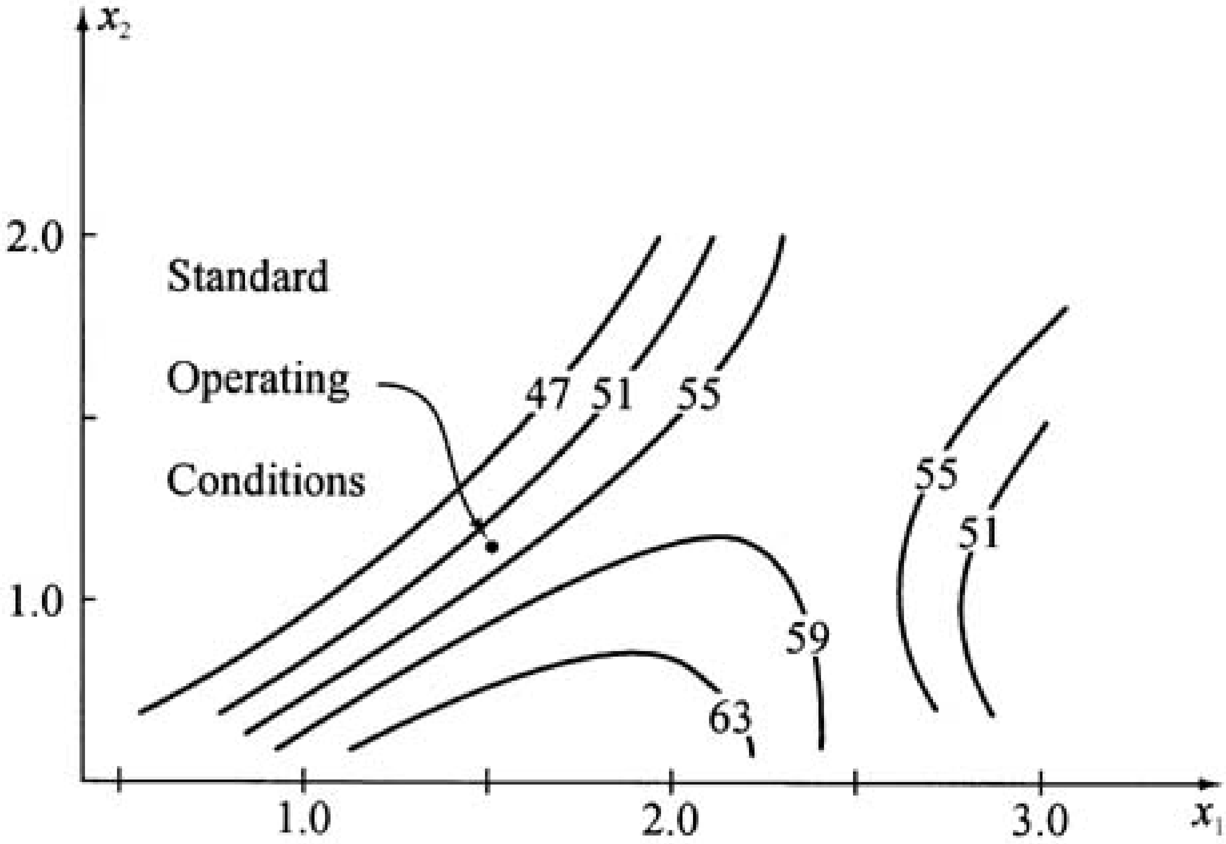
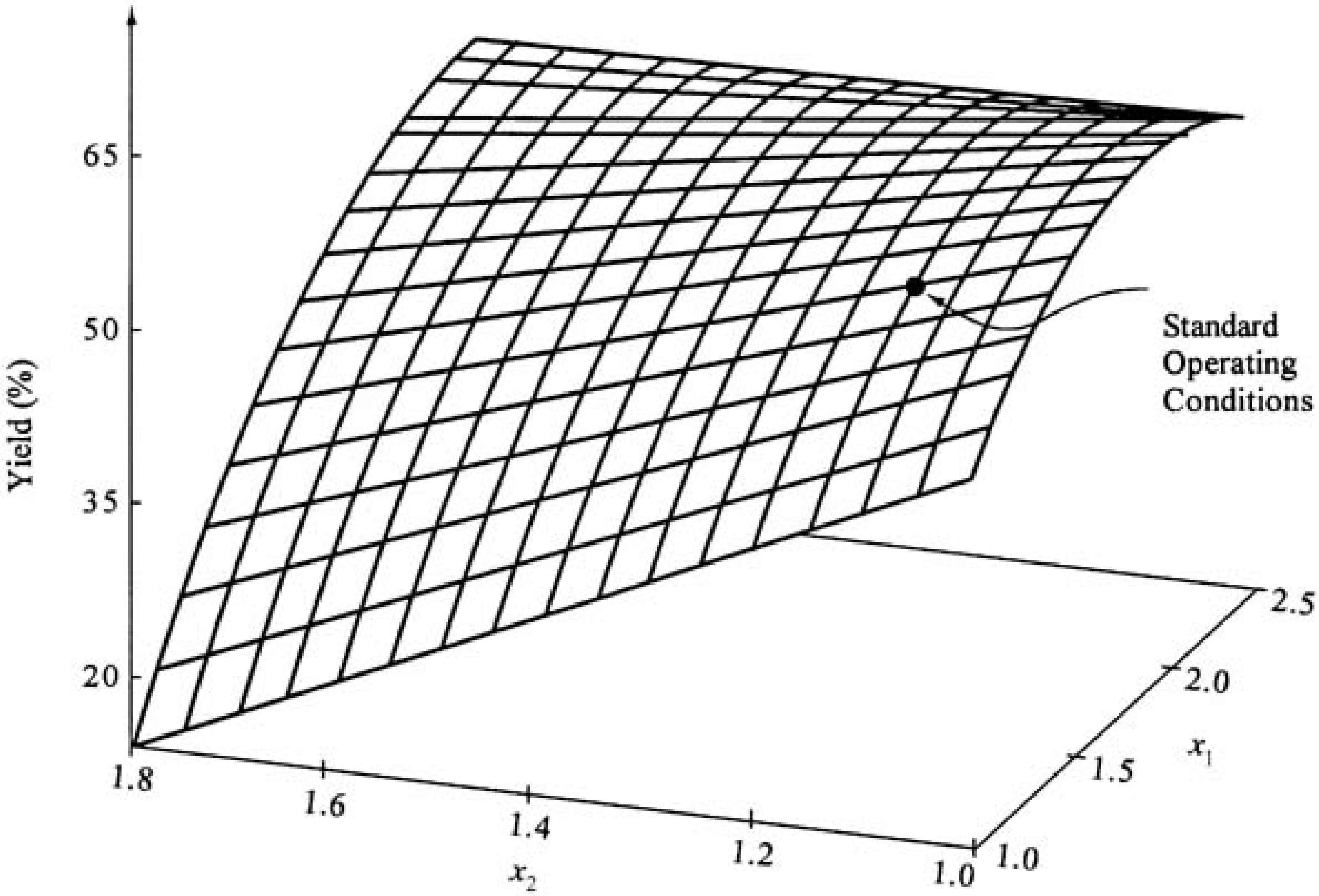


Fig. 4.19 A perspective graph of fitted yield when $x3 = 1.75$. (From *Statistics for Engineering Problem Solving* by S. B. Vardeman © 1994. Reprinted with permission of Brooks/Cole, an imprint of the Wadsworth Group, a Division of Thomson Learning. Fax 800-730-2215.)



4.5 Special Statistical Tools for Chemical Applications

The statistical methods discussed thus far are of a quite general nature, routinely finding application beyond the bounds of the chemical industry. In this section, we will briefly highlight two statistical methodologies whose most important applications are to chemical problems. That is, we will touch on some of the ideas of mixture experiments and the role of statistics in mechanistic modeling.

Mixture Experiments

In many situations in industrial chemistry, some important measured property of a product is a function of the proportions in which a set of *p* ingredients or components is represented in a mixture leading to the product. For example, case studies in the literature have covered subjects ranging from octanes of gasoline blends, discussed by Snee;21 to strengths of different formulations of ABS pipe compound, treated in Koons and Wilt;22 to aftertaste intensities of different blends of artificial sweeteners used in an athletic sport drink, discussed by Cornell;23 to moduli of elasticity of different rocket propellant formulations, considered by Kurotori.24 For experimenting in such contexts, special statistical techniques are needed. These tools have been discussed at length by Cornell,25,26 and our purpose here is not to attempt a complete exposition, but only to whet the reader's appetite for further reading in this area.

The goal of mixture experimentation is to quantify how proportions $x1$, $x2$, $x3$,…, xp of ingredients 1 through *p* affect a response *y*. Usually, the hope is to fit some kind of approximate equation involving some parameters $\underset{\sim}{b}$, say

$$y \approx f(x_1, x_2, \ldots, x_p | \underset{\sim}{b})$$

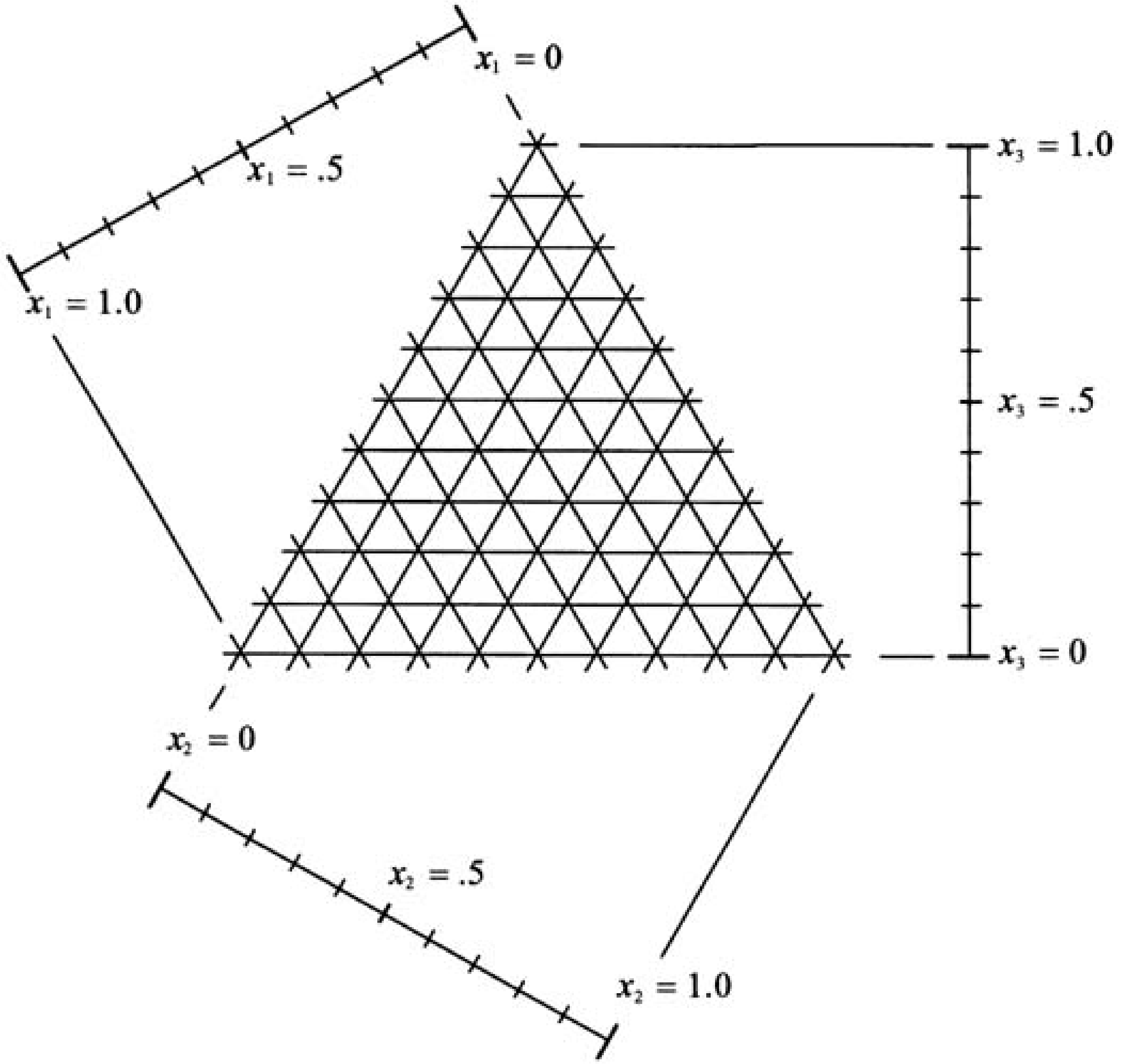
to a set of *n* data points ($x1$, $x2$,…, xp, y), for the purpose of using the fitted equation to guide optimization of *y*, that is, to find the "best" blend. The logic of data collection and equation fitting is complicated in the mixture scenario by the fact that

$$x_1 + x_2 + \cdots + x_p = 1 \qquad (4\text{-}15)$$

The linear constraint (4-15) means that (*p* way) factorial experimentation is impossible, and that special measures must be employed in order to use standard regression analysis software to do least squares equation fitting. We will briefly describe in turn some approaches to experimental design, equation fitting, and presentation of results for the mixture problem under its fundamental constraint (4-15).

In the case of $p = 3$ (a three-component mixture problem), the set of all possible combinations of values for $x1$, $x2$, and $x3$ satisfying (4-15) can be conveniently represented as an equilateral triangular region. Figure 4.20 shows such a region and the so-called simplex coordinate system on the region. The corners on the plot stand for cases where the "mixture" involved is actually a single pure component. Points on the line segments bounding the figure represent two-component mixtures, and interior points represent genuine three-component mixtures. For example, the center of the simplex corresponds to a set of conditions where each component makes up exactly one-third of the mixture.

Fig. 4.20 The set of points with $x1 + x2 + x3 = 1$ and a simplex coordinate system. (From *Statistics for Engineering Problem Solving* by S. B. Vardeman © 1994. Reprinted with permission of Brooks/Cole, an imprint of the Wadsworth Group, a Division of Thomson Learning. Fax 800-730-2215.)

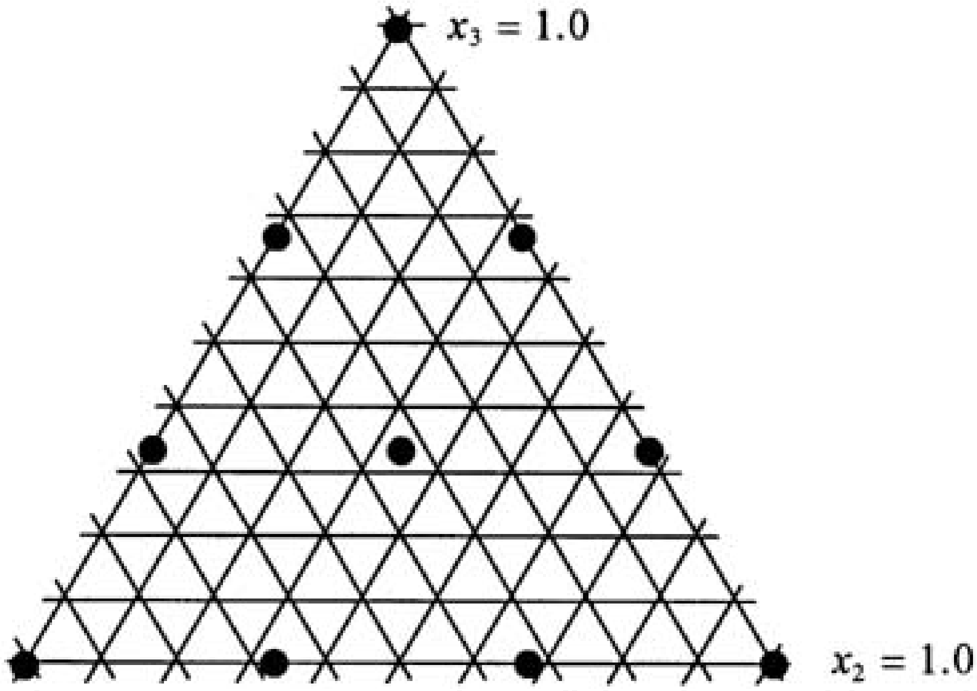


One standard mixture (experimental) design strategy is to collect data at the extremes (corners) of the experimental region along with collecting data on a regular grid in the experimental region. Figure 4.21 shows a $p = 3$ example of such a so-called simplex lattice design, and Table 4.10 lists the ($x1$,$x2$,$x3$) points involved. (As in the cases of the data in Tables 4.74.9, the order used in the listing in Table 4.10 is not one that would be used in sequencing data collection runs. Instead, a randomly chosen order often is employed.)

TABLE 4.10 ($x1$,$x2$,$x3$) Points in a Particular $p = 3$ Simplex Lattice Design

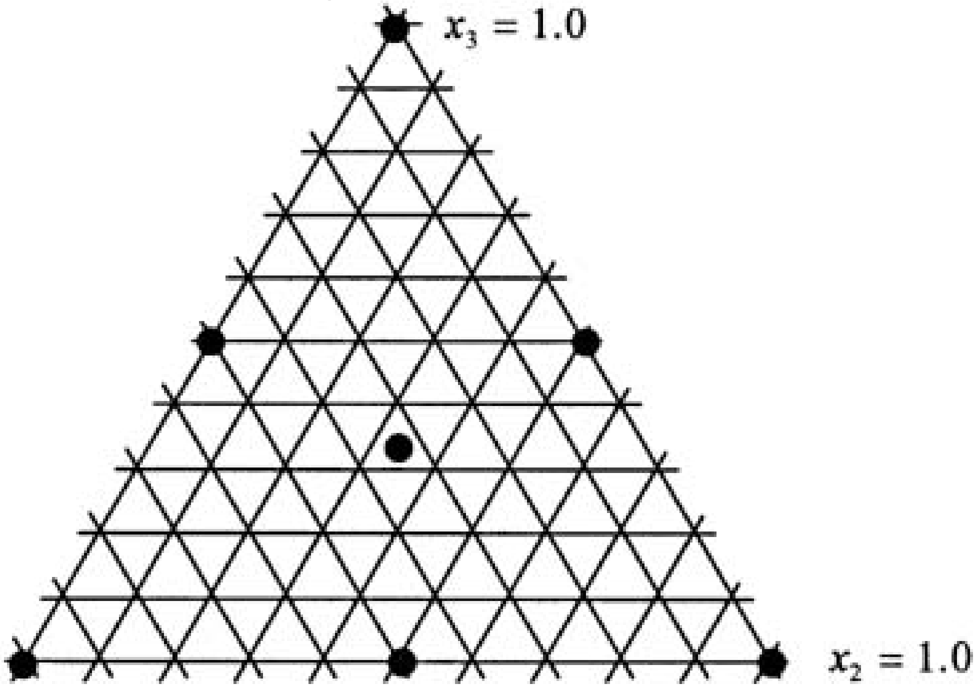
x_1	x_2	x_3
1	0	0
0	1	0
0	0	1
$\frac{1}{3}$	$\frac{2}{3}$	0
$\frac{2}{3}$	$\frac{1}{3}$	0
$\frac{1}{3}$	0	$\frac{2}{3}$
$\frac{2}{3}$	0	$\frac{1}{3}$
0	$\frac{1}{3}$	$\frac{2}{3}$
0	$\frac{2}{3}$	$\frac{1}{3}$
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Fig. 4.21 A $p = 3$ simple lattice design.



Another standard mixture experiment strategy is the so-called simplex centroid design, where data are collected at the extremes of the experimental region and for every equal-parts two-component mixture, every equal-parts three-component mixture, and so on. Figure 4.22 identifies the blends included in a $p = 3$ simplex centroid design.

Fig. 4.22 A $p = 3$ simplex centroid design.



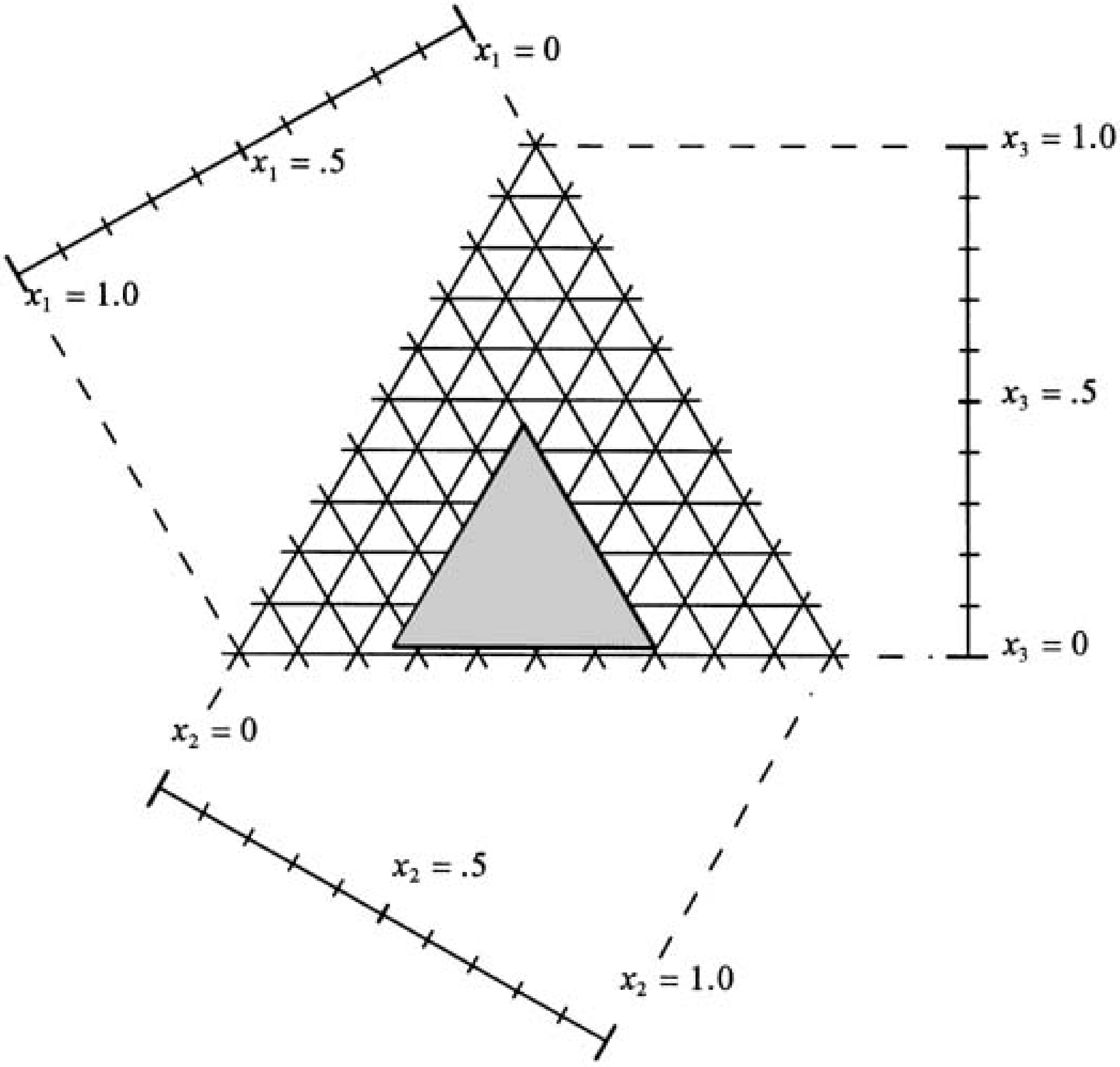
Often, the space of practically feasible mixtures is smaller than the entire set of $x_1, x_2, \ldots, x_p$ satisfying (4-15). For example, in many contexts, "pure" mixtures do not produce viable product. Concrete made using only water and no sand or cement obviously is a useless building product. One common type of constraint on the proportions $x_1, x_2, \ldots, x_p$ that produces quite simple experimental regions is that of lower bounds on one or more of the individual proportions. Cornell,25 for example, discusses a situation where the effectiveness in grease stain removal of a $p = 3$ bleach mixture was studied; and past experience with the product indicated that the proportions by weight of bromine, $x1$, of powder, $x2$, and of HCl, $x3$, needed to satisfy the constraints:

$$x_1 \geqslant 0.30, \ x_2 \geqslant 0.25, \quad \text{and} \quad x_3 \geqslant 0.02 \qquad (4\text{-}16)$$

for effective action of the product (i.e. the mixture needed to be at least 30% bromine, at least 25% powder, and at least 2% HCl by weight.)

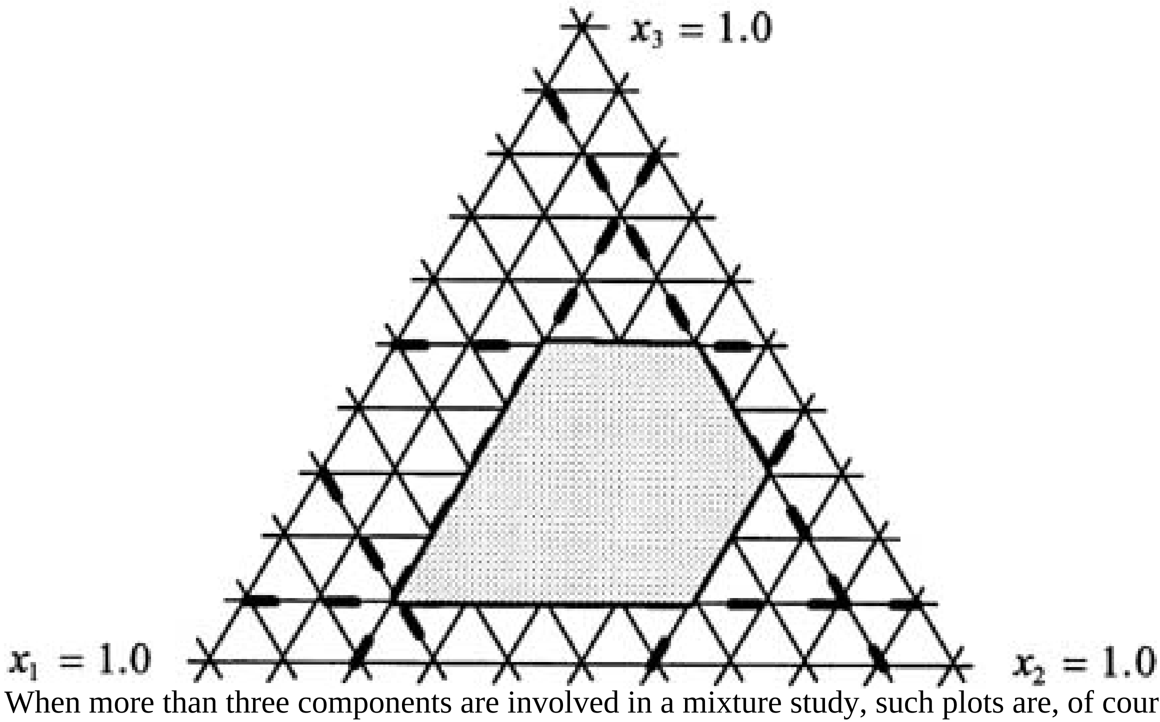
The effect of adding the lower bound constraints (4-16) to the basic mixture constraint (4-15) can be pictured as in Fig. 4.23. There, a triangular subregion of the basic $p = 3$ simplex depicts the feasible ($x1$, $x2$, $x3$) points. The choice of experimental mixtures for such an experimental region can be made by direct analogy to or rescaling of designs such as the simplex lattice and simplex centroid designs illustrated above to cover the entire simplex. (It is common to refer to the rescaling process as the use of pseudo-components.)

Fig. 4.23 The $p = 3$ simplex and a set of feasible bleach mixtures.



Constraint systems more complicated than simple lower bounds produce irregularly shaped experimental regions and less obvious methods of choosing ($x1$, $x2$,…, xp) points to cover the experimental region. When $p = 3$, it is possible to sketch the region of feasible points on a simplex plot and use it to help guide the choice of mixture experiment strategy. Figure 4.24 illustrates the kind of region that can arise with other than exclusively lower bound constraints.

Fig. 4.24 An irregularly shaped experimental region in a $p = 3$ mixture study.



When more than three components are involved in a mixture study, such plots are, of course, no longer possible, and other more analytic methods of identifying candidate experimental mixtures have been developed. For example, McLean and Anderson27 presented an algorithm for locating the vertices of an experimental region defined by the basic constraint (4-15) and any combination of upper and or lower bound constraints

$$0 \leqslant a_i \leqslant x_i \leqslant b_i \leqslant 1$$

on the proportions xi. Cornell25,26 discusses a variety of algorithms for choosing good mixture experiment designs under constraints, and many of the existing algorithms for the problem have been implemented in the MIXSOFT software package developed by Piepel.28

Empirical polynomial descriptions of (approximately) how a response y depends upon proportions $x1$, $x2$,…, xp are popular mixture analysis tools. The process of fitting polynomials to mixture experiment data in principle uses the same least squares notion illustrated in the fitting of a parabola to the data of Table 4.2. However, the mechanics of using standard multiple regression analysis software in the mixture context is complicated somewhat by the basic constraint (4-15). For example, in view of (4-15) the basic ($p = 1$ parameter) linear relationship

$$y \approx b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p \qquad (4\text{-}17)$$

is in some sense "overparameterized" in the mixture context, in that it is equivalent to the (p parameter) relationship

$$y \approx b_1x_1 + b_2x_2 + \cdots + b_px_p \qquad (4\text{-}18)$$

if one identifies the coefficients in (4-18) with the sums of the corresponding coefficients in (4-17) and the coefficient *b*0. As a result, it is the "no intercept" relationship (4-18) that is typically fit to mixture data when a linear relationship is used. In a similar way, when a second-order or (multivariable) quadratic relationship between the individual proportions and the response variable is used, it has no intercept term and no pure quadratic terms. For example, in the *p* = 3 component mixture case, the general quadratic relationship typically fit to mixture data is

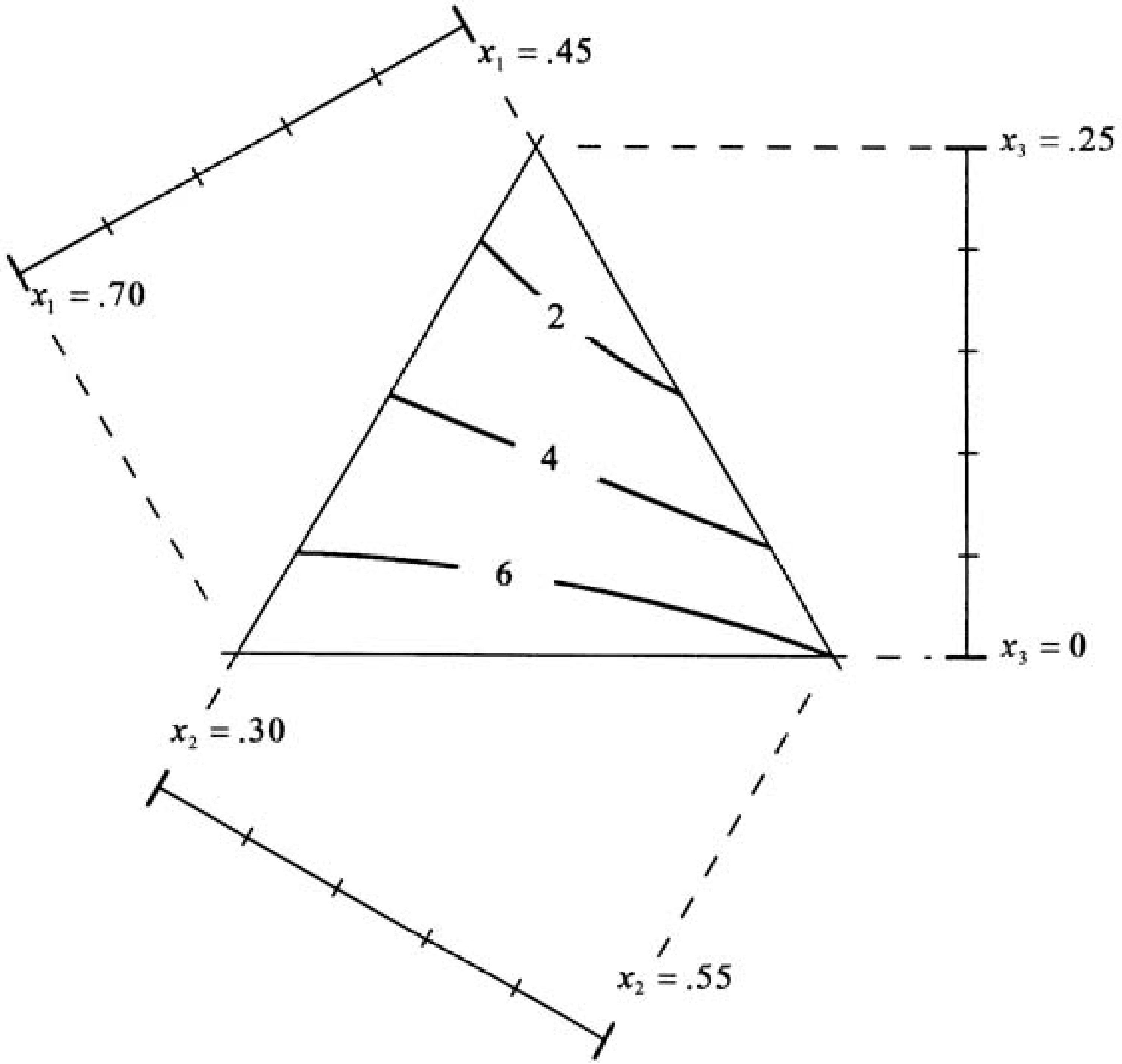
$$\begin{aligned} y \quad &\approx \quad b_1x_1 + b_2x_2 + b_3x_3 + b_4x_1x_2 \\ &\quad + b_5x_1x_3 + b_6x_2x_3 \end{aligned} \qquad (4\text{-}19)$$

(Any apparently more general relationship involving an intercept term and pure quadratic terms can by use of (4-15) be shown to be equivalent to (4-19) in the mixture context.) Relationships of the type of (4-19) are often called Scheffé models, after the first author to treat them in the statistical literature. Other more complicated equation forms are also useful in some applications, but we will not present them in this chapter. The interested reader is again referred to Cornell25,26 for more information on forms that have been found to be tractable and effective.

We should point out that the ability to fit equations of the form (4-18) or like (4-19), or of an even more complicated form, is predicated on having data from enough different mixtures to allow unambiguous identification of the parameters $\underline{b}$. This requires proper data collection strategy. Much of the existing statistical research on the topic of mixture experiment design has to do with the question of wise allocation of experimental resources under the assumption that a particular type of equation is to be fit.

One's understanding of fitted polynomial (and other) relationships often is enhanced through the use of contour plots made on coordinate systems like that in Fig. 4.25. (This is even true for *p* 3 component mixture scenarios, but the use of the idea is most transparent in the three-component case.) A plot like Fig. 4.25 can be a powerful tool to aid one in understanding the nature of a fitted equation and finding regions of optimum fitted response.

Fig. 4.25 A contour plot made on the *p* = 3 simplex. (From *Statistics for Engineering Problem Solving* by S. B. Vardeman © 1994. Reprinted with permission of Brooks/Cole, an imprint of the Wadsworth Group, a Division of Thomson Learning. Fax 800-730-2215.)



The mixture experiment counterpart to conventional screening/fractional factorial experimentation also is possible. So-called axial designs have been developed for the purpose of providing screening-type mixture data for use in rough evaluation of the relative effects of a large number of mixture components on a response variable. The same kind of sequential experimental strategy illustrated in the process improvement example is applicable in mixture contexts, as well as contexts free of a constraint such as (4-15).

Mechanistic Model Building

The kinds of equations most easily fit to multi-factor data using standard (least squares) regression analysis techniques are polynomial equations such as (4-6), (4-14), (4-18), and (4-19). These are particularly convenient because they are linear in their parameters, *b*. But they are probably best thought of as empirical "mathematical French curve" descriptions of the relation of a response, *y*, to the explanatory variables, *x*. Polynomial equations function as extremely useful summaries of observed patterns in one's data, but they do not typically provide direct insight into chemical mechanisms that produce those patterns, and the fitted parameters, *b*, do not often have direct physical meanings. Their use is particularly appropriate where there is little *a priori* knowledge of mechanisms involved in a process that might aid in its description, and/or no such knowledge is really essential to achieving one's goals.

Sometimes, however, it is desirable (on the basis of possible reaction kinetics or for other reasons) to posit theoretical descriptions of process outputs in terms of explanatory variables. That is, physicochemical principles often lead (through differential or integral equation descriptions of a system) to equation forms for a response that, like

$$y \quad = \quad \frac{K_1K_AK_BP_AP_B}{(1+K_AP_A+K_BP_B)^2}$$

$$y \quad = \quad C_0\exp(-Kt)$$

and

$$y = \frac{K_1 x}{1 + K_2 x}$$

are nonlinear in the parameters. Although such equations or models may be less tractable than empirical polynomial equations, the parameters involved more often than not *do have* direct physical interpretations. Further, when such a model can be verified as being an adequate description of a process (thereby confirming scientific understanding) and the parameters involved are estimated from process data, such mechanistic models can provide much safer extrapolations beyond an experimental region than the cruder empirical polynomial models.

The process of research in chemical systems is one of developing and testing different models for process behavior. Whether empirical or mechanistic models are involved, the discipline of statistics provides data-based tools for discrimination between competing possible models, parameter estimation, and model verification for use in this enterprise. In the case where empirical models are used, techniques associated with "linear" regression (linear least squares) are used, whereas in mechanistic modeling contexts "nonlinear" regression (nonlinear least squares) techniques most often are needed. In either case, the statistical tools are applied most fruitfully in iterative strategies.

Reilly and Blau29 Chapter 16 of Box et al.19 provide introductions to the general philosophy of using statistical methods in mechanistic modeling contexts, as well as a number of useful references for further reading.

Fairly sophisticated and specialized statistical software is needed in the practical application of nonlinear regression methods to mechanistic modeling for industrial chemistry applications. The techniques implemented in such software are discussed in Seber and Wild,32 Bates and Watts,30 Bard,31 and Riley and Blau.29

4.6 Modern Business Process Improvement and the Discipline of Statistics

The modern global business environment is fiercely competitive in all sectors, including the chemical sector. It is by now widely recognized that corporate survival in this environment depends upon constant improvement in *all* business processes, from billing to production. Companies have adopted a variety of programs and focuses aimed at facilitating that improvement. A decade ago, efforts organized around a *Total Quality Management* banner (with liberal references to emphases of consultants like W. E. Deming, J. M. Juran, and A. Feigenbaum) were popular. More recently, programs keyed to ISO 9000[33] certification criteria and Malcolm Baldridge Award[34] criteria have become prominent. And currently probably the most visible programs are the so-called *Six Sigma* programs.

In one sense there is nothing new under the sun, and all successful business process improvement programs (including those in the chemical sector) must in the end reduce to organized problem-solving disciplines. So it is not surprising that programs quite different in name are often very alike in fundamental content. And as they must necessarily make use of empirical information (data), they must have significant statistical components. To make this connection to statistics slightly more explicit, we proceed to provide a few additional details on the Six Sigma movement. (Further material on the subject is easy to find using an internet search engine, as there are many consultants eager to sell their advice and Six Sigma training. The American Society for Quality at www.asq.org offers many entries into the subject. And a search at amazon.com for "Six Sigma" books already produced 51 hits in March 2002. Fashions change quickly enough that it seems pointless to provide more detailed recommendations for follow up on the subject.)

The phrase "Six Sigma" originated at Motorola in the late 1980s. Programs there and at General Electric in the mid 1990s are widely touted as important contributors to the company profits and growth in stock values. The name is now commonly used in at least three different ways. "Six Sigma" refers to

1. a goal for business process performance
2. a discipline for improvement to achieve that performance
3. a corporate program of organization, training, and recognition conceived to support the process improvement discipline

As a goal for business process improvement, "Six Sigma" is equivalent to "$C_{pk} = 2$." What is perhaps confusing to the uninitiated is that this goal has connections (through normal distribution tail area calculations) to small ("parts per million") fractions defective relative to two-sided specifications on y. Six Sigma proponents often move between the "small process variation" and "parts per million" understandings with little warning.

Six Sigma process improvement disciplines are typically organized around the acronym "MAIC." The first step in a MAIC cycle is a Measure step, wherein one finds appropriate process responses to observe, identifies and validates measurement systems and collects baseline process performance (process monitoring) data. The second step is an Analyze step. This involves summarizing the initial process data and drawing appropriate inferences about current process performance. The third step in a MAIC cycle is an Improve step, where process knowledge, experimentation, and more data analysis are employed to find a way to make things work better. Finally, the 4-step cycle culminates in a Control (process monitoring) effort. The object here is to see that the newly improved performance is maintained after a project team moves on to other problems.

Six Sigma corporate organization, training and recognition programs borrow from the jargon and culture of the martial arts. People expert in the process improvement paradigm are designated "black belts," "master black belts," and so on. These people lead project teams and help train new initiates ("green belts") in the basics of the program and its tools (including statistical tools). The emphasis throughout is on completing projects with verifiable large dollar impact.

Having made the point that improvement in all business activities is of necessity data-driven, it is hopefully obvious that the emphases and methods of the subject of statistics are useful beyond the lab and even production. Of course, for broad implementation, it is the most elementary of statistical methods that are relevant.

4.7 Conclusion

We have tried in this chapter to give readers the flavor of modern applied statistical methods and to illustrate their usefulness in the chemical industry. Details of their implementation have of necessity been reserved for further more specialized reading, for which the interested reader is encouraged to consult the references given in this chapter.

4.8 References

Albin, S., "The Lognormal Distribution for Modeling Quality Data When the Mean is Near Zero," J. Qual. Technol., 22, 105110 (1990).

Chambers, J., Cleveland, W., Kleiner, B., and Tukey, P., Graphical Methods of Data Analysis, Duxbury, Boston, MA, 1983.

Odioso, R., Henke, A., Stauffer, H., and Frech, K., "Direct Hydration of Olefins,", Ind. Eng. Chem., 53(3), 209211 (1961).

Devore, J., Probability and Statistics for Engineering and the Sciences (3rd ed.), Brooks/Cole, Pacific Grove, CA, 1991.

Vardeman, S. B., and Jobe, J. M., Basic Engineering Data Collection and Analysis, Duxbury/Thomson Learning, Pacific Grove, CA, 2001.

Vardeman, S. B., Statistics for Engineering Problem Solving, PWS-Kent, Boston, MA, 1994.

Neter, J., Kutner, M., Nachsteim, C., and Wasserman, W., Applied Linear Statistical Models, (4th ed.), McGraw-Hill, New York, 1996.

Hahn, G., and Meeker, W., Statistical Intervals: A Guide for Practitioners, Wiley, New York, 1991.

Wadsworth, H., Stephens, K., and Godfrey, B., Modern Statistical Methods for Quality Control and Improvement, Wiley, New York, 1986.

Burr, I., Statistical Quality Control Methods, Dekker, New York, 1976.

Duncan, A., Quality Control and Industrial Statistics, (5th ed.), Irwin, Homewood, IL, 1986.

Grant, E., and Leavenworth, R., Statistical Quality Control, (7th ed.), McGraw-Hill, New York, 1996.

Ott, E., and Schilling, E., Process Quality Control, McGraw-Hill, NY, 1990.

Vardeman, S., and Jobe, J. M., Statistical Quality Assurance Methods for Engineers, Wiley, New York, 1999.

Vander Wiel, S., Tucker, W., Faltin, F., and Doganaksoy, N., "Algorithmic Statistical Process Control: Concepts and an Application,", Technometrics, 34(3), 286297 (1992).

Wernimont, G., "Statistical Quality Control in the Chemical Laboratory,", Qual. Eng., 2, 5972 (1989).

Hicks, C., and Turner, K., Fundamental Concepts in the Design of Experiments, (5th ed.), Oxford University Press, New York, 1999.

Mason, R., Gunst, R., and Hess, J., Statistical Design and Analysis of Experiments, Wiley, New York, 1989.

Box, G., Hunter, W., and Hunter, J. S., Statistics for Experimenters, Wiley, New York, 1978.

Box, G., "George's Column,", Qual. Eng., 2, 497502 (1990).

Snee, R., "Developing Blending Models for Gasoline and Other Mixtures,", Technometrics, 23, 119130 (1981).

Koons, G., and Wilt, M., "Design and Analysis of an ABS Pipe Compound Experiment," in Experiments in Industry: Design, Analysis and Interpretation of Results, R. Snee, L. Hare, and J. R. Trout (eds.), pp. 111117, ASQC Quality Press, Milwaukee, WI, 1985.

Cornell, J., "A Comparison between Two Ten-Point Designs for Studying Three-Component Mixture Systems,", J. Qual. Technol., 18, 115 (1986).

Kurotori, I., "Experiments with Mixtures of Components Having Lower Bounds,", Ind. Qual. Control, 22, 592596 (1966).

Cornell, J., How to Run Mixture Experiments for Product Equality, Vol. 5 in the ASCQ "Basic References in Quality Control" series, American Society for Quality Control, Milwaukee, WI, 1983.

Cornell, J., Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data (2nd ed.), Wiley, New York, 1990.

MacLean, R., and Anderson, V., "Extreme Vertices Design of Mixture Experiments,", Technometrics, 8, 447454 (1966).

Piepel, G., Mixsoft and Misoft User's Guide, Version 1.0, Mixsoft-Mixture Experiment Software, Richland, WA, 1989.

Reilly, P., and Blau, G., "The Use of Statistical Methods to Build Mathematical Models of Chemical Reacting Systems,", Can. J. Chem. Eng., 52, 289299 (1974).

Bates, D., and Watts, D., Nonlinear Regression Analysis and Its Applications, Wiley, New York, 1988.

Bard, Yonathan, Nonlinear Parameter Estimation, Academic Press, New York, 1974.

Seber, G., and Wild, C., Nonlinear Regression, Wiley, New York, 1989.

International Organization for Standardization (www.iso.ch).

National Institute of Standards and Technology (www.quality.nist.gov).

5.1 Introduction

There is an increased emphasis on chemical process safety as a result of highly publicized accidents. Public awareness of these accidents has provided a driving force for industry to improve its safety record. There has been an increasing amount of government regulation.

The chemical industry is one of the safest industries, but its safety record in the eyes of the public has suffered. Perhaps this is because sometimes when there is an accident in a chemical plant it is spectacular and receives a great deal of attention. The public often associates the chemical industry with environmental and safety problems, which results in a negative image of the industry.

Some of the important changes in this chapter since the ninth edition was published involve the area of governmental regulations. These are discussed in the section titled "Regulations." A great deal of information on regulations is available on the Internet and will be discussed briefly in this chapter.

The Internet also provides considerable information on incidents, good industry practice, and design guidelines. The best practices in industry are briefly discussed in this chapter. Details are readily available from resources listed in the references section at the end of the chapter.

Hazards from combustion and runaway reactions play a leading role in many chemical process accidents. Knowledge of these reactions is essential for the control of process hazards. Much of the damage and loss of life in chemical accidents results from a sudden release of material at high pressures, which may or may not result in fire; so it is important to understand how sudden pressure releases can occur. They can be due, for example, to ruptured high pressure tanks, runaway reactions, flammable vapor clouds, or pressure developed from external fire. Fires can cause severe damage from thermal radiation. Chemical releases from fires and pressure releases can form toxic clouds that can be dangerous to people over large areas. Static electricity often is a hidden cause in accidents. It is important to understand the reactive nature of the chemicals involved in a chemical facility.

Chemical process safety involves both the technical and the management aspects of the chemical industry, and this chapter addresses both. It is not enough to be aware of how to predict the effect of process hazards and how to design systems to reduce the risks of these hazards. It also is important to consider how chemical process safety can be managed. Technical people at all levels have a degree of management responsibility, and can contribute to the overall management of safe plants.

Loss of containment due to mechanical failure or misoperation is a major cause of chemical process accidents. The publication *One Hundred Largest Losses: A Thirty Year Review of Property Damage Losses in the Hydrocarbon-Chemical Industry*1 cites loss of containment as the leading cause of property loss in the chemical process industries.

5.2 Inherently Safer Plants*

The design of chemical plants to be more nearly "inherently safe" has received a great deal of attention in recent years. This is due in part to the worldwide attention to safety issues in the chemical industry brought on by the gas release at the Union Carbide plant in Bhopal, India, in December 1984. This and the fairly frequent occurrence of other chemical plant incidents has raised the issue of chemical plant safety to a very high level of visibility and concern. The major factors that should be considered in the planning, design, and operation of chemical plants are described below. The reader is referred to the list of recommended supplementary reading at the end of the chapter for more detailed information.

*Portions of this section have been reproduced from "Opportunities in the Design of Inherently Safer Chemical Plants," by Stanley M. Englund in Advances in Chemical Engineering Vol. 15, edited by Bodansky, O. and Latner, A., copyright (1990). Elsevier Science (USA), with permission from the publisher.31

The term, "inherent," means "belonging by nature, or the essential character of something." An "inherently safe" plant is safe by its nature and by the way it is constituted.

No facilities can be completely "inherently safe," but they can be made "inherently safer" by careful examination of all aspects of plant design and management. It is possible to identify "inherently unsafe" equipment and management practices and try to avoid them. It has been found that the basic or root cause of most chemical process accidents can be traced to some failure of a management system. Human error plays an important role in many process accidents and is often closely related to problems with management systems. Major efforts are under way to address the potential problems of human error. By being aware of modern advances in design and management techniques, and putting these techniques into practice, it is possible to have facilities that are safer than ever.

The term "intrinsic" has a meaning similar to that of "inherent," but the common usage of "intrinsic" in the chemical industry usually means a protection technique related to electricity. Intrinsic safety is based on the restriction of electrical energy to a level below which sparking or heating effects cannot ignite an explosive atmosphere (Lees,2 p. 513).

Responsibility for Safety in Design and Operation

It is very important that responsibility for the safe design and operation of a plant be clearly defined in the early stages. This means that competent and experienced people should be made responsible and held accountable for decisions made from the start of plant design on through plant start-up and operation.

Review of Design Alternatives

Hazards should be considered and eliminated in the process development stage where possible. This would include considerations of alternative processes, reduction or elimination of hazardous chemicals, site selection, etc. By the time the process is developed, the process designers already have major constraints imposed on them.

Hazards should also be identified and removed or reduced early in the design. Adding protective equipment at the end of the design or after the plant is operating can be expensive and not entirely satisfactory.

Allowing time in the early stages of design for critical reviews and evaluation of alternatives would involve studies such as an early hazard and operability (HAZOP) study, using flowsheets, before final design begins.4 Fault tree analysis," quantitative risk assessment (QRA), checklists, audits, and other review and checking techniques can also be very helpful. These techniques are extensively discussed in the technical literature and will not be discussed in detail here.

Emergency Planning

Emergency planning is primarily for the protection of plant personnel and people in nearby areas and the environment that could be affected by plant problems. It should be considered early in the design and should be coordinated with the existing site emergency plan.

Emergency planning includes tornado and storm shelters, flood protection, earthquakes, proximity to public areas, and safe exit routes. It also includes planning for the effect that an emergency in the "new process" would have on other plants, and the effect that an emergency in another plant would have on the new process. The effects of potential spills on waterways and aquifers should be considered.

Emergency response planning is discussed fully in Chapter 6.

Placement of Process and Storage Areas

The Bhopal plant of Union Carbide was built originally 1.5 miles from the nearest housing (see below under "Case Histories"). Over time, a residential area grew up next to the plant. This demonstrates the need to prevent hazardous plants from being located close to residential areas and to prevent residential areas from being established near such plants.4 If possible, the cost of a plant should include an adequate buffer zone unless other means can be provided to ensure that the public will not build adjacent to the plant. The nature and size of this buffer zone depend on many factors, including the amount and type of chemicals stored and used.

Storage of Hazardous Materials

The best way to minimize releases of a hazardous or flammable material is to have less of it around. In the Flixborough disaster2 on June 1, 1974, the process involved the oxidation of cyclohexane to cyclohexanone by air (with added nitrogen) in the presence of a catalyst (see below under "Case Histories"). The cyclohexanone was converted to caprolactam, which is the basic raw material for Nylon 6. The reaction product from the final reactor contained approximately 94 percent unreacted cyclohexane at 155°C and over 20 psig. The holdup in the reactors was about 240,000 lb, of which about 80,000 lb escaped. It is estimated that about 20,00060,000 lb actually was involved in the explosion. The resulting large unconfined vapor cloud explosion (or explosionsthere may have been two) and fire killed 28 people and injured 36 at the plant and many more in the surrounding area, demolished a large chemical plant, and damaged 1,821 houses and 167 shops. The very large amount of flammable liquid well above its boiling and flash points contributed greatly to the extreme severity of the disaster (Lees,2 p. 863).

The results of the Flixborough investigation make it clear that the large inventory of flammable material in the process plant contributed to the scale of the disaster. It is concluded that "limitations of inventory (of flammable materials) should be taken as specific design objectives in major hazard installations." It should be noted that reduction of inventory may require more frequent and smaller shipments, and this would entail more chances for errors in connecting and reconnecting. These possibly "negative" benefits should also be analyzed. Quantitative risk analysis of storage facilities has revealed solutions that may run counter to intuition.5 For example, contrary to popular opinion, reducing inventories in tanks of hazardous materials does little to reduce risk in situations where most of the exposure arises from the number and extent of valves, nozzles, and lines connecting the tank. Removing tanks from the service altogether, on the other hand, helps.

A large tank may offer greater safety than several small tanks of the same aggregate capacity because there are fewer associated nozzles and lines. Also, a large tank is inherently more robust, or can economically be made more robust by deliberate overdesign, than is a small tank and therefore is more resistant to external damage. On the other hand, if the larger tank has larger connecting lines, the relative risk may be greater if release rates through the larger lines increase the risk more than the tank's inherently greater strength reduces it. In the transportation of hazardous materials, the benefits of head shields and shelf couplers have been shown through quantitative risk assessment. Maintaining tank car integrity in a derailment is often the most important line of defense in the transportation of hazardous materials.

Liquefied Gas Storage

Usually, leaks of liquefied gases are much less serious if such gases are stored refrigerated at low temperatures and pressures than if they are stored at ambient temperatures under pressure. A leak of a volatile liquid held at atmospheric temperature and pressure results in only a relatively slow evaporation of the liquid. Escape of a refrigerated liquefied gas at atmospheric pressure gives some initial flashoff, and then it evaporates at a rate that is relatively slow but faster than the first case, depending on weather conditions.

Loss of containment of a liquefied gas under pressure and at atmospheric temperature, however, causes immediate flashing of a large proportion of the gas, followed by slower evaporation of the residue. This is usually a more serious case than if a refrigerated tank is used. The hazard from a gas under pressure is normally much less in terms of the amount of material stored, but the physical energy released if a confined explosion occurs at high pressure is large.

Use of Open Structures

There are many examples of serious fires and explosions that probably resulted in part from handling moderate to large quantities of flammable or combustible liquids and liquefied flammable gases inside enclosed structures. If a sufficient quantity of flammable mixture should ignite inside an ordinary chemical processing building, it is highly probable an explosion will occur that will seriously damage the building. For this reason, processing equipment is often installed in a structure without walls, usually called an "open structure." This permits effective ventilation by normal wind currents and aids the dispersion of any vapors that do escape. If ignition of gas occurs within the structure, the absence of walls minimizes the pressure developed from the combustion and the probability of flying shrapnel from a shattered structure.6

Substantial explosion damage will be done to a building by combustion of a surprisingly small quantity of a flammable gasair mixture. If there is an explosion of a flammable gasair mixture in a building where the flammable gas mixture occupies a space equal to only 1 or 2 percent of the building volume, the building may be seriously damaged if it does not have adequate explosion venting. This is because most buildings will suffer substantial structural damage from an internal pressure appreciably less than 1 psi (0.07 bar). Thus, a building does not need to be "full" or even close to "full" of a flammable mixture for a building explosion to occur that can cause considerable damage.

In 1950, a serious hydrocarbon explosion occurred in an enclosed chemical processing unit of The Dow Chemical Company. It was instrumental in causing Dow to establish a policy of using open structures for chemical processes that use substantial quantities of flammable liquids and liquefied flammable gases, and combustible liquids above their flash points.

Need to Understand Reactive Chemicals Systems

The main business of most chemical companies is to manufacture products through the control of reactive chemicals. The reactivity that makes chemicals useful can also make them hazardous. Therefore, it is essential that a process designer understand the nature of the reactive chemicals involved in his or her process.

Usually reactions are carried out without mishaps, but sometimes chemical reactions get out of control for many reasons, including:

1. Wrong raw material used.
2. Operating conditions changed.
3. Time delays unanticipated.
4. Equipment failure.
5. Reactants contaminated.
6. Materials of construction wrong.
7. Misoperation.
8. External fire.

Such mishaps can be worse if the chemistry is not fully understood. A chemical plant can be "inherently safer" only if knowledge of the reactive chemicals systems is carefully used in its design.

Reactive Hazard Evaluations

Reactive hazard evaluations should be made on all new processes, and on all existing processes on a periodic basis. There is no substitute for experience, good judgment, and good data in evaluating potential hazards. Reviews should include:

1. Process chemistry:
1. reactions
2. potential pressure buildup
3. intermediate streams
4. side reactions
5. heat of reaction
2. Reactive chemicals test data for evi-dence of:
1. flammability characteristics
2. shock sensitivity
3. exotherms
4. other evidence of instability
3. Planned operation of process, especially:
1. upsets
2. delays
3. critical instruments and controls
4. mode of failure
5. redundancy
6. worst credible case scenarios

Worst-Case Thinking

At every point in the operation, the process designer should conceive of the *worst* possible combination of circumstances that could *realistically* exist, such as:

1. air leakage
2. deadheaded pumps
3. instrument failure
4. loss of agitation
5. loss of compressed air
6. loss of cooling water
7. plugged lines
8. power failure
9. raw material impurities
10. wrong combination or amount of reactants
11. wrong valve position

An engineering evaluation should then be made of the worst-case consequences, with the goal that the plant will be safe even if the worst case occurs. A HAZOP study could be used to help accomplish "worst-case thinking." When the process designers know what the worst-case conditions are, they should:

1. Try to avoid worst-case conditions.
2. Be sure adequate redundancy exists.
3. Identify and implement lines of defense:
1. preventive measures;
2. corrective measures;
3. sometimes, as a last resort, containment or, possibly, abandoning the process if the hazard is unacceptable

It is important to note that the worst case should be something that is realistic, not something that is conceivable but which may be unreasonable.

Reactive Chemicals Testing

Much reactive chemical information involves thermal stability and the determination of:

1. the temperature at which an exothermic reaction starts
2. the rate of reaction as a function of temperature
3. heat generated per unit of material

The information can be obtained by a variety of types of laboratory tests and by thermodynamic calculations, although reliable kinetic data cannot usually be obtained solely by calculations.

Losses from Dust Explosions

Most organic solids, most metals, and some combustible inorganic salts can form explosive dust clouds. In order to have a dust explosion, it is necessary to satisfy certain conditions:

1. suitably sized dust particles
2. sufficient source of ignition energy
3. dust concentration within explosive limits
4. explosible dust
5. oxidizer must be present

If an explosive dust in air that meets the above criteria occurs in a process, an explosion should be considered as inevitable. The process designer should take into account the possibility of dust explosions and design accordingly.

In dust explosions the combustion process is very rapid. The flame speed is high compared with that in gas deflagrations. Detonations normally do not occur in dust explosions in industrial plants.

The sequence of events in a serious industrial dust explosion is often as follows:

1. A primary explosion occurs in part of a plant, causing an air disturbance.
2. The air disturbance disperses dust and causes a secondary explosion, which is often more destructive than the primary explosion.

If the occurrence of a flammable (explosive) dust is inevitable in a particular process, several design alternatives or combinations of alternatives are available:

1. containment (maximum pressure of a dust explosion is usually below 120150 psig)
2. explosion venting to a safe place
3. inerting (most organic dusts are non-flammable in atmospheres containing less than about 10% oxygen)
4. suppression

A fundamental solution to the dust explosion problem is to use a wet process so that dust suspensions do not occur. However, the process must be wet enough to be effective. Some dusts with a high moisture content can still be ignited.

Dust concentrations in major equipment may be designed to be below the lower flammable limit, but this often cannot be depended on in actual operation. Dust concentrations cannot be safely designed to be above an upper flammable limit because such a limit is ill-defined.2

For a large number of flammable dusts, the lower explosion limit lies between 0.02 and 0.06 kg/m3. The upper explosion limit is in the range of 26 kg/m3, but this number is of limited importance.

A small amount of flammable gas or vapor mixed in with a flammable dust can cause an explosive mixture to be formed even if both are at concentrations below the explosive range by themselves. These mixtures are called "hybrid" mixtures. The ignition energy to ignite a hybrid mixture is often less than that required for the flammable dust by itself.

Venting is only suitable if there is a safe discharge for the material vented. Whenever an explosion relief venting device is activated, it may be expected that a tongue of flame containing some unburned dust will first be ejected. The unburned dust will be ignited as it flows out of the vent and can produce a large fireball that will extend outward, upward, and downward from the vent. It is essential for protection of personnel that venting is to an open place not used by people. If a duct must be used, the explosion pressure in the enclosure will be increased considerably. Therefore, particular attention must be paid to the design of the enclosure in which the explosion could take place.

The NFPA 68 guide issued in 19987 has nomographs, which can be used to select relief areas required for combustible dusts when test data on the dusts are available. The nomographs in NFPA 68 are considered by many to be the preferred way to design dust explosion relief devices.

Relief venting to reduce dust explosion pressure requires the equipment to be protected to have a certain minimum strength. If the enclosure strength is too low, the enclosure will be damaged or destroyed before the explosion relief device can function. NFPA 687 states that the strength of the enclosure should exceed the vent relief pressure by at least 0.35 psi. For industrial equipment such as dryers and baghouses, it is often desirable to have considerably more strength built into the structure to reduce the size of the vent area required. Also, the supporting structure for the enclosure must be strong enough to withstand any reaction forces developed as a result of operation of the vent.

Inerting is a very good preventive measure against dust explosions. The maximum oxygen concentration at which dust explosions are "just not possible" cannot be predicted accurately, as it depends on the nature of the combustible material; testing is usually required. It has been found that in an atmosphere of 10 percent oxygen and 90 percent nitrogen, most combustible organic dusts are no longer explosive. To allow a safety margin, it is good industrial practice to maintain oxygen concentrations below 8 percent. For metal dusts, the allowable oxygen content is about 4 percent.7

Inerting leads to the possibility of asphyxiation by operating personnel if they were exposed to the inert gas. Strict precautions must be taken to prevent exposure of personnel to inerting atmospheres.

Explosion suppression systems are designed to prevent the creation of unacceptably high pressure by explosions within enclosures that are not designed to withstand the maximum explosion pressure.8 They can protect process plants against damage and also protect operating personnel in the area. Explosion suppression systems restrict and confine the flames in a very early stage of the explosion. Suppression systems require more maintenance than do relief venting devices. Explosion suppression systems are made by only a few manufacturers and are quite expensive. This may be the reason why this type of safeguard has not been as widely used in industry as one might expect, although its effectiveness has been proved by much practical experience.

Explosion suppression is a proven technology and should be considered as a candidate for explosion protection. The NFPA has published a standard reference on explosion-suppression protection.9 Manufacturers should be consulted on design, installation, and maintenance.

Even with explosion suppression, it is common for the explosion pressure to reach one atmosphere before it is suppressed. The added pressure surge from the injection of the suppressing agent must also be considered. Therefore, sufficient mechanical strength is always required for enclosures protected by explosion suppression.

Substitution of Less Hazardous Materials

It may be possible to substitute a less hazardous material for a hazardous product. For example, bleaching powder can be used in swimming pools instead of chlorine.4 Benzoyl peroxide, an initiator used in Rubber polymerization reactions, is available as a paste in water, which makes it much less shock-sensitive than the dry form. Other substitutions that have been used to make transportation, storage, and processing safer include:

1. Shipping ethylene dibromide instead of bromine.
2. Shipping ethyl benzene instead of ethylene.
3. Storing and shipping chlorinated hydrocarbons instead of chlorine.
4. Storing and shipping methanol instead of liquefied methane.
5. Replacing flammable refrigerants by halogenated hydrocarbons that are known to have acceptable environmental effects.
6. Storing and shipping carbon tetrachloride instead of anhydrous hydrochloric acid. The CCl_4 is burned with supplemental fuel to make HCl on demand at the user's site.
7. Using magnesium hydroxide slurry to control pH instead of concentrated sodium hydroxide solutions, which are corrosive to humans and relatively hazardous to handle.
8. Using pellets of flammable solids instead of finely divided solids to reduce dust explosion problems.

The use of substitutes may appear to be more costly. The added safety provided by substitutes may make their use worthwhile and can in some cases actually lower the true cost of the project when the overall impact on the process, surrounding areas, and shipping is considered.

Substitutes should be employed only if it is known that overall risk will be reduced. Inadequately tested processes may introduce unrecognized health, safety, and environmental problems.

Catastrophic Failure of Engineering Materials1012

Uniform corrosion of metals can usually be predicted from laboratory tests or experience. Corrosion allowances, which will require thicker metal, can be called for in the design of equipment when uniform corrosion rates are expected.

The most important materials failure to avoid in the design of metal equipment is *sudden catastrophic failure*. This occurs when the material fractures under impulse instead of bending. Catastrophic failure can cause complete destruction of piping or equipment, and can result in explosions, huge spills, and consequent fires. Causes of some of the more common types of catastrophic failures are:

1. low temperature brittleness
2. stress corrosion cracking
3. hydrogen embrittlement
4. high penetration rates involving pitting and corrosion
5. fatigue failure
6. creep
7. mechanical shock
8. thermal shock
9. high rates of temperature change in brittle materials
10. zinc embrittlement of stainless steel
11. caustic embrittlement
12. nitrate stress corrosion

Redundant Instrumentation and Control Systems13

Computer-controlled chemical plants have become the rule rather than the exception. As a result, it is possible to measure more variables and get more process information than ever, and chemical plants can be made "inherently safer" than ever before. However, it must be kept in mind that instruments and control components *will* fail. It is not a question of *if* they will fail, but *when* they will fail, and what the consequences will be. Therefore, the question of redundancy must be thoroughly considered. *The system must be designed so that when failure occurs, the plant is still safe*.

Redundant measurement means obtaining the same process information with two like measurements or two measurements using different principles. Redundant measurements can be *calculated* or *inferred* measurements. Two like measurements would be two pressure transmitters, two temperature measurements, two level measurements, and so on. An example of inferred measurement would be using a pressure measurement and vapor pressure tables to check an actual temperature measurement.

A continuous analog signal that is continuously monitored by a digital computer is generally preferable to a single point or single switch, such as a high level switch or high pressure switch. A continuous analog measurement can give valuable information about what the value is *now* and can be used to compute values or compare with other measurements. Analog measurements may make it possible to predict future values from known trends. Analog inputs may be visual, and one can see what the set point is and what the actual value is. The software security system should determine who changes set points, and should not be easy to defeat.

A single point (digital) signal only determines whether switch contacts are open or not. It can indicate that something has happened, but not that it is going to happen. It cannot provide information to anticipate a problem that may be building up or a his-tory about why the problem happened. Single point signals are easy to defeat. Some single point measurements are necessary, such as fire eyes, backup high level switches, and so on.

As a rule, it is best to avoid:

1. both pressure transmitters on the same tap
2. both temperature measuring devices in the same well

3. both level transmitters on the same tap or equalizing line
4. any two measurements installed so that the same problem can cause a loss of both measurements

It is a good idea to use devices that use different principles to measure the same variable, if possible.

An alarm should sound if any time redundant inputs disagree. In many cases the operating personnel will have to decide what to do. In some cases the computer control system will have to decide by itself what to do if redundant inputs disagree.

The more hazardous the process, the more it is necessary to use multiple sensors for flow, temperature, pressure, and other variables.

Since it must be assumed that all measuring devices will fail, they should fail to an alarm state. If a device fails to a nonalarm condition, there can be serious problems. If a device fails to an alarm condition, but there is really not an alarm condition, it is also serious, but generally not as serious as if it fails to a nonalarm condition, which can provide a false sense of security.

Usually it is assumed that two devices measuring the same thing will not fail independently at the same time. If this is not acceptable, more than two devices may be used. If this is assumed, one can consider the effects of different levels of redundancy:

Number of Inputs	Consequence
One	Failure provides no information on whether there is an alarm condition or not.
Two	Failure of one device shows that there is a disagreement, but without more information, it cannot be determined whether there is an alarm condition or not. More information is needed; the operator could "vote" if there is time.
Three	Failure of one device leaves two that work; there should be no ambiguity on whether there is an alarm condition or not.

Pressure Relief Systems

The design of relief systems involves, in general, the following steps:

1. Generate scenario. What could reasonably happen that could cause high pressures? This could be fire, runaway reactions, phase changes, generation of gases or vapors, leaks from high pressure sources, and so on.
2. Calculate the duty requirementsthe pounds per hour of material that has to be vented, and its physical condition (temperature, pressure, ratio of vapor to liquid, physical properties). This is a rather involved calculational procedure.
3. Calculate the relief area required based on the duty, inlet and outlet piping, and downstream equipment. This is also a rather involved calculational procedure.
4. Choose the relief device to be specified from vendor information.

A group of chemical companies joined together in 1976 to investigate emergency relief systems. This later resulted in the formation of The Design Institute for Emergency Relief Systems (DIERS), a consortium of 29 companies under the auspices of the AlChE. DIERS was funded with $1.6 million to test existing methods for emergency relief system design and to "fill in the gaps" in technology in this area, especially in the design of emergency relief systems to handle runaway reactions.14 DIERS completed contract work and disbanded in 1984.

Huff was the first to publish details of a comprehensive two-phase flow computational method for sizing emergency relief devices, which, with refinements, has been in use for over a decade.1518 The most significant theoretical and experimental finding of the DIERS program was the ease with which two-phase vaporliquid flow can occur during an emergency relief situation. *The occurrence of two-phase flow during runaway reaction relief almost always requires a larger relief system than does single-phase flow vapor venting*. The required area for two-phase flow venting can be from two to much more than two times larger to provide adequate relief than if vapor-only venting occurs.15 Failure to recognize this can result in drastically undersized relief systems that will not provide the intended protection.

Two-phase vaporliquid flow of the type that can affect relief system design occurs as a result of vaporization and gas generation during a runaway reaction or in many liquid systems subjected to fire (especially tanks that are nearly full). Boiling can take place throughout the entire volume of liquid, not just at the surface. Trapped bubbles, retarded by viscosity and the nature of the fluid, reduce the effective density of the fluid and cause the liquid surface to be raised. When it reaches the height of the relief device, two-phase flow results. Fauske and Leung19 described test equipment that can be used to help determine the design of pressure relief systems for runaway reactions that often result in two- or three-phase flow.

Safe and Rapid Isolation of Piping Systems and Equipment

It should be possible to easily isolate fluids in equipment and piping when potentially dangerous situations occur. This can be done using emergency block valves (EBVs). An EBV is a manually or remotely actuated protective device that should be used to provide manual or remote shutoff of uncontrolled gas or liquid flow releases. EBVs can be used to isolate a vessel or other equipment, or an entire unit operation. Manual valves are often used on piping at block limits where it is unlikely that there would be a hazard to personnel if an accident occurs. Remotely controlled EBVs are recommended on tanks and on piping in areas where it may be hazardous for personnel in the case of an accident, or where a quick response may be necessary.

EBVs used on tanks should be as close as possible to the tank flange and not in the piping away from the tank. In cases where EBVs may be exposed to fire, the valve and valve operating mechanism must be fire tested.

In one design case involving large quantities of highly flammable materials, a HAZOP showed that suitably located EBVs were one of the most important features that could be installed to reduce the possibility of loss of containment and serious fire and explosions.

Piping, Gaskets, and Valves

Piping.20

All-welded pipe and flanges should be used in "inherently safer" chemical plants. Since flanges are a potential source of leaks, as few flanges as possible should be used. One, of course, has to be realistic: if it is necessary to clean out pipes, flanges must be provided at appropriate places to make it possible. Also, enough flanges must be provided to make maintenance and installation of new equipment reasonably easy.

Threaded piping should be avoided for toxic and flammable materials. It is very difficult to make threaded fittings leakproof, especially with alloys such as stainless steel. Where threaded piping is necessary, use schedule 80 pipe as a minimum. Pipe nipples should never be less than schedule 80.

Pipe support design should be given special attention. It may be desirable to increase pipe diameter to provide more pipe strength and rigidity and make it possible to have greater distance between supports. Normally, in chemical plants, it is not desirable to use piping less than 1/2 in diameter and preferably not less than 1 in. in diameter, even if the flow requirements permit a smaller pipe, except for special cases. Pipe smaller than 1/2 has insufficient strength and rigidity to be supported at reasonable intervals. Tubing, properly supported, should normally be used for anything smaller than 1/2 Tubing is not as fragile as pipe in small sizes. It can be bent so that the number of fittings required is reduced. If it is necessary to use smaller pipe, or small tubing, special provisions should be made for its support and mechanical protection. Also, consideration should be given to using schedule 80 or schedule 160 pipe if small pipe is required to provide extra mechanical strength, even if the fluid pressure does not require it.

Gaskets.21

Gaskets are among the weakest elements of most chemical plants. Blown-out or leaky gaskets have been implicated in many serious incidents. A leak at a flange can have a torch effect if it is ignited. A fire of this type was considered as a possible cause of the Flixborough disaster.2 (See below under "Case Histories.")

Modern technology makes it possible to greatly reduce the incidence of gasket failure by the use of spiral wound gaskets. These are sold by several manufacturers, including Flexitallic, Parker Spirotallic, Garlock, and Lamons. A spiral wound gasket is considered the safest gasket type available because of the metal inner and outer rings which contain the filler material at the ID and OD. This sturdy construction of a spiral wound gasket does not permit blowout of the gasket material, which is the potential for gaskets made of sheet material.

Bolting with spiral wound gaskets is very important. *Use of plain carbon steel bolts (such as A307 Grade B) with spiral wound gaskets is not permitted*. They are not strong enough. High strength alloy bolts such as A193-B7 (contains Cr and Mo) with A194 heavy hex nuts should be used. To properly seal spiral wound gaskets, it is necessary to tighten the bolts to specified torque limits, which are generally higher than those of conventional gaskets. Compared with conventional gaskets, spiral wound gaskets require better flange finish, heavier flanges, and better flange alignment.

Valves

It is desirable and inherently safer to use fire-tested valves whenever it is necessary to isolate flammable or combustible fluids in a pipeline or tank or other type of equipment. Fire-tested valves were formerly called firesafe valves. They were not really firesafe, and the name has been changed to fire-tested valves. Fire-tested valves should be considered for handling most fluids that are highly flammable, highly toxic, or highly corrosive and that cannot be permitted to escape into the environment.

Fire-tested valves should be used to isolate reactors, storage vessels, and pipelines. They can be used wherever EBVs are required. With the popularity of automated plants, quarter-turn valves are very popular and are used in most installations. The only common quarter-turn valves that are fire-tested are ball valves and high performance butterfly valves. For special purposes, there are other special fire-tested valves.

Plastic Pipe and Plastic-Lined Pipe

Plastic-lined pipe is excellent for many uses, such as highly corrosive applications, or where sticking is a problem, and where ease of cleaning is a factor. It is often the cheapest alternative. However, if there is a fire, there may be "instant holes" at each flange because the plastic will melt away, leaving a gap. Therefore, plastic-lined pipe should not ordinarily be used for flammable materials that must be contained in case of a fire. An exception to this is a firesafe plastic-lined pipe system such as that made by the Resistoflex Corp., which provides a metal ring between each flange that will make plastic-lined pipe firesafe. The pipe will probably have to be replaced after a fire, but the contents of the pipe will be contained during a fire.

In general, all types of solid plastic or glass-reinforced plastic pipe should be avoided, if possible, for use with flammable liquids. Compared with metal, plastic piping:

1. will melt and burn easier
2. is more fragile and can be easily mechanically damaged
3. is harder to support adequately

Plastic pipe should be used with appropriate judgment.

Avoidance of Inherently Unsafe Equipment

Some equipment items can be regarded as "inherently unsafe" for use in flammable or toxic service and should be avoided if possible. These items include a variety of devices.

Glass and transparent devices, such as sight glasses, bull's-eyes, sightports, rotameters, and glass and transparent plastic piping and fittings, may be hazardous. Glass devices are sensitive to heat and shock. Transparent plastic devices may be resistant to shock, but are not resistant to high temperatures. If they fail in hazardous service, severe property damage and personnel injury can result. The guidelines to follow are:

1. If broken, would they release flammable material?
2. If broken, would they expose personnel to toxic or corrosive materials?

Flexible or expansion joints in piping should be eliminated wherever possible. Flexible joints and expansion joints are any corrugated or flexible transition devices designed to minimize or isolate the effects of:

1. thermal expansion
2. vibration
3. differential settling
4. pumping surges
5. wear
6. load stresses
7. other unusual conditions

The need for flexible joints sometimes can be eliminated by proper design so that solid pipe will be able to handle misalignment, thermal expansion, and so on, by bending slightly. In many cases electronic load cells can be used to accurately weigh large reactors and process tanks that may have pipes attached to them, with no flexible or expansion joints. This is done by cantilevering the pipes to the equipment using sufficient runs of straight horizontal unsupported pipe with 90°C elbows to take up movements and vibration without interfering significantly with the operation of the load cells.

Flexible joints should not be used as a correction for piping errors or to correct misalignment.

Pumps for Hazardous Service

A wide variety of excellent pumps are available in the chemical industry. It is sometimes a problem to choose the best from the large number available. This discussion will be limited to centrifugal pumps. Assuming that one has sized the pump, decided on a centrifugal pump, and chosen a suitable list of vendors, the main choices involve (1) metallurgy, (2) seal-less pumps versus conventional centrifugal pumps, and (3) the type of seals for conventional pumps.

Metallurgy.22

Cast iron should not be used for flammable or hazardous service. The minimum metallurgy for centrifugal pumps for hazardous or flammable materials is cast ductile iron, type ASTM A 395, having an ultimate tensile strength of about 60,000 psi.

Seal-less Pumps.23

The most common maintenance problem with centrifugal pumps is with the seals. Seals can be essentially eliminated with seal-less pumps. Seal-less pumps are manufactured in two basic types: canned motor and magnetic drive. Magnetic drive pumps have thicker "cans," which hold in the process fluid and the clearances between the internal rotor and "can" are greater compared with canned motor pumps. Seal-less pumps are becoming very popular and are widely used in the chemical industry. Mechanical seal problems account for most of the pump repairs in a chemical plant, with bearing failures a distant second. The absence of an external motor (on canned pumps) and a seal is appealing to those experienced with mechanical seal pumps. However, it cannot be assumed that just because there is no seal, seal-less pumps are always safer than pumps with seals, even with the advanced technology now available in seal-less pumps. Seal-less pumps must be used with considerable caution when handling hazardous or flammable liquids. A mistreated seal-less pump can rupture with potentially serious results. The "can" can fail if valves on both sides of the pump are closed, and the fluid in the pump expands because of heating up from a cold condition with the pump off, or if the pump is started up under these conditions. If the pump is run dry for even a short time, the bearings can be ruined. Seal-less pumps, especially canned motor pumps, produce a significant amount of heat because nearly all the electrical energy lost in the system is absorbed by the fluid being pumped. *If this heat cannot be properly dissipated, the fluid will heat up with possibly severe consequences*. Considerable care must be used in installing a seal-less pump to be sure that misoperations cannot occur.

Properly installed and maintained, seal-less pumps, especially magnetic drive pumps, offer an economical and safe way to minimize hazards and leaks of hazardous liquids.

5.3 Technical Management of Chemical Process Safety

Although understanding and using sound technology is important in designing and operating safe process plants, technology is not enough. As the chemical process industries have developed more sophisticated ways to improve process safety, safety management systems have been introduced to make process safety engineering activities more effective. The following is a brief summary of a recent approach to the technical management of chemical process safety.24

Twelve process safety elements have been identified as important in the context of plant design, construction, start-up, operation, maintenance, modification, and decommissioning. This does not include personnel safety, transportation issues, chronic releases to the environment, or community response, which are separate and important issues. Process safety management must deal with each element. Even the best companies, with the best safety records, have room for improvement.

Accountability

Accountability begins with a clear, explicit and reasonably specific statement of a company's expectations, objectives, and goals. Example: "Process safety audits must be conducted based on the relative risk involved."

Process Knowledge and Documentation

Capturing process knowledge is a foundation upon which many aspects of a process safety program are built. This is sometimes referred to as "company memory." Preserving this knowledge and making it available within a company are important for process safety for a number of reasons, including:

1. keeping a record of design conditions and materials of construction to help ensure that operations remain faithful to the original intent
2. providing a basis for understanding how and why the process should be operated a certain way
3. making it possible to pass information from older to younger workers
4. providing a baseline for process changes
5. recording causes of incidents and accidents and corrective action for future guidance

Process Safety Reviews for Capital Projects

The need for process safety reviews of capital projects, whether new or revised facilities, has been a recognized feature of engineering organizations. These reviews must be comprehensive and systematic. It is best to do reviews early to avoid costly modifications later. Not only do process hazards need to be identified and addressed by the design, but broader issues should be considered, such as:

1. Is the company prepared to accept the hazards and risks of this project?
2. Is the location appropriate?

Process Risk Management

Process risk management involves the systematic identification, evaluation, and control of potential losses that may arise in existing operating facilities from events such as fires, explosions, toxic releases, runaway reactions, or natural disasters. If risks are not identified, they cannot be considered. Whether resulting losses are measured in terms of direct costs, impacts on employees and/or the public, property or environmental damage, lost business, or various liabilities, the possibility of experiencing such losses is considered a risk. When risks have been identified, it is possible to plan for their reduction, and it can be determined whether the remaining, or residual, risk is acceptable. Risk can never be entirely eliminated. It is the purpose of this component of risk management to manage the risks that remain after implementation of risk controls.

Management of Change

Changes to process facilities are necessary for many reasons. Change includes all modifications to equipment, procedures, and organizations that may affect process safety. Normal operation of a process should be within defined safe limits; operation outside these limits should require review and approval. Replacement personnel should be trained before moves of people are made. There should be no difference between things to consider for changes intended to be permanent and changes that have a limited life, that is, are "temporary" in nature. A hazard will proceed to an incident whether the change is permanent or temporary. In any operation, situations arise that were not foreseen. To ensure that these deviations from normal practice do not create unacceptable risks, it is important to have a variance procedure incorporated into the management system.

Process and Equipment Integrity

Equipment used to handle or process hazardous materials should control the risk of releases and other accidents. Management systems should ensure that equipment is designed properly and constructed and installed according to the design. Components of the process and equipment integrity element include: reliability engineering, materials of construction, inspection procedures, installation procedures, preventive maintenance, and maintenance procedures. It is very important that a program of tests and inspections be available to detect impending failures and mitigate their potential before they can develop into more serious failures.

Incident Investigation

Incidents can be defined broadly as unplanned events with undesirable consequences. Incident investigation is the management process by which the underlying causes of incidents are uncovered and steps are taken to prevent similar incidents. Almost always, it has been found that process safety incidents are a result of some failure of the management system, which, had it functioned properly, could have prevented the incident. Incident investigations should identify all underlying causes and management system failures. It is not enough to discover the immediate cause, such as operator error, but the investigation should go deeper and find out why the operator made the error and possibly discover training and equipment deficiencies. The incident investigation should report facts and conclusions and make recommendations.

Training and Performance

Training is an essential part of any process safety management program. Proper training of personnel is an absolute requirement for keeping complex process equipment and machinery operating safely. Good training programs tell the student not just what to do, but also how to do it and why it is important. Teaching should be a high-priority task. There should be specific criteria for instructor selection. It is not enough to know the subject, although that is important; the teacher should have good presentation skills. The effectiveness of training should be evaluated regularly to determine the effectiveness of the instructors and the performance of the students.

Human Factors (Ergonomics)

Human factors, or ergonomics, refers to human characteristics that must be considered in designing technical systems and equipment so that they can be used safely and efficiently. Anything that increases the difficulty of an operator's job may result in errors if shortcuts are taken. Such shortcuts may work under normal conditions, but are unacceptable in upset or abnormal conditions. The process safety management system should address the humanhardware interface, in both automated and manual processes. Safety reviews and HAZOP studies should include examination of the humanhardware interface. Even in a nearly perfectly designed situation, operators will make occasional errors. Consideration should be given to involving specialized experts in human error assessment, especially in processes involving high potential risks and extensive operator control.

Company Standards, Codes, and Regulations

Company standards, codes, and regulations are intended to communicate minimum acceptable safe practice and to ensure that all operating locations within the company share a common approach to process safety. All U.S. chemical plants are also subject to federal government regulations, and in some states there are specific environmental and toxic laws that apply. (See the section "Regulations" below.) Plants in many countries outside the United States have specific laws that apply to those countries. Many companies also have internal standards and guidelines to ensure consistency in decision-making by design engineers and plant personnel. There should be clear documentation so that it is known which codes, standards, and guidelines are to be followed.

Audits and Corrective Actions

Audits are methodical, independent, and usually periodic examinations of local installations, procedures, and practices. Audits help ensure compliance with a sound process safety program and that risks are being properly managed. The frequency of audits should be determined by the hazards of the facility. Corrective actions are the steps taken by a company in response to the recognition of a process safety deficiency, either through an audit or by other means. It is important that an audit team have expertise in facility operations, safety disciplines such as fire protection, and management systems. Corrective action includes the process of addressing identified deficiencies, weaknesses, or vulnerabilities. It also includes processes for corrective action planning and follow-up.

Enhancement of Process Safety Knowledge

Organizations with strong process safety programs can contribute to advancing the latest process safety technology. They should share nonproprietary results of safety research and support process safety-oriented research and development programs of professional and trade associations and colleges. Improved process safety knowledge can produce a competitive advantage by improved yields, better quality, increased productivity, and less downtime. Consideration should be given to encouraging technical staff participation in professional and trade association programs. The Center for Chemical Process Safety of the American Institute of Chemical Engineers, which is supported by many large chemical and chemical-related companies, is an example of a technical group that sponsors conferences, training programs, and publications in the area of chemical process safety.

5.4 Process Safety Management Systems

The public, customers, employees, and local and federal government regulatory agencies all demand that companies take the actions needed to reduce the possibility of the release of hazardous materials. As the chemical process industries have developed better ways to improve process safety, safety management systems have been introduced to help implement improvements in process safety engineering activities. Management systems for chemical process safety are comprehensive sets of policies, procedures, and practices designed to ensure that barriers to major incidents are in place, in use, and effective.24,38 The following are some definitions that are useful in discussions on managing process safety.

1. *Accident*. An unplanned event with undesirable consequences.
2. *Consequences*. Damage from a scenario; for example, the number of people exposed to a chemical cloud.
3. *Event*. The most elementary action in an accident; for example, an operator response or action.
4. *Hazard*. The result of combining a scenario and a consequence; for example, a chemical cloud is formed, and people are exposed. Also, the inherent potential of a material or activity to harm people, property, or the environment.
5. *Incident*. An occurrence that may be either an accident or a near miss.
6. *Near miss*. Extraordinary event that could have resulted in an accident but did not.
7. *Probability*. The likelihood of the occurrence of an event or a measure of degree of belief, the values of which range from 0 to 1.
8. *Quantitative risk analysis (QRA)*. The systematic development of numerical estimates of the expected frequency and/or consequence of potential accidents associated with a facility or an operation. Often used interchangeably with "quantitative risk assessment."
9. *Risk*. The result of combining scenario, consequence, and likelihood; for example, the likelihood (probability) of a cloud being formed and people being exposed, given the process specifics. Or, a measure of potential economic loss or human injury in terms of the probability of the loss or injury occurring and the magnitude of the loss or injury if it occurs. Or, a combination of the expected frequency (events/year) and/or consequence (effects/event) of a single accident or a group of accidents.
10. *Risk assessment*. The systematic evaluation of the risk associated with potential accidents at complex facilities or operations. Or, the process by which the results of a risk analysis are used to make decisions, either through relative ranking of risk reduction strategies or through comparison with risk targets. The terms "risk analysis" and "risk assessment" are often used interchangeably in the literature.
11. *Risk management*. A part of chemical process safety management. Risk analysis (often used interchangeably with risk assessment) is a part of risk management.
12. *Scenario*. The end result of a series of events; for example, the release of a chemical cloud.

Some Tools for Evaluating Risks and Hazards

Dow Fire and Explosion Index

The Dow Fire and Explosion Index (F&EI), developed by The Dow Chemical Company, is an objective evaluation of the potential of a facility for a fire, an explosion, or a reactive chemical accident. Its purpose is to quantify damage from incidents, identify equipment that could contribute to an incident, and suggest ways to mitigate the incident; it also is a way to communicate to management the quantitative hazard potential. It is intended for facilities handling flammable, combustible, or reactive materials whether stored, handled, or processed. The goal of the F&EI evaluation is to become aware of the loss potential and to identify ways to reduce the potential severity in a cost-effective manner. It does not address frequency (risk) except in a general way. The number is useful mainly for comparisons and for calculations of damage radius, maximum probable property damage, and business interruption loss, and to establish frequency of reviews. The method of carrying out an F&EI evaluation is available to the public from the American Institute of Chemical Engineers, 345 E. 47th St, New York, NY 10017.

Failure Modes and Effects Analysis (FMEA)

FMEA is a systematic, tabular method for evaluating the causes and effects of component failures. It represents a "bottom-up" approach, in contrast with a fault tree, where the approach is "top-down." In large part, HAZOP is a well-developed form of FMEA.2

Fault Tree

A fault tree is a logical model that graphically portrays the combinations of failures that can lead to a particular main failure or accident of interest. A fault tree starts with a top event, which is usually a hazard of some kind. The possibility of the hazard must be foreseen before the fault tree can be constructed. A fault tree helps reveal the possible causes of the hazard, some of which may not have been foreseen.2

Safety Audit

A safety audit is a method of reviewing the actual construction and operation of a facility. Often, safety audits are conducted by a small interdisciplinary team. At least some of the members of the team are not connected with the plant. The audit may be carried out before startup and also is repeated later at intervals of, typically, one to five years.

Chemical Exposure Index

The Chemical Exposure Index is a technique for estimating the relative toxic hazards of chemicals, developed by The Dow Chemical Company. It provides for the relative ranking of toxic chemicals in a given facility, including factors relating to toxicity, quantity volatilized, distance to an area of concern, and physical properties. A description of the method can be found in *Guidelines for Safe Storage and Handling of High Toxic Hazard Materials*, Center for Chemical Process Safety, American Institute of Chemical Engineers.50

5.5 Evaluation of Hazards and Risks HAZOP2,5053

HAZOP stands for "Hazard and Operability Studies," a set of formal hazard identification and elimination procedures designed to identify hazards to people, processes, plants, and the environment. The techniques aim to stimulate the imagination of designers and operations people in a systematic way so they can identify potential hazards. In effect, HAZOP studies assume that there will be mistakes, and provide a systematic search for these mistakes. In some European countries, HAZOP studies are mandatory and attended by observers from regulatory authorities to ensure that the studies are carried out correctly.

The examination of accidents52 during 1988 at a large U.S. chemical company revealed that the accidents could be classified as follows:

1. Spills: 52 percent
2. Emissions: 30 percent
3. Fires: 18 percent

Of the fires, about 50 percent occurred during construction, 25 percent were due to pump seal failure, and the remaining 25 percent resulted from engineering and operational oversights that a HAZOP study possibly could have prevented.

Of the emissions, 37 percent were due to piping failure, with lined pipe being the largest contributor. Operational and procedural issues accounted for 53 percent of the remainder.

Of the spills, 11 percent were due to equipment failures. Piping failures (especially lined pipe and gaskets) accounted for 30 percent, and 56 percent were caused by various types of operational errors, noncompliance with procedures, or nonexistent procedures.

Material handling was a factor in many spills and emissions.

The most frequent type of operational error was a valve being left in an improper position, either open or closed.

HAZOP studies probably could have reduced the number and seriousness of the problems experienced.

Some investigations have shown that a HAZOP study will result in recommendations that are 40 percent safety-related and 60 percent operability-related. Thus, HAZOP is far more than a safety tool; a good HAZOP study also results in improved operability of the process or plant, which can mean greater profitability.

The HAZOP technique can be used to identify human error potential. From a practical point of view, human error and its consequences can occur at all levels of a management structure as well as in the operation of a particular plant or process. Carried out correctly, Technica54 states that a HAZOP study will identify at least 7075 percent of potential operational and safety problems associated with a particular design process, including human error.

The HAZOP technique also can be used for the evaluation of procedures. Procedures may be regarded as a "system" designed to "program" an operator to carry out a sequence of correct actions. Deviations from intent are developed, with the emphasis on "operator action deviation" rather than "physical property deviation." It is the procedure, not the hardware, that is the object of study, but hardware modifications may be recommended to cover potential problems identified from procedure deviations.

Operating Discipline

HAZOP can be an important part of establishing the operating discipline of a plant, which can be defined as the total information required to understand and operate the facility. By recording recommendations and actions in detail on all parts of the process, the rationale behind the way the process is designed and intended to operate and key details of the process will be available, which are rarely on record from other sources. This can be especially helpful when plant changes are made, and when new plants based on the same design are built.

Risk Analysis and Assessment

Risk is defined (see above) as the combination of expected frequency (events/year) and consequence (effects/event) of accidents that could occur as a result of an activity. Risk analysis evaluates the likelihood and consequence of an incident. Risk assessment is an extension of risk analysis that includes making judgments on the acceptability of the risk. It may be qualitative or quantitative, or it can range from simple, "broad brush" screening studies to detailed risk analyses studying large numbers of incidents and using highly sophisticated frequency and consequence models. Throughout the published literature, the terms "risk analysis" and "risk assessment" are used interchangeably.

Quantitative Risk Analysis (QRA)

Quantitative Risk Analysis (QRA) models the events, incidents, consequences, and risks, and produces numerical estimates of some or all of the frequencies, probabilities, consequences, or risks.38,55 QRA can be done at a preliminary level or a detailed level, and in all cases may or may not quantify all events, incidents, consequences, or risks.56 QRA is the art and science of developing and understanding numerical estimates of the risk associated with a facility or an operation. It uses highly sophisticated but approximate tools for acquiring risk understanding.

QRA can be used to investigate many types of risks associated with chemical process facilities, such as the risk of economic losses or the risk of exposure of members of the public to toxic vapors. In health and safety applications, the use of QRA can be classified into two categories:

1. Estimating the long-term risk to workers or the public from chronic exposure to potentially harmful substances or activities.
2. Estimating the risk to workers or the public from episodic events involving a one-time exposure, which may be acute, to potentially harmful substances or activities.

QRA is fundamentally different from many other chemical engineering activities (e.g., chemistry, heat transfer), whose basic property data are capable of being theoretically and empirically determined and often established experimentally. Some of the basic "property data" used to calculate risk estimates are probabilistic variables with no fixed values, and some of the key elements of risk must be established by using these probabilistic variables. QRA is an approach for estimating the risk of chemical operations by using the probabilistic approach; it is a fundamentally different approach from those used in many other engineering activities because interpreting the results of QRA requires an increased sensitivity to uncertainties that arise primarily from the probabilistic character of the data.

5.6 Combustion Hazards

Introduction

The enchanting flame has held a special mystery and charm the world over for thousands of years. According to Greek myth, Prometheus the Titan stole fire from the heavens and gave it to mortalsan act for which he was swiftly punished. Early people made use of it anyway. Soon the ancients came to regard fire as one of the basic elements of the world. It has since become the familiar sign of the hearth and a mark of youth and blood, as well as the object of intense curiosity and scientific investigation.

Suitably restrained, fire is of great benefit; unchecked or uncontrolled, it can cause immense damage. We respond to it with a powerful fascination coupled with an inbred respect and fear. A good servant but a bad master is Thoreau's "most tolerable third party."[25]

Fire[26]

Fire or combustion is normally the result of fuel and oxygen coming together in suitable proportions and with a source of heat. The consumption of a material by a fire is a chemical reaction in which the heated substance combines with oxygen. Heat, light, smoke, and products of combustion are generated. The net production of heat by a fire involves both heat-producing and heat-absorbing reactions, with more heat being produced than is absorbed. Energy in the form of heat is required:

1. To produce vapors and gases by vaporization or decomposition of solids and liquids. Actual combustion usually involves gases or vapors intimately mixed with oxygen molecules.
2. To energize the molecules of oxygen and flammable vapors into combining with one another and so initiating a chemical reaction.

The amount of energy required to cause combustion varies greatly. Hydrogen and carbon disulfide can be ignited by tiny sparks, or simply may be ignited by static generated as the gases or vapors discharge from pipes into air. Other materials, such as methylene chloride, require such large amounts of energy to be ignited that they sometimes are considered nonflammable. Fire also can result from the combining of such oxidizers as chlorine and various hydrocarbon vapors; oxygen is not required for a fire to take place.

There are exceptions to the general rule that a solid must vaporize or decompose to combine with oxygen; some finely divided materials such as aluminum powder and iron power can burn, and it is generally accepted that they do not vaporize appreciably before burning.

Products of Combustion

Heat, light, smoke, and asphyxiating toxic gases are produced by fire. In a hot, well-ventilated fire, combustion usually is nearly complete. Almost all the carbon is converted to carbon dioxide, and all the hydrogen to steam, and oxides of various other elements such as sulfur and nitrogen are produced.

This is not the case in most fires, where some of the intermediate products, formed when large complex molecules are broken up, persist. Examples are hydrogen cyanide from wool and silk; acrolein from vegetable oils; acetic acid from timber or paper; and carbon or carbon monoxide from the incomplete combustion of carbonaceous materials. As the fire develops and becomes hotter, many of these intermediates, which are often toxic, are destroyed (e.g. hydrogen cyanide is decomposed at a significant rate at 538°C).

Small airborne particles of partially burnt carbonaceous materials form smoke, which is often thickened by steam, when there is only partial combustion of fuel.

Solids

Ordinarily, combustible solids do not combine directly with oxygen when they burn. They give off vapor and gaseous decomposition products when they are heated, and it is the vapors or gases that actually burn in the characteristic form of flames. Thus, before a solid can be ignited, it usually must be heated sufficiently for it to give off flammable concentrations of vapors. Glowing, which is combustion in the solid state, is characteristic of materials in the final stages of a fire's decay when flammable gases have been burned away, or when the production of gases and vapors has been suppressed.

Solids with larger surface areas, in relation to their volume, burn more readily than those that are more compact when exposed to heat and oxygen in the air. Common materials such as textiles in the form of fibers or fabrics, foamed rubber, foamed plastics, thin sheets of plastic, paper, corrugated cardboard, combustible dusts, dry grass and twigs, and wood shavings are examples of materials with large surface areas in relation to their volume. In a well-established fire, materials with relatively small surface areas, such as chunks of coal or logs, burn readily.

Combustion is self-propagating; burning materials produce heat which causes more of the solid to produce flammable vapors until either the fuel or oxygen is exhausted, or until the fire is extinguished in some other way.

Dusts

Most combustible solids can produce combustible dusts. Combustible dusts are particularly hazardous; they have a very high surface area to volume ratio. When finely divided as powders or dusts, solids burn quite differently from the original material in the bulk. Dust and fiber deposits can spread fire across a room or along a ledge or roof beam very quickly. Accumulations of dust can smoulder slowly for long periods, giving little indication that combustion has started until the fire suddenly flares up, possibly when no one suspects a problem.

Many combustible dusts produced by industrial processes are explosible when they are suspended as a cloud in air. Even a spark may be sufficient to ignite them. After ignition, flame spreads rapidly through the dust cloud as successive layers are heated to ignition temperature. The hot gases expand and produce pressure waves that travel ahead of the flame. Any dust lying on surfaces in the path of the pressure waves will be thrown into the air and could cause a secondary explosion more violent and extensive than the first.

Liquids

A vapor has to be produced at the surface of a liquid before it will burn. Many common liquids give off a flammable concentration of vapor in air without being heated, sometimes at well below room temperature. Gasoline, for example, gives off ignitable vapors above about -40°C, depending on the blend. The vapors are easily ignited by a small spark or flame. Other liquids, such as fuel oil and kerosene, need to be heated until sufficient vapor is produced.

Many liquids can be formed into mists that will burn at temperatures where the vapor pressure is insufficient to produce a flammable mixture of the vapor and air.

For any flammable vapor there are maximum and minimum concentrations of the vapor in air beyond which it cannot burn. When the mixture of vapor in air is too weak, there is insufficient fuel for burning; when the mixture is too strong, there is insufficient oxygen for burning.

If the density of a flammable vapor is greater than that of air, as is normally the case, flammable concentrations may collect at low levels, such as at floor level or in basements, and can travel a considerable distance to a source of ignition, from which flames will then flash back.

Gases

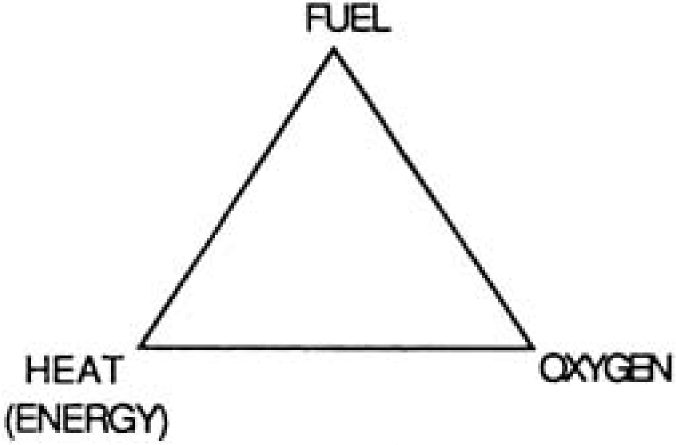
Flammable gases usually are very easily ignited if mixed with air. Flammable gases often are stored under pressure, in some cases as a liquid. Even small leaks of a liquefied flammable gas form relatively large quantities of gas, which is ready for combustion.

The Fire Triangle

The well-known "fire triangle" (see Fig. 5.1) is used to represent the three conditions necessary for a fire:

1. Fuel.
2. Oxidizer: oxygen or other gaseous oxidizer such as chlorine; or liquid oxidizer such as bromine; or solid oxidizer such as sodium bromate.
3. Energy, usually in the form of heat.

Fig. 5.1 The fire triangle.



If one of the conditions in the fire triangle is missing, fire does not occur; if one is removed, fire is extinguished. Usually a fire occurs when a source of heat contacts a combustible material in air, and then the heat is supplied by the combustion process itself.

The fire triangle indicates how fires may be fought or prevented:

1. Cut off or remove the fuel.

2. Remove the heat or energyusually by putting water on the fire.

Remove the supply of oxygenusually by foam or inert gas.

Flammability

No single factor defines flammability, but some relevant parameters include:

1. Flash pointoften considered the main index of flammability; low flash points usually mean increased flammability.
2. Flammability limits; wide limits mean increased flammability.
3. Autoignition temperature; low temperature means increased flammability.
4. Ignition energy; low ignition energy means increased flammability.
5. Burning velocity; high velocity means increased flammability.

A combustion process is an exothermic reaction initiated by a source of ignition that produces more energy than it consumes. The speed at which the reaction proceeds through the mixture of reactants depends on the concentration of the flammable gas or vapor. This speed is lower at higher ("rich") as well as a lower ("lean") concentrations of the flammable gas than at the stoichiometric mixture. There are lower and upper limits beyond which the reaction cannot propagate through the gas mixture on its own. Some definitions follow:

1. *Flammability limits*. The range of flammable vaporair or gasair mixtures between the upper and lower flammable limits. Flammability limits are usually expressed in volume percent. Flam-mability limits are affected by pressure, temperature, direction of flame propagation, oxygen content, type of inerts, and other factors. The precise values depend on the test method.
2. *Upper flammability limit*. The maximum concentration of vapor or gas in air above which propagation of flame does not occur on contact with a source of ignition.
3. *Lower flammability limit*. The minimum concentration of vapor or gas in air or oxygen below which propagation of flame does not occur with a source of ignition.

The concentrations at the lower and upper flammability limits are roughly 50 percent and 200400 percent, respectively, of the stoichiometric mixture. The maximum flammability usually (not always) occurs at the stoichiometric mixture for combustion.2,27

4. *Flammable limits for mixtures of flammable gases and vapors*. For mixtures of several flammable gases and vapors, the flammable limits can be estimated by application of Le Chatelier's equation, if the flammable limits of the components are known:2

$$L = \frac{1}{\sum_{i=1}^{n} \frac{y_i}{L_i}} \qquad U = \frac{1}{\sum_{i=1}^{n} \frac{y_i}{U_i}}$$

L = lower flammability limit of the fuel mixture, vol. %,
Li = lower flammability limit of fuel component *i*, vol. %,
U = upper flammability limit of the fuel mixture, vol. %,
Ui = upper flammability limit of fuel component *i*, vol. %,
yi = concentration of fuel component *i*, vol. %.

This equation is empirical and is not universally applicable, but is useful and a reasonable approximation when actual mixture data are not available.

It is possible for a mixture to be flammable even though the concentration of each constituent is less than its lower limit.

5. *Methods of measurement*. Flammability limits are determined by measuring the volume percent of a flammable gas in an oxidizing gas that will form a flammable mixture, thus identifying the lower and upper flammable limits as well as the critical oxygen concentration (the minimum oxidizer concentration that can be used to support combustion).
6. *Uniformity of lower limits on a mass basis*. Concentrations of vapors and gases usually are reported in volume percent. As molecular weight increases, the lower limit usually decreases. On a mass basis, the lower limits for hydrocarbons are fairly uniform at about 45 mg/L at 0°C and 1 atm. Many alcohols and oxygen-containing compounds have higher values; for example, on a mass basis, ethyl alcohol in air has a lower limit of 70 mg/L.27
7. *Effect of temperature on flammable limits*. The higher the temperature at the moment of ignition, the more easily the combustion reaction will propagate. Therefore, the reference temperature (initial temperature) of the flammable mixture must be stated when flammable limits are quoted. There are not a lot of data for flammable limits under different conditions of initial temperature. The behavior of a particular mixture under different conditions of initial temperature usually must be determined by tests.
8. *Burning in atmospheres enriched with oxygen*. The flammability of a substance depends strongly on the partial pressure of oxygen in the atmosphere. Increasing oxygen content affects the lower flammability limit only slightly, but it has a large effect on the upper flammability limit. Increasing oxygen content has a marked effect on the ignition temperature (reduces it) and the burning velocity (increases it).

At the lower explosive limits of gasair mixtures, there is already an excess of oxygen for the combustion process. Replacing nitrogen by additional oxygen will influence this limit very little.8

9. *Burning in chlorine*. Chemically, oxygen is not the only oxidizing agent, though it is the most widely recognized and has been studied the most. Halogens are examples of oxidants that can react exothermically with conventional fuels and show combustion behavior. The applicability of flammability lim-its applies to substances that burn in chlorine. Chlorination reactions have many similarities to oxidation reactions. They tend not to be limited to thermodynamic equilibrium and often go to complete chlorination. The reactions are often highly exothermic. Chlorine, like oxygen, forms flammable mixtures with organic compounds. As an example: a chlorineiron fire occurred in a chlorine pipeline, causing a chlorine gas release. Chlorine had liquefied in the lines because of the very cold weather, and the low spot was steam-traced. Steam had been taken from the wrong steam line, using 400 psig steam instead of 30 psig steam. The 400 psig steam was hot enough to initiate the reaction. This serves as a reminder that steel and chlorine can react. The allowable temperature for safe use depends upon the state of subdivision of the iron.
10. *Burning in other oxidizable atmospheres*. Flames can propagate in mixtures of oxide of nitrogen and other oxidizable substances. For example, Bodurtha27 reports that the flammability limits for butane in nitric oxide are 7.5 percent (lower) and 12.5 percent (upper).
11. *Flame quenching*. Flame propagation is suppressed if the flammable mixture is held in a narrow space. There is a minimum diameter for apparatus used for determination of flammability limits. Below this diameter the flammable range measurements are narrower and inaccurate.

If the space is sufficiently narrow, flame propagation is suppressed completely. The largest diameter at which flame propagation is suppressed is known as the quenching diameter. For an aperture of slotlike cross section there is critical slot width.

The term "quenching distance" sometimes is used as a general term covering both quenching diameter and critical slot width, and sometimes it means only the latter.

There is a maximum safe gap measured experimentally that will prevent the transmission of an explosion occurring within a container to a flammable mixture outside the container. These data refer to a stationary flame. If the gas flow is in the direction of the flame propagation, a smaller gap is needed to quench the flame. If the gas flow is in the opposite direction, a larger gap will provide quenching. If the gas velocity is high enough, the flame can stabilize at the constriction and cause local overheating. These quenching effects are important in the design of flame arrestors.

12. *Heterogeneous mixtures*.28 In industry, heterogeneous (poorly mixed) gas phase mixtures can lead to fires that normally would be totally unexpected. It is important to recognize that heterogeneous mixtures can ignite at concentrations that normally would be nonflammable if the mixture were homogeneous. For example, one liter of methane can form a flammable mixture with air at the top of a 100-L container although the mixture only would contain 1.0 percent methane by volume if complete mixing occurred at room temperature, and the mixture would not be flammable. This is an important concept because "layering" can occur with any combustible gas or vapor in both stationary and flowing mixtures.

Heterogeneous gas phase mixtures can lead to unexpected fires if a relatively small amount of flammable gas is placed in contact with a large amount of air without adequate mixing, even though the average concentration of flammable gas in the mixture is below the flammable limit. *Heterogeneous mixtures are always formed at least for a short time when two gases or vapors are first brought together*.

13. *Effect of pressure*. Flammability is affected by initial pressure. Normal variations in atmospheric pressure do not have any appreciable effect on flammability limits.

A decrease in pressure below atmospheric usually narrows the flammable range. When the pressure is reduced low enough, a flame or an explosion can no longer be propagated throughout the mixture.

An increase in pressure above atmospheric usually (not always) widens the flammability range, especially the upper limit.

14. *Explosions in the absence of air*. Gases with positive heats of formation can be decomposed explosively in the absence of air. Ethylene reacts explosively at elevated pressure, and acetylene reacts explosively at atmospheric pressure in large-diameter piping. Heats of formation for these materials are +52.3 and +227 kJ/g/mol, respectively. Explosion prevention can be practiced by mixing decomposable gases with more stable diluents. For example, acetylene can be made nonexplosive at a pressure of 100 atm by including 14.5 percent water vapor and 8 percent butane.

Ethylene oxide vapor will decompose explosively in the absence of oxygen or air under certain conditions when exposed to common sources of ignition if heated to high enough temperatures. One way to prevent the decomposition reaction is to use methane gas to blanket the ethylene oxide liquid. It has also been found that liquid ethylene oxide will undergo a deflagration in the absence of oxygen with a very rapid pressure increase if ignited at a temperature and pressure above a certain level. Fortunately, the conditions required for propagation of the decomposition of liquid phase ethylene oxide are outside the current normal handling and processing ranges for the pure liquid. Propagation has not been observed below 80°C at from 14 to 100 atm pressure.29 Ethylene oxide also can undergo explosive condensation when catalyzed by a small amount of caustic.30

Inert Gases

The addition of inert gases to a mixture of flammable gases and air affects flammability limits. Carbon dioxide causes a greater narrowing of the flammable range than does nitrogen. Water vapor is an acceptable inert gas if the temperature is high enough to exclude much of the oxygen, which requires a temperature of 9095°C. Because water vapor and carbon dioxide have a higher heat capacity than nitrogen, they are somewhat more effective as inerting agents than nitrogen. Some halogen-containing compounds also can be used for inerting materials at relatively low concentrations. An example of this is the use of Freon-12 (CCl2F2). Caution must be used with halohydrocarbons because of the possibility of the halocarbons themselves burning, especially at high pressures. Environmental considerations are making the use of halogenated hydrocarbons for inerting increasingly undesirable. Materials are being developed that are considered environmentally acceptable. Figures 5.2 and 5.328 show flammability envelopes for methane and *n*-hexane for various airinert mixtures at 25°C and 1 atm. All flammable envelopes are similar to Figs 5.2 and 5.3 except in minor detail. The lower limit is virtually insensitive to added inerts. The upper limit, however, decreases linearly with added inert until the critical concentration of inert is reached beyond which no compositions are flammable. In these graphs, *C*st means the stoichiometric composition.

Fig. 5.2 Limits of flammability of various methaneinert gasair mixtures at 25°C and atmospheric pressure. (*Courtesy Bureau of Mines*)

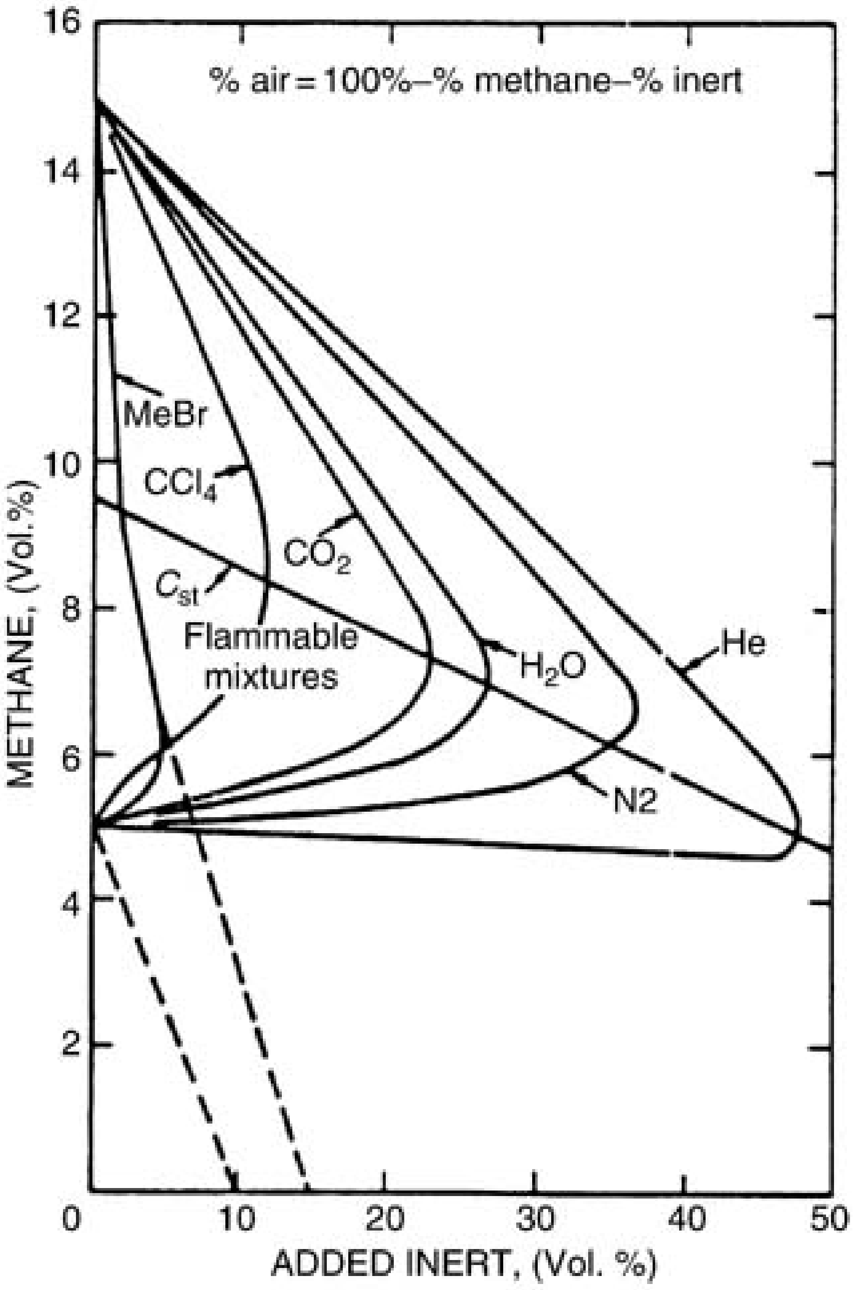
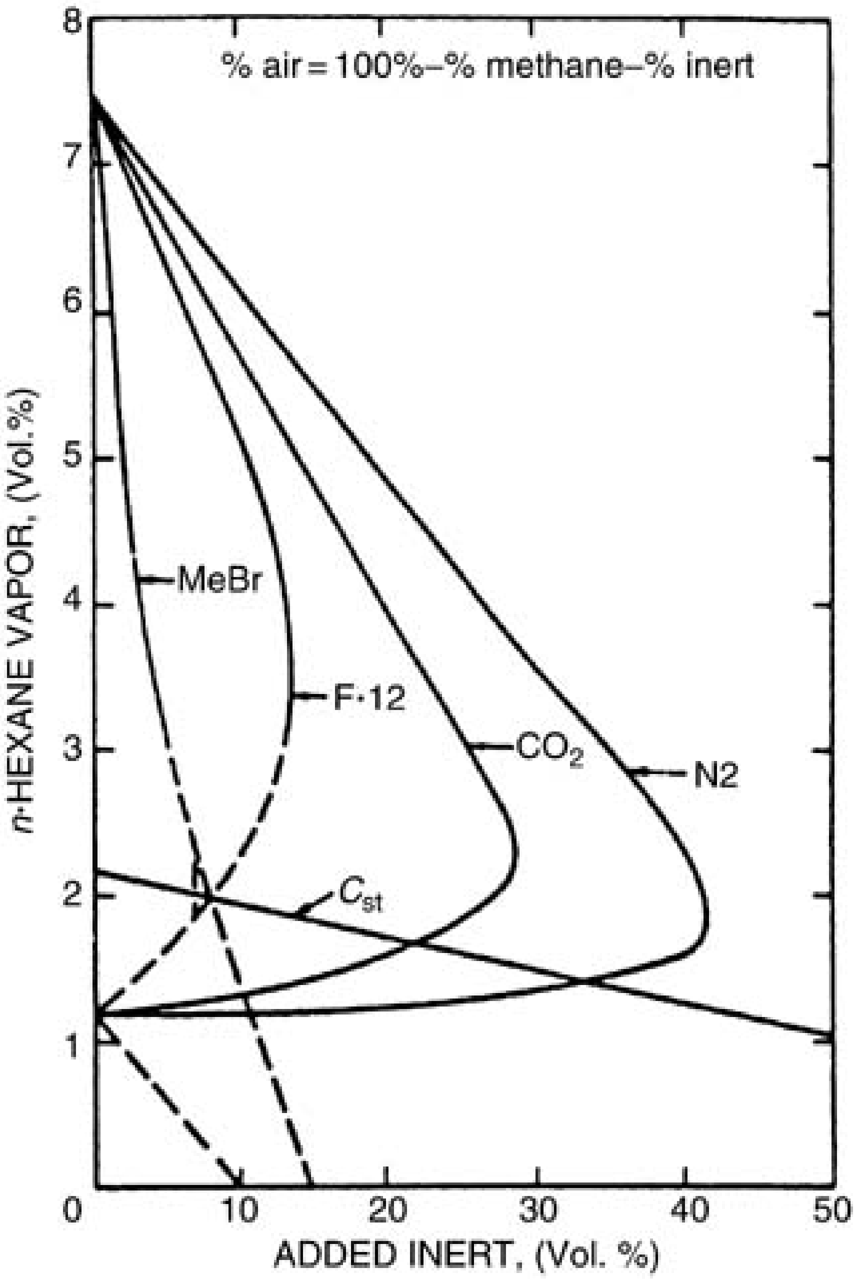


Fig. 5.3 Limits of flammability of various *n*-hexane-inert gas mixtures at 25°C and atmospheric pressure. (*Courtesy Bureau of Mines*)



The limits of flammability are dictated by the ability of a system to propagate a flame front. Propagation does not occur until the flame front reaches about 12001400 K. Since the typical terminal temperature for hydrocarbons at stoichiometric conditions is about 2300 K, it can be seen that having only one-half the fuel or oxidizer present will produce about one-half the flame temperature, which is too low to propagate flame.

A useful rule to remember is that the lower flammable limits of most flammable vapors are close to one-half the stoichiometric composition, which can be calculated. Another easy rule to remember is that about 10 percent oxygen or less in air (assuming the rest is mostly nitrogen) will not support combustion of most flammable hydrocarbon vapors.

The flammability limits of hydrocarbon-type fuels in oxygen and inert gas atmospheres are a function of the inert gas and any fuel or oxygen in excess of that required by the stoichiometry of the combustion process. In systems where fuel content is fixed, inert material having a high heat capacity will be more effective at flame suppression than inert material having a low heat capacity.

Many of the flammable limits reported in the literature are somewhat too narrow, and certain gas compositions regarded as being nonflammable are in fact flammable when given the proper set of circumstances. *In other words, take data on flammability limits from the literature with a grain of salt*. It is best not to design closely on the basis of most available data on flammability limits.

The use of inert gases can cause some serious hazards that must be recognized if inerts are to be used effectively and safely. Considerations in the use of inert gases include:

1. *An inert atmosphere can kill if a person breathes it*. Precautions should be taken to ensure that personnel cannot be exposed to the breathing of inert atmospheres.
2. *Some products need at least a small amount of oxygen to be stored safely*. This includes styrene and some other vinyl monomers, which must have some oxygen in them to make the usual polymerization inhibitor for styrene (*t*-butyl catechol, or TBC), effective. If pure nitrogen, for example, is used to blanket styrene, the inhibitor will become ineffective. TBC customarily is added to styrene monomer to prevent polymer formation and oxidation degradation during shipment and subsequent storage; it functions as an antioxidant and prevents polymerization by reacting with oxidation products (free radicals in the monomer). If sufficient oxygen is present, polymerization is effectively prevented (at ambient temperatures); but in the absence of oxygen, polymerization will proceed at essentially the same rate as if no inhibitor were present. The styrene may polymerize and can undergo an uncontrolled exothermic reaction, which may generate high temperatures and pressures that can be very hazardous. The inhibitor level of styrene must be maintained above a minimum concentration at all times. The minimum concentration of TBC in styrene for storage is about 45 ppm.
3. *To be effective, inert atmospheres must be maintained within certain composition limits*. This requires the proper instrumentation and regular attention to the system.
4. *Inerting systems can be quite expensive and difficult to operate successfully*. Before the use of inert systems, alternatives should be explored, such as the use of nonflammable materials or operating well outside, preferably below, the flammability range.

Mists and Foams

If the temperature of a liquid is below its flash point, flammable concentrations of vapor cannot exist, but conditions still can exist for flammability if mists or foams are formed. A suspension of finely divided drops of a flammable liquid in air has many of the characteristics of a flammable gasair mixture and can burn or explode. A mist may be produced by condensation of a saturated vapor or by mechanical atomization. Normally, the diameter of drops in a condensed mist is less than 0.01 mm, whereas in a mechanical spray it usually is greater than 0.1 mm.

The commonly accepted fallacy that liquids at temperatures below their flash points cannot give rise to flammable mixtures in air has led to numerous accidents. Flash points are measured under stagnant conditions in carefully controlled laboratory experiments, but in the real world one works with a wide variety of dynamic conditions that can produce mists and foams.

Flammable mistvaporair mixtures may occur as the foam on a flammable liquid collapses.28 Thus, when ignited, many foams can propagate flame. An additional hazard can arise from the production of foams by oxygen-enriched air at reduce pressures. Air confined over a liquid can become oxygen enriched as pressure is reduced because oxygen is more soluble than nitrogen in most liquids. Thus, the presence of foams on combustible liquids is a potential explosion hazard.

The lower flammability limit for fine mists (<0.01 mm diameter) of hydrocarbons below their flash point, plus accompanying vapor, is about 48 g of mist/m3 of air at 0°C and 1 atm. Mist can occur in agitated vessels under some conditions, especially when an agitator blade is at or near the liquidvapor interface in the vessel.

Work on condensed oil mists (drop diameter mostly less than 0.01 mm) has demonstrated that they have flammability characteristics similar to those the mixture would have if it were wholly in the vapor phase at the higher temperature necessary for vaporization. The flammability characteristics are affected by drop size. For larger drop sizes (above 0.01 mm) the lower limit of flammability decreases as drop diameter increases. For mists, the amount of inert gas needed to suppress flammability is about the same as that needed to suppress an equivalent vaporair mixture of the same material if it were vaporized at a somewhat higher temperature.

A useful rule is that mists of flammable or combustible liquids in air can burn or explode at temperatures below their flash points.

Ignition

Flammable gases and vapors can be ignited by many sources. In the design and operation of processes, it is best not to base fire and explosion safety on the presumption that ignition sources have been excluded. Bodurtha27 reported that of 318 natural gas fires and explosions, the sources of ignition of 28 percent were unknown. All reasonable measures should be taken to eliminate possible sources of ignition in areas in which flammable materials are handled.

Autoignition

If the temperature of a flammable gasair mixture is raised in a uni-formly heated apparatus, it eventually reaches a value at which combustion occurs in the bulk gas. This temperature is defined as the spontaneous ignition temperature (SIT) or autoignition temperature (AIT). The gasair mixture that has the lowest ignition temperature is called by various names, such as the minimum AIT, the minimum spontaneous ignition temperature, and the self-ignition temperature.27 Usually the AIT reported in the literature is the minimum AIT.

The AIT of a substance depends on many factors, such as:

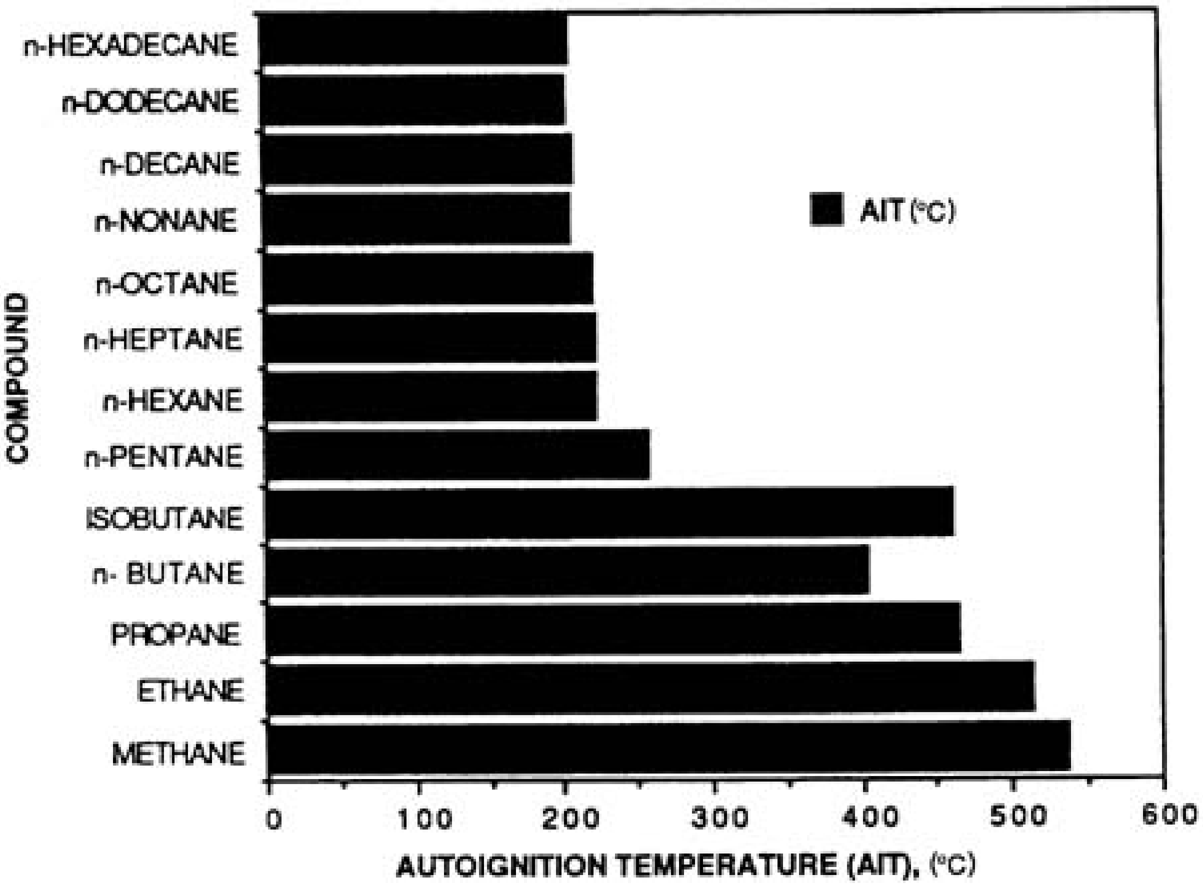
1. ignition delay
2. energy of ignition source
3. pressure
4. flow effects
5. surfaces
6. concentration of vapors
7. volume of container
8. oxygen content
9. catalytic materials
10. flow conditions

Thus, a specific AIT applies only to the experimental conditions employed in its determination. Usually the values quoted are obtained in clean laboratory equipment.

The AIT of a substance may be reduced below ideal laboratory conditions by as much as 100200°C for surfaces that are insulated with certain types of insulation, or are contaminated by dust.

Mixtures that are fuel-rich or fuel-lean ignite at higher temperatures than do those of intermediate compositions. Also, in a homologous series of organic compounds, the AIT decreases with increasing molecular weight, as shown in Fig. 5.4.

Fig. 5.4 Autoignition temperatures of paraffin hydrocarbons at 1 atm. (*Data from Bodurtha 1980*.27)



Ignition Delay

Ignition of a flammable mixture raised to or above the temperature at which spontaneous combustion occurs is not instantaneous; the time delay between the moment of exposure to high temperature and visible combustion is called the ignition delay. This time delay decreases as the ignition temperature increases. The time delay may be as little as a fraction of a second at higher temperatures, or several minutes close to the AIT.

Environmental Effects

It has been found that the AIT becomes lower with increasing vessel size in the range of 35 ml to 12 L. An increase in pressure usually decreases AITs, and a decrease in pressure raises AITs. Usually oxygen enrichment of the air tends to decrease the minimum AIT, and a decrease in oxygen content increases the minimum AIT. Low-temperature oxidation can result in "cool flames," which may grow into ignition.

Catalytic Materials

Ignition may occur where the temperature is less than the minimum AIT. Catalytic materials, such as metal oxides, can promote oxidation on their surfaces, leading to a high local temperature and subsequent ignition. There is a recorded reactive chemical case31 in which a solvent at 80°C was being pressurized with a gas phase consisting of a high oxygen concentration. The solvent has a flash point in oxygen of greater than 130°C and normally is considered not to be a flammability hazard. There was an ignition, causing the vessel to rupture its main gasket with major damage to the facility. It was found that a mist had been formed in the vessel by the agitator, and that the source of ignition probably was a trace of palladium catalyst remaining from a previous run.

From this incident, several important lessons can be learned:

1. Ignition of a flammable mixture can result from totally unexpected contamination by trace amounts of catalysts if the oxidizer and fuel are present.
2. Mists of oxidizable liquids may form that can burn or explode at temperatures outside the "normal" flammable range.
3. It can be dangerous to perform experiments with pure oxygen, or air enriched with oxygen, especially under pressure and at elevated temperatures, when oxidizable materials are present.
4. The real criterion regarding flammable mixtures in air should be whether a flammable atmosphere can exist under the given process conditions, rather than whether a flammable liquid is at a temperature below its flash point.

Cleaning Up Spills of Flammable or Combustible Liquids

It is customary to clean up small spills of many liquid materials with sand or other noncombustible absorbent material. Some absorbing agents, such as untreated clays and micas, will cause an exothermic reaction with some liquids, especially monomers, which might ignite the liquid if it is flammable or combustible. Before any material is provided to be used to soak up spills of oxidizable material, tests should be made to determine if the material can cause fires with potential spills.

Ignition Caused by Insulation

Ignition of combustible materials that have been absorbed into commonly used insulating materials is possible at temperatures lower than the AIT for nonabsorbed material. All oxidizable materials oxidize to some extent in air at ambient temperatures, usually at a very low rate. When an absorbent material is absorbed into insulation, it is "spread" over a large area, increasing its access to oxygen. Because the absorbent is an insulator, heat from oxidation is retained rather than dissipated, and the temperature will rise if the heat is produced faster than it can be dissipated. The rate of oxidation increases as the material temperature increases, which produces more heat, compounding the hazard. If the temperature rises enough, the material will ignite ("spontaneous combustion"). This is similar to the classic oily rag and wet haystack phenomenon, which has caused many fires in homes and on farms. In the wet haystack phenomenon, fermentation by microorganisms will create heat. Some air is necessary; too much air will remove too much heat to allow the combustion temperature to be reached. For equipment operating above about 200°C containing combustible liquids with high boiling points, insulation should be impervious to the material handled. To date, only a closed cell foamed glass provides the required degree of protection where oxidizable liquid materials are used above 200°C. Insulation based on glass fiber, silicate, or alumina materials is known to cause hazardous situations and should not be used in this service.

Laboratory tests and actual fires show that Dowtherm A® (a heat-transfer fluid consisting of a eutectic mixture of biphenyl oxide and biphenol) can be ignited if it is soaked in glass fiber insulation and in contact with air at temperatures considerably below the normal AIT. This is also true for stearic acid soaked in glass fiber insulation. Table 5.1 shows the reduction in AIT of Dowtherm A® and of stearic acid soaked in glass fiber insulation.

TABLE 5.1 Reduction in AITs Caused by Liquids Soaking in Glass Fiber Insulation

Material	Normal AIT (°C)	Ignition Temperature in Glass Fiber Insulation (°C)
Dowtherm A®	621	260290
Stearic acid	395	260290

Ignition of this type generally occurs only with materials having a high boiling point. Usually materials with low boiling points will vaporize and cannot remain soaked in hot insulation. There are exceptions. For example, ethylene oxide has a fairly low boiling point, but if it leaks into insulation, a polymer can be formed that has a high boiling point and can autoignite insulation at low temperatures.

Ignition Caused by Impact

Solids and liquids can be ignited by impact. Impact tests are made by having a weight fall freely through a known distance and impacting the sample. Impact can occur, for example, if containers are accidentally dropped. The interpretation of the data from impact tests can be difficult.

Ignition Caused by Compression of Liquids

Liquids can be ignited by sudden compression. This can happen when there is water hammer caused by the pressure surge from quick-acting valves and from the compression in liquid pumps. Sudden compression can occur with liquids, for example if a tank car is bumped rapidly and the liquid goes to one end very quickly, possibly trapping some vapor bubbles that compress and create local hot spots that can cause ignition.

Ignition Caused by Rubbing Friction

Solids can be ignited by frictional sources when rubbed against each other or against another material. The frictional heat produced may be enough to ignite other materials, such as lubricants, that are nearby. A common example of this occurs when bearings run hot, causing oil or grease to vaporize and possibly ignite.

Ignition Caused by Glancing Blows

Friction can cause ignition in other ways. Sparks may occur when two hard materials come in contact with each other in a glancing blow (the blows must be glancing to produce friction sparks). These kinds of sparks are not directly related to frictional impact. Hand and mechanical tools are the most likely sources of friction sparks that occur outside of equipment. The need for nonsparking tools is somewhat controversial; Bodurtha27 states that it is extremely unlikely that anyone would be using tools in a flammable atmosphere, and it is usually more prudent to control the atmosphere than the tools. Sparkproof tools are not really sparkproof in all situations.

Ignition Caused by Static Electricity

Static electricity is a potential source of ignition wherever there is a flammable mixture of dusts or gases (see next section).

Ignition Caused by Compression of Gases

If a gas is compressed rapidly, its temperature will increase. Autoignition may occur if the temperature of the gas becomes high enough (this is more or less the principle of diesel engines).

An advancing piston of high-pressure gas can compress and heat trapped gas ahead of it. For a perfect gas, the temperature rise due to adiabatic compression is given by

$$\frac{T_2}{T_1} = \left(\frac{P_2}{P_1}\right)^{(k-1)/k}$$

where $T1$ and $T2$ are the initial and final gas absolute temperatures, $P1$ and $P2$ are the initial and final absolute pressures, and k is the ratio of heat capacity at constant pressure to the heat capacity at constant volume. For air and many other diatomic gases, $k = 1.4$. Many hydrocarbons have k values of between 1.1 and 1.2. The value of k is a function of temperature and pressure.

Energy Levels for Ignition

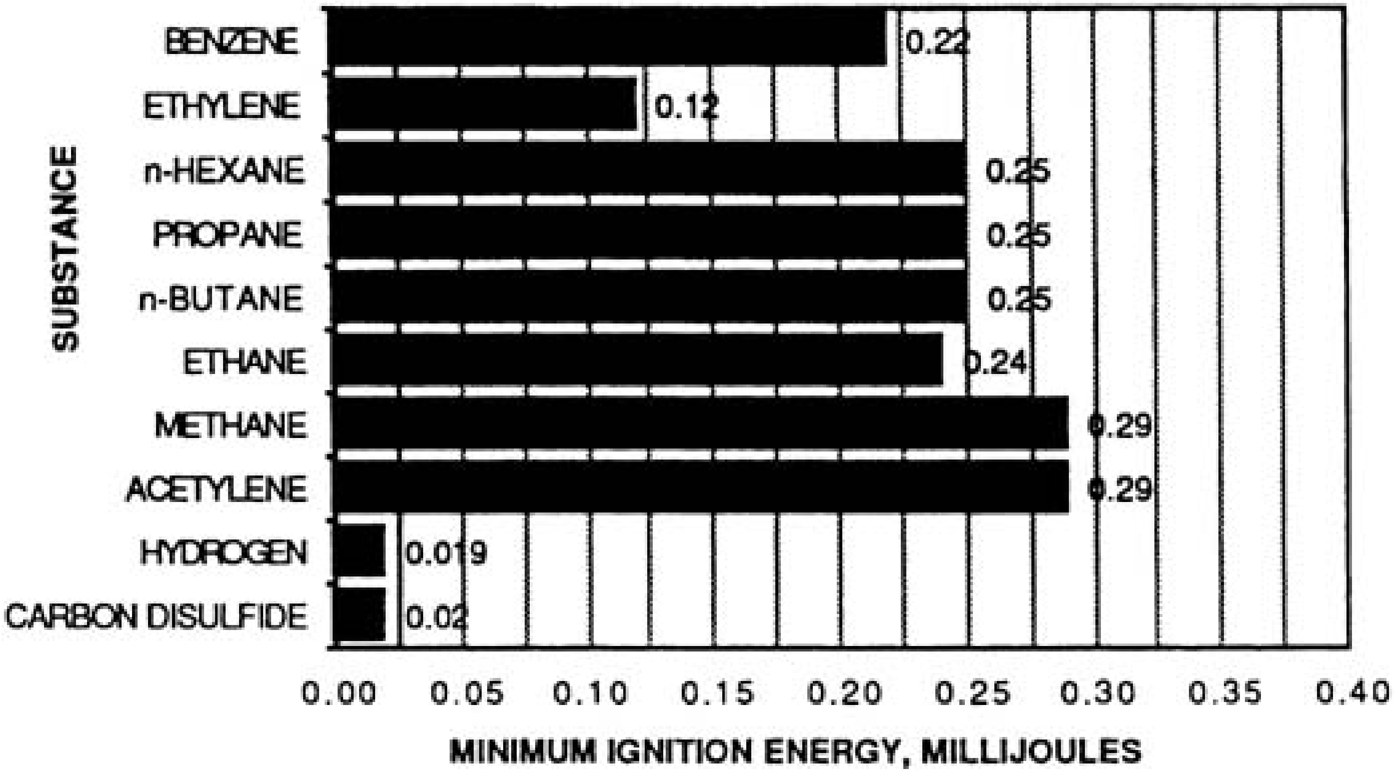
If a flammable gas mixture is to be ignited by a local source of ignition, there is a minimum volume of mixture required to cause a continuing flame throughout the mixture.

For example, to ignite a methaneair mixture in a cold container, a hot patch of 18 mm2 at 10001100°C is required in order to heat enough volume of gas to produce a continuing flame,2 even though the auto-oxidation temperature for methane is 540°C. Ignition of a flammable gasair mixture by electrical discharge can occur only if the electrical discharge is of sufficient energy.

Minimum Ignition Energy

There is a minimum ignition energy, which usually occurs near the stoichiometric mixture. The minimum ignition energy for some representative substances in air is shown in Fig. 5.5.2 The energy required to cause ignition frequently is reported in millijoules (mJ). One joule is 0.24 calorie, so 1 mJ is 0.00024 calorie, which is a very small amount of energy.

Fig. 5.5 Minimum ignition energy for selected substances. (*Less 1980*.2)



A person typically has capacitance of 200 picofarads (pF), and if charged to 15 kilovolts (kV) could initiate a discharge of 22.5 mJ. This is enough to ignite many flammable mixtures. The energy in ordinary spark plugs is 2030 mJ.

The hazard of an explosion should be minimized by avoiding flammable gasair or dustair mixtures in a plant. It is bad practice to rely solely on elimination of sources of ignition, as it is nearly impossible to ensure this.

Effect of Oxygen-Enriched Atmospheres

The minimum spark energy to cause ignition varies greatly with the amount of oxygen in oxygen-enriched air. Stull30 showed that with a composition of 10 percent methane in air, about 0.5 mJ of spark energy is required to initiate a reaction at the lower flammable limit. If the air is enriched with oxygen, the minimum spark energy decreases. If the flammable material is combined with 100 percent oxygen, the spark energy required is only about 1 percent of the required energy in air at 21 percent oxygen! This demonstrates the extremely small amount of energy required to initiate the reaction, as well as the additional ease with which oxygen-enriched atmospheres are initiated. Table 5.2 compares initiation energies of some common substances in air and in pure oxygen.

TABLE 5.2 Comparison of Initiation Energies of Some Common Substances in Air and Pure Oxygen32

Flammable Substance	Relative Amount of Energy Required to Initiate Combustion (mJ) In Air	 In Pure Oxygen
Methane	0.3	0.003
Hydrogen	0.019	0.0012
Acetone	1.15	0.0024
Diethyl ether	0.2	0.0013

Effect of Pressure

An increase in pressure decreases the amount of energy required to cause ignition. In a mixture of propane, oxygen, and nitrogen, doubling the pressure decreases the minimum energy required to cause ignition by a factor of about 5.

If no other data are available for determination of hazards, minimum ignition energies at ambient temperatures and pressures should be considered as approximately:

1. 0.1 mJ for vapors in air
2. 1.0 mJ for mists in air
3. 10.0 mJ for dusts in air

5.7 Static Electricity

Introduction

Many apparently mysterious fires and explosions have eventually been traced to static. In spite of the large amount of information about static electricity, it remains a complex phenomenon not often understood and appreciated. Static electricity is a potential source of ignition whenever there is a flammable mixture of gas or dust.

When two different or similar materials are in contact, electrons can move from one material across the boundary and associate with the other. If the two materials in contact are good conductors of electricity and are separated, the excess electrons in one material will return to the other before final contact is broken. But if one or both of the materials are insulators, this flow will be impeded. If the separation is done rapidly enough, some excess electrons will be trapped in one of the materials. Then both materials are "charged." Electric charges can build up on a nonconducting surface until the dielectric strength is exceeded and a spark occurs. The residual charge could ignite flammable mixtures.

The two materials or phases in initial contact may be:

1. a single liquid dispersed into drops
2. two solids
3. two immiscible liquids
4. a solid and a liquid
5. a solid and a vapor or gas
6. a liquid and a vapor or gas

The important thing to keep in mind is that whenever there is contact and separation of phases, a charge may develop that could be disastrous. Three conditions must be met before an explosion caused by static electricity can take place:

1. An explosive mixture must be present.
2. An electric field must have been produced due to the electrostatic charge that had been generated and accumulated in a liquid or solid.
3. An electric field must be large enough to cause a spark of sufficient energy to ignite the mixture.

In designing preventive measures, all three factors should be controlled.

Static electricity is essentially a phenomenon of low current but high voltage and high resistance to current flow. A low-conductivity liquid flowing through a pipeline can generate a charge at a rate of 10-910-6 ampere (A). A powder coming out of a grinding mill can carry a charge at a rate of 10-810-4 A. At a charging rate of 10-6 A, the potential of a container insulated from earth can rise at a rate of 1000 V/s and a voltage of 10,000 volts or higher can readily be obtained in this way.

Several electrostatic voltages and energies commonly encountered are typified by the following examples:

1. A person walking on dry carpet or sliding across an automobile seat can generate up to 5000 V in dry weather. An individual having a capacitance of 100 pF, a reasonable figure, could generate a spark energy of 1.25 mJ. This is far more than is needed to ignite some flammable vaporair mixtures.
2. A person can accumulate dangerous charges up to about 20,000 V when humidity is low.
3. A truck or an automobile traveling over pavement in dry weather can generate up to about 10,000 V.
4. Nonconductive belts running over pulleys generate up to 30,000 V. The voltage generated by a conveyor belt can be as high as 106 V; the system can in effect act as a Van der Graaf generator.
5. The energy in the spark from an ordinary spark plug is 2030 mJ.

The capacitance and the energy for ignition of people and of common objects are important. The capacitance of a human being is sufficient to ignite various flammable gas mixtures at commonly attained static voltages.

Hazard Determinants

Capacitance

The capacitance of an object is the ratio of the charge of the object to its potential. The capacitance gets larger as the object gets larger. With a given charge, the voltage gets higher as the capacity of the object gets smaller. For a sphere, capacitance is given by

$$C = Q(10^{-3})/V$$

The energy stored in a capacitor is27

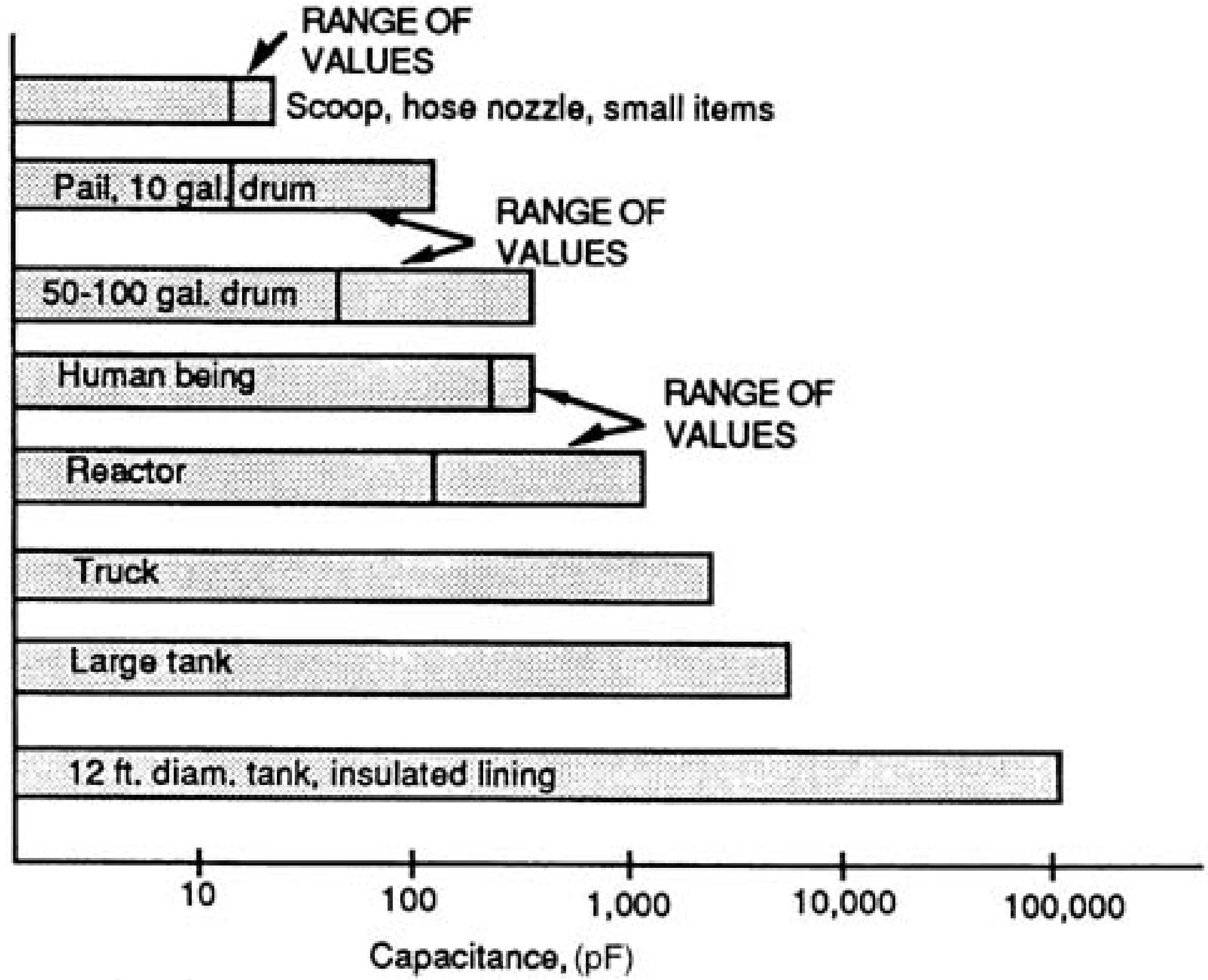
$$W = 0.5CV^2(10-3) = 500Q^2/C$$

C = capacitance, pF (1 pF = picofarad = 10-12 farad)
Q = charge, microcoulombs (1 coulomb = 1 A/s = charge on 6.2 × 1018 electrons)
V = voltage in kilovolts
W = energy, millijoules (mJ)

This energy may be released as a spark when the voltage gets high enough. The minimum sparking potential for charged electrodes is about 350 V and occurs at a spacing of 0.01 mm. Sparks from an equally charged nonconductor are less energetic and may contain only part of the stored energy. These comparatively weak sparks are not likely to ignite dust clouds but can ignite flammable gases.

The energy that can be stored by capacitance of an object can be compared with the minimum ignition energies of flammable gasair mixtures and of dustair mixtures to determine the probability that a spark discharge may have sufficient energy to cause ignition. If the charged object is a poor conductor, the calculation of energy available to produce a spark may not be possible because the charge often is not uniformly distributed, and the resistance to flow of current is high. Figure 5.6 shows some typical values of electrical capacitance.33

Fig. 5.6 Some typical values of electrical capacitance. (*Data from Eichel 1967*.33)



Relaxation Time

When a liquid is flowing in closed metal pipes, static electricity is not a hazard. When the liquid enters a tank, it may become a hazard. Charges caused by liquid separation during pumping, flow, filtration, and other effects such as splashing and agitation can accumulate on the surface of the liquid in the tank and cause sparking between the liquid surface and the tank or conducting objects in the tank. The charge thus generated can be dissipated by relaxation or via discharge through a spark or corona discharge. The relaxation time is the time required for 63 percent of the charge to leak away from a charged liquid through a grounded conductive container. The half-time value is the time required for the free charge to decay to one-half of its initial value. The half-time is related to the relaxation time by the relationship

$$T_{\mathrm{h}} = T_{\mathrm{r}} \times 0.693$$

*T*h = half-time
*T*r = relaxation time

Relaxation times vary from small fractions of a second up to minutes and even hours for some highly purified hydrocarbons that have very low conductivity. It is important to recognize that a large charge can accumulate in the liquid even in a grounded container. In fact, it was reported that the majority of accidents attributed to static electricity in the petroleum industry have been with liquid in grounded containers.2

Relaxation time can be calculated as follows:

$$T_{\mathrm{r}} = E(E_0/k)$$

*T*r = the relaxation time, in seconds; the time for 63 percent of the charge to leak away
E = relative dielectric constant, dimensionless
*E*0 = absolute dielectric constant in a vacuum, = 8.85 × 10-14 to less than 1 × 10-18
K = liquid conductivity, Siemens per centimeter (S/cm) Siemens (S) are also called mhos

Example: Benzene in a large tank could have a specific conductivity as low as 1 × 10-18 mho/cm and as high as 7.6 × 10-8 S/cm. The corresponding relaxation times for the two conductivities can be calculated as follows. Pure benzene has a dielectric constant of 2.5 to less than 1 × 10-18. Using the above equation:

1. *T*r = (8.85 × 10-14) (2.5)/(7.6 × 10-8) = 2.91 × 10-6 s
2. *T*r = (8.85 × 10-14) (2.5)/(1 × 1 × 18) = 2.21 × 105 s (this is in excess of 60 hr)

Benzene typically contains some water and has a higher conductivity than in the above example and has a much lower relaxation time.

The purity of a liquid has a great effect on its relaxation time, and thus its static hazard potential. In actual practice, relaxation times of a few seconds to an hour are encountered, depending on the purity and dryness of the liquid. This emphasizes the dangers of open sampling of tank contents soon after filling. If it is likely that the liquid being used has a low conductivity, it is important that enough time elapses between activities that can produce a static charge, such as loading a tank, and any activity that could cause a spark, such as sampling from the top of the tank.

In case (2) in the above example, a conductivity of 10-18 S/cm is so low that there may be little charge separation and little charge formation, and there may be no hazard even though the calculated relaxation time is extremely long. Materials with a half-time value of less than 0.012 s have been reported not to cause a hazard. A useful rule to remember is that the concept of relaxation is very important because it is possible for liquid in a tank to retain an electric charge for a long time if the liquid is a poor conductor, even if the tank is grounded. The specific conductivity, and therefore the relaxation time, is greatly affected by impurities. For example, the specific conductivity of benzene can vary from as long as 1 × 10-18 to about 7.6 × 10-8 S/cm, depending on its purity. It can vary significantly with the amount of water or other materials dissolved in the benzene.27

Resistivity

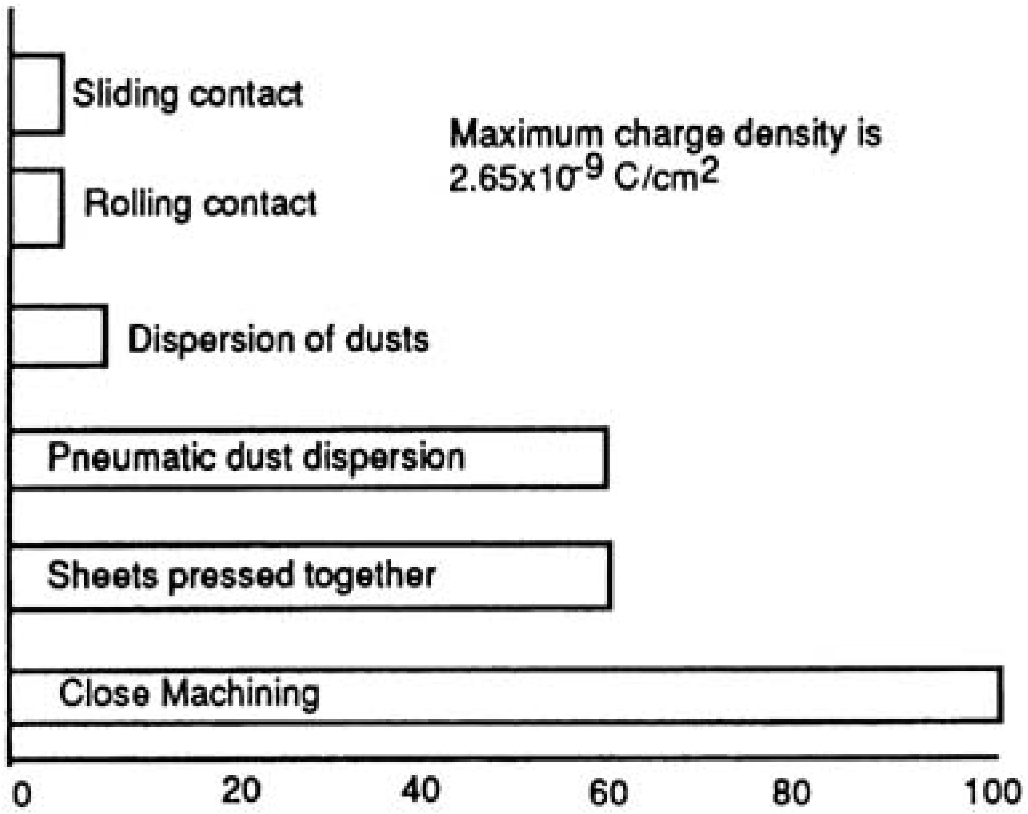
The extent of charge separation is dependent on the resistivity of the liquid. Some materials have a sufficiently high conductivity to render them safe in terms of static buildup. If the resistivity is low, charge separation is easy, but so is charge recombination through the liquid. If the resistivity is high, there may be appreciable charge separation without immediate recombination, leading to a high charge. If the resistivity is extremely high, there may be no charge separation, and there is no buildup of a charge. If the conductivity of a liquid falls in the hazardous range, it is possible to modify it by the use of a very small amount of an additive. Additives usually are a combination of a polyvalent metal salt of an acid such as carboxylic or sulfonic acid and a suitable electrolyte. Additives of this type can impart a conductivity of 10-8 S/m (Siemens per meter) in a 0.1 percent solution in benzene.2

A useful rule to remember is that when the resistivity of a liquid exceeds 1015 ohm centimeters (Ω·cm), or is less than 1010 Ω·cm, static generation or accumulation is negligible. Between these limits, the net generation of charges increases with the *maximum charge generation at* 103 Ω·cm. Styrene, for example, a commonly used monomer, has a resistivity of 4 × 1013 at 20°C,34 and therefore is capable of building up a potentially hazardous charge.

Static Charge Development

Static electrification of solids can occur in various ways. Different operations will produce the percentages of the theoretical maximum charge density shown in Fig. 5.7.33

Fig. 5.7 Percentage of maximum theoretical charge produced by various operations. (*Data from Eichel 1967*.33)



It should be noted that pure gases do not generate significant static electricity in transmission through pipes and ducts. Gases contaminated with rust particles or liquid droplets produce static, but this is not a problem in a closed, grounded piping system. *If these gases impinge on an ungrounded, conductive object, dangerous charges can accumulate on that object*. Wet steam, which contains water droplets, can develop charges. *If the water droplets contact an ungrounded conductor, that object can develop a static charge*.

Flammable gases may ignite when discharged to air during thunderstorms, even without a direct lightning hit. Dry hydrogen and occasionally other gases may ignite when they are discharged to air in normal weather. This may be so because the electric field developed by the ejected gases can develop a corona discharge which can cause ignition. The minimum ignition energy of hydrogen is only 0.02 mJ. A toroidal ring developed by the National Aeronautics and Space Administration (NASA) is reported to prevent unwanted discharge and subsequent ignition of a vent-stack outlet.27

Humidification

The conductivity of electrical nonconductors, such as plastics, paper, and concrete, depends on their moisture content. Relatively high moisture in these materials increases conductivity and therefore increases dissipation of static electricity. With relative humidity of 6070 percent or higher, a microscopic film of moisture covers surfaces, making them more conductive.

Humidification can and often should be practiced to reduce the hazard of static electricity, but should not be relied on entirely to remove all possibility of static discharge.

In winter, cold air brought into a building and heated to normal room temperature is extremely dry, often less than 510 percent relative humidity. When processing solid materials that can develop a static charge, this air should be humidified to reduce static hazards as well as improve the comfort of personnel.

Filling Liquid Containers

A fire during top loading of a flammable liquid into a tank constitutes a serious problem if there could be a flammable mixture in the vapor space. Static electricity can be generated by splashing if the liquid is top-loaded, so it is normal practice to fill with a dip pipe positioned so the tip of the dip pipe is near the bottom of the tank. This may not be sufficient to prevent static charge buildup, as a charge may be generated in the bottom of the tank before the pipe tip is fully submerged, and it is possible for the liquid to acquire a charge before it reaches the tank.

Product filters using cotton, paper, felt, or plastic elements are prolific generators of static electricity. It is considered that at least 30 s is necessary to dissipate this charge, although with dry nonconductive liquids, it may require as long as 500 s.

Loading a less volatile liquid into a tank where there was previously a more volatile liquid is particularly hazardous because the more volatile liquid may form a flammable mixture, and the less flammable material is often a poor conductor and will not readily dissipate static charge. This type of loading accounts for 7080 percent of severe losses at terminals.2 This appears to occur most often when the compartments are one-fourth to one-third full, and when the temperature is close to -1°C.

Inerting the tank while it is being filled will reduce the possibility of ignition by static electricity and is highly recommended when it is possible and practical. However, this is not always practical. In any case, if inerts are to be used, they must be added carefully, as the following example illustrates. Two firemen were fatally injured when an explosion occurred as they were attempting to use portable CO2 fire extinguishers to inert a tank truck. The source of ignition was believed to be a spark from the horn of the extinguisher to the latch on the tank truck. It was found that the voltage on the horn increased as the carbon dioxide "snow" passed down the horn to the outlet side.

Grounding and bonding lines, although very important, will not immediately dissipate the charge on the surface of a nonconducting liquid in a tank. A relaxation time for charge to be a dissipated should be allowed after filling or other operations to permit static charge on the liquid surface to dissipate to the dip pipe or tank shell. *The minimum time is 1 min, but longer periods are advisable with some liquids that have extremely low conductivity. Bottom loading may reduce the static electricity hazard but does not eliminate it*.35

5.8 Explosions

Development of Pressure

Exothermic reactions can led to high temperatures and in the case of large fires to large loss of property and severe damage from radiant energy. However, in many plant accidents it is the sudden generation of pressure that leads to severe damage, injury, and deaths. Hence, it can be stated that "pressure blows up plants, not temperature." Of course, temperature and pressure are closely related, but it is the pressure effect that is of concern in this section.

The word "deflagration" can be defined in several ways. One definition is "a reaction that propagates to the unreacted material at a speed less than the speed of sound in the unreacted substance."27 Another definition of deflagration is from Latin meaning "to burn down, or to burn rapidly with intense heat and sparks given off."30 A deflagration may be an explosion, but not all deflagrations are explosions (a violently burning fire may be a deflagration, but that is not an explosion). On the other hand, not all explosions are deflagrations (a steam boiler may explode, but that is not a deflagration).

An explosion is a sudden and violent release of energy. Usually it is the result, not the cause, of a sudden release of gas under high pressure. The presence of a gas is not necessary for an explosion. An explosion may occur from a physical or mechanical change, as in the explosion of a steam boiler, or from a chemical reaction. The explosion of a flammable mixture in a process vessel may be either a deflagration or a detonation, which differ fundamentally. Both can be very destructive. Detonations are particularly destructive, but are unlikely to occur in vessels.

A detonation is a reaction that propagates to unreacted material at a speed greater than the speed of sound in the unreacted material; it is accompanied by a shock wave and extremely high pressures for a very short time. It is debatable whether the flammable range is the same as the detonable range. Detonation limits normally are reported to be within the flammable limits, but the view is widely held that separate detonation limits do not exist.

Unconfined vapor clouds can both deflagrate and detonate, with a deflagration being much more likely. A detonation is more destructive, but a deflagration also can produce a damaging pressure wave. A deflagration can undergo transition to a detonation in a pipeline, but this is most likely in vessels.

If a flammable mixture may be present in process equipment, precautions should be taken to eliminate ignition sources. However, it is prudent to assume that, despite these efforts, a source of ignition will at some time occur.

Deflagration

The conditions for a deflagration to occur are that the gas mixture is within the flammable range and that there is a source of ignition or that the mixture is heated to its AIT.

For the burning of hydrocarbonair mixtures:

$$\frac{P_{2\mathrm{MAX}}}{P_1} = \frac{N_2 T_2}{N_1 T_1} = \frac{M_1 T_2}{M_2 T_1}$$

T = absolute temperature

M = molecular weight of gas mixture

N = number of moles in gas mixture

P = absolute pressure

1,2 = initial and final states

2MAX = final state maximum value

The maximum pressure rise for a deflagration of flammable mixtures is approximately as follows for initial absolute pressures of 140 bar, for initial temperatures of 0300°C, and for relatively small volumes of a few cubic meters:

$\frac{P_2}{P_1}$ = approximately 8 for hydrocarbonair mixtures

$\frac{P_2}{P_1}$ = approximately 16 for hydrocarbonoxygen mixtures

For conventionally designed pressure vessels:

$$\frac{P_b}{P_1} = \text{approximately } 4\text{–}5$$

*P*b = vessel bursting pressure

*P*1 = normal design pressure

*P*2 = pressure caused by deflagration

Therefore, in the absence of explosion relief, the deflagration explosion of a hydrocarbonair mixture is easily capable of bursting a vessel if it is operating near its design pressure when the deflagration takes place. For reactions operating at or near atmospheric pressure, such as many drying and solids processing operations, it may be practical to construct facilities that will withstand the maximum explosion pressure of most dustair and flammable gasair mixtures.

Detonations

Detonation of a gasair mixture may occur by direct initiation of detonation by a powerful ignition source or by transition from deflagration. This transition occurs in pipelines but is most unlikely in vessels. Two useful rules are:

1. Almost any gas mixture that is flammable is detonable if initiated with a sufficiently energetic source.
2. Detonation of a gasair mixture is possible in pipelines but is unlikely in vessels.

Bartknecht8 states that the range of detonability is narrower than the range of flammability. For example, the range of detonability of hydrogen in air is 1859 vol. percent, compared with the flammability of 475 vol. percent. With flammable gases in air, if the length-to-diameter ratio of a pipe or vessel is more than about 10:1, and the pipe diameter is above a critical diameter, 1225 mm, a detonation is possible.

Detonation Pressure

In the case of the burning of a flammable mixture of gases in a pipe with one end closed, a series of pressure waves traveling at the speed of sound moves through the unburned gas. Later waves traveling through the unburned gas, which has been heated by compression from the earlier waves, speed up because of the higher temperature and overtake the first wave, and a shock wave develops. Flame follows the shock wave and catches up with it, forming a detonation wave. A stable detonation wave may develop, which moves with supersonic speed relative to the unburned mixture, and peak incident (side-on) pressures are of the order of *30 times* the initial absolute pressure.

Reflected Pressure

Reflected pressure increases the pressure on a rigid surface if the shock wave impinges on the surface at an angle to the direction of the propagation of the wave. The maximum ratio of reflected pressure to incident (side-on) pressure when a strong shock wave strikes a flat surface head-on is 8:1. Furthermore, acceleration from a suddenly applied force of the detonation wave can double the load that a structure "feels." Table 5.3 shows overpressure that can be expected from typical detonations.36

TABLE 5.3 Overpressure from Detonations36

	Pressure (MPa)	Pressure (lb/in.2)
Incident overpressure	3.5	510
Maximum reflected pressure (wave strikes surface head-on)	28	4100
Load the structure feels (due to acceleration)	56	8100

(MPa means pressure in megapascals.)

Thus, the stable detonation wave may cause enormously high pressures at closed ends of pipes, bends, and tees, where the greatest destruction from a gaseous detonation may occur.

Geometry

The following are some factors to consider when detonation is possible:

1. Large length-to-diameter ratios promote the development of detonations; vessels should be designed with the lowest length-to-diameter ratio practicable if a detonation is possible.
2. Equipment such as tanks (not including pipelines) designed to withstand 3.5 MPa (about 500 psig) usually will be adequate to contain a detonation, with other safeguards, for flammable gases in air at atmospheric pressure.
3. Dished heads survive detonations better than do flat heads because of the more unfavorable incidence of flat heads.
4. If turns in a process line are necessary, two 45° bends or a long sweep elbow will greatly reduce reflected pressure compared with a single 90° elbow.
5. Restrictions such as orifices in pipelines may intensify a detonation by promoting pressure piling, which results when there are interconnected spaces such

that the pressure rise in one space causes a pressure rise in a connected space. The enhanced pressure in the latter then becomes the starting pressure for a further explosion.

6. Detonation may be extinguished when it enters a wider pipe from a smaller one, but the detonation may be regenerated somewhere along the longer pipe.
7. Flame arresters, if properly designed, can arrest detonations.

Explosion Violence

The *rate of pressure rise* is a measure of the violence of an explosion. The maximum rate of pressure rise for confined explosions is greatly affected by the volume of the vessel, with the influence of vessel volume on the rate of pressure rise being given by the following equation:

$$(\mathrm{d}p/\mathrm{d}t)_{\max}(V^{1/3}) = \text{a constant} = K_G$$

(d*p*/d*t*)max = maximum rate of pressurerise, bar/s
V = vessel volume, m3
*K*G = a specific material constant, (bar)(m)(s)-1

This is the *cubic law*, which states that for a given flammable gas, the product of the maximum pressure rise and the cube root of the vessel volume is a specific material constant, *KG*.

The cubic law allows the prediction of the course of an explosion of a flammable gas or vapor in a large vessel, based on labora-tory tests. It is valid only for the following conditions:

1. the same optimum concentration of the gasair mixture
2. same shape of reaction vessel
3. the same degree of turbulence
4. the same ignition source

Thus, to characterize an explosion, it is not enough to quote the maximum rate of pressure rise: the volume, vessel geometry, turbulence, and ignition energy must also be stated. Table 5.4 lists the *KG* values for some common flammable gases measured under laboratory conditions.

TABLE 5.4 *KG* Values of Gases, Spark-Ignited with Zero Turbulence,
Ignition Energy •10 J, *P*max = 7.4 bar8

Flammable Gas	KG (bar)(m)/s
Methane	55
Propane	75
Hydrogen	550

(From Bartknecht, *Explosion Course, Prevention Protection*, by Permission.)

It can be seen that the violence of an explosion with propane is about 1.5 times higher than one with methane, and one with hydrogen is about 10 times higher than one with methane. The explosive behavior of propane is representative of many flammable organic vapors in air. Some important relationships among pressure, temperature, turbulence, and vessel shape are discussed below.

1. *Explosion pressure is primarily the result of temperature reached during combustion, not a change in moles*. With complete combustion of propane in air there is a negligible change in moles of gas:

$$\{\ldots.\mathrm{air}\ldots.\}$$

$$C_3H_8 + 5O_2 + 18.8\,N_2$$

$$= 3CO_2 + 4H_2O + 18.8\,N_2$$

Number of moles at start = 24.8.
Number of moles after complete combustion = 25.8.

Therefore, explosion pressure usually develops principally from an increase in temperature, not an increases in gas moles, during the combustion process of many materials.

Peak explosion pressure at constant volume occurs near the stoichiometric concentration in air. If only a small part of the total volume of a container is filled by an explosive gasair mixture at atmospheric pressure, and the remainder of the vessel contains air, an explosion of this mixture can create enough pressure to severely damage containers that are designed to withstand only slight pressuresuch as buildings and low-pressure storage tanks.

2. *Initial pressure affects maximum explosion pressure and rate of pressure rise*. If the initial pressure is increased above atmospheric pressure, there will be a proportional increase in the maximum explosion pressure and in the rate of pressure rise. Reducing the initial pressure will cause a corresponding decrease in maximum explosion pressure until finally an explosion reaction can no longer be propagated through the gas mixture.
3. *Initial temperature affects maximum explosion pressure and rate of pressure rise*. The maximum explosion pressure decreases when the starting temperature increases at the same starting pressure because of the lower density and thus smaller mass of material within a confined volume at higher temperatures. The maximum rate of pressure rise, (d*p*/d*t*)max, increases as the initial temperature rises because the burning velocity increases with an increase in initial temperature.
4. *Initial turbulence increases the rate of pressure rise*. Initial turbulence greatly increases the rates of explosion-pressure rise.8,27 It has been found that with pentane and methane mixtures in air, (d*p*/d*t*)max can be five to nine times more with high initial turbulence than with no turbulence. The maximum explosion pressure is raised by about 20 percent. The course of explosions of flammable gases with a low normal speed of combustion, such as methane, is influenced by turbulence to a much higher degree than is the course of explosions with a high speed of combustion, such as hydrogen. Test data usually are obtained in equipment with a high degree of turbulence.
5. *Effect of vessel shape and increased initial pressure*. The maximum explosion pressure in confined vessels is not significantly affected by the volume or shape of the vessel in confined explosions for vessels that approximate the "cubic shape," that is, with a ratio of diameter to length (or vice versa) of about 1:1 to 1:1.5. In closed elongated vessels with central ignition, spherical ignition of the flame front will cause the flame to proceed swiftly in an axial direction. In the process, it compresses the unburned gases ahead of it, causing the violence of the explosion to increase, and pressure oscillations may occur.

5.9 Boiling Liquid Expanding Vapor Explosions (BLEVES)

Among the most damaging of accidents is a Boiling Liquid Expanding Vapor Explosion (BLEVE, pronounced BLEV-ee). This occurs when a pressure vessel containing liquid is heated so that the metal loses strength and ruptures. Typically, this happens when the vessel failure results from overheating upon exposure to fire. The failure usually is in the metal contacting the vapor phase; the metal in this area heats to a higher temperature because there is no liquid heat sink to keep the metal temperature from rising rapidly, as there is where metal contacts a liquid phase. A BLEVE can occur with both flammable materials and nonflammable materials, such as water. In all cases the initial explosion may generate a blast wave and missiles. If the material is flammable, it may cause a fire or may form a vapor cloud that then gives rise to a secondary explosion and fireball. Kletz states that BLEVEs can cause as many casualties as can unconfined vapor cloud explosions.2

The best known type of BLEVE involves liquefied petroleum gas (LPG). Once a fire impinges on the shell above the liquid level, the vessel usually fails within 1020 min. In the case of a BLEVE involving a flammable material, the major consequences are, in order of decreasing importance:

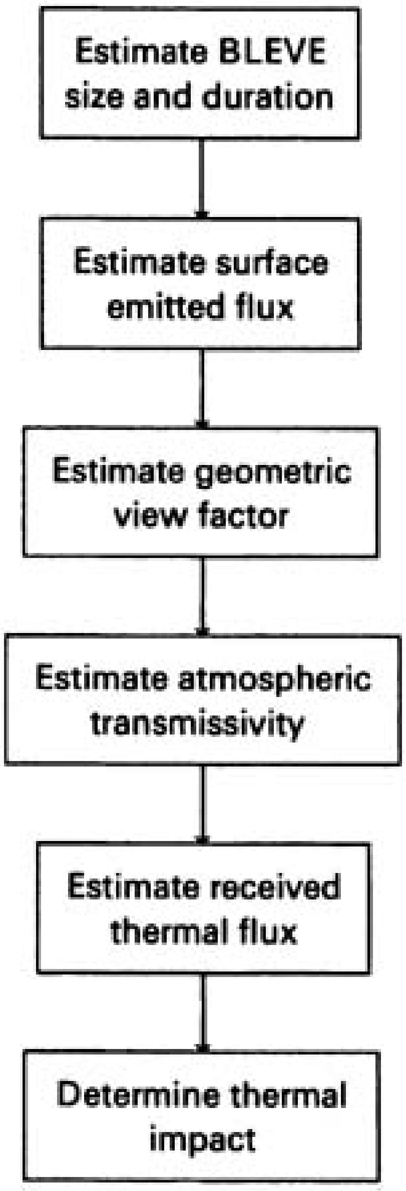
1. Thermal radiation from the resultant fireball
2. Fragments produced when the vessel fails
3. Blast wave produced by the expanding vapor/liquid

For example, a BLEVE of a propane sphere with a diameter of 50 ft, holding about 630,000 gal, could cause damage as far away as 13,600 ft, and radiation damage and fragmentation damage would each extend to about 3,000 ft.

In a fire, a tank containing liquid is most vulnerable in the shell at the vapor space because very little heat can be absorbed by the vapor, and the metal in the vapor space can heat up rapidly to a temperature where it will weaken rapidly. The metal contacting the liquid phase will heat up much less rapidly because the liquid can absorb significant amounts of heat, keeping the shell temperature down in that area for a significant amount of time. Thus, there is a dilemma: a partly full vessel may BLEVE sooner than will a full vessel, but a full vessel will have more fuel for the resulting fireball and fire than will a partly empty vessel.

Significant equipment and building damage from radiation is possible from a BLEVE. Wooden structures may be ignited if the radiant heat density at the structure's location exceeds the threshold value for ignition of wood. Severe damage from fragmentation can be expected in the area where 50 percent or more of the fragments may fall, or typically about 300 ft from the vessel.

A BLEVE can lead to shock waves, projectiles, and thermal radiation. The effects of a shock wave and projectiles were dealt with earlier; by far the most serious consequence of a BLEVE is the radiation received from the fireball. The following calculational procedure is used to determine thermal impact (details are available in CPQRA38):



5.10 Damage Estimates37

Damage estimates deal with the consequences of explosions and thermal radiation to both people and property. Physical models for explosions and thermal radiation generate a variety of incident outcomes: shock wave overpressure estimates, fragment velocities, and radiant flux. These models rely on the general principle that severity of outcome is a function of distance from the source of release. In addition to estimating the damage resulting from an explosion, it is also necessary to estimate how the consequences of these incident outcomes depend on the object of the study. To assess effects on human beings, damage estimates may be expressed as deaths or injuries. If physical property is the object, the damage estimates may be expressed as monetary losses.

Explosion Consequences

A principal parameter characterizing an explosion is the overpressure. Explosion effect modeling generally is based on TNT explosions to calculate the overpressure as a function of distance, Although the effect of a TNT explosion differs from that of a physical or a chemical explosion (particularly in the near-field), the TNT model is the most popular because a large data base exists for TNT explosions. Several kinds of energy may be released in an explosion; three basic types are: (1) physical energy, (2) chemical energy, and (3) nuclear energy. Nuclear energy is not considered here. Physical energy may take such forms as pressure energy in gases, strain energy in metals, or electrical energy. Chemical energy derives from a chemical reaction. Examples of explosions involving chemical energy are runaway exothermic reactions, including decomposition and polymerization. Table 5.5 summarizes the effects of explosion overpressure on structures. With respect to human casualties, heavy building damage usually is equated to a fatal effect, as the people inside the buildings probably would be crushed. People outside of buildings or structures are susceptible to direct blast injury (blast overpressure) and indirect blast injury (missiles or whole body translation).

TABLE 5.5 Effect of Explosion Overpressure on Structures

Pressure (psi)	Damage
0.02	Annoying noise (137 dB if of low, 1015 Hz frequency)
0.03	Breaking of large glass windows under strain
0.04	Loud noise (143 dB), sonic boom, glass failure
0.10	Breakage of small glass windows under strain
0.15	Typical pressure for glass breakage
0.30	"Safe distance" (probability 0.95 of no serious damage below this value); projectile limits; some damage to house ceilings; 10% window glass broken
0.40	Limited minor structural damage
0.51.0	Large and small windows usually shattered; occasional damage to window frames
0.70	Minor damage to house structures
1.00	Partial demolition of houses; houses made uninhabitable
12.00	Corrugated asbestos shattered; corrugated steel or aluminum panels, fastenings fail, followed by buckling; wood panel fastenings of standard housing fail; panels blown in
1.30	Steel frames of clad buildings slightly distorted
2.00	Partial collapse of walls and roofs of houses
23.00	Concrete or cinder blocks shattered if not reinforced
2.30	Lower limit of serious structural damage
2.50	50% destruction of brickwork of houses
3.00	Heavy machines (300 lb), industrial buildings suffered little damage; steel frame buildings distorted and pulled away from foundation
34.00	Frameless, self-framing steel panel building demolished; rupture of oil storage tanks
4.00	Cladding of light industrial buildings ruptured
5.00	Wooden utility poles snapped
57.00	Nearly complete destruction of houses
7.00	Loaded railcars overturned
78.00	Brick panels, 812 in. thick, not reinforced, fail by shearing or flexure
9.00	Loaded train boxcars completely demolished
10.00	Probable total destruction of buildings; heavy machine tools (7,000 lb) moved and badly damaged; very heavy machine tools (12,000 lb) survive
300.00	Limit of crater lip.

(Courtesy of Center for Chemical Process Safety, American Institute of Chemical Engineers.38)

Relatively high blast overpressures (>15 psig) are necessary to produce a human fatality from a direct blast. Instead, the major threat is produced by missiles or by whole body translation. Fatalities arising from whole body translation are mainly due to head injury from decelerative impact. Injury to people due to fragments usually results from either penetration by small fragments or blunt trauma from large fragments. TNO39 suggested that projectiles with a kinetic energy of 100 J can cause fatalities. Table 5.6 shows damage to people (physiological damage) as a function of overpressure.

TABLE 5.6 Physiological Damage as a Result of Overpressure

Effect	Peak Overpressure (psi)
Knock down	1.0
Ear drum damage	5.0
Lung damage	15
Threshold for fatalities	35
50% fatalities	50
99% fatalities	65

Radiation Consequences

The effect of thermal radiation on people and objects is determined by one of two approaches:

1. Simple tabulations based on experimental results.
2. Theoretical models based on the physiology of the skin burn response.

Data on time to pain threshold40 are summarized in Table 5.7. For comparison, solar radiation intensity on a clear, hot summer day is about 320 Btu/hr ft2 (1 kW/m2). Other criteria for thermal radiation damage are shown in Table 5.8.38

TABLE 5.7 Time to Pain Threshold for Varying Levels of Radiation40

Radiation Intensity (Btu/hr/ft2)	Radiation Intensity (kW/m2)	Time to Pain Threshold (s)
500	1.74	60
740	2.33	40
920	2.90	30
1500	4.73	16
2200	6.94	9
3000	9.46	6
3700	11.67	4
6300	19.87	2

(Courtesy American Petroleum Institute.)

TABLE 5.8 Effects of Thermal Radiation

Radiation Intensity (kW/m2	Observed Effect
37.5	Sufficient to cause damage to process equipment.
25.0	Minimum energy required to ignite wood at indefinitely long exposures.
12.5	Minimum energy required for piloted ignition of wood, melting of plastic tubing.
9.5	Pain threshold reached after 6 seconds; second-degree burns after 20 seconds.
4.0	Sufficient to cause pain to personnel if unable to reachcover within 20 seconds; however, blistering of the skin (second degree burns) is likely; 0% lethality.
1.6	Will cause no discomfort for long exposure.

(Courtesy of Center for Chemical Process Safety, American Institute of Chemical Engineers.38)

The effect of thermal radiation on structures depends on whether they are combustible or not, and the nature and duration of the exposure. Thus, wooden materials will fail because of combustion, whereas steel will fail because of thermal lowering of the yield stress.

Unconfined Vapor Cloud Explosions (UVCE)

When a large amount of volatile material is released rapidly to the atmosphere, a vapor cloud forms and disperses. If the cloud is ignited before it is diluted below its lower flammability limit, an uncontrolled vapor cloud explosion will occur. This is one of the most serious hazards in the process industries. Both shock waves and thermal radiation will result from the explosion, with the shock waves usually the more important damage producers. UVCEs usually are modeled by using the TNT model.38 The energy of the blast wave generally is only a small fraction of the energy available from the combustion of all the material that constitutes the cloud; the ratio of the actual energy released to that available frequently is referred to as the "explosion efficiency." Therefore, the TNT weight equivalent of a UVCE includes an explosion efficiency term, which typically is an empirical factor ranging from 1 percent to 10 percent. The explosion effects of a TNT charge are well documented.

Physical Explosions

A physical explosion usually results from the production of large volumes of gases by nonchemical means. The gases necessary for a physical explosion may be those already existing, such as compressed nitrogen released suddenly from a ruptured cylinder, or steam released explosively from a crack in a steam drum.

The following are some settings and situations in which physical explosions have been known to take place:

1. steam boilers
2. hydraulic overfill of tanks or pipes with external applied pressure (as in pressure testing)
3. compressed air tanks
4. deadheaded pumps
5. thermal expansion of tanks or pipes
6. liquid cryogenic fluids on water (such as liquid methane on water)
7. water suddenly mixed with sulfuric acid (also may cause a chemical explosion)
8. BLEVE with superheated liquid (flammable or nonflammable) (see next section)
9. explosion of grinding wheel at too high a speed
10. liquid water in molten MgCl2 solution at high temperatures
11. implosions due to vacuum
12. overpressured refrigerant systems
13. molten metals exploding violently on contact with water
14. some molten metals exploding when mixed with each other
15. the mixing of two immiscible liquids whose boiling points are not widely separated

Steam boilers are commonly used in power plants and industries of all kinds. They generally are taken for granted now, but in the second half of the nineteenth century boilers blew up with alarming regularity. Records indicate that from 1870 to 1910 there were at least 10,000 boiler explosions in the United States and adjacent areas of Canada and Mexico; that is, more than one recorded explosion every 36 hours! By 1910, the rate had jumped to between 1,300 and 1,400 per year. On October 8, 1894, in the Henry Clay Mine in Shamokin, Pennsylvania, 27 boilers disintegrated almost simultaneously! Mainly because of the incorporation of the ASME Boiler Code into laws, boiler explosions have decreased dramatically.41

When a pressurized vessel ruptures, the resulting stored energy is released. This energy can cause a shock wave and accelerate vessel fragments. If the contents are flammable, ignition of the released gas could produce additional effects. There is a maximum amount of energy in a bursting vessel that can be released, and it is released in the following proportions:36

Distribution of Energy When Vessel Ruptures

Type of Failure	Strain Energy	Kinetic Energy of Fragments	Shock Wave Energy
Brittle failure	<10%	•20%	up to 80%
Plug ejection	small	up to 6080%	2040%

The relative distribution of these energy components will change over the course of the explosion, but most of the energy is carried by the shock wave with the remainder going to fragment kinetic energy. To estimate the damage resulting from the shock wave from a physical explosion, the TNT model is used widely. To determine the TNT equivalent of a physical explosion, the total energy in the system must be estimated. For a physical explosion, if the expansion occurs isothermally, and ideal gas laws apply, then the TNT equivalent of the explosion can be calculated. This energy then can be used to estimate overpressure at any distance from the explosion. The analogy of the explosion of a container of pressurized gas to a point source explosion of TNT is not appropriate in the near-field. Prugh42 suggests a correction method using a virtual distance *R*v from an explosion center.

In addition to shock wave effects, a major hazard of a ruptured gas-filled vessel is from projectiles. To estimate damage from projectiles, both the initial velocity and the range are required. A simplified method for calculating the initial velocity uses the following equation:43

$$u = 2.05(PD^3/W)^{0.5}$$

u = initial velocity, ft/s
P = rupture pressure, psig
D = fragment diameter, in.
W = weight of fragments, lbs

Clancey44 gives the following values for initial velocity for the majority of fragments from a TNT explosion:

1. thin case: 8,000 ft/s
2. medium case: 6,000 ft/s
3. thick case: 4,000 ft/s

Once the initial velocity has been determined, the maximum range of the fragment, ignoring air resistance, can be estimated from

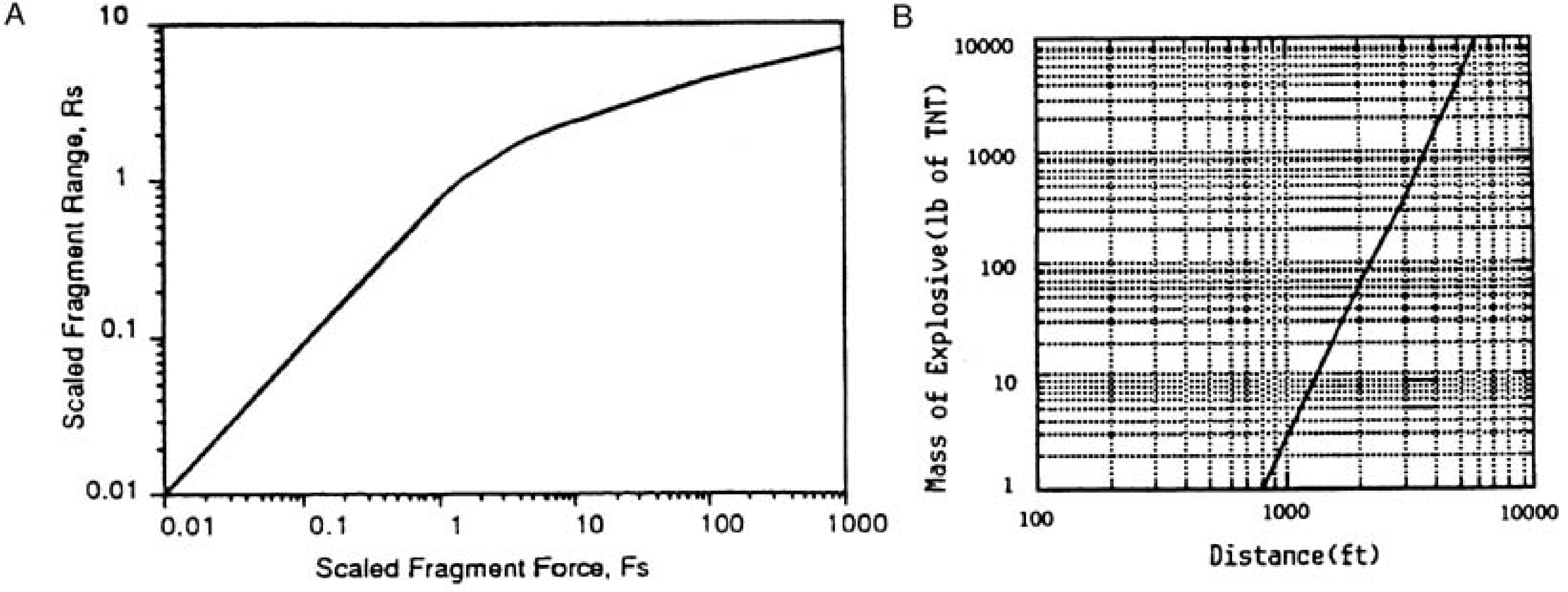
$$R_{\max} = \frac{u^2}{g}$$

where *R*max is the maximum range of fragments and *g* is the acceleration of gravity.

If the above values for typical velocity are substituted into the above equation, a maximum range of 5 × 105 ft is possible. Therefore, it is clearly necessary to include air resistance. To include air resistance, a value of *C*D, the drag coefficient, must be estimated. The drag coefficient ranges from 0.48 for a sphere to 2 for flow perpendicular to a flat strip, and for most fragments ranges from 1.5 to 2.0.

If one knows the air density, drag coefficient, exposed area of the fragment, mass of the fragment, and the initial velocity, the maximum range *R* can be calculated with the aid of Fig. 5.8.45 Although this technique gives the maximum range, most fragments do not travel the maximum distance but fall at distances between 0.3 and 0.8 of the maximum.

Fig. 5.8 (a) Scaled fragment range vs. scaled force. (*Baker et al.* 1983.45) (b) Maximum horizontal range of blast fragments. (*Clancy*.44)



where: $F_s = \dfrac{r_o C_D A_D u^2}{Mg}$ $R_s = \dfrac{r_o C_D A_D R}{M}$

and: r_o = density of air lb/cu ft C_D = drag coefficient

A_D = exposed area of fragment, ft2 R = max range of fragment, ft

M = mass of fragment, lb U = fragment initial velocity, fps

$g = 32.17 \text{ ft. lb.}_m/\text{lb}_f s^2$

$$g = 32.17\ \mathrm{ft.\ lb._m/lb_f s^2}$$

The energy required to impart an initial velocity of *u* to a fragment is

$$E = \frac{1}{2}mu^2$$

m is the mass of the fragment (lb)
u the initial velocity (ft/s)

Example. A high pressure vessel containing air at 600 bar has ruptured, leading to 15 fragments of approximately equal mass (85 lb), one of which was found as far as 400 ft from the vessel. This fragment has a drag coefficient of 1.5 and an exposed area of 3 ft2. Assuming that 20 percent of the explosion energy went to energy of the fragments, estimate the energy of the explosion in weight equivalent TNT. The air density is 0.081 lb/ft3.

Procedure. Assuming that the fragment found at 400 ft is at the maximum range for the fragments, the scaled fragment range *R*s can be calculated:

$$\begin{aligned} R_s &= \frac{r_0 C_D A_D R}{M} \\ &= \frac{0.081 \times 1.5 \times 3 \times 400}{85} \\ &= 1.7 \end{aligned}$$

From Fig. 5.8. we obtain a scaled force (*F*s) of approximately 5. The initial velocity of the fragment then can be calculated as

$$\begin{aligned} u &= \sqrt{\frac{MgF_s}{r_0 C_D A_D}} \\ &= \sqrt{\frac{85 \times 32.17 \times 5}{0.081 \times 1.5 \times 3}} \\ &= 194\,\mathrm{ft/s} \end{aligned}$$

The energy required to give the fragment this initial velocity is

$$\begin{aligned} E &= \tfrac{1}{2}(85)(194)^2 \\ &= 1.6 \times 106\,\mathrm{lb\,ft^2/s^2} \\ &= 64\,\mathrm{BTU} \end{aligned}$$

Since there were 15 fragments, the total energy of the explosion that went into fragment kinetic energy is 15 × 635.8 BTU = 9537 BTU. If only 20 percent of the explosion energy went into fragment kinetic energy, then the total explosion energy is 47,680 BTU, which is the equivalent of 23.8 lb of TNT. Using the method of Clancy44, 2.4 lb TNT can provide a maximum range of 950 ft for projectiles (Fig. 5.8 (b)).

5.11 Mechanical Heat

Mechanical motion in fluids becomes kinetic energy and may become heat in devices with rotating parts. Mechanical heat input from rotating agitators, pump impellers, and other mechanical equipment must be taken into account in the design of process equipment, particularly in systems contain-ing reactive chemicals. This section will pro-vide some guidelines for the analysis of individual cases involving pumps and agitated tanks.4648

Some useful rules are as follows:

1. A deadheaded pump is a pump operating full of liquid and with inlet and outlet valves closed.
2. Almost all deadheaded centrifugal pumps with motors of three horsepower or larger are headed for trouble if left deadheaded. (Depending on the horsepower, a few minutes may be too long.)
3. The heat input from the rotating impeller in a deadheaded centrifugal pump is always a large value relative to the heat sink of the fluid and the pump.
4. It is not necessary for there to be a chemical reaction in a pump for an explosion to take place. Deadheaded pumps containing only water or brine have blown up.
5. An agitator or a circulating pump left on in a vessel of a reactive chemical may heat up the contents enough to cause a runaway reaction.
6. All centrifugal pumps with motors larger than 3 hp should be protected in some way to prevent deadheading. A temperature alarm in the casing is a minimum form of protection. A better way may be to have the high-temperature alarm wired to the process control computer, to both alarm and shut off the pump. Other systems are available and may be used; they may include (but are not limited to) a relief valve on the pump, a minimum flow valve, and a flow orifice in the recirculating line. A relief valve on a pump relieving back to the pump inlet may not eliminate the problem of heat buildup in a deadheaded pump and usually should be avoided unless other protective measures are used such as a high-temperature device.
7. An ammeter on the pump motor usually is not a reliable means of detecting deadheaded conditions. The low power factors often experienced with pump motors, and the nature of pump curves, often make it difficult to distinguish between normal running and deadheaded conditions using an ammeter.
8. For mechanical heat equivalent, the following are recommended: (a) For pumps, use 50 percent of the connected motor horsepower for centrifugal pumps that are deadheaded, unless better information is available. (b) For agitators, use 100 percent of the vendor rated shaft input horsepower for the input shaft (total power less drive and bearing inefficiencies) for the actual material in the vessel.

5.12 Vacuum49

Ask any chemical engineers who have had some plant experience what they know about vacuum, and they probably will smile and tell a tale about some piece of equipment that tried to turn itself inside out. Usually no one was hurt, and often there is no massive leakagebut not always!

The design for the internal pressure condition of vessels usually is straightforward and well understood. Under vacuum conditions, equipment is subject to external pressure from the atmosphere; and the design for external pressures is more difficult than that for internal pressures. The devious ways in which external pressure can be applied often may be overlooked.

The following are some obvious causes of vacuum collapse:

1. liquid withdrawal by pump or gravity draining
2. removal of gas or vapor by withdrawing with a blower, fan, or jet
3. siphoning of liquids.

Less obvious causes include:

1. condensation of vapor
2. cooling of hot gas
3. combination of cooling and condensation of a mixture of gas and condensable vapor.

Sometimes obscure causes of vacuum collapse include:

1. Absorption of a gas in a liquid; for example, ammonia in water, carbon dioxide in water, hydrogen chloride in water.
2. Reaction of two or more gases to make a liquid or solid; for example, ammonia plus hydrogen bromide to form ammonium bromide.
3. Reaction of a gas and a solid to form a solid; for example, corrosion in a tank, air plus Fe or FeO forming Fe_2O_3 in the presence of water.
4. Reaction of a gas and a liquid to give a liquid; for example, chlorination, hydrogenation, ethylation.
5. Sudden dropping of finely divided solids in a silo, creating a momentary vacuum that can suck in the sides of the silo.
6. Flame arrestors plugging; for example:
    1. In styrene service, vapor may condense in flame arrestors, and the liquid formed is low in inhibitor; the liquid may polymerize and plug off the arrestor. Possible solutions: clean the arrestor frequently or use a PVRV (pressure-vacuum-relief valve).
    2. Liquid service in cold weather: vapor may condense in a flame arrestor and the liquid formed may freeze and plug the arrestor. Possible solution: heat and insulate the arrestor to prevent condensation.
7. Maintenance and testing. It is not a good idea to apply vacuum on a vessel during maintenance or testing without full knowledge of the external pressure rating unless a suitable vacuum relief device is in place and operable.

Protective Measures for Equipment

If equipment may be subject to vacuum, consideration should be given to designing the equipment for full vacuum. This may eliminate the need for complicated devices such as vacuum relief valves and instruments; if they are used, designing the equipment for full vacuum will prevent collapse of the vessel if the instruments or relief valves fail or plug.

A disadvantage of this approach is that it usually is expensive. However, when the total cost of a suitably instrumented vessel not designed for vacuum is compared with the cost of a vessel designed for vacuum but without the extra equipment, the difference may be small or negligible, and the vessel designed for vacuum will be inherently safer. If a vessel is designed for vacuum, precautions should be taken to ensure that internal or external corrosion will not destroy the integrity of the vessel.

5.13 Regulations*

Regulations are a major consideration in the design and operation of chemical facilities. This section provides a description of the significant process requirements. Details of the regulations are available on the Internet or from government agencies, such as the U.S. Department of Labor, or from publications such as those produced by the Thompson Publishing Group and by Primatech, Inc.

*This section was prepared with the help of William Carmody, Midland, Michigan. Carmody has had more than 30 years experience in chemical and manufacturing operations for The Dow Chemical Company, Midland, Michigan and six years in Safety and Loss Consulting for Midland Engineering Limited, Midland, Michigan. He has developed entire PSM programs and has conducted many Process Hazard Analyses.

Abbreviations used in Government Regulations information:
CFR Code of Federal Regulations
EPA Environmental Protection Agency
EPCRA Emergency Planning and Community Right To Know Act
HAZWHOPER Hazardous Waste Operations & Emergency Response
HHC Highly Hazardous Chemicals
MSDS Material Safety Data Sheet
NIOSH National Institute for Occupational Safety and Health
OSHA Occupational Safety and Health Administrates
PHA Process Hazard Analysis
PPA Pollution Prevention Act
PSM Process Safety Management
RCRA Resource Conservation and Recovery Act
RMP Risk Management Plans
SARA Superfund Amendments and Reauthorization Act
TRI Toxics Release Inventory

Process Safety Management

On February 24, 1992, the U.S. Department of Labor, Occupational Safety and Health Administration (OSHA) promulgated a final rule, 29 CFR Part 1910.119, "Process Safety Management of Highly Hazardous Chemicals."

OSHA administrates regulations whose objectives are primarily involved with protecting workers. This can be regarded as "inside the fence line." This is a safety issue and is addressed in this section. The rule requires employers to effectively manage the process hazards associated with chemical processes to which the rule applies. OSHA is responsible for the Process Safety Management (PSM) program that is used to prevent or minimize the consequences of catastrophic releases of toxic, reactive, flammable, or explosive chemicals. Standard Number CFR 1910.119 contains requirements for preventing or minimizing the consequences of catastrophic releases of toxic, reactive, flammable, or explosive chemicals. It establishes procedures for PSM that will protect employees by preventing or minimizing the consequences of chemical accidents involving highly hazardous chemicals. The requirements in this standard are intended to eliminate or mitigate the consequences of such releases.

PSM applies to a process involving a chemical at or above the specified threshold quantities listed in 1910.119, Appendix A, and also listed in Table 5.9. The requirements of the rule are also applicable to processes that involve a flammable liquid or gas on-site, in one location, in a quantity of 10,000 lb or more, except for hydrocarbon fuels used solely for workplace consumption as a fuel, or flammable liquids stored in atmospheric pressure tanks.

TABLE 5.9 List of Highly Hazardous Chemicals, Toxics, and Reactive Chemicals (Mandatory)

Standard Number: 1910.119 Appendix A (on the Internet)

This is a listing of toxic and reactive highly hazardous chemicals which present a potential for a catastrophic event at or above the threshold quantity.

Chemical Name	CAS[a]	Threshold Quantity[b]
Acetaldehyde	75070	2,500
Acrolein (2-Popenal)	107028	150
Acrylyl chloride	814686	250
Allyl chloride	107051	1,000
Allylamine	107119	1,000
Alkylaluminum	Varies	5,000
Ammonia, anhydrous	7664417	10,000
Ammonia solutions (greater than 44% ammonia by weight)	7664417	15,000
Ammonium perchlorate	7790989	7,500
Ammonium permanganate	7787362	7,500
Arsine (also called arsenic hydride)	7784421	100
Bis(chloromethyl) ether	542881	100
Boron trichloride	10294345	2,500
Boron trifluoride	7637072	250
Bromine	7726956	1,500
Bromine chloride	13863417	1,500
Bromine pentafluoride	7789302	2,500
Bromine trifluoride	7787715	15,000
3-Bromopropyne (also called propargyl bromide)	106967	100
Butyl hydroperoxide (tertiary)	75912	5,000
Butyl perbenzoate (tertiary)	614459	7,500
Carbonyl chloride (see Phosgene)	75445	100
Carbonyl fluoride	353504	2,500
Cellulose Nitrate (concentration greater than 12.6% nitrogen)	9004700	2,500
Chlorine	7782505	1,500
Chlorine dioxide	10049044	1,000
Chlorine pentrafluoride	13637633	1,000
Chlorine trifluoride	7790912	1,000
Chlorodiethylaluminum (also called diethylaluminum chloride)	96106	5,000
1-chloro-2, 4-dinitrobenzene	97007	5,000
Chloromethyl methyl ether	107302	500
Chloropicrin	76062	500
Chloropicrin and methyl Bromide mixture	None	1,500
Chloropicrin and methyl Chloride mixture	None	1,500
Commune hydroperoxide	80159	5,000
Cyanogen	460195	2,500
Cyanogen chloride	506774	500
Cyanuric fluoride	675149	100
Diacetyl peroxide (concentration greater than 70%)	110225	5,000

Diazomethane	334883	500
Dibenzoyl peroxide	94360	7,500
Diborane	19287457	100
Dibutyl peroxide (tertiary)	1100504	5,000
Dichloro acetylene	7572294	250
Dichlorosilane	4109960	2,500
Diethylzinc	557200	10,000
Diisopropyl peroxydicarbonate	105646	7,500
Dilauroyl peroxide	105748	7,500
Dimethyldichlorosilane	75785	1,000
Dimethylhydrazine, 1,1-	57147	1,000
Dimethylamine, anhydrous	124403	2,500
2,4-dinitroaniline	97029	5,000
Ethyl methyl ketone peroxide (also methyl ethyl ketone peroxide; concentration greater than 60%)	1338234	5,000
Ethyl nitrite	109955	5,000
Ethylamine	75047	7,500
Ethylene fluorohydrin	371620	100
Ethylene oxide	75218	5,000
Ethyleneimine	151564	1,000
Fluorine	7782414	1,000
Formaldehyde (formalin)	50000	1,000
Furan	110009	500
Hexafluoroacetone	684162	5,000
Hydrochloric acid, anhydrous	7647010	5,000
Hydrofluoric acid, anhydrous	7664393	1,000
Hydrogen bromide	10035106	5,000
Hydrogen chloride	7647010	5,000
Hydrogen cyanide, anhydrous	74908	1,000
Hydrogen fluoride	7664393	1,000
Hydrogen peroxide (52% by weight or greater)	7722841	7,500
Hydrogen selenide	7783075	150
Hydrogen sulfide	7783064	1,500
Hydroxylamine	7803498	2,500
Iron, pentacarbonyl	13463406	250
Isopropylamine	75310	5,000
Ketene	463514	100
Methacrylaldehyde	78853	1,000
Methacryloyl chloride	920467	150
Methacryloyloxyethyl isocyanate	30674807	100
Methyl acrylonitrile	126987	250
Methylamine, anhydrous	74895	1,000
Methyl bromide	74839	2,500
Methyl chloride	74873	15,000
Methyl chloroformate	79221	500
Methyl ethyl ketone peroxide (concentration greater than 60%)	1338234	5,000
Methyl fluoroacetate	453189	100
Methyl fluorosulfate	421205	100
Methyl hydrazine	60344	100
Methyl iodide	74884	7,500
Methyl isocyanate	624839	250
Methyl mercaptan	74931	5,000
Methyl vinyl ketone	79844	100
Methyltrichlorosilane	75796	500
Nickel carbonly (nickel tetracarbonyl)	13463393	150
Nitric acid (94.5% by weight or greater)	7697372	500
Nitric oxide	10102439	250
Nitroaniline (para)Nitroaniline	100016	5,000
Nitromethane	75525	2,500
Nitrogen dioxide	10102440	250
Nitrogen oxides (NO; NO(2); N204; N203)	10102440	250
Nitrogen tetroxide (also called nitrogen peroxide)	10544726	250
Nitrogen trifluoride	7783542	5,000
Nitrogen trioxide	10544737	250
Oleum (6580% by weight; also called fuming sulfuric acid)	8014947	1,000
Osmium tetroxide	20816120	100
Oxygen difluoride (fluorine monoxide)	7783417	100
Ozone	10028156	100
Pentaborane	19624227	100
Peracetic acid (concentration greater 60% acetic acid; also called peroxyacetic acid)	79210	1,000
Perchloric acid (concentration greater than 60% by weight)	7601903	5,000
Perchloromethyl mercaptan	594423	150
Perchloryl fluoride	7616946	5,000
Peroxyacetic acid (concentration greater than 60% acetic acid; also called peracetic acid)	79210	1,000
Phosgene (also called carbonyl chloride)	75445	100
Phosphine (Hydrogen phosphide)	7803512	100
Phosphorus oxychloride (also called phosphoryl chloride)	10025873	1,000
Phosphorus trichloride	7719122	1,000
Phosphoryl chloride (also called phosphorus oxychloride)	10025873	1,000

Propargyl bromide	106967	100
Propyl nitrate	62734	2,500
Sarin	107448	100
Selenium hexafluoride	7783791	1,000
Stibine (antimony hydride)	7803523	500
Sulfur dioxide (liquid)	7446095	1,000
Sulfur pentafluoride	5714227	250
Sulfur tetrafluoride	7783600	250
Sulfur trioxide (also called sulfuric anhydride)	7446119	1,000
Sulfuric anhydride (also called sulfur trioxide)	7446119	1,000
Tellurium hexafluoride	7783804	250
Tetrafluoroethylene	116143	5,000
Tetrafluorohydrazine	10036472	5,000
Tetramethyl lead	75741	1,000
Thionyl chloride	7719097	250
Trichloro (chloromethyl) silane	1558254	100
Trichloro (dichlorophenyl) silane	27137855	2,500
Trichlorosilane	10025782	5,000
Trifluorochloroethylene	79389	10,000
Trimethyoxysilane	2487903	1,500

[a] Chemical abstract service number.

[b] Threshold quality in pounds (amount necessary to be covered by this standard).

Process means any activity involving a highly hazardous chemical including any use, storage, manufacturing, handling, or the on-site movement of such chemicals, or combination of these activities. For purposes of this definition, any group of vessels that are interconnected and separate vessels which are located such that a highly hazardous chemical could be involved in potential release shall be considered a single process.

The PSM elements required by 29 CFR Part 1910.119 are briefly described in the following:

Employee Participation

Employers must develop a written plan of action for how they will implement employee participation requirements. Employers must consult with employees, affected contractors, and their representatives on the conduct and development of process hazard analyses and on other elements of the standard. They must have access to information developed from the standard, including process hazard analyses.

Process Safety Information

Employers must compile considerable documented process safety information on the hazards of chemicals used in a covered process as well as information on the process technology and equipment before conducting the process hazard analyses required by the standard.

Process Hazard Analysis (PHA)

Employers must perform an analysis to identify, evaluate, and control hazards on processes covered by this standard. The process hazard analysis shall be appropriate to the complexity of the process and shall identify, evaluate, and control the hazards involved in the process. The OSHA standard specifies a number of issues that the analysis must address, as well as requirements for who must conduct the analysis, how often it must be performed, and response to its findings. Methodologies that are appropriate include:

1. what-if
2. checklists
3. what-if/checklist
4. Hazard and Operability Study (HAZOP)
5. Failure Mode and Effects Analysis (FMEA)
6. fault tree analysis

The selection of a PHA methodology or technique will be influenced by many factors including the amount of existing knowledge about the process. All PHA methodologies are subject to certain limitations. The team conducting the PHA needs to understand the methodology that is going to be used. A PHA team can vary in size from two people to a number of people with varied operational and technical backgrounds. Some team members may only be a part of the team for a limited time. The team leader needs to be fully knowledgeable in the proper implementation of the PHA methodology that is to be used and should be impartial in the evaluation. The other full- or part-time team members need to provide the team with expertise in areas such as process technology, process design, operating procedures, and practices.

Operating Procedures

Employers must develop and implement written operating instructions for safely conducting activities involved in each covered process consistent with the process safety information. The written procedures must address steps for each operating phase, operating limits, safety and health considerations, and safety systems and their functions. Included must be normal operation, startup, shutdown, emergency operations, and other operating parameters.

Training

The proposal requires training for employees involved in covered processes. Initial training requires all employees currently involved in each process, and all employees newly assigned, be trained in an overview of the process and its operating procedures. Refresher training shall be provided at least every three years, and more often if necessary, to each employee involved in the process. After training, employees must ascertain that workers have received and understood the training.

Contractors

Employers must inform contract employees prior to the initiation of the contractor's work of the known potential fire, explosion, or toxic release hazards related to the contractor's work and the process. Contract employees and host employers must ensure that contract workers are trained in the work practices necessary to perform their jobs safely and are informed of any applicable safety rules of the facility work and the process.

Pre-Startup Safety Review

Employers must perform a pre-startup safety review for new facilities and for modified facilities when the modification is significant enough to require a change in the process safety information. The safety review shall confirm that prior to the introduction of highly hazardous chemicals to a process:

1. Construction and equipment is in accordance with design specifications.
2. Safety, operating, maintenance, and emergency procedures are in place and are adequate.
3. For new facilities, a process hazard analysis has been performed and recommendations have been resolved or implemented before startup.
4. Modified facilities meet the requirements contained in management of change.

Mechanical Integrity

Employers must ensure the initial and on-going integrity of process equipment by determining that the equipment is designed, installed, and maintained properly. The standard requires testing and inspection of equipment, quality assurance checks of equipment, spare parts and maintenance materials, and correction of deficiencies. The following process equipment is targeted in this proposal: pressure vessels and storage tanks; piping systems (including valves); relief and vent systems and devices; emergency shutdown systems; controls, and pumps.

Hot Work Permit

Employers must have a hot work program in place and issue a permit for all hot work operations conducted on or near a covered process.

Management of Change

Employers must establish and implement written procedures to manage changes (except for "replacements in kind") to process chemicals, technology, equipment, and procedures; and, changes to facilities that affect a covered process. Employees involved in operating a process and maintenance and contract employees whose tasks will be affected by a change in the process shall be informed of, and trained in, the change prior to startup of the process or affected part of the process. The procedures shall ensure that the necessary time period for the change and authorization requirements for the proposed change are addressed.

Incident Investigation

Employers must investigate each incident that resulted in, or could reasonably have resulted in a catastrophic release of highly hazardous chemical in the workplace. An incident investigation shall be initiated as promptly as possible, but not later than 48 hr following the incident. A report shall be prepared at the conclusion of the investigation.

Although not stressed by the regulations, the objective of the incident investigation should be the development and implementation of recommendations to ensure the incident is not repeated. This objective should apply not only to the process involved, but also to all similar situations having the same potential. In major incidents, the Chemical Safety Board's investigation of reports serves as a vehicle to communicate to a much broader audience than the organizations that had the incident.

Emergency Planning and Response

Employers must establish and implement an emergency action plan for the entire plant in accordance with the provisions of OSHA's emergency action plan to meet the minimum requirements for emergency planning. This is the only element of PSM that must be carried out beyond the boundaries of a covered process.

Compliance Audits

Employers must certify that they have evaluated compliance with the provisions of this section at least every three years to verify that procedures and practices developed under the standard are adequate and are being followed. The compliance audit shall be conducted by at least one person knowledgeable in the process. The employer shall determine and document an appropriate response to each of the findings of the compliance audit, and document that deficiencies have been addressed.

Trade Secrets

Employers must make all information necessary to comply with the requirements of this section available to those persons responsible for compiling the process safety information, developing process hazard analyses, developing the operating procedures, those involved in incident investigations, emergency planning, and response and compliance audits without regard to possible trade secret status of such information. Nothing in this paragraph shall preclude the employer from requiring the persons to whom the information is made available to enter into confidentiality agreements not to disclose the information.

The above elements outline the programs required by PSM. These programs are performance-type standards. They spell out programs and choices and are not limited to specific details. These elements have served to organize and guide the process safety programs of all who are covered by it. They have served to bring direction to training and publications involving process safety. The AIChE's Center for Chemical Process Safety has publications and training programs to support most of these elements.

Risk Management Plans (RMPs)

The EPA is charged primarily with the responsibility to protect the public and the environment. One could regard this as "outside the fence line." RMPs are required by the Environmental Protection Agency (EPA). Since protecting the public and the environment is mainly an environmental issue rather than a safety issue, this subject will be covered only briefly in this section.

Congress enacted Section 112(r) of the 1990 Clean Air Act (CAA) to address the threat of catastrophic releases of chemicals that might cause immediate deaths or injuries in communities. It requires owners and operators of covered facilities to submit RMPs to the EPA. The final RMP rule was published in 40 CFR 68 in the Federal Register on June 20, 1996. RMPs must summarize the potential threat of sudden, large releases of certain dangerous chemicals and facilities' plans to prevent such releases and mitigate any damage.

Operators of facilities that are subject to the EPA's RMP must perform offsite consequence analyses to determine whether accidental releases from their processes could put nearby populations at risk. In performing a consequence analysis it is assumed that all or part of a hazardous substance escapes from a process at a given facility. It is then estimated how far downwind hazardous gas concentrations may extend.

Facilities that must prepare and submit RMPs must estimate the offsite consequences of accidental releases. This can be done using tables (such as those provided in CAA 112(r) Offsite Consequence Analysis) or a computerized model. There are a number of commercially available computer models. Submitters are expected to choose a tool that is appropriate for their facility.

The owners and operators of stationary sources producing, processing, handling, or storing of extremely hazardous substances have a general duty to identify hazards that may result from an accidental release This includes agents that may or may not be identified by any government agency which may cause death, injury, or property damage. In other words, just because a substance is not listed is not an excuse to fail to consider its hazards.

This section with its emphasis on Process Safety does not cover the considerable other safety, design, and operating requirements of other chemical-related regulations. Many of these requirements also include national codes as guidelines or as adopted regulations. Examples of these are in the American Society of Mechanical Engineers (ASME) *2001 Boiler Pressure Vessel Code*, the National Fire Protection Association (NFPA) which covers a wide range of fire safety issues and the American Petroleum Institute (API) *Recommended Practice 520, Sizing, Selection, and Installation of Pressure Relieving Devices in Refineries*.

An extremely hazardous substance is any agent that may or may not be listed by any government agency which, as the result of short-term exposures associated with releases to the air, cause death, injury, or property damage due to its toxicity, reactivity, flammability, volatility, or corrosivity.

Toxics Release Inventory

Two statutes, the Emergency Planning and Community Right-to-Know Act (EPCRA) and section 6607 of the Pollution Prevention Act (PPA), mandate that a publicly accessible toxic chemical database be developed and maintained by the U.S. EPA. This database, known as the Toxics Release Inventory (TRI), contains information concerning waste management activities and the release of toxic chemicals by facilities that manufacture, process, or otherwise use these materials. The TRI of 1999 is a publicly available database containing information on toxic chemical releases and other waste management activities that are reported annually by manufacturing facilities and facilities in certain other industry sectors, as well as by federal facilities. The TRI program is now under the EPA's Office of Environmental Information. This inventory was established under the EPCRA of 1986 which was enacted to promote emergency planning, to minimize the effects of chemical accidents, and to provide the public. As of November 2001, there were 667 toxic chemicals and chemical compounds on the list.

5.14 HAZWOPER

The Hazardous Waste Operations and Emergency Response Standard (HAZWOPER), 29 CFR Part 1910.120, applies to five distinct groups of employers and their employees. This includes any employees who are exposed or potentially exposed to hazardous substancesincluding hazardous wasteand who are engaged in one of the following operations as specified by 1910.120:

1. Clean-up operations
2. Corrective actions
3. Voluntary clean-up operations
4. Operations involving hazardous wastes
5. Emergency response operations for releases of, or substantial threats of release of, hazardous substances regardless of the location of the hazard.

In addition, with the passage of the Pollution Prevention Act (PPA) in 1991, facilities must report other waste management amounts including the quantities of TRI chemicals recycled, combusted for energy recovery, and treated on- and offsite.

More Information

For more information on Regulations, the books, magazine articles, and Internet references in the Reference section can be very helpful. Following the requirements of the many aspects of Regulations can be quite complicated and involve a lot of detail. There is a considerable amount of good assistance available which help can make the subject manageable.

5.15 The Principal Reason for Most Chemical Process Accidents

Ask any group of people experienced in chemical plant operations what causes most chemical process accidents, and you will get a variety of answers including: operator error, equipment failure, poor design, act of God, and bad luck. However, in the opinion of representatives of many of the large chemical and oil companies in the United States, these answers are generally incorrect. The Center for Chemical Process Safety, an organization sponsored by the American Institute of Chemical Engineers, includes representatives of many of the largest chemical and oil companies in the United States, and states that "It is an axiom that process safety incidents are the result of *management system failure*." Invariably, some aspect of a process safety management system can be found that, had it functioned properly, could have prevented an incident (or reduced the seriousness of it). "It is a rare situation where an 'Act of God' or other uncontrollable event is the sole cause of an incident. Much more common is the situation where an incident is the result of multiple causes, including management system failures. Therefore, it is more appropriate to presume that management system failures underlie every incident so that we may act to uncover such failures and then modify the appropriate management systems, rather than presume that if an 'Act of God' appears to be the immediate cause, investigation should cease because there is nothing that can be done to prevent such future incidents."24

For example, consider a case where a small amount of hazardous material is spilled while a sample is being taken from a process line. It is not enough to look into the situation and conclude that this is an example of an operator error where procedures were not followed, and then simply to recommend that the employee be instructed to follow procedures in the future. Further investigation may reveal deficiencies in the training system or in the equipment. Still more investigation may reveal deficiencies in the management system that plans resources for training or that provides for proper equipment for sampling. It then may be appropriate to change the management system to prevent repetition of the incident.

Levels of Causes

There are several levels of causes of accidents, usually (1) the immediate cause, (2) contributing causes to the accident or to the severity of the accident, and (3) the "root cause." The root cause is what really caused the accident, and when this is determined, it may be possibly to *prevent* future similar accidents. With the 2020 hindsight that is available after an accident, the root cause usually can be found. The purpose of the discussion in the next section is to illustrate how knowledge about the root causes of some important accidents can help to keep them from happening again. It will be noted that the root cause is rarely the fault of one person, but instead is the result of a management system that does not function properly.

Following are brief analyses of several case histories that have been of landmark importance in the industrial world, and that have affected the chemical industry all over the world.

5.16 Case Histories

Flixborough, England 19742

On June 1, 1974, an accident occurred in the Nypro plant in Flixborough, England, in a process where cyclohexane was oxidized to cyclohexanone for the manufacture of caprolactam, the basic raw material for the production of Nylon 6. The process consisted of six reactors in series at 155°C and 8.8 bar (130 psig) containing a total of 120 tons of cyclohexane and a small amount of cyclohexanone. The final reactor in the process contained 94 percent cyclohexane. There was a massive leak followed by a large unconfined vapor cloud explosion and fire that killed 26 people, injured 36 people, destroyed 1,821 houses, and damaged 167 shops. It was estimated that 30 tons of cyclohexane was involved in the explosion. The accident occurred on Saturday; on a working day, casualties would have been much higher.

The accident happened when the plant had to replace one of six reactors and rushed to refit the plant to bypass the disabled reactor. Scaffolding was jerry-rigged to support a 20-in. pipe connecting reactor four with reactor six, which violated industry and the manufacturer's recommendations. The reactor that failed showed stress crack corrosion. The only drawings for the repair were in chalk on the machine shop floor. Both ends of the 20-in. pipe had expansion joints where they attached to the reactors. The pipe was supported on scaffolding-type supports and was offset with a "dog-leg" to fit the reactors, which were at different levels to promote gravity flow. The safety reviews, if any, were insufficient.

Immediate Cause

A pipe replacing a failed reactor failed, releasing large quantities of hot cyclohexane forming a vapor cloud that ignited.

Contributing Causes to the Accident and the Severity of the Accident:

1. The reactor failed without an adequate check on why (metallurgical failure).
2. The pipe was connected without an adequate check on its strength and on inadequate supports.
3. Expansion joints (bellows) were used on each end of pipe in a "dog-leg" without adequate support, contrary to the recommendations of the manufacturer.
4. There was a large inventory of hot cyclohexane under pressure.
5. The accident occurred during startup.
6. The control room was not built with adequate strength, and was poorly sited.
7. The previous works engineer had left and had not been replaced. According to the Flixborough Report, "There was no mechanical engineer on site of sufficient qualification, status or authority to deal with complex and novel engineering problems and insist on necessary measures being taken."
8. The plant did not have a sufficient complement of experienced people, and individuals tended to be overworked and liable to error.

Root Cause

Management systems deficiencies resulted in:

1. A lack of experienced and qualified people.
2. Inadequate procedures involving plant modifications.
3. Regulations on pressure vessels that dealt mainly with steam and air and did not adequately address hazardous materials.
4. A process with a very large amount of hot hydrocarbons under pressure and well above its flash point installed in an area that could expose many people to a severe hazard.

This accident resulted in significant changes in England and the rest of the world in the manner in which chemical process safety is managed by industry and government. One of the conclusions reached as a result of this accident, which has had a wide effect in the chemical industry, is that "limitations of inventory (or flammable materials) should be taken as specific design objectives in major hazard installations."

The use of expansion joints (bellows, in this case) which were improperly installed may have been a principal reason for the accident. This provides additional reasons not to use expansion joints (except in special exceptional circumstances).

Bhopal, 1985 (C&EN Feb. 11, 1985; Technica 198954)

On December 3 and 4, 1985, a chemical release causing a massive toxic gas cloud occurred at the Union Carbide India, Ltd, plant in Bhopal, India. (Union Carbide is now a part of The Dow Chemical Company.) The process involved used methyl isocyanate (MIC), an extremely toxic chemical, to make Sevin, a pesticide. According to various authoritative reports, about 1,7002,700 (possibly more) people were killed, 50,000 people were affected seriously, and 1,000,000 people were affected in some way. The final settlement may involve billions of dollars. It was one of the worst industrial accidents in history. The accident occurred when about 120240 gallons of water were allowed to contaminate an MIC storage tank. The MIC hydrolyzed, causing heat and pressure, which in turn caused the tank rupture disk to burst.

Equipment designed to handle an MIC release included a recirculating caustic soda scrubber tower and a flare system designed for 10,000 lb/hr, which would be moderate flows from process vents. It was not designed to handle runaway reactions from storage. The design was based on the assumption that full cooling would be provided by the refrigeration system. The actual release was estimated to be 27,000 lb over 2 hr, with the tank at 43°C. At the time of the release the refrigeration had been turned off. The flare tower was shut down for repairs. A system of pressurized sprinklers that was supposed to form a water curtain over the escaping gases was deficient, in that water pressure was too low for water to reach the height of the escaping gas.

There have been conflicting stories of how the water got into the tank, including operator error, contamination, and sabotage.

Immediate Cause

The immediate cause was hydrolysis of MIC due to water contamination. The exact source of the water has not been determined.

Contributing Causes

1. Flare tower was shut down for repair.
2. Scrubber was inadequate to handle a large release.
3. Chilling system was turned off. (It also was too small.)
4. MIC tank was not equipped with adequate instrumentation.
5. Operating personnel lacked knowledge and training.
6. The inventory of MIC was large.
7. There was a lack of automatic devices and warning systems; it has been reported that safety systems had to be turned on manually.
8. When the plant was built, over 20 years before the accident, there were very few people near it. At the time of the accident, a shanty town had grown up near the plant with a density of 100 people per acre, greatly increasing the potential exposure of people to toxic releases. There was no emergency action plan to notify neighbors of the potential for toxic releases or of what to do if there was a release, nor was there a functioning alarm system.

Root Cause

The root cause of the accident appears to be a management system that did not adequately respond to the potential hazards of MIC. There was probably a greater inventory of MIC than was needed. The main process expertise was in the United States. Local management does not appear to have understood the process or the consequences of changes made. This includes plant design, maintenance and operations, backup systems, and community responsibility. (Union Carbide has provided legal arguments alleging that sabotage caused the release). There appears to be enough blame to go around for all those involved in any way in the plant, including government units.

This accident has become widely known. It is an objective of many chemical process safety programs and government actions to "avoid another Bhopal"that is, to avoid a severe release of toxic chemicals (usually referring to toxic chemicals in the air). Almost every chemical company in the world has been affected by this incident in one way or another, in the design and operation of chemical plants, in community action programs, and in the activities of such organizations as the American Institute of Chemical Engineers, the Chemical Manufacturers Association, and many governmental units.

Phillips Explosion, 198957

On October 23, 1989, at approximately 1300, an explosion and fire ripped through the Phillips 66 Company's Houston Chemical Complex in Pasadena, Texas. At the site, 23 workers were killed, and more than 130 were injured. Property damage was nearly $750 million. Business interruption cost is not available but is probably a very large figure.

The release occurred during maintenance operations on a polyethylene reactor. Two of the six workers on the maintenance crews in the immediate vicinity of the reactor leg where the release occurred were killed, together with 21 other employees of the facility. Debris from the plant was found six miles from the explosion site. Structural steel beams were twisted like pretzels by the extreme heat. Two polyethylene production plants covering an area of 16 acres were completely destroyed.

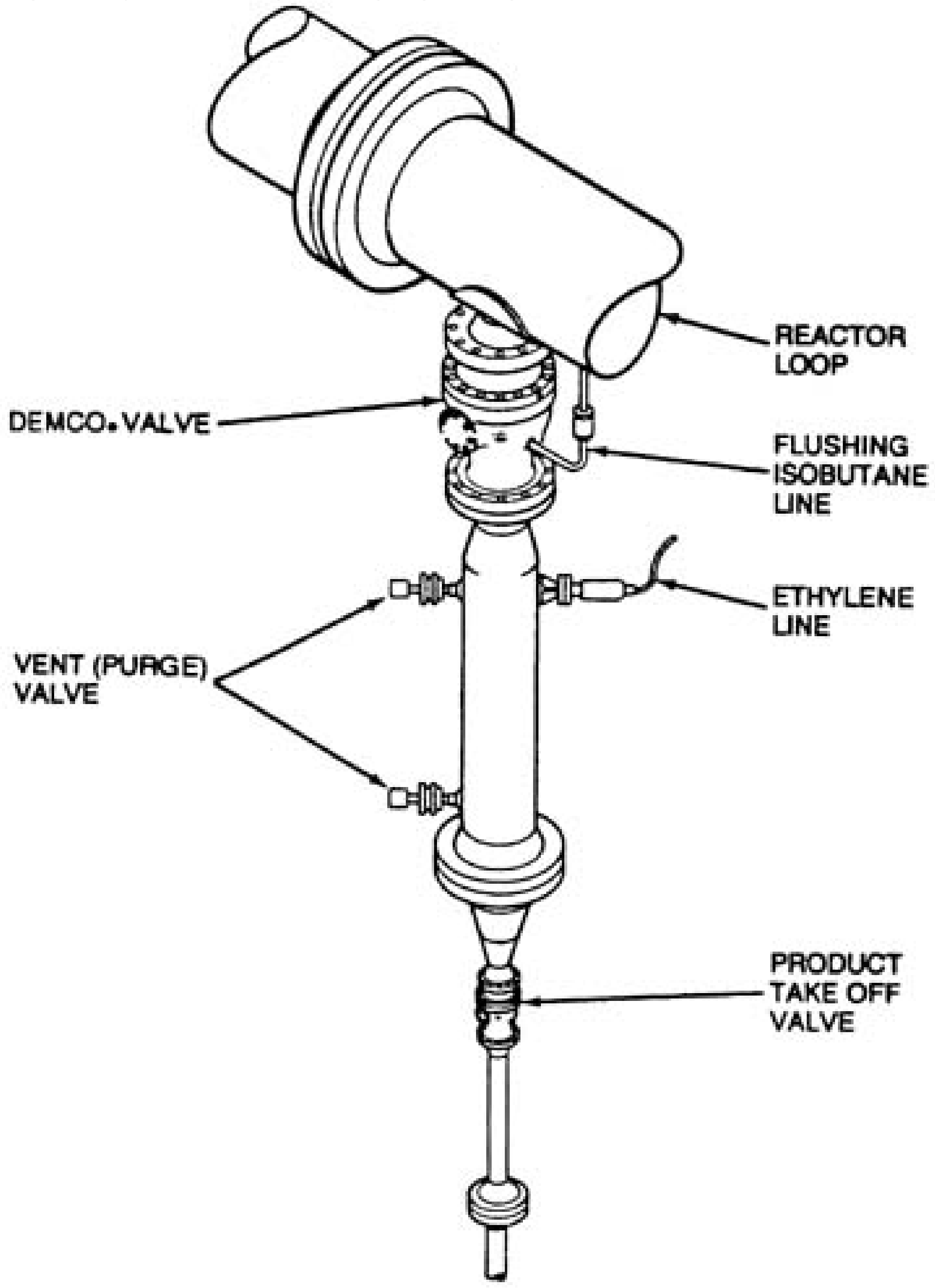
The Phillips complex produces high-density polyethylene, which is used to make milk bottles and other containers. Prior to the accident, the facility produced approximately 1.5 billion pounds of the material per year. It employed 905 company employees and approximately 600 daily contract employees. The contract employees were engaged primarily in regular maintenance activities and new plant construction.

The accident resulted from a release of extremely flammable process gases that occurred during regular maintenance operations on one of the plant's polyethylene reactors. It is estimated that within 90120 s more than 85,000 lb of flammable gases were released through an open valve. A huge flammable vapor cloud was formed that came into contact with an ignition source and exploded with the energy of 4,800 lb of TNT. The initial explosion was equivalent to an earthquake with a magnitude of 3.5 on the Richter scale. A second explosion occurred 1015 min later when two isobutane tanks exploded.

Each explosion damaged other units, creating a chain reaction of explosions. One witness reported hearing ten separate explosions over a 2-hr period.

In the process used by Phillips at this site to produce high-density polyethylene, ethylene gas is dissolved in isobutane and, with various other chemicals added, is reacted in long pipes under elevated pressure and temperature. The dissolved ethylene reacts with itself to form polyethylene particles that gradually come to rest in settling legs, where they are eventually removed through valves at the bottom. At the top of the legs there is a single ball valve (DEMCO® brand) where the legs join with other reactor pipes. The DEMCO valve is kept open during production so that the polyethylene particles can settle into the leg. A typical piping settling leg arrangement is shown in Fig. 5.9.

Fig. 5.9 Typical piping settling leg arrangement.



In the Phillips reactor, the plastic material frequently clogged the settling legs. When this happened, the DEMCO valve for the blocked leg was closed, the leg disassembled, and the block removed. During this particular maintenance process, the reactor settling leg was disassembled and the block of polymer removed. While this maintenance process was going on, the reaction continued, and the product settled in the legs that remained in place. If the DEMCO valve were to open during a cleaning-out operation, there would be nothing to prevent the escape of the gas to the atmosphere.

After the explosion it was found that the DEMCO valve was open at the time of the release. The leg to be cleaned had been prepared by a Phillips operator. The air hoses that operated the DEMCO valve were improperly connected in a reversed position such that a closed DEMCO valve would be opened when the actuator was in the closed position. In addition, the following unsafe conditions existed:

1. The DEMCO valve did not have its lockout device in place.
2. The hoses supplied to the valve actuator mechanism could be connected at any time even though the Phillips operating procedure stipulated that the hoses should never be connected during maintenance.
3. The air hoses connecting the open and closed sides of the valve were identical, thus allowing the hoses to be cross-connected and permitting the valve to be opened when the operator intended to close it.
4. The air supply valves for the actuator mechanism air hoses were in the open position so that air would flow and cause the actuator to rotate the DEMCO valve when the hoses were connected.
5. The DEMCO valve was capable of being physically locked in the open position as well as in the closed position. The valve lockout system was inadequate to prevent someone from inadvertently opening the DEMCO valve during a maintenance procedure.

Established Phillips corporate safety procedures and standard industry practice require backup protection in the form of a double valve or blind flange insert whenever a process or chemical line in hydrocarbon service is opened. According to OSHA, Phillips had implemented a special procedure for this maintenance operation that did not incorporate the required backup. Consequently, none was used on October 23.

The consequences of the accident were exacerbated by the lack of a water system dedicated to fire fighting, and by deficiencies in the shared system. When the process water system was extensively damaged by the explosion, the plant's water supply for fighting fires was also disrupted. The water pressure was inadequate for fire fighting. The force of the explosion ruptured waterlines and adjacent vessels containing flammable and combustible materials. The ruptured water lines could not be isolated to restore water pressure because the valves to do so were engulfed in flames. Of the three backup diesel pumps, one had been taken out of service and was unavailable, and another soon ran out of fuel. It was necessary to lay hoses to remote sitessettling ponds, a cooling tower, a water treatment plant, and so on. Electric cables supplying power to regular service pumps were damaged by fire, and those pumps were rendered inoperable. Even so, the fire was brought under control within 10 hr.

In the months preceding the explosion, according to testimony, there had been several small fires, and the alarm had sounded as many as four or five times a day. There had been a fatality at the same plant doing a similar operation about three months before this incident. Some of the employees in the area where the release occurred may not have heard the siren because of the ambient noise level, and may not have known of the impending disaster. Employees in the immediate area of the release began running as soon as they realized the gas was escaping.

The large number of fatalities was due in part to the inadequate separation between buildings in the complex. The site layout and the proximity of normally high-occupancy structures, such as the control and finishing building, to large-capacity reactors and hydrocarbon vessels contributed to the severity of the event.

The distances between process equipment were in violation of accepted engineering practices and did not allow personnel to leave the polyethylene plants safely during the initial vapor release; nor was there sufficient separation between reactors and the control room to carry out emergency shutdown procedures. The control room, in fact, was destroyed by the initial explosion. Of the 22 victims' bodies recovered at the scene, all were located within 250 ft of the vapor release point.

OSHA's investigation revealed that a number of company audits had identified unsafe conditions but largely had been ignored. Thus, a citation for willful violations of the OSHA "general duty" clause was issued to Phillips with proposed penalties of $5,660,000. In addition, proposed penalties of $6,200 were issued for other serious violations. A citation for willful violations with proposed penalties of $724,000 was issued to Fish Engineering and Construction, a Phillips maintenance contractor. Other financial penalties have been proposed. In the investigation it became apparent that Fish had become accustomed to tolerating safety and health violations at the site by its personnel and Phillips personnel, as well as participating in those violations by knowing about them and not taking direct positive action to protect its employees.

Since 1972, OSHA has conducted 92 inspections in the Dallas region at various Phillips locations; 24 were in response to a fatality or a serious accident.

OSHA determined that Phillips had not acted upon reports by its own safety personnel and outside consultants who had pointed out unsafe conditions.

OSHA also had conducted 44 inspections of the Fish Company, seven of them in response to a fatality or a serious accident.

One of the major findings by OSHA was that Phillips had not conducted a process hazard analysis or equivalent (such as HAZOP) in its polyethylene plants.

Immediate Cause

There was a release of flammable process gases during the unplugging of Number 4 Reactor Leg on Reactor 6 while undergoing a regular maintenance procedure by contractor employees. The unconfined flammable vapor cloud was ignited and exploded with devastating results.

The immediate cause of the leak was that a process valve was opened by mistake while the line was open. The valve was open to the atmosphere without a second line of defense such as another valve or a blind flange.

Contributing Causes to the Accident and the Severity of the Accident

1. Procedures to require backup protection in the form of a double valve or a blind flange insert were not used. The lockout system was inadequate.
2. Air hoses were improperly connected in the reversed position.
3. The air hoses for the open and closed side of the valve were identical, allowing the hoses to be cross-connected.
4. The DEMCO valve actuator mechanism did not have its lockout device in place.
5. There was not a water system dedicated to fire fighting, and there were deficiencies in the system shared with the process.
6. The site layout and proximity of high-occupancy structures contributed to the severity.
7. There was inadequate separation of buildings within the complex. Especially, there was inadequate spacing between the reactors and the control room.

Root Causes

The root causes of the accident and its extreme severity appear to be failures of the management system, as shown by the following:57

1. According to OSHA, Phillips had not conducted a process hazard analysis or equivalent (such as HAZOP) in its polyethylene plants.
2. It was found by OSHA that the contractor, Fish Engineering, had a history of serious and willful violations of safety standard, which Phillips had not acted upon. The same contractor also had been involved in a fatal accident at the same facility three months earlier.
3. A report by OSHA stated that Phillips had not acted upon reports issued previously by the company's own safety personnel and outside consultants. Phillips had numerous citations from OSHA since July 1972. OSHA discovered internal Phillips documents that called for corrective action but which were largely ignored.
4. Safe operating procedures were not required for opening lines in hazardous service.
5. An effective safety permit system was not enforced with Phillips or contractor employees, especially line opening and hot work permits.
6. Buildings containing personnel were not separated from process units in accordance with accepted engineering principles, or designed with enough resistance to fire and explosion.
7. The fire protection system was not maintained in a state of readiness:
    1. One of the three diesel-powered water pumps had been taken out of service.
    2. Another of the three diesel-powered water pumps was not fully fueled, and it ran out of fuel during the fire fighting.
    3. Electric cables supplying power to regular service fire pumps were not located underground and were exposed to blast and fire damage.

5.17 Summary

As the tragic case histories unfold, the significance of the process safety consideration presented in this chapter becomes chillingly apparent, and the necessity for inherently safe process design is revealed. The case histories also reveal significant flaws in management systems, which tends to be true for most chemical process accidents. Even an ideally safe process can be transformed into one with a high potential for disaster if a valid system is not in place to ensure that the inherently safe process design retains its integrity.

Because there is always risk when equipment, instrumentation, and human activity are involved, there is no method of making a plant completely safe. However, facilities can be made "inherently safer" by careful examination of all aspects of design and management, using modern techniques that are now available. If we are to improve our process safety performance and our public image, "inherently safe" process design coupled with "inherently safe" process management is imperative.

In addition to the information presented in this chapter and in the publications it has cited, references 5872 are recommended as appropriate source material.

5.18 References

M & M Protection Consultants, One Hundred Largest Losses: A Thirty Year Review of Property Damage in the Hydrocarbon-Chemical Industry, 12th ed., 222 S. Riverside Plaza, Chicago, IL 60606, 1989.

Lees, F. P., Loss Prevention in the Process Industries, London, Butterworths, 1980.

Englund, S. M., "Opportunities in the Design of Inherently Safer Chemical Plants," in Advances in Chemical Engineering, J. Wei et al. (Eds.), Academic Press, New York, 1990.

Kletz, T, Cheaper, Safer Plants or Wealth and Safety at Work, Institution of Chemical Engineers, Rugby, England, 1985.

Schaller, L. C., Du Pont, Wilmington, DE 19898, *Plant Operations Progress*, No. 1 (Jan. 1990).

Howard, W., Consultant, Monsanto Chemical Company, Personal Communication, 1981.

NFPA 68, National Fire Protection Association, Batterymarch Park, Quincy, MA, 1998.

Bartknecht, W., Explosions Course Prevention Protection, Springer-Verlag, Berlin, 1981.

NFPA 69 National Fire Protection Association, Batterymarch Park, Quincy, MA, 1986.

Liening, G., "Prolonging Service Life of Metals and Alloys in Chemical Processing Plants," Chem. Proc., 22 (Sept. 1986).

Liening, G., The Dow Chemical Co., Midland, MI, personal communication (Sept. 5, 1986).

Perry's Chemical Engineers' Handbook, 6th ed., pp. 2348, McGraw-Hill, New York, 1984.

Grinwis, D., Process Systems Associate; Larsen, Paul, Associate Instrument Engineering Consultant; and Schrock, Luther, Process Consultant, The Dow Chemical Co., Midland, MI and Freeport, TX, personal-communication (Oct. 1986).

Fisher, H., Chem. Eng. Prog., 3 (Aug. 1985).

Huff, J. E, "A General Approach to the Sizing of Emergency Pressure Relief Systems," Reprints of Int. Symp. on Loss Prev. and Safety Promotion in the Process Ind., Heidelberg, Germany, 1977, p. IV 223, DECHEMA, Frankfurt (1977).

Huff, J. E., Institute of Chemical Engineers Symposium Series, No. 85, p. 109 (1984).

Huff, J. E., "Emergency Venting Requirements for Gassy Reactions from Closed System Tests," Plant/Operations Progress, 3(1), 5059 (Jan. 1984).

Huff, J. E., "The Role of Pressure Relief in Reactive Chemical Safety," International Symposium on Preventing Major Chemical Accidents, Washington, DC, sponsored by the Center for Chemical Process Safety of the American Institute of Chemical Engineers, The United States Environmental Protection Agency, and the World Bank, (Feb. 3, 1987).

Fauske, Hans K., and Leung, J., "New Experimental Technique for Characterizing Runaway Chemical Reactions," Chem. Eng. Progr., 39 (Aug. 1985).

Jackson, B. L., Piping Specialist, The Dow Chemical Co., Midland, MI, personal communication (Sept. 17, 1986).

Alexander, S., and King, R., Materials Engineering Dept., Michigan Div., The Dow Chemical Co., Midland MI, personal communication (Feb. 14, 1988).

Cromie, J., Engineering Associate, The Dow Chemical Co., Midland, Michigan, personal communication (1986).

Reynolds, J. A., Union Carbide Corp., "Canned Motor and Magnetic Drive Pumps," Chem. Proc., 7175 (Nov. 1989).

Technical Management of Chemical Process Safety, Center for Chemical Process Safety, American Institute of Chemical Engineers, 345 E. 47th St., New York, NY, 1989.

Cloud, M. J., "Fire, the Most Tolerable Third party," Mich. Nat. Res., 18 (MayJune, 1990).

Fire Safety Data, Fire Protection Association, 140 Aldersgate St., London EC1A 4HX, 1988.

Bodurtha, F. T., Engineering Dept., Du Pont, Industrial Explosion Prevention and Protection, McGraw-Hill, New York, 1980.

Zabetakis, M. G., "Flammability Characteristics of Combustible Gases and Vapors," Bulletin 627, U.S. Dept. of the Interior, Bureau of Mines, Washington, DC (1965).

Cawse, J. N., Pesetsky, B., and Vyn, W. T., "The Liquid Phase Decomposition of Ethylene Oxide," Union Carbide Corporation, Technical Center, South Charleston, WV 25303 (no date available).

Stull, D. R., Fundamentals of Fire and Explosion, p. 50, Corporate Safety and Loss Dept., The Dow Chemical Co., Midland, MI, American Institute of Chemical Engineers, New York, 1976.

Kohlbrand, "Case History of a Deflagration Involving an Organic Solvent/Oxygen System below Its Flash Point," 24th Annual Loss Prevention Symposium, Sponsored by the American Institute of Chemical Engineers, San Diego, CA (Aug. 1922, 1990).

Mackenzie, J., "Hydrogen Peroxide without Accidents," Chem. Eng., 84ff (June 1990).

Eichel, F. G., "Electrostatics," Chem. Eng., 154167 (Mar. 13, 1967).

Boundy, R. H., and Boyer, R. F., Styrene, American Chemical Society Monograph Series, p. 63, Reinhold Publishing Co., New York, 1952.

NFPA 77, National Fire Protection Association, Batterymarch Park, Quincy, MA, 1986.

Bodurtha, F. T., "Industrial Explosion Control Course," Center for Professional Advancement, Chicago, IL (Sept. 1416, 1987).

Webley, P., Director, Massachusetts Institute of Technology Practice School, Midland Station, Midland, MI, personal communication (May 1990).

CPQRA (Chemical Process Quantitative Risk Analysis), Center for Chemical Process Safety of the American Institute of Chemical Engineers, 1989.

TNO, "Methods for the Calculation of the Physical Effects of the Escape of Dangerous Materials: Liquids and Gases" ("The Yellow Book"), Apeldoorn, The Netherlands, 1979.

API RP 521, 2nd ed., American Petroleum Institute, Washington, DC, 1982.

Walters, S., "The Beginnings," Mech. Eng., 4, 3846 (1984).

Prugh, R. W., "Quantitative Evaluation of BLEVE Hazards," AICHE Loss Prevention Symposium, Paper No. 74e, AICHE Spring National Meeting, New Orleans, LA (Mar. 610, 1988).

Moore, C. V., "The Design of Barricades for Hazardous Pressure Systems," Nucl. Eng. Des., 5, 15501566 (1967).

Clancey, V. J., "Diagnostic Features of Explosion Damage," Sixth Int. Meeting of Forensic Sciences, Edinburgh, 1972.

Baker, W. E., Cox, P. A., Westine, P. S., Kulesz, J. J., and Strehlow, R. A., Explosion Hazards and Evaluation, Elsevier, New York, 1983.

Ludwig, E. E., Applied Process Design for Chemical and Petrochemical Plants, 2nd ed., Vol. 1, Gulf Publishing, Houston, TX, 1977.

Brasie, W. C., Michigan Division, Process Engineering, The Dow Chemical Co., Midland, MI, personal communication (Mar. 9, 1983).

Brasie, W. C., Michigan Division, Process Engineering, The Dow Chemical Co., Midland, MI, personal communication (Oct. 6, 1982).

Allen, W. T., Process Engineering, The Dow Chemical Co., Midland, MI, personal communication (May 1988).

AIChE. Guidelines for Safe Storage and Handling of High Toxic Hazard Materials, Center for Chemical Process Safety, American Institute of Chemical Engineers, 345 E. 47th St., New York, NY, 1988.

Knowlton, R. E., Hazard and Operability Studies, Chemetics International Co., Ltd., Vancouver, BC, Canada V6J, 1C7 (Feb. 1989).

NUS Corp., HAZOP Study Team Training Manual. Predictive Hazard Identification Techniques for Dow Corning Facilities, Gaithersburg, MD (July 1989).

Technica Consulting Scientists and Engineers, London, England, 1990.

Technica, Inc., "HAZOP Leaders Course," 1989, Columbus, OH, Nov. 610; course leaders David Slater and Frederick Dyke.

Arendt, J. S., Lorenzo, A. F., and Lorenzo, D. K., "Evaluating Process Safety in the Chemical Industry," A Manager's Guide to Quantitative Risk Assessment, Chemical Manufacturers Association, DC, 1989.

Burk, Art., Principal Safety Consultant. Du Pont, Newark, DE, personal communication (Feb. 20, 1990).

OSHA (Occupational Safety and Health Administration), U.S. Department of Labor, The Phillips Company Houston Chemical Complex Explosion and Fire (Apr. 1990).

Bartknecht, W., Dust Explosions Course, Prevention, Protection. Springer-Verlag, Berlin, 1989.

Beveridge, H. J. R., and Jones, C. G., "Shock Effects of a Bursting-Disk in a Relief Manifold," Institution of Chemical Engineers Symposium, Series No. 85.

Burgess, D., and Zabetakis, M. G., "Fire and Explosion Hazards Associated with Liquefied Natural Gas," U.S. Bureau of Mines, Report RI 6099 (1962).

Burk, A., Principal Safety Consultant. Du Pont, Newark. DE, presentation on "Process Hazards Analysis and Quantitative Risk Assessment" (July 20, 1989).

"CEFIC Views on the Quantitative Assessment of Risks From Installations in the Chemical Industry," European Council of Chemical Manufacturers' Federations, Bruxelles (Apr. 1986).

Condensed Chemical Dictionary, Van Nostrand Reinhold Co., New York, 1983.

Guidelines for Vapor Release Mitigation, Center for Process Safety, American Institute of Chemical Engineers, 345 E. 47th St., New York, NY (Sept. 1987).

Hymes, I., "The Physiological and Pathological Effects of Thermal Radiation," UKAEA Safety and Reliability Directorate, Report SRD R275, Culcheth, UK.

Klein, H. H., "Analysis of DIERs Venting Tests: Validation of a Tool for Sizing Emergency Relief Systems for Runaway Chemical Reactions," Plant Op. Progr., 5(1), 110 (Jan. 1986).

Nazario, F. N., Exxon Research and Engineering Co., Chem. Eng., pp. 102109 (Aug. 15, 1988).

Pieterson, C. M., and Huaerta, S. C., "Analysis of the LPG Incident in San Juan Ixhatepec, Mexico City, 19 Nov. 1984," TNO Report B4-0222, P.O. Box 342 7300 AH, Apeldoorn, The Netherlands (1985).

Roberts, A. F., "Thermal Radiation Hazards from Releases of LPG from Pressurized Storage," Fire Safety J., 4(3), 197 (1981).
Townsend, D. I., and Tou, J. C., "Thermal Hazard Evaluation by an Accelerating Rate Calorimeter," Thermochimica Acta, 37, 30 (1980).
Welty, J. R., Engineering Heat Transfer, SI Version. John Wiley & Sons, New York, 1978.
Wensley, J. H., "Improved Alarm Management through Use of Fault Tolerant Digital Systems," Instrument Society of America, International Conference and Exhibit, Houston, TX (Oct. 1316, 1986).

5.19 Additional Reading References

American Petroleum Institute, API Recommended Practice 520, Sizing, Selection, and Installation of Pressure Relieving Devices in Refineries, Part I, Sizing and Selection, American Petroleum Institute, 19952000.

American Society of Mechanical Engineers, 2001 Boiler Pressure Vessel Code, ASME International, 22 Law Drive, Fairfield, NJ 07007-2900, USA, 2001.

Bartknecht, W., Explosions Course: Prevention, Protection, Springer-Verlag, 1993.

Bretherick, L., Handbook of Reactive Chemical Hazards, 5th ed., Butterworths, London, 1995.

Crowl, D., and Bollinger, R., Inherently Safer Chemical Processes: A Life Cycle Approach (Center for Chemical Process Safety (CCPS)). Am. Inst of Chemical, Engineers, 1997.

DIERS (Design Institute for Emergency Relief Systems), American Institute of Chemical Engineers, 3 Park Ave, New York, N.Y., 10016-5991, U.S.A, http://www.diers.net/

Englund, S. M., "Design and Operate Plants for Inherent Safety," Chemical Engineering Progress, 8591 (Part 1) (Mar. 1991), and 7986 (Part 2) (May 1991).

Englund, S. M., "Process and Design Options for Inherently Safer Plants," in Prevention and Control of Accidental Releases of Hazardous Gases, Van Nostrand, 1993.

Englund, S. M., "Chemical Process Safety," in Perry's Chemical Engineers' Handbook, 7th ed., D. W. Green (Ed.), published in McGraw-Hill, New York, 1997.

Hendershot, D., Smaller Is SaferSimplifying Chemical Plant Safety, Safe Workplace, National Council on Compensation Insurance, 750 Park of Commerce Drive, Boca Raton, FL 33487, 2000. Hendershot is a senior technical fellow in the Process Hazard Assessment Department of the Rohm and Haas Company, Bristol, PA, and has written extensively on Process Safety.

Hendershot, D., "Chemistrythe Key to Inherently Safer Manufacturing Processes," Presented before the Division of Environmental Chemistry, American Chemical Society, Washington, D.C., August 21, 1994.

Kletz, T., What Went Wrong?: Case Histories of Process Plant Disasters, Gulf Publishing Company, Houston, TX, May, 1998. Kletz is well known for his many publications and for bringing the term, "Inherently Safer Plants," into popular usage.

Kletz, T., Process Plants: A Handbook of Inherently Safer Design, Taylor and Francis, Philadelphia, PA, 1998.

Lees, F., Loss Prevention in the Process Industries: Hazard Identification, Assessment, and Control, Butterworths, London, 1996.

Loss Prevention Committee, Safety, Health Division, AIChE, Proceedings of the 29th Annual Loss Prevention Symposium(Serial), December, 1995

Publications by National Fire Protection Association (NFPA), 1 Batterymarch Park, Quincy, MA 02269. For a more complete list, see http://www.nfpa.org/Codes/CodesAndStandards.asp

NFPA 30 Flammable and Combustible Liquids Code (2000).

NFPA 69 Standard on Explosion Prevention Systems (1997).

NFPA 68 Guide for Venting of Deflagrations (1998).

NFPA 325 Guide to Fire Hazard Properties of Flammable Liquids

PHA Software, "PHAWorks," PSMSource (reference tool for OSHA's 1910.119) Primatech Inc., 50 Northwoods Blvd., Columbus, Ohio 43235, http://www.primatech.com

Smith, K. E., and Whittle, D. K., "Six Steps to Effectively Update and Revalidate PHAs," Chem. Eng. Progr 7077 (Jan. 2001).

Thompson Publishing Group, Chemical Process Safety Report, 1725 K St.. NW, Suite 700, Washington, D.C. 2006.

Internet References and WEB pages

American Institute of Chemical Engineers, Center for Chemical Process Safety. http://www.aiche.org/ccps/

American Society of Mechanical Engineers, 2001 Boiler Pressure Vessel Code. ASME International, 22 Law Drive, Fairfield, NJ 07007-2900, USA Phone: 1-800-843-2763 or 1-973-882-1167 http://www.asme.org/

CCPS (Center for Chemical Process Safety), American Institute of Chemical Engineers 3 Park Ave, New York, NY, 10016-5991, U.S.A. ccps@aiche.org

Chemical Safety Board (Incident Reports) http://www.acusafe.com/Incidents/frame-incident.htm

Manufacturers Chemical Association: http://es.epa.gov/techinfo/facts/cma/cma.html

Manufacturers Chemical Association (Responsible Care): http://es.epa.gov/techinfo/facts/cma/cmacommo.html

OSHA Regulations & Compliance Links http://www.osha.gov/comp-links.html

OSHA Regulations (Standards29 CFR) http://www.osha-slc.gov_OshStd_toc/OSHA_Std_toc.html

RMP Regulations http://www.epa.gov/swercepp/acc-pre.html

Publications by CCPS (Center for Chemical Process Safety), American Institute of Chemical Engineers, 3 Park Ave, New York, N.Y., 10016-5991. This is not a complete list. For a complete list, see on the Internet, http://www.aiche.org/pubcat/seadtl.asp?Act=C&Category=Sect4&Min=30

Guidelines for Chemical Process Quantitative Risk Analysis, 2nd ed., 2000.

Guidelines for Engineering Design for Process Safety, 1993.

Guidelines for Auditing Process Safety Management Systems, 1993.

Guidelines for Chemical Reactivity Evaluation and Application to Process Design, 1995.

Guidelines for Consequence Analysis of Chemical Releases, 1999.

Guidelines for Evaluating the Characteristics of Vapor Cloud Explosions, Flash Fires, and BLEVEs, 1994.

Guidelines for Consequence Analysis of Chemical Releases.

Guidelines for Hazard Evaluation Procedures, 2nd ed. with Worked Examples, 1992.

Guidelines for Implementing Process Safety Management Systems, 1994.

Guidelines for Investigating Chemical Process Incidents, 1992.

Guidelines for Safe Storage and Handling of Reactive Materials, 1995.

Guidelines for Technical Management of Chemical Process Safety, 1989.

6.1 Introduction

Prevention, Prediction, and Preparation

The preceding chapter explored many technical aspects of chemical process safety and some safety management systems that form the foundation of a comprehensive emergency preparedness program. Clearly, the first step in preparing for emergencies is to identify and mitigate the conditions that might cause them. This process starts early in the design phase of a chemical facility, and continues throughout its life. The objective is to *prevent* emergencies by eliminating hazards wherever possible.

Although hazard elimination is the goal, experience has taught us that guaranteed, failure-free designs and devices have so far eluded human kind, despite astonishing advances in knowledge and technology. Even the most "inherently safe" chemical facility must prepare to control potentially hazardous events that are caused by human or mechanical failure, or by natural forces such as storms or earthquakes.

The process of careful, structured analysis and evaluation used to eliminate hazards during design and construction will also allow a chemical facility to accurately *predict* unplanned events that may create emergencies, and to effectively *prepare* to manage them should they occur. A comprehensive emergency preparedness program has all of these elements: *prevention*, *prediction*, and *preparation*.

The fundamental need to predict and prepare for a failure of some kind is familiar to everyone. Fortunately, most of the failures that we encounter create little more than inconveniences in our lives. Others have much more serious potential. Such a failure can trigger an emergency, a term that Webster defines as "an unforeseen combination of circumstances or the resulting state that calls for immediate action." If the immediate action is ineffective, the emergency will escalate to a full-blown crisis.

Certainly most if not all of us in our personal experiences have had many opportunities to reconfirm the wisdom of the admonition, "Plan for the best, but prepare for the worst." As a result, we prepare ourselves for human and mechanical failure in a variety of ways, some so simple and familiar that we scarcely are aware that we are managing a personal emergency preparedness program.

To varying degrees, each of us has assessed our personal vulnerability to a specific emergency and the potential consequences to family and property. In some cases the required immediate action may be minor, but elsewhere the same combination of circumstances will demand significant resources to avoid a crisis of major proportions. The loss of household electric power in an urban condominium may only turn out the light, but on a farm, where electricity pumps the well water, milks the cows, refrigerates the produce, and irrigates the land, loss of power can bring disaster to the unprepared.

In an industrial environment, the consequences of human or mechanical failure can be far greater, even threatening the lives of employees and neighbors. Therefore, emergency prevention and preparedness efforts must have high priority, receiving continuous attention from every employee, including those at every level of management and supervision.

In the chemical industry, emergency preparedness programs have long been recognized as vital elements in protecting people, property, and the environment from harm. Few chemical facilities are without an emergency response plan. Still, when the alarms sound and the emergency is real, the response often does not proceed as planned. Too often, preparation for the unlikely event has been inadequate. As a result, many chemical facility managers are taking a fresh and critical look at their existing emergency preparedness programs in cooperation with their communities, and placing a still higher priority on being truly prepared for emergencies. This is consistent with the aggressive approach to other aspects of safety management that has made the chemical industry one of the safest industries.

Although the commitment to *prevent* the events that lead to injuries and emergencies of all kinds remains the first priority in safety management, we must "prepare for the worst." In this chapter we will explore how chemical facilities and their neighbors can better *predict* and *prepare* for unplanned events that threaten lives, property, and the environmentevents that call for immediate and coordinated action.

Need for Emergency Preparedness Programs

The need for more effective management of emergency preparedness programs by chemical facilities and their host communities became painfully clear to the world in the 1980sa decade marred by tragic events linked to the manufacture and distribution of chemical products. These events revealed serious deficiencies, not only in training people to react effectively during an emergency, but also in managing the systems employed to identify, evaluate, and mitigate hazards that may cause an emergency.

The chemical industry has vigorously responded to these problems in a variety of ways. In some cases well before the crisis of the 1980s, individual chemical units took aggressive action to increase the reliability of their operations, to communicate these actions to their neighbors, and to involve the communities in the process of emergency response planning. This proactive interaction at the local level proved fruitful in promoting the active partnership needed to solve mutual problems and to respond to public concerns. In 1986, these individual initiatives were institutionalized by the industry in the Chemical Manufacturers Association's (now the American Chemistry Council) "Community Awareness/Emergency Response (CAER)" program, which has in turn become a major element of the more recent "Responsible Care" program.

The proven success of these activities provides a powerful reason for giving high priority to open and cooperative management of an emergency preparedness program. By implementing such a program, a chemical facility builds important bridges to its neighbors and fosters positive perceptions. Involving the host community provides an opportunity to demonstrate a sincere commitment to the protection of people and the environment, and a route to the mutually profitable solution of real problems through meaningful two-way communication with the public.

Much of what had been voluntary prior to 1986 became law in the United States with the enactment of the Superfund Amendments and Reauthorization Act (SARA). The Emergency Planning and Community Right-to-Know portions of the Act (known as Title III, or EPCRA) place specific organizational, planning, communication, and training responsibilities on the public and private sectors, as do the accidental release provisions of the Clean Air Act amendments enacted in 1990. Other laws, such as OSHA's 29 CFR 1910.114, Process Safety Management of Highly Hazardous Chemicals, also require emergency response plans. In addition, various state and local regulations must be considered, understood, and reflected in a complete emergency preparedness program.

These laws and regulations, which will be covered later in this chapter, should be viewed as describing only the minimum requirements. Beyond the legal and selfprotective reasons for chemical facilities to improve their management of emergency response programs lie other important driving forces. One of these is the moral and ethical responsibility to employees and the public to work toward elimination of events destructive to the quality of life. In practice, full recognition of this responsibility requires actions in prevention, prediction, and preparation that go beyond the letter of the law.

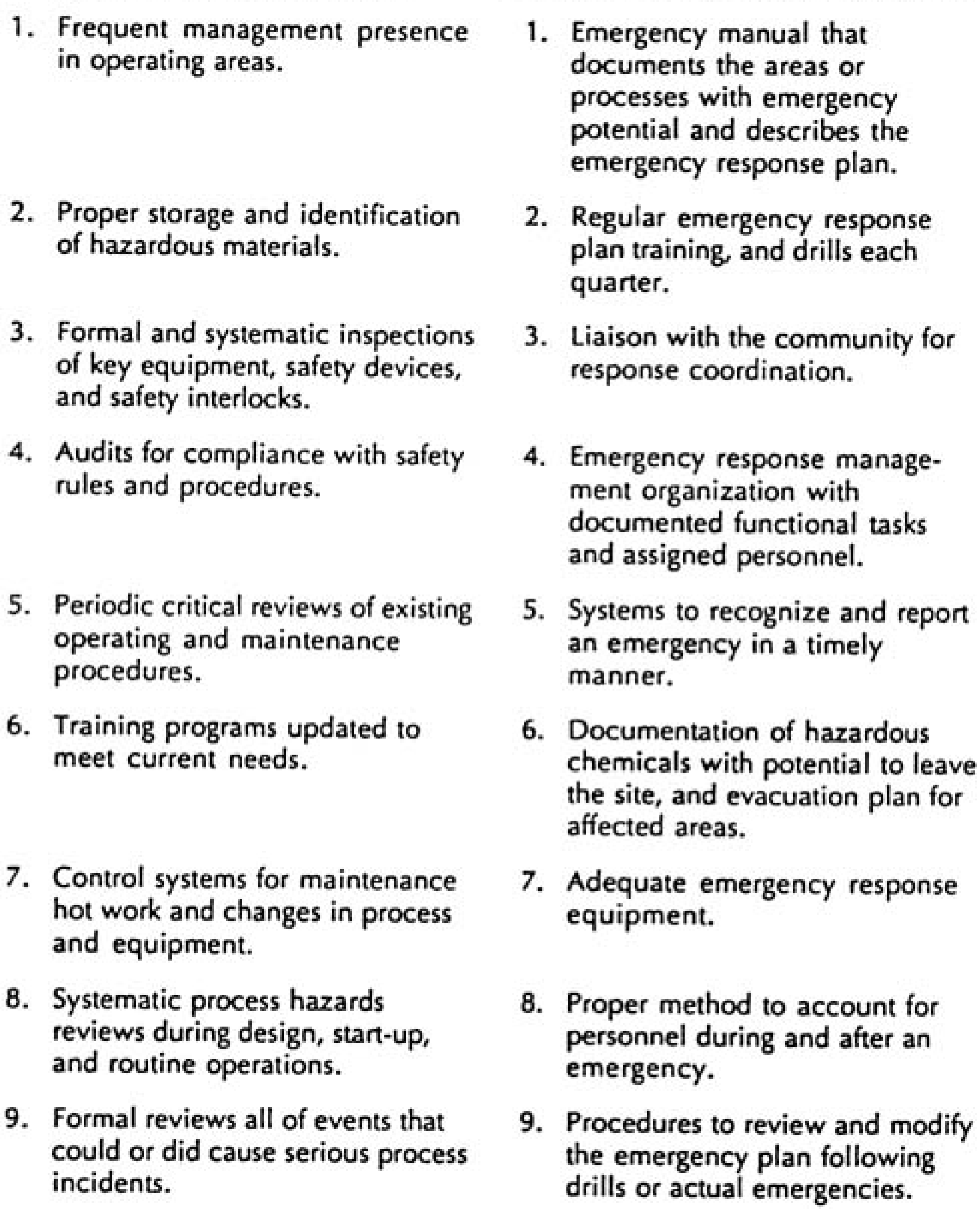
The financial benefits of good emergency preparedness program management are important as well. Emergencies are always expensive. Uncontrolled emergencies can become financially devastating crises. Nervous communities, fearful of the dangers dimly perceived through industrial fences, can and have put some chemical operations out of business. Creating and maintaining a comprehensive emergency preparedness program does not come free. However, the investment is proving to be money well spent by most chemical facilities.

6.2 Preventing and Predicting Emergencies: Getting Started

The objective of a comprehensive emergency preparedness program is the protection of people, property, and the environment from unplanned hazardous events. Organizations handling hazardous materials recognize that the process of creating an effective program starts with the identification, evaluation, reduction, and control of hazards (commonly called risk management), and proceeds through the preparation, drilling, and maintenance of plans and procedures designed to contain an emergency situation should one occur.

To assure an effective emergency preparedness program, chemical facilities need to make periodic, formal assessments of their vulnerability to and preparedness for emergencies. Managers must be involved in developing and monitoring key indicators that will help in assessing an organization's ability to prevent or deal with an emergency. Awareness of any program weaknesses revealed by examining these indicators leads to corrective action to ensure that the potential for incidents decreases, and emergency preparedness improves. Sample checklists including some of these key indicators are shown in Fig. 6.1.

Fig. 6.1 Emergency preparedness checklists.



EMERGENCY PREVENTION

1. Frequent management presence in operating areas.
2. Proper storage and identification of hazardous materials.
3. Formal and systematic inspections of key equipment, safety devices, and safety interlocks.
4. Audits for compliance with safety rules and procedures.
5. Periodic critical reviews of existing operating and maintenance procedures.
6. Training programs updated to meet current needs.
7. Control systems for maintenance hot work and changes in process and equipment.
8. Systematic process hazards reviews during design, start-up, and routine operations.
9. Formal reviews all of events that could or did cause serious process incidents.

EMERGENCY RESPONSE PLANNING

1. Emergency manual that documents the areas or processes with emergency potential and describes the emergency response plan.
2. Regular emergency response plan training, and drills each quarter.
3. Liaison with the community for response coordination.
4. Emergency response management organization with documented functional tasks and assigned personnel.
5. Systems to recognize and report an emergency in a timely manner.
6. Documentation of hazardous chemicals with potential to leave the site, and evacuation plan for affected areas.
7. Adequate emergency response equipment.
8. Proper method to account for personnel during and after an emergency.
9. Procedures to review and modify the emergency plan following drills or actual emergencies.

Central to all emergency preparedness programs is a written emergency response plan (ERP). An ERP obviously is a key element of emergency preparedness, yet it is only one procedural part of a systematic process that includes the following sequential steps:

1. identify and evaluate hazards
2. mitigate hazards wherever possible
3. identify and evaluate remaining hazards
4. identify and evaluate resources
5. develop emergency procedures and ERP
6. train facility personnel
7. communicate plans to the public
8. integrate with community ERPs
9. conduct and critique drills
10. review and revise ERPs
11. do it again

The creation and maintenance of an ERP requires the allocation of valuable resources, as well as strong leadership from a manager. No one else in the organization has the authority to commit the resources required. Only the power of the manager's office can overcome the organizational inertia that is often encountered.

Many people do not like to "prepare for the worst." They may harbor sincere doubts about the value of planning for events that they consider unlikely to occur. Some people are complacent after years free of serious accidents, and honestly believe that "It can't happen here." Others may be concerned that an open discussion of potentially disastrous events will needlessly upset employees and neighbors. These and other "start-up" problems could make trouble for a manager initiating (or reviving) an emergency preparedness program.

Managers need to find ways to stimulate employee interest and enthusiasm in the planning process. This is best done through the involvement of employees. One strategy that minimizes problems is the early involvement of those employees who serve as emergency responders in their communities. Some of them have witnessed the tragic consequences of failing to heed early danger signs and being unprepared for events. Thus, they can be willing and able catalysts in the emergency response planning process.

Open communication with all employees early in the process also is important. Sharing objectives and encouraging contributions stimulates thoughtful discussion, leads to more complete hazard recognition and mitigation, and ensures a greater chance of effective action when the emergency alarms do sound.

Armed with facts, employees acting as informal ambassadors to the community can reduce the potential for public alarm over a facility's preparations to deal with serious but low-probability events.

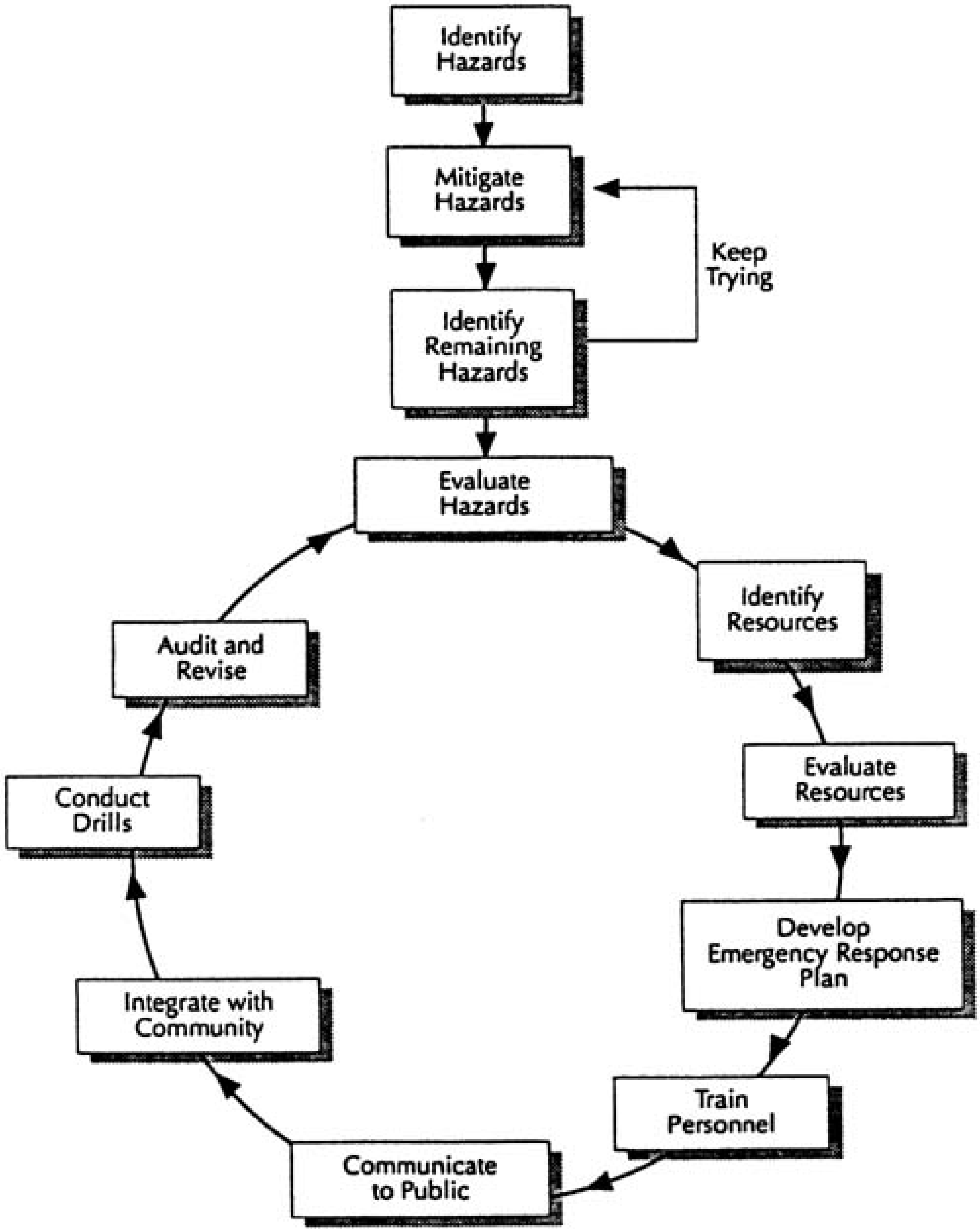
< previous page

next page >

6.3 Hazard Identification and Mitigation

As shown in Fig. 6.2, emergency preparedness begins with the identification and mitigation of hazards. When properly done, the systematic analysis and evaluation of chemical process hazards stimulate actions that eliminate the potential for many emergency situations and pinpoint the situations that remain. The direction of these analytical and corrective efforts is an integral part of managing a comprehensive emergency preparedness program.

Fig. 6.2 Elements of a comprehensive emergency preparedness program.



Process Safety Management Team

Hazard identification and mitigation are the responsibility of a facility's line organizationthe same people who are responsible for all other organizational performance parameters. However, a line organization often needs support in identifying and mitigating complex chemical process hazards, support that can be effectively provided through the formation of a standing team functioning under the manager's direction. This team, called a Process Safety Management Team (PSMT), includes representatives of each unit of a facility from various levels of the organization. They meet on a regular basis (usually monthly) and report frequently to the manager on the status of their activities.

Identifying Hazards: PSR Teams

Supported by the PSMT, the line organization develops a structured approach for performing process hazard analyses on a repetitive basis. One effective way to do this is to organize knowledgeable facility personnel into ad hoc Process Safety Review (PSR) teams. These teams, supplemented by outside specialists as required, are responsible for studying all processes, identifying all potential hazards, and recommending appropriate corrective or control measures.

The makeup of a PSR team is critical to the success of this process. The effectiveness of the team depends on the skills, knowledge, and cooperative effort of its members, and the leadership ability of its chairperson. Each member must be familiar with the process being studied, and have at least a working knowledge of the basic engineering principles and chemistry involved. The team should include supervisors from operations and maintenance and a technical support person. Knowledgeable process operators and maintenance personnel would be valuable team members as well. Others who could contribute to the team's work include design engineers, specialists in electrical and instrument systems, safety engineers, reaction kinetics consultants, and equipment vendors. The selection and training of PSR team members must be carefully monitored. Most managers make this the responsibility of the standing PSMT, which also may assist a PSR team in choosing the most appropriate hazards evaluation method for a specific process.

Review Methods

There are many structured methods for carrying out effective reviews of process hazards. The four most commonly used methods are:

1. what if/checklist
2. failure mode and effect analysis
3. hazard and operability study (HAZOP)
4. fault tree analysis

What If/Checklist

The most frequently used method of process hazard review, the what if/checklist, is effective in reviews of relatively uncomplicated processes from raw materials to final product. The team formulates and answers "What if?" questions at each handling or processing step to evaluate the effects of component failures or procedural errors. They use a checklist to ensure that all important subjects are addressed. This method should be used as the first step in all process hazard reviews.

Failure Mode and Effect Analysis

When the team studies a specific item of equipment, such as a reaction vessel, they often use the failure mode and effect analysis method. Its semi-quantitative approach assists in prioritizing hazards.

Hazard and Operability Study (HAZOP)

The HAZOP procedure systematically questions every part of a process to discover how deviations from the intention of the design can occur, and to determine if the consequences of such deviations are hazardous.

Fault Tree Analysis

Fault tree analysis, the most complex of the commonly used methods, is employed to determine the possible causes of a preselected undesired event. Through the use of logic diagrams and failure rate data, the team can make a quantitative evaluation of the frequency of the undesired event.

For additional information on such methods, refer to Chapter 5.

Recommendations and Reports

Regardless of the method used, the PSR team's most important responsibility is to alert management to serious hazards that may have been overlooked or given inadequate attention. To fulfill this responsibility, a PSR team must take the following steps:

1. Identify the hazards that could cause explosion, fire, release of toxic materials, serious injury, or inappropriate exposure to chemicals.
2. Evaluate the magnitude of the hazards for the areas of probable involvement; the consequences of an event in terms of injuries, environmental harm, and property damage; and, qualitatively or quantitatively, the probability of the hazards' occurrence.
3. Develop practical recommendations to eliminate or control the hazards identified.

The PSR team reviews in detail up-to-date reference material such as:

1. architectural drawings
2. equipment layout drawings
3. process schematics
4. instrument diagrams
5. chemical and physical characteristics of process materials
6. equipment design specifications
7. operating procedures
8. process conditions
9. emergency shutdown procedures

When a facility has more than one chemical process, the PSMT develops a priority order for reviews and recommends a review frequency to the facility manager. The frequency usually ranges from once every two to three years for high-hazard-class processes such as explosives manufacture or acetylene purification, to once every five to seven years for low-hazard-class processes such as alcohol purification, steam generation, and operations involving combustible materials.

Changes not anticipated in the original design of equipment often pose serious problems. Some examples of such changes are:

1. introduction of different raw materials
2. changes in temperatures, pressures, speeds
3. deterioration of equipment

At the completion of each PSR, the team prepares a written report that defines needs, makes recommendations to remedy problems, and recommends priorities for the correction of deficiencies. Following review and acceptance of the report, the facility manager assigns responsibility for corrective action to the appropriate operating personnel. The PSMT then assists the manager in monitoring the status of the recommendations from all PSRs.

Mitigating Hazards: Release Detection and Mitigation

The release of flammable or toxic chemicals from uncontrolled pressure relief vents or as a result of equipment failure may present a serious threat to employees or neighbors who fail to guard against exposure. Every chemical facility must address this potential problem and prepare to protect people from these hazards by working toward reducing the potential hazard and its consequences. Typically, the line organization and PSR teams are responsible for hazard mitigation. They conduct a hazard study that includes the following activities:

1. Detailed appraisal of the potential for a accidental release of toxic gas or vapor.
2. Evaluation of instruments and other methods for detecting such leaks.
3. Provisions for rapid alerting of threatened personnel, and for communicating with emergency responders.
4. Identification of buildings in which people might be trapped by such a release.
5. Assessment of the capability of buildings or rooms to prevent the infiltration of gas or vapor.
6. Evaluation of plans for building evacuations, including the provision and maintenance of appropriate personal protective equipment.

Early *detection* of a chemical leak is necessary in order to limit its effect on people and the environment. The most fundamental method for detecting a chemical release is the systematic patrol of operating units by personnel trained to recognize potentially hazardous vapors using odor or visual observations. The frequency and scope of the patrols will vary with the nature of the process equipment and materials; however, every facility should have documented patrol procedures in place.

When particularly hazardous vapors are present, the patrol procedures should include special provisions to protect the patrollers from the fumes. For example, in facilities producing or consuming hydrogen sulfide, operators may carry emergency respiratory equipment to escape any unexpected fumes that they may encounter, and patrol in pairs or individually, under constant surveillance.

Many chemical facilities supplement operator patrols with an instrumented detection system. Such a system may be a necessary resource when a hazards-study concludes that the system will substantially increase the available escape or emergency response time, or where:

1. The harmful substance is odorless or deadens the sense of smell at hazardous concentrations.
2. The harmful substance is toxic at concentrations undetectable by smell.
3. Large numbers of people may be exposed quickly.
4. Ventilation systems might draw toxic fumes into a building before other means of detection could trigger protective action.

At the core of an instrumented leak detection system is a gas detector. There are many kinds of detectors on the market with varying degrees of sensitivity and selectivity. All require careful regular testing and maintenance. An engineering study of many site-specific factors is required before one makes a choice and designs an appropriately instrumented system. Some generally accepted guidelines are:

1. Early warning of a leak is enhanced if the sensors can be placed near the process equipment subject to leakage.
2. Air movement characteristics are critical in achieving reliable detection.
3. Most detectors respond to several gases or fumes, so the possible presence of all airborne substances affecting the detection system must be considered.
4. When it is necessary to monitor work areas or the intakes to ventilation systems, a highly sensitive system is desirable.

Most detection systems are designed not only to report the presence of hazardous fumes through instrument readouts but to sound an alarm and automatically initiate corrective or protective action. In an office or shop, for example, the system can be designed to shut down all ventilating fans and close exterior air inlet dampers.

When process safety reviews have identified chemical releases as potential sources of facility emergencies, the organization must provide the training and materials needed to ensure a prompt and appropriate reaction to *mitigate* the hazards. Some countermeasures that are effective in limiting the spread of a hazardous material spill or release should be included in the design of chemical process equipment, and described in emergency response procedures. The following discussion of release mitigation is largely excerpted from *Guidelines for Vapor Release Mitigation*, prepared by R. W. Prugh and R. W. Johnson for the Center for Chemical Process Safety of the AIChE (copyright 1988 by the American Institute of Chemical Engineers, reproduced by permission of the Center for Chemical Process Safety of the AIChE).

"Water, steam, and air curtains and water sprays are primarily effective in dispersing and/or diluting vapors with air to reduce the severity of effects of a hazardous vapor release. In some cases, vapors can be partially 'knocked down' or absorbed after release."

"Ignition source control and deliberate ignition are also possible vapor release countermeasures." "For areas around processes handling flammable vapors, ignition source control is practiced to reduce the probability of vapor ignition if a leak occurs." "Administrative controls are exercised on plants where flammable materials are processed. Such controls may include hot work permits, restricted smoking areas, not allowing lighters or matches on the site, and electrical grounding and bonding procedures."

"Deliberate ignition is a countermeasure against spills of highly toxic materials which are also flammable, such as hydrogen sulfide, hydrogen cyanide, and methyl mercaptan. Igniting nontoxic flammable materials such as hydrocarbons may present hazards outweighing possible advantages." In any case, deliberate ignition must be carefully planned and executed so that the resulting fire is truly controllable.

"Practical methods for combating vapor from liquid leaks are dilution, neutralization, or covering. All three reduce the vaporization rate of the pool. Water dilution is effective for spills of water-miscible or water-soluble material. Spraying water into the spill reduces the vapor pressure by reducing the concentration of the liquid." "For acidic spills, limestone or soda ash is often used" to react with the spilled liquid to produce a less volatile salt or ester. "A foam cover can be effective in reducing vaporization from spills," and "dense liquids can be covered with lighter immiscible nonreactive liquids" to accomplish the same thing.

6.4 Preparing for Emergencies: Identifying and Evaluating Resources

The process of hazard identification, evaluation, and mitigation, when sustained over a period of time and coupled with other good safety management practices, can prevent most process-related emergencies. However, the potential for various low-probability process-related events will remain. Should one occur, there must be resources available to promptly bring the event under control. These resources include designated personnel, plans, systems, and facilities that are needed for effective action and communication. The facility line organization is responsible for providing and maintaining these resources.

Personnel

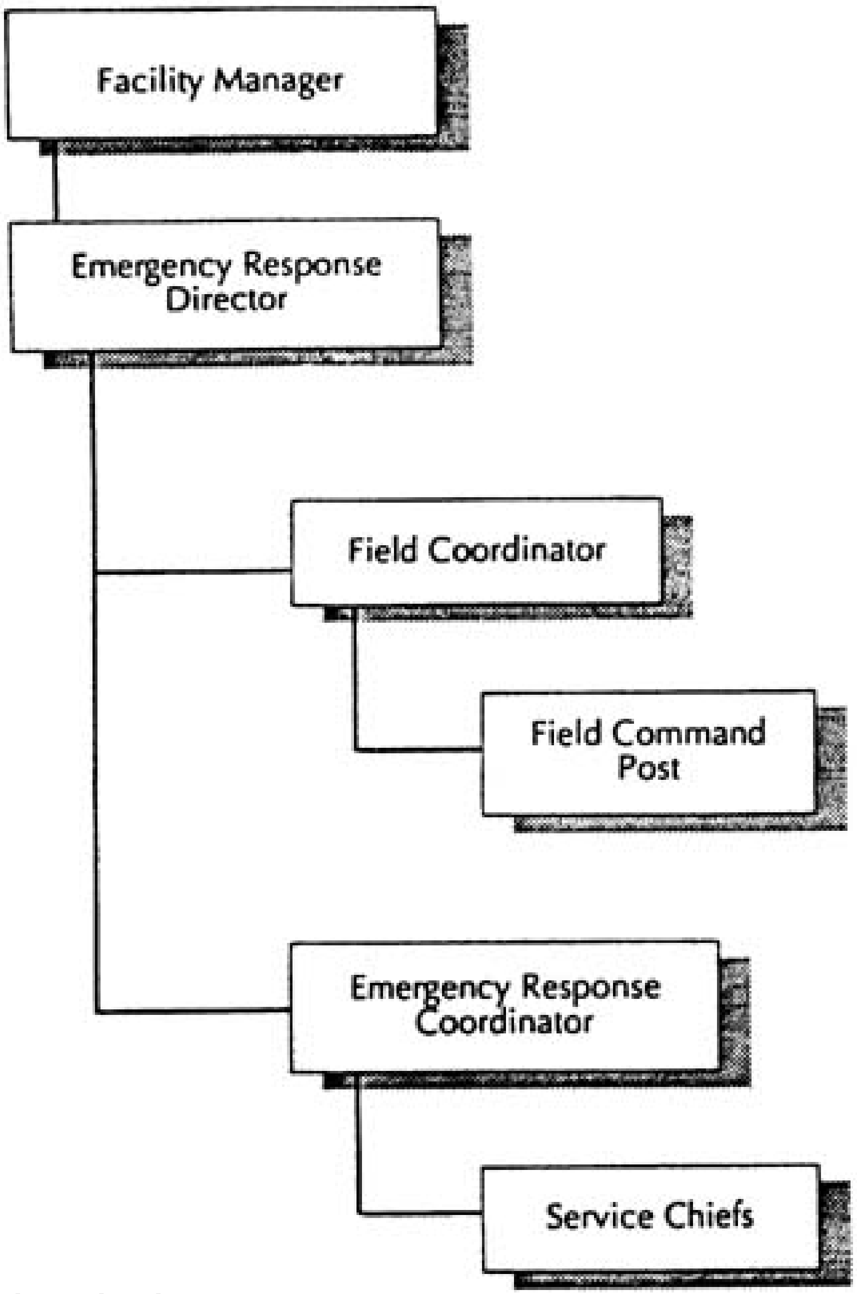
Emergency Preparedness Team

As with the PSMT and the PSR team, the efforts of the line will benefit from the support of a standing Emergency Preparedness Team functioning under the direction of the facility manager. This team, which includes representatives of each unit of the facility, leads the development and maintenance of an ERP for the facility, and monitors specific procedures and training for handling emergency situations of all kinds. It should meet on a regular basis (usually monthly), and report frequently to the manager on the status of its activities.

Emergency Management Organization

Emergencies demand rapid, well-coordinated decisions, communications, and action to bring them under control as swiftly as possible. There must be a formal Emergency Management Organization (EMO) whose purpose is to achieve this objective. The structure of the organization is not critical as long as it is capable of rapid assembly, carries out its assigned responsibilities, and meets the needs of the facility. A generic EMO is shown in Fig. 6.3; the EMO is described below, in the section on "Developing an Emergency Response Plan."

Fig. 6.3 Generic emergency management organization.



Fire Brigades

Most chemical facilities of any size have established fire brigades as a key resource in their emergency preparedness programs. By virtue of its training and familiarity with the equipment and the physical layout of a site, the brigade can be a major factor in *preventing* incidents as well as in minimizing injuries and property damage due to fire, explosion, or other causes.

The size and the structure of a fire brigade vary, depending upon the hazards present, the size of the facility, and the level of internal fire-fighting capability desired; and large, multi-process facilities usually have several unit or area brigades. Each brigade has a designated captain and an alternate. The training of brigade members must be commensurate with the duties and the functions that they are expected to perform, and thus depends on the fire prevention plan option selected by site management. (Fire prevention plans are discussed later in this section.)

It is best to have a written statement establishing the brigade and its duties, specifying its size and organizational structure, and outlining the type, amount, and frequency of training provided. In the United States, OSHA has promulgated minimum standards for fire brigades, which may be found in 29 CFR 1910.156.

Plans

Work Unit Plans

In a large facility, each work unit has emergency procedures for its own area of operations, and these work unit plans form the foundation of the facility ERP. Thus, the facility's emergency response can be no more effective than the recognition and response capabilities of the employees in the facility work units.

The primary objectives of these unit plans are (1) to control and contain any emergency condition within the unit, and (2) to provide protection for unit personnel and equipment from events originating outside the unit. Unit plans specify who has authority to take emergency action, and how escalation to a full facility emergency occurs. Subjects that must be covered include:

1. unit emergency command and coordination
2. communications within the unit
3. communications with other units
4. emergency assessment
5. unit evacuation and personnel accounting
6. emergency shutdowns
7. communications with the facility management
8. criteria and procedures for securing resources from outside the unit

Unit plans clearly assign primary responsibility for initial emergency assessment and reaction to the lead unit employee on site at the time of the event. Among the factors to be considered by that individual in making the initial assessment are:

1. previous experience with similar situations

2. how long the situation has existed
3. what might occur "If …"
4. properties and hazards of materials involved
5. other complicating situations

Using the information gained from the assessment, unit personnel take whatever immediate action is required to protect people and property on and off the site.

Work unit plans should contain guidelines for assuring uniformity in the assessment of and reaction to unusual events. An effec-tive approach is to develop an emergency classification system that includes criteria for classification and guidance for the appropriate response actions for each level of emergency that is defined.

For example, an emergency judged to be controllable within the unit, with no evacuation outside the unit and no impact in the community, is a Level One Emergency. Small releases of hazardous materials usually fall into this category. Appropriate response items to consider would include:

1. protection of unit personnel
2. shutdown or isolation of affected equipment
3. notification on- and off-site
4. containment and cleanup

Similarly, other events are categorized by their potential or actual severity. Preplanned response checklists are prepared for each one. An incident that requires response by facility personnel outside the affected unit but has no impact on the community is a Level Two Emergency. A Level Two Emergency requires activation of the facility ERP. If some community impact becomes probable, the emergency is a Level Three Emergency. That level of emergency triggers additional response requirements, including interaction with the community.

The unit plan must be formal and in writing. Because the emergency procedures of each individual operating area or work unit form the foundation of the facility ERP, they should be reviewed by the Emergency Preparedness Team for completeness and consistency across the site before they are included in the facility ERP.

Fire Prevention Plan

One of the most basic resources that all chemical facilities must have is a fire prevention plan. To be fully effective, this plan must be in writing, and it must be reviewed with all employees on an established schedule.

A comprehensive fire prevention plan includes the following subjects:

1. Major facility fire hazards.
2. Storage and handling practices for combustible and flammable materials.
3. Identification of potential ignition sources.
4. Procedures to maintain systems and equipment installed to control ignition sources.
5. Names and titles of those responsible for the system and equipment maintenance and the control of fuel source hazards.
6. Procedures to minimize accumulations of flammable and combustible waste materials.
7. Plans for communicating to all employees the fire hazards and their specific responsibilities in the event of fire.

Fire response options open to facility managers range from a plan to evacuate all employees without attempting fire fighting on any scale, to full involvement of an established fire brigade in fighting advanced-stage structural fires. The fire prevention plan must specify which option applies to the facility. The option selected determines the type and extent of education and training required, as well as the type of equipment needed on the site.

Fire prevention and response information may be covered in other written facility documents (such as job descriptions, rules, or procedures) and communicated to employees on the same schedule as the fire prevention plan.

Evacuation Plans

Many emergency events require the evacuation of at least some facility personnel, if only for precautionary reasons. This presents other needs, which must be addressed in the evacuation plan. Personnel who are directed to evacuate their normal work stations need visible wind direction indicators to help them move away from a dangerous gas cloud drifting downwind of the release point. If the building is enveloped in the cloud, however, they will need personal protective equipment in order to evacuate the building safely. When there are processes that must be rapidly shut down and/or isolated as evacuation proceeds, the plan must include written procedures and appropriate means to ensure incident-free action.

Headcount Plan

The evacuation plan also must establish gathering points for all personnel in order to identify those who are missing. The procedure for accounting for personnel, called the headcount, must work effectively and rapidly in the first minutes of the emergency to determine if anyone is missing or known to be injured. Communications to the fire brigades or other first responders must be quick and accurate so that search and rescue operations and medical aid can be successfully implemented.

Each work unit in a facility must have its own plan for headcounts as a part of its emergency preparedness program. This plan must be in writing, and it must be practiced frequently enough to ensure good execution in a time of real emergency. It should include:

1. a designated assembly point for each employee
2. an alternate assembly point, should the primary location be inaccessible
3. a procedure for counting at each assembly point
4. a designated unit headcount coordinator and backup
5. primary and alternate phone numbers where coordinator may be reached
6. phone numbers for reaching the personnel chief, who coordinates and summarizes the site-wide accounting

Headcount procedures must include guidelines for accounting for visitors to the facility at the time of the emergency. These visitors may include truck drivers, vendors, and contractors. Some guidelines are:

1. All visitors must sign in and out, preferably at one designated gate.
2. Visitors will have a designated host employee responsible for their safety whenever possible.
3. Everyone temporarily on the facility must report to the assembly point of the unit he or she is in at the time of the emergency.
4. Unit procedures must include reporting these "extra" people by name to head-count headquarters.
5. The personnel chief of the EMO must have access to the log at the visitors' gate.

Alarm Systems

Another essential resource that every chemical facility must have is an effective alarm system, which initiates action by endangered personnel and emergency responders.

A satisfactory facility emergency alarm system must meet three additional requirements. It must, at a minimum, immediately alert *all* the people on the site. (A single signaling device should be adequate for small sites; however, large facilities may require many devices placed to reach all occupied areas of the site.) The system must function even when the facility has a general power failure. Alarm activation controls should be located so that an emergency condition is unlikely to prevent access to them.

On most older chemical facilities, the basic component of the emergency warning system is the fire alarm. It is essential that there be different signals to distinguish between an actual fire and other threatening events, as the action taken in response to a toxic gas release, for example, may be quite different from that taken in response to a fire. On multi-process sites, the signals also should identify the work unit involved.

Depending upon the number of people at a facility and their familiarity with response and evacuation procedures, it may be necessary to supplement signals with verbal instructions delivered by public address equipment, radios, or automated telephone systems. For reliability at any time of day, a system using verbal instructions must be located in a regularly manned job station such as a control room or gate house.

Any alarm system used must communicate clearly the nature of the emergency event and its severity. In a Level One Emergency some people may have to be evacuated in order to ensure their safety, but by and large the emergency can be contained within the unit. In that case, information usually may be given to unaffected employees through normal telephone communications.

In the event of a Level Two or a Level Three Emergency, each facility must have an easily recognizable signal that clearly communicates that a major incident has occurred. Some facilities have alarm boxes that will automatically sound this signal when pulled. Others have boxes that require manual operation. Still others rely on a telephone message to trigger the signal.

In some facilities, the "major" emergency alarm is a steady blast on the facility steam whistle. Others use sirens or special bells. Whatever is chosen, the alarm must reach everyone on the site, triggering activation of the facility ERP and the initiation of a headcount procedure.

It is essential that every employee on a chemical facility knows how, where, and when to turn in an appropriate and effective emergency alarm. The specifics will vary considerably, but the procedure for turning in an alarm is similar in all cases.

Employees first need to know how to report a fire. On most facilities, this is done by using a fire alarm box or by telephoning a central station such as a guard house from which the fire alarm can be sounded. Chemical facilities need to have well-developed training and drill programs in place to be sure that response procedures and equipment are maintained.

When fire alarms are supplemented by verbal instructions, the facility personnel who are contacted need to know the location and the nature of the fire, the action under way to control it, and the actions required (if any) by those not involved. There are many commercially available communications systems that deliver this information efficiently, even on large sites.

Facilities for Protection and Communication

Safe Havens

When accidental releases of a toxic gas occur or threaten to occur, the immediate protection of on-site personnel is of paramount importance. One resource available for this purpose is a building or room that by the nature of its construction and its heating and ventilating characteristics can prevent the infiltration of intolerable concentrations of the toxic substance. The best location for a safe haven is determined by an engineering study. That location then is documented in the work unit and facility ERPs.

Safe havens may be rare on some sites; however, many buildings can provide personnel with temporary protection until the incident has been analyzed and a decision made on the need for evacuation. To qualify as a designated temporary safe haven, a building must be reasonably well sealed against air infiltration, with adjustable ventilation systems that can reduce or close off exhaust vents and outside makeup air. Emergency procedures should state how long a building can be considered a safe haven under specified exposure conditions.

Main Emergency Control Center

Using a list of safe havens developed by the work units, the Emergency Preparedness Team helps management select one building or room to be used as a control center in case of a major emergency. This Main Emergency Control Center (MECC) should be as remote from potential hazards as possible in order to serve as a reasonably safe haven to those involved in directing, coordinating, and communicating activities for the duration of an emergency.

MECCs need to have basic information readily available and maintained in an up-to-date, standby condition. Such a collection would include:

1. copies of the facility ERP
2. facility maps and diagrams
3. process material isolation points
4. fire control maps and diagrams
5. maps of the surrounding area
6. aerial photos of the facility and the surrounding area
7. names, addresses, phone numbers for:
    1. all facility employees
    2. off-site company people to be notified
    3. groups and organizations who may be notified
    4. community officials who may be notified
8. Material Safety Data Sheets (MSDSs) for facility materials
9. MSDSs for the materials of neighboring chemical facilities
10. copies of mutual aid agreements
11. highlights of the facility's history, products, and performance

The MECC should have the following basic equipment:

1. adequate telephones and lines
2. unlisted telephone and/or a hotline
3. two-way radios
4. fume path projector
5. FAX machine
6. regular radio with tape recorder
7. regular television with recorder and playback video cassette recorders
8. chart pads and stands
9. battery-powered lighting
10. personal protective equipment
11. hand-held tape recorders

Alternate Control Centers

Recognizing the unpredictable nature of a developing emergency, it is advisable to select at least one alternate on-site main control center. Both the MECC and the alternate on-site center should have backup power supplies.

The presence of substantial quantities of explosive or toxic materials on-site may justify establishing yet a third location off-site, in case of a complete evacuation of the facility. This could be a fixed location in the nearby community, or a mobile unit such as a truck or van properly equipped for managing an emergency. Some of the supplies and equipment for the alternate off-site main control center may have to be packed in a readily transported kit rather than being on standby at the alternate off-site location.

Media Headquarters

In an emergency, it is essential to maintain regular contact with the public and the media. A separate location for communications will allow that contact to continue without interfering with the operations of the MECC or overloading its communications system. There, the public affairs chief and the facility manager can receive media representatives and provide periodic updates on the emergency situation. The media headquarters should be equipped with several direct outside telephones and the facilities to make the reporters comfortable for the duration of the emergency. Permanently mounted facility plot plans or aerial photographs are helpful resources, as are handouts describing in general terms the facility and its products.

6.5 Developing an ERP

Following the identification and evaluation of the resources available to the facility, the Emergency Preparedness Team, under the direction of the manager, develops the written ERP for the entire facility. The Emergency Preparedness Team is responsible for ensuring that the written facility ERP informs all employees of their roles in an emergency and the hazards to which they may be exposed.

An ERP must be tailored to a specific facility, reflecting its unique conditions and individual needs. Among the variables that affect the details of a plan are:

1. materials used, produced, or stored
2. nature of the operations
3. available employee skills
4. geographic location
5. proximity to other facilities
6. available emergency resources
7. mutual aid agreements

No two ERPs will be exactly alike, nor is there one best outline. The plan of a production facility will differ from that of a warehouse. However, every plan must include:

1. initial alarm procedures
2. emergency escape procedures and routes
3. emergency operating and shutdown procedures for critical operations
4. accounting procedures for all on-site personnel
5. rescue and medical duties for specified personnel
6. procedures for communicating the emergency inside and outside the organization
7. structure, duties, and resources of the EMO, including the names or titles of people with detailed knowledge of the plan and its assigned duties
8. reference material

Plan Design

The ERP is a working plan, applicable to any event with emergency potential occurring at any time and at any location on the facility. It should be written concisely, with diagrams and checklists used wherever possible, so that it may be effectively used for guidance during an actual emergency. To accomplish this, it is helpful to divide the ERP into several major sections, which are in turn divided by subject.

For example, an ERP may be broken down into three sections: the Emergency Management Organization, Action Plans, and Reference Material.

The first section documents the structure, duties, resources, and communications systems for the facility EMO, and the conditions that will trigger the EMO's response. It includes plans for site-wide notification and response.

The second section includes detailed action plans for each particular type of emergency. This section includes summaries or outlines of the emergency procedures developed by each work unit of a facility, and may contain considerable process-oriented information.

The third section consists of reference material on plan philosophy, training, plan maintenance, drills, and similar supportive data that usually are not needed at the time of an emergency. Members of the EMO and others with assigned response duties should be able to quickly locate and refer to the appropriate action guidelines, to determine that key functions are being performed.

Using this approach, the outline of an ERP for a multiprocess chemical facility would look like this:

1. Section I
    1. Table of Contents
    2. EMO Structure and Tasks
    3. Notification: On-Site
    4. Accounting for Personnel
    5. Headcount Center
    6. EMO Center
    7. Emergency Scene
    8. Notification: Off-site
2. Section II
    1. Table of Contents
    2. Work Unit 1
        1. Emergency types
        2. Action checklists
        3. Responsibilities
        4. Communications
        5. Shutdown procedures
    3. Work Unit 2 (same as above)
3. Section III
    1. Table of Contents
    2. Plan Philosophy and Objectives
    3. Training Programs
    4. Drills and Tests
    5. Return to Normal
    6. Appendices
        1. Special situations (e.g. bomb threats)
        2. Hazardous material data
        3. (Others as required)

EMO Structure

Perhaps the most essential element of an ERP is the establishment of the organization that will manage an emergency response once the event has developed beyond a Level One Emergency. The EMO must be capable of rapid assembly, and must be able to carry out all of its assigned responsibilities.

The EMO is composed of the facility manager, the emergency response coordinator, the field coordinator, and various service chiefs. The number of service chiefs and their duties will vary. There need not be a separate individual in charge of each service, and small facilities may combine two or more services under one individual, as long as each function is performed.

The facility manager, in cooperation with the Emergency Preparedness Team, assigns people, by their title or function at the facility, to the EMO, with designated alternates for each EMO function. The duties and responsibilities of each EMO function are defined and documented in the ERP. Because there must be a functional EMO in place regardless of when a facility emergency occurs, designated shift personnel take specific EMO responsibilities until the primary members can reach the site.

All the positions established for the EMO should have written position descriptions, which may be placed in the third section of the ERP as an appendix to the plan. In the first section of the ERP, the EMO summary need only include a concise checklist for each position, with diagrams showing relationships and major communication flow paths.

Facility Manager

In the EMO, the facility manager has the overall responsibility for protecting people on-site, facility property, the environment, and the public during and after an emergency. With the assistance of the public affairs and communications chiefs, the manager usually serves as the spokesperson for the facility and the company, communicating with representatives of the media and other concerned audiences.

Emergency Response Coordinator

The designated emergency response coordinator, who may also serve as the chairperson of the Emergency Preparedness Team, directs all emergency control activities from the MECC. All other service chiefs on the EMO report to the coordinator. Using information from the emergency scene and from the service chiefs, the coordinator makes the key decisions on what should be done, and coordinates activities on and off the site. The coordinator reports to the manager, who should be available nearby for overall guidance and counsel.

Field Coordinator

The job of the field coordinator is to correct the emergency situation as rapidly as possible with minimum risk to those in or near the affected area. He or she establishes a field command post as close to the scene of the emergency as can be safely done. The post often is in or near a radio-equipped building or vehicle, thus permitting rapid establishment of communications between the command post and the emergency response coordinator.

Ideally, the field coordinator is familiar with the operations and materials involved in the emergency; so she or he often is the highest-ranking supervisor of the affected area who is available at the time of the event. That supervisor may remain as field coordinator for the duration of the emergency, or may be relieved by another designated member of management. In the latter case, the area supervisor may become a service chief with greater hands-on involvement, with the field coordinator concentrating on marshaling required resources and maintaining effective communications. The choice depends to a great extent on the size and complexity of the facility, and the resources available for the EMO.

Public Affairs Chief

Working in close coordination with the facility manager, the public affairs chief releases appropriate information to the news media, regulatory personnel, government officials, and other public groups and individuals. No information is to be made public by anyone other than the facility manager without specific direction from the public affairs chief. The objective is to provide full and accurate statements in a timely fashion, so that public attention is focused on facts and useful information rather than on rumors and speculation.

The public affairs chief also establishes and monitors the media headquarters, which is isolated from the EMCC to avoid interference with operations there. The public affairs chief's function includes assisting the manager in the preparation of formal statements and background information to be distributed to reporters. It also may include arranging with local radio and television stations to make periodic announcements during an emergency so that the public and the employees not on the site are properly informed. The public affairs chief also arranges to monitor and perhaps record the broadcasts of local radio and television stations in order to determine what additional statements or clarifications to the public may be required.

Communications Chief

The communications chief establishes and maintains communications capability with appropriate people on and off the site. This chief must be familiar with the various communicating systems available, including telephones, public address systems, two-way radios, and messengers. The job includes recommending and coordinating revisions or additions to the communicating systems during an emergency, and assisting the facility manager and the public affairs chief with any communications as needed.

Engineering Chief

The engineering chief's primary responsibility is to maintain electrical power for vital services. These services include on-the-scene portable lighting, continuous fire pump operation, and a steady supply of nitrogen, steam, and process cooling water. Another key duty is to assemble repair groups capable of isolating damaged sections of pipelines, electrical lines, and other necessary equipment in order to contain problems and maintain or restore operations outside the affected area. Because these repair groups must be drawn largely from site personnel such as electricians, welders, riggers, and pipefitters, the position of engineering chief should be filled by someone from the facility maintenance organization.

Emergency/Fire Chief

The emergency/fire chief is responsible for fire-fighting and fume control activities. The best person for this job has good knowledge of and access to the fire-fighting, rescue, and fume control resources available on and off the site. On a small facility this may be the captain of the fire brigade. The job includes direction of designated facility personnel, and coordination with any outside forces brought in to bring the fire or fume condition under control. The environmental chief may assist in determining optimum courses of action based upon actual or threatened adverse effects on air and water emissions from the facility.

Medical Chief

The medical chief ensures that the proper medical care is provided to people on the facility who have been injured or exposed to toxic materials. When facilities have medical professionals on the site, one of them carries out this function. This chief's responsibilities include establishing field stations to treat affected personnel, and, in cooperation with the transportation chief, providing transportation for injured people to other medical facilities.

The medical chief also participates in discussions with community officials and appropriate facility personnel regarding actual or potential medical problems for people outside the site boundaries. Representatives from nearby local hospitals, ambulance services, fire departments, police forces, and emergency management groups also may be involved in these discussions. The medical chief communicates any action or contingency plans developed in the discussions to the emergency response coordinator and the public affairs chief.

Environmental Chief

The environmental chief oversees all activities designed to minimize adverse effects upon the quality of air and water as a result of an emergency. The function includes coordinating air and water quality monitoring on and off the site during and after the emergency, and assisting the emergency/fire chief in selecting the optimum approach to abating a fire or fume condition. The environmental chief also provides assistance in projecting the path and concentration of a fume release, using computer modeling (if available) or maps with plastic overlays, and developing an effective plan of action.

Personnel Chief

The personnel chief is responsible for accounting for all personnel on the facility at the time of the emergency. Unit emergency programs must include training on how this is accomplished. Unit supervisors initiate the process with headcounts at designated rally points. The success of rescue and medical activities depends on how quickly and thoroughly this information is obtained and reported to the personnel chief.

The personnel chief also coordinates communication with relatives of injured or deceased employees, and makes certain that this is completed before any names are released publicly. This requires close co-ordination with the public affairs chief and the manager.

The personnel chief need not be located at the MECC, but it is imperative that the personnel chief be in close touch with other members of the EMO. Telephones in the headcount center should have answering and message recording equipment to capture any unit reports arriving before the headquarters is staffed.

Security Chief

The security chief makes sure that entry to and egress from the facility are properly controlled. This involves securing gates; limiting entry to authorized personnel; registering all who pass through gates; meeting visitors, including representatives of the media, and escorting them to proper locations; and controlling all traffic on the site. The function also includes communicating with local police so that access to the facility is maintained, and crowd control procedures are enforced. Usually, the individual serving as security/chief at the time of the emergency also is responsible for initiating procedures to summon key facility personnel and urgently needed outside agencies.

Transportation Chief

The transportation chief coordinates and controls all transportation on the facility. This includes directing the assembly of available vehicles and crews, and identifying needs beyond site capabilities such as cranes, trackmobiles, and bulldozers that must be obtained from outside organizations. The function also includes providing suitable transportation for facility employees who monitor the effects of emergencies beyond site boundaries, or interact with community officials at an off-site location.

6.6 Training Personnel

Having a written facility ERP that is supported by established work unit emergency plans and procedures it is an important part of the manager's job to ensure that unplanned events will be promptly controlled with minimum risk to people, property, and the environment. There is much important work to be done, however, before the ERP is anything more than a paper resource. Facility personnel must be trained to use the ERP effectively. There must be frequent drills to test the plan and the people against the standards established by management as well as those established by law.

Within the facility, there are three groups of people who require training. First, there are members of the fire brigade, who must be trained to fight fires at the level specified by facility management. Next are the employees who have been assigned active roles in controlling emergencies of all types. This group includes members of the EMO and their alternates, plus designated support people such as headcount coordinators, guards, and emergency repair personnel. Finally, everyone else on the facility must have a basic understanding of the ERP, and must know how to respond when specific alarms sound.

The facility manager is responsible for ensuring that the appropriate training and retraining are done in a timely and effective manner. An employee should be trained when he or she is hired, at least annually thereafter, and when the employee's work area changes or the plan is revised. Most managers make this primarily a line organization function. They expect facility supervisors to use all available resources and means of education to accomplish the tasks. Some specialized training assistance, however, must be provided, particularly for the members of a fire brigade.

Fire Brigade Training

Training programs for fire brigades have two major objectives. One is to inform the brigade members of new hazards at their facility and innovations in fire-fighting techniques and equipment. The other objective is to provide hands-on training for developing skills in emergency operations and using equipment, including:

1. portable fire extinguishers
2. hoses and accessories
3. portable lighting
4. forcible entry tools
5. ladders
6. salvage equipment
7. first-aid supplies
8. replacement parts
9. personal protective equipment
10. transportation equipment

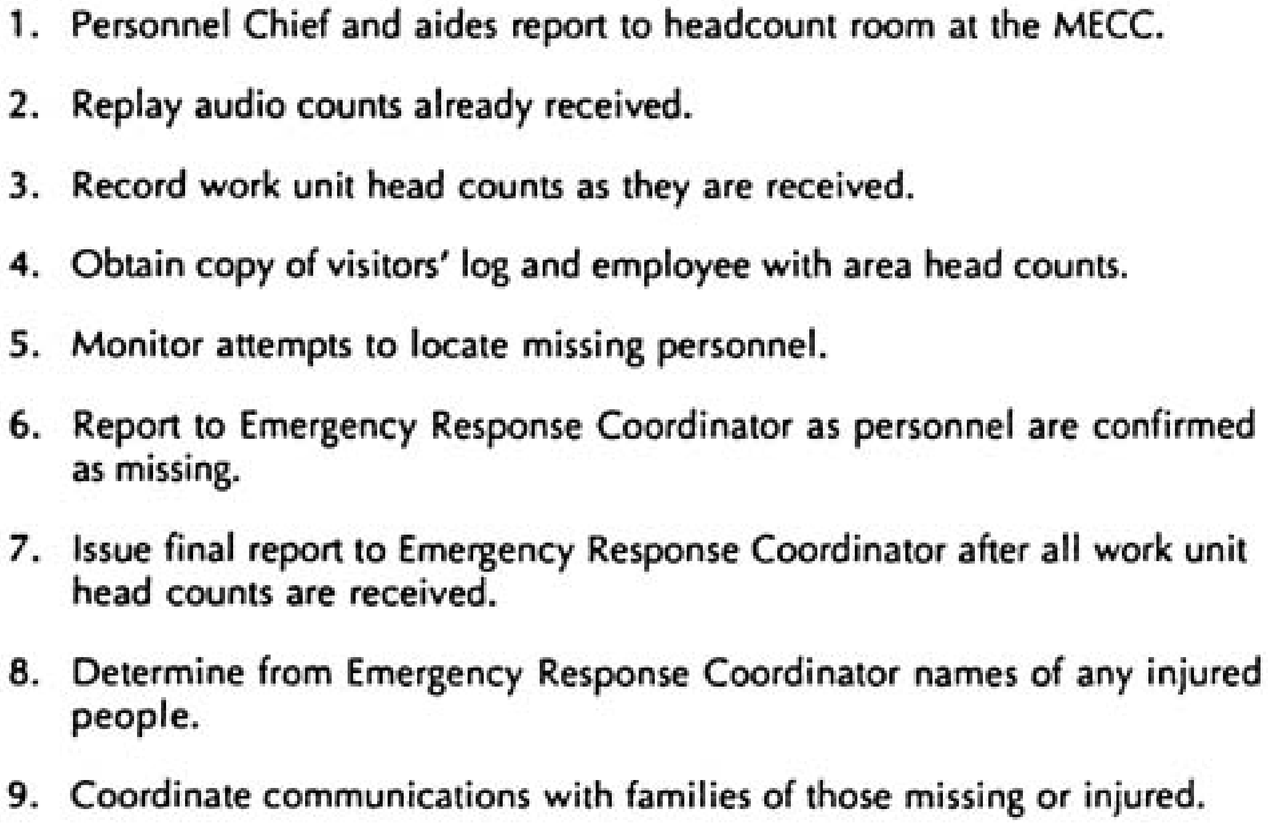
A comprehensive training program for fire brigades must include classroom and hands-on training.

Outside resources can provide valuable assistance in the education and training of brigade members. Local fire departments and state fire schools are usually enthusiastic partners in such efforts. Often they are the key to securing adequate resources at a reasonable cost. In industrialized communities, mutual aid agreements may include cooperative training provisions with other chemical facilities, which provide opportunities for even more effective use of available resources. All mutual aid agreements should require cross-training in special hazards at the other facilities.

EMO Training

Training for personnel assigned to the facility EMO can be led by the emergency response coordinator, who meets periodically with each member of the EMO to review and refine position descriptions and the associated functional checklists. An example of such a checklist is shown in Fig. 6.4. The meetings may be followed by limited drills involving only the people and responsibilities included in the individual EMO function. Some facility managers assign specific emergency response training duties to each member of the EMO. For example, the engineering chief organizes, equips, and trains the repair groups who will be called on to physically stabilize a situation at the time of an emergency. The emergency/fire chief could be given the responsibility for maintaining a trained force of fire and fume fighters, which includes competent leadership on all shifts.

Fig. 6.4 Personnel chief functional checklist.



1. Personnel Chief and aides report to headcount room at the MECC.
2. Replay audio counts already received.
3. Record work unit head counts as they are received.
4. Obtain copy of visitors' log and employee with area head counts.
5. Monitor attempts to locate missing personnel.
6. Report to Emergency Response Coordinator as personnel are confirmed as missing.
7. Issue final report to Emergency Response Coordinator after all work unit head counts are received.
8. Determine from Emergency Response Coordinator names of any injured people.
9. Coordinate communications with families of those missing or injured.

Employee Training

There are many ways to be sure that all employees understand the ERP. The Emergency Preparedness Team can assist in the assessment of existing training programs, alerting the organization to training weaknesses and suggesting or providing creative ways to overcome those deficiencies. Unit supervisors must periodically review the ERP in scheduled group safety meetings or with individuals. Key plan elements can be reproduced on wallet-size cards, desktop displays, or telephone stick-ons. Individual or group discussions of how to react in given situations can detect weaknesses in procedures, training, or understanding.

Training programs for the three groups of employees discussed here should be documented, reviewed regularly, and included in a reference section of the facility ERP. But no matter how comprehensive the training programs may be, their effectiveness is unknown until a drill of the ERP is conducted and its systems and procedures are tested.

Facility Drills

To evaluate the effectiveness and completeness of an ERP, a facility must conduct periodic announced and unannounced Level Two and Level Three emergency drills. These are in addition to the more frequent Level One unit drills that are held to ensure that the more limited response procedures of a work unit are complete and well understood. Major internal emergency drills should be held at least four times each year, and scheduled to involve each working shift at least once a year.

To gain the maximum benefit from the drills, assigned observers should witness all aspects of response activity, and gather soon after to participate in a verbal evaluation of actions taken. These observers should include members of the Emergency Preparedness Team. This process is enhanced by capturing on-scene action on videotape or in still photographs. Critiquing can be extremely valuable in identifying necessary plan changes, training needs, and resource deficiencies of all kinds. The process of critiquing must be controlled and managed. The leader must be sure to identify the strengths as well as the weaknesses that were revealed in the drill. The objective is to stimulate actions for positive change, not to assign blame.

Drills should be held on weekends and at night occasionally to test segments of the organization that work at times when all of the specialized resources of the facility are not immediately available. Initially, limited scale drills can test segments of the ERP, such as manning the MECC with the shift personnel that are available and carrying out a headcount without involving daytime employees.

As the proficiency of the organization increases, drill scenarios can become more complex. Complicating factors approaching worst real-life conditions should be introduced periodically, including:

1. telephone switchboard overload
2. absence of key EMO members
3. arrival of major TV network anchorperson
4. simulated mass casualties
5. two-way radio failure
6. evacuation of primary and backup MECCs
7. major community impact

The Emergency Preparedness Team, with approval of the manager, designs the drill scenarios, monitors the organization's performance, provides leadership for critiquing, and recommends corrective actions. The team also develops and monitors a plan for involving the community in the important task of integrating a facility's ERP into the public emergency preparedness programs of the region.

6.7 Involving the Community

Of major concern to the chemical industry is the public perception that facility managers have little concern for the welfare of their neighbors. Managers themselves have contributed to such false impressions by failing to interact with their communities in a consistent and meaningful way.

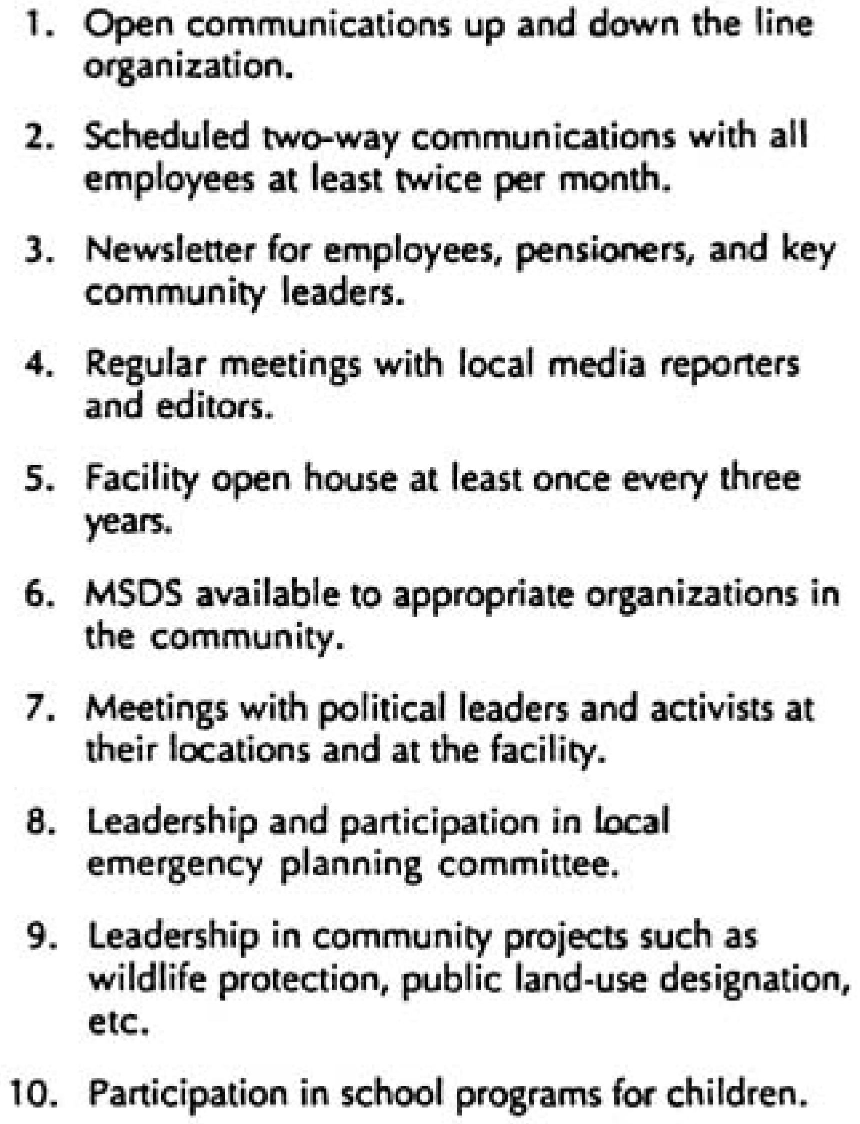
The public clearly wants to know more about the risks presented by a chemical facility. Increasingly, the public wants to help decide which risks are acceptable and which are not.

Recognizing that industrial facilities exist only with the consent of their host communities, most industrial organizations are assigning a high priority to building stronger bridges with their neighbors. The chemical industry in particular is finding that the involvement of the community in the process of emergency preparedness planning presents an excellent opportunity for constructive two-way communications. This has proved particularly productive if these communications are part of an aggressive and continuous risk communications process.

Communications

Each chemical facility needs to create opportunities for sharing information with its surrounding community; it should not wait for these opportunities to occur. Facility managers in particular should be active in the process, and should monitor the performance of their organizations. Some items that should be on a manager's checklist for risk communications are shown in Fig. 6.5.

Fig. 6.5 Risk communication checklist.



1. Open communications up and down the line organization.
2. Scheduled two-way communications with all employees at least twice per month.
3. Newsletter for employees, pensioners, and key community leaders.
4. Regular meetings with local media reporters and editors.
5. Facility open house at least once every three years.
6. MSDS available to appropriate organizations in the community.
7. Meetings with political leaders and activists at their locations and at the facility.
8. Leadership and participation in local emergency planning committee.
9. Leadership in community projects such as wildlife protection, public land-use designation, etc.
10. Participation in school programs for children.

The process should start with employee communications. In its employees, a facility has an important, informal communication link with its community that is often ignored. Employees deserve to know at least as much about their facility as their neighbors, and they deserve to know it first. They should be aware through communications (if not through actual involvement) of the facility's entire emergency preparedness program, from prevention to preparation. The employees also should be familiar with the products made and their end uses. They should understand the potential hazards of the processes and materials with which they work, and how to protect themselves and the public from those hazards.

Other communication channels should be developed and regularly used. Scheduled meetings with representatives of community emergency service groups are useful for exchanging information and objectives concerning emergency preparedness, for promoting the sharing of resources, for gaining familiarity with one another's physical facilities and people, for identifying problems, and for recommending action for their solution.

In the 1980s, some chemical facilities and their communities formalized this approach and broadened participation in their meetings to include public officials and representatives of regulatory agencies and the media. Operating as Hazardous Material Advisory Councils (HMACs) and meeting regularly with established leadership and agendas, these groups quickly became key resources in organizing a community's efforts to better understand the potential for hazardous material incidents and to protect against them. Typically the responsibilities of an HMAC include coordinating a regional risk assessment, assisting the development of a community response plan specific to hazardous materials, and assisting with educational programs for various segments of the public.

The Responsible Care program of the American Chemistry Council embraces HMACs as one good way to communicate relevant and useful information that responds to public concerns for safety, health, and the environment. However, managers are finding many other ways to interact with the public to achieve a fuller measure of community awareness and involvement in affairs of mutual interest. Among the many options from which a manager may choose are the following:

1. hosting facility tours featuring emergency prevention and mitigation procedures
2. speaking at community meetings (service clubs, schools, governing bodies, etc.)
3. sending newsletters to selected neighbors
4. preparing informative brochures or newspaper inserts
5. appearing on local TV or radio

An open and sincere comprehensive risk communication process led by the facility manager creates a better-informed public that is able to understand real risks (vs. perceptions) and is likely to respond effectively in case of an actual emergency.

Integrating Plans

Most communities have long had written ERPs designed for natural events such as floods and windstorms; some communities have had written plans dealing with emergencies created by people, such as bomb threats and civil disturbances; but, until recently, few had specific plans for responding to emergencies involving hazardous materials. As a result, the consequences of accidental chemical releases have been in many cases tragically magnified by the undisciplined reactions of people near the release source. It has been reported, for example, that when the alarms sounded at Bhopal, residents of the nearby homes ran toward the plant rather than taking action to protect themselves from the enveloping fumes.

It is not enough to train the personnel of a chemical facility to implement an ERP effectively. Appropriate people in the community, especially near neighbors, need to understand the elements of an ERP that are designed to protect them and the role they play in making the plan work. There must be a continuous effort to integrate the facility ERP into community emergency planning at local and district levels. Drills involving external resources that test all the plan elements against the standards mutually established with the community and those imposed by laws and regulations are necessary to ensure successful implementation of the plan.

Off-Site Warning

Designing an effective off-site warning system presents some major challenges. Despite excellent ongoing communications between a chemical facility and its neighbors, there is no positive way to ensure that the general public will respond quickly and appropriately to a warning alarm of any kind. Thus, it is essential that the selected warning system be developed with the close cooperation of the community. Even then, it is difficult to predict such factors as the inclination of people to be warned and the degree of public confidence in the validity of an alarm.

No off-site warning system will assure complete coverage of the intended audience. Best results are achieved by combining two or more systems for sequential alertingthe first to trigger preplanned immediate action by the public at greatest risk, followed by other communications that provide further information and guidance to a larger audience. Some of the systems most commonly used are:

1. facility fixed-sound sources, such as sirens and whistles
2. mobile alerting by police or fire personnel, either from vehicles with loudspeakers or door-to-door
3. fixed public address systems in the community or in the facility
4. automated telephone calling
5. alert radios energized by a special signal to produce a warning tone followed by broadcast messages
6. strobe lights in situations where the noise level is a problem
7. local radio stations and the emergency broadcast system
8. local TV stations

More sophisticated and less commonly used warning systems include helicopters equipped with loudspeakers, modified cable TV installations, and computer networks between a chemical facility and community emergency response groups.

Local Emergency Plans

Existing plans for a coordinated response to emergencies in a community vary greatly in content and organization, but the plans have two common objectives. They are to:

1. define authority and responsibilities of various emergency service participants
2. describe the interaction between those participants, government, and industry

In creating their plans, most communities draw on the Integrated Emergency Management System (IEMS) developed by the Federal Emergency Management Agency (FEMA).

A local plan has many of the same elements as a chemical facility ERP. It includes:

1. an emergency management organization, with designated functional responsibilities
2. the location of the emergency operating center and its resources
3. guidelines for classifying emergencies
4. activation and declaration checklists
5. communications requirements and available systems
6. evacuation and sheltering plans
7. methods for securing added resources
8. descriptions of local hazards

Most local plans are written to be nonspecific as to the cause of the emergency, with various appendices describing the details of response to specific events. These appendices are based upon the results of risk assessments made by the community with the cooperation of industry.

One such appendix should relate to emergencies caused by fixed facility or transportation incidents in which hazardous materials are involved. Chemical facilities must provide substantial support to the community in preparing this portion of the local emergency plan, and provide resources and training leadership that are not available elsewhere in the community. Where a Hazardous Material Advisory Council exists, there is an effective forum for doing this. In any case, a chemical facility manager should seek ways to help the community prepare for and recover from incidents of this nature.

Local Emergency Planning Committees

An important contribution to community and industry cooperation in emergency preparedness was the passage in 1986 of the Superfund Amendments and Reauthorization Act (SARA), which contained an emergency planning and community right-to-know provision. Title III, or EPCRA, as this portion of SARA is commonly called, is intended to encourage and support hazardous materials emergency planning efforts at the state and local level, and to provide citizens and local agencies with appropriate information concerning potential hazards in their communities.

The major portions of Title III require:

1. a statewide organization for planning emergency action and receiving hazardous chemical information
2. notification to the community of emergency releases of chemicals
3. reports of hazardous chemical inventories and copies of MSDSs to be furnished to the community
4. an annual inventory of hazardous chemical releases to the environment

Drills and Critiques

The optimum frequency of major drills involving personnel outside a chemical facility is dependent upon a number of variables:

1. location of the facility
2. dependence upon community emergency agencies
3. size and complexity of the facility
4. site and off-site risk assessments
5. population patterns

An important element of emergency preparedness is the establishment of an appropriate major drill frequency in cooperation with off-site agencies. A reasonable goal is to hold one such drill each year. The scenario might include an on-site, internally generated hazardous material emergency one year and a transportation emergency somewhere in the adjacent community the following year.

For facilities and communities just beginning to test their plans, desktop or simulated drills are effective for identifying procedural problems that need to be corrected before they proceed to full-scale drills. In these simulations, staffing of the appropriate emergency center would occur, but the emergency service groups would not actually mobilize at the scene of the incident.

As people gain confidence in the completeness and the effectiveness of the emergency response plan, it becomes important to measure the performance of all who are involved. Monitors record and later report on all aspects of response actions, including:

1. elapsed times before critical actions occur
2. actions and coordination of responding groups
3. actions of uninvolved personnel
4. alarm and communication effectiveness
5. emergency control center management
6. control at the emergency scene
7. accounting for personnel
8. medical aid for simulated casualties
9. off-site notifications
10. handling media representativesreal or simulated

Following each drill there must be an organized critique that provides the information needed to strengthen the plan and/or its implementation. All the people actively involved should be represented at the critique, and a written report of conclusions and recommendations should be widely distributed. It then is the responsibility of the facility Emergency Preparedness Team and the local emergency planning committee to coordinate and assist in solving any problems identifieda process that begins emergency preparedness activities again: identifying hazards; evaluating and strengthening resources; modifying the emergency plan; training people; communicating and integrating plans; and testing them once again.

6.8 Laws, Regulations, and Support

Laws

A number of legal requirements must be incorporated in a facility's ERP. Emergency prevention, preparedness, and response planning are regulated at the federal, state, and, occasionally, local levels. At the federal level, these laws include

1. Clean Air Act (CAA)
2. Clean Water Act (CWA)
3. Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA)
4. Emergency Planning and Community Right-to-Know Act (EPCRA, or SARA Title III)
5. Energy Reorganization Act (Nuclear Regulatory Commission)
6. Hazardous Materials Transportation Act (HMTA)
7. Occupational Safety and Health Act (OSHA)
8. Resource Conservation and Recovery Act (RCRA)
9. Toxic Substances Control Act (TSCA)
10. U.S. Coast Guard requirements
11. Environmental Protection Agency (EPA) regulations

A list of these laws and their *Code of Federal Regulations* (CFR) citations appears in the bibliography at the end of the chapter. States may have their own laws and regulations that also govern emergency response planning. References to these laws may be found in the Bureau of National Affairs (BNA) *Environment Reporter*.

These laws and their regulations are enforced by all levels of governmental agencies. A knowing or willful violation has serious implications for both companies and individuals, who may be held civilly or criminally liable for noncompliance. The penalties can be severe, ranging from daily-assessed fines to imprisonment. Thus, it is prudent to know the regulations that apply to the facility, and to ensure that the regulatory requirements are met.

The regulations regarding emergency planning and response are comprehensive, covering every aspect from prevention to reporting. RCRA's regulations cover the entire process, from planning to training to formal reports on the facility's response to an emergency involving hazardous waste. Other regulations focus on a specific aspect of emergency response, or part of the facility's operations. For example, the TSCA requires that spills or releases that contaminate the environment be reported orally and in writing within a certain time frame. All but one of the laws put the burden of planning on the facility. EPCRA, the exception, requires state and local agencies to prepare an ERP for the community. Facilities that meet criteria specified in EPCRA regulations have to assist in the development of the plan; however, they are not responsible for creating it.

The OSHA Process Safety Management regulation (29CFR 1910.119) and the EPA's Risk Management Plan regulation require significant attention to emergency planning and response. Inevitably, some of the regulations and requirements overlap. Most chemical facilities are subject to more than one law, and could be expected to prepare separate plans for specific parts of their facility. RCRA's Contingency Plan, for example, must be developed and maintained apart from other emergency response plans. The key to managing all the requirements and satisfying the regulations in an efficient, coordinated manner is first to understand the requirements and how they apply to the facility, and then to look for the common denominators among the requirements. The finished product, or master plan, will satisfy all the common denominators that apply, and will avoid duplication of effort. It also can be used as the basis for plans that must be maintained separately or that have requirements in addition to the common denominators.

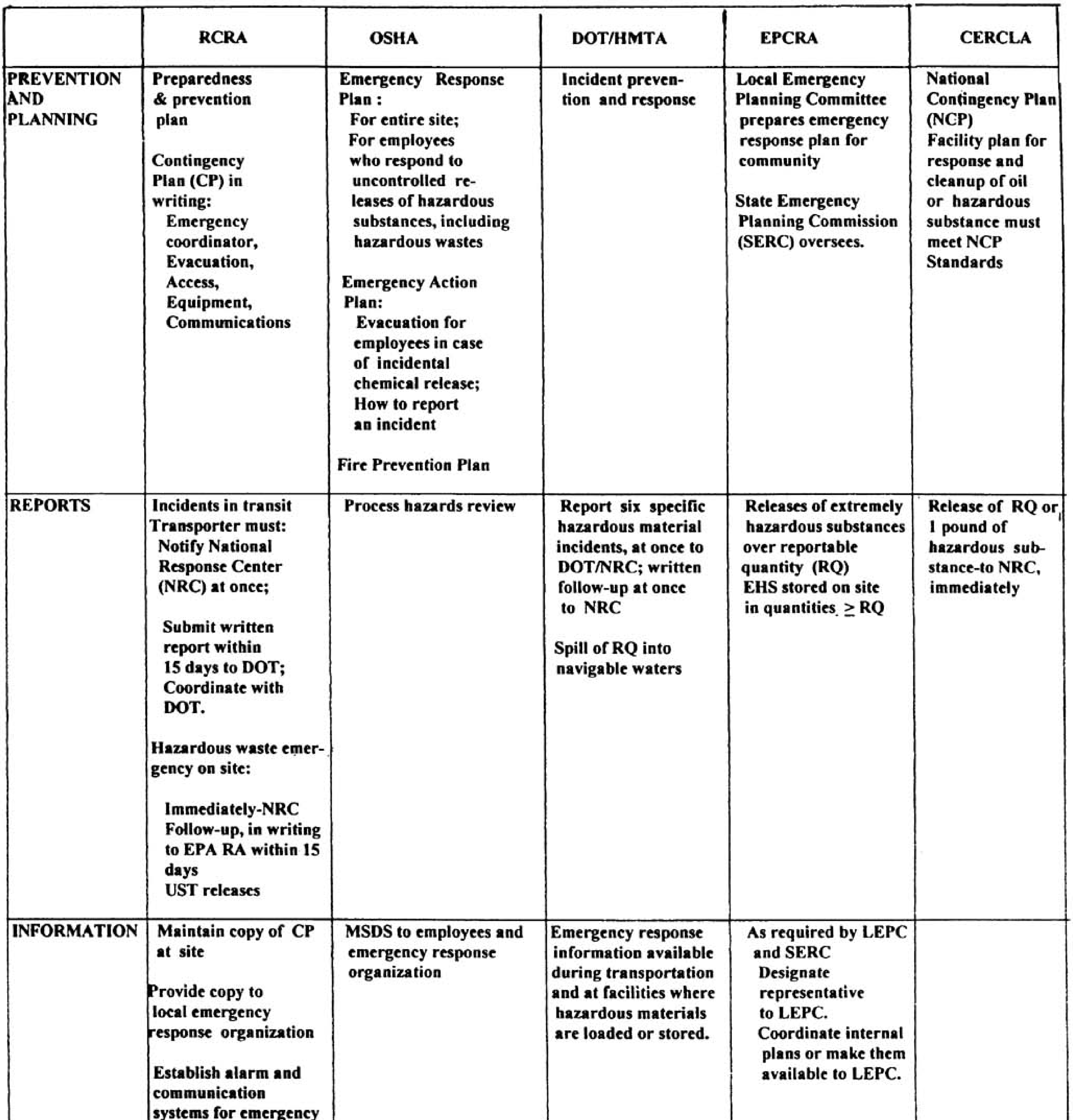
Meeting the Requirements

Regulations governing emergency response planning can be broken down into four general categories:

1. preparation
2. plans
3. reports and other communications
4. drills and evaluations

Figure 6.6 shows a matrix-type summary of the major federal laws and their requirements for emergency planning and response. Such a matrix is very helpful in determining what the requirements are and how they apply to a facility. It could be further tailored to cover only the requirements that apply to a specific site or operation.

Fig. 6.6 Emergency prevention, planning, and response.



	RCRA	OSHA	DOT/HMTA	EPCRA	CERCLA
PREVENTION AND PLANNING	Preparedness & prevention plan Contingency Plan (CP) in writing: Emergency coordinator, Evacuation, Access, Equipment, Communications	Emergency Response Plan : For entire site; For employees who respond to uncontrolled releases of hazardous substances, including hazardous wastes Emergency Action Plan: Evacuation for employees in case of incidental chemical release; How to report an incident Fire Prevention Plan	Incident prevention and response	Local Emergency Planning Committee prepares emergency response plan for community State Emergency Planning Commission (SERC) oversees.	National Contingency Plan (NCP) Facility plan for response and cleanup of oil or hazardous substance must meet NCP Standards
REPORTS	Incidents in transit Transporter must: Notify National Response Center (NRC) at once; Submit written report within 15 days to DOT; Coordinate with DOT. Hazardous waste emergency on site: Immediately-NRC Follow-up, in writing to EPA RA within 15 days UST releases	Process hazards review	Report six specific hazardous material incidents, at once to DOT/NRC; written follow-up at once to NRC Spill of RQ into navigable waters	Releases of extremely hazardous substances over reportable quantity (RQ) EHS stored on site in quantities $\geq$ RQ	Release of RQ or 1 pound of hazardous substance-to NRC, immediately
INFORMATION	Maintain copy of CP at site Provide copy to local emergency response organization Establish alarm and communication systems for emergency	MSDS to employees and emergency response organization	Emergency response information available during transportation and at facilities where hazardous materials are loaded or stored.	As required by LEPC and SERC Designate representative to LEPC. Coordinate internal plans or make them available to LEPC.	

	communication systems for emergency notification.				
TRAINING, DRILLS, AND EVALUATION	Emergency response must be documented and records retained Initial and annual review Evacuation drill	Initial and annual refresher training for employees involved in emergency response; varies with roles– All employees trained in Emergency Action Plan; initially and with every change to plan. Training in MSDS information			

CWA	CAA	TSCA	NUREG 054	USCG	EPA
Spill Prevention Control and Countermeasure (SPCC) Plan Shows how facility will: prevent, respond, follow-up to oil spills in harmful quantities. Must be in writing	State Iplementation Plan (SIP) must include an Emergency Air Pollution Episode plan Accident Prevention Program for EHS States may require prevention and emergency response plan		Emergency response plan, including: Emergency planning zones; Prevention, mitigation, and limitation of core damage and consequences of release. Subject to annual review		Risk management plan
Release of RQ of oil or hazardous substance to NRC, immediately.	Announce uncontrolled releases of pollutants over certain set levels Releases $\geq$ RQ to NRC	Emergency incidents of environmental contamination At once to EPA, Written follow-up within 10 days	Notify state and local officials: Change in condition Protective action recommendations Notify community within 15 minutes. 100% notification within 15 miles.	All spills into navigable waters of oil, hazardous substances, $\geq$ RQ, at once, to US Coast Guard or to NRC	Worst case scenarios
Keep copy of SPCC at site available to EPA RA during normal working hours. Submit SPCC amendments to EPA RA and state water pollution control agency					Emergency response plan
Train : SPCC responders Employees who operate, maintain equipment			Annual training On-site Off-site Annual graded emergency response		Training required for all employees in emergency response plan

equipment			Annual graded emergency response exercises		

Prevention and Preparation

Some regulations require that a facility conduct a risk assessment and/or other preparatory activities. The RCRA calls it a preparedness and prevention plan. A facility subject to the RCRA must determine how structures, processes, and operations can be changed in order to minimize the possibility of an emergency involving hazardous waste. The facility also has to determine the communications and alarm systems that will be used in the event of such an emergency. The CWA includes prevention in its requirements for the Spill Prevention, Containment, and Countermeasure (SPCC) Plan. The 1990 amendments to the CAA added an accident prevention plan for extremely hazardous substances.

Plans

At the very least, a facility is required to develop a plan describing how it will respond to an incident that threatens human health and/or the environment. Generally, the plan includes notification, evacuation, protection of employees, and control of the incident. This emergency response plan usually must be in writing. For example, the OSHA requires a minimum of three plans: emergency response, emergency action, and fire prevention. The CAA requires that the state implementation plan have an emergency air pollution episode plan.

Communications

There are two aspects to emergency communications: the actual equipment used to communicate information about the incident and the types of communications or information-sharing required. The RCRA has specific requirements for the types of emergency communication equipment (alarm systems, phone or radio communications) that must be present. Under the EPCRA, facilities must provide information about their operations and substances used or stored on site when the Local Emergency Planning Committee (LEPC) or State Emergency Response Commission (SERC) requests it. If the facility uses or stores extremely hazardous substances (EHSs) in reportable quantities, it must appoint a representative to the LEPC. Several laws require that a copy of the ERP be made available to employees and representatives of government agencies during working hours. The OSHA requires that facilities provide material safety data sheets (MSDS) for all hazardous substances present on the site.

Reports

Reports are another important communications aspect of the ERP. Most laws insist on prompt notification of the proper agencies immediately after an incident occurs. These reporting requirements can be complex, particularly in view of the fact that many laws have their own lists of hazardous substances and reportable quantities. The CERCLA requires that releases of a reportable quantity of what it defines as a hazardous substance must be reported immediately to the National Response Center (NRC). The EPCRA requires facilities that store and/or release reportable quantities of substances on its EHS list to report that information to the LEPC and the SERC. EPA's Risk Management Plan requires prior disclosure of possible "worst case" incident scenarios.

Written follow-up reports are often a requirement. The HMTA has identified six specific hazardous material incidents that must be reported immediately and again in writing. The RCRA gives managers 15 days after a hazardous waste emergency to submit a written report to the EPA.

Training, Drills, Audits, and Evaluations

An emergency plan is relatively useless unless the employees affected by it are trained in its use. The RCRA, OSHA, HMTA, CWA, and the Energy Reorganization Act require annual and refresher training. In addition, the facilities must keep records of the training, and must make them available to the appropriate agency when they are requested. Some laws go so far as to require practice drills. Nuclear power plants must conduct on- and off-site training, and go through annual graded emergency response exercises. The plan and the response executed according to the plan then are evaluated so that the plan can be improved.

Sources of Assistance

Seeing all the requirements together can be overwhelming. Fortunately, there are agencies, associations, and programs that can assist in the preparation of a comprehensive emergency prevention and response plan.

The federal government and the agencies responsible for the laws that govern emergency response planning provide 800-number hotlines and manuals that describe various aspects of emergency prevention, planning, and response.

The volunteer or professional emergency responders in the community have valuable practical experience that can be put to work in developing the facility ERP. Working with them also establishes a forum for communications and understanding with the community.

The American Chemistry Council (ACC) Community Awareness and Emergency Response (CAER) program provides comprehensive guidelines for the development and implementation of an ERP. The CAER program has been expanded to include all aspects of the chemical industry in an initiative called Responsible Care.

Other services of the ACC include CHEMTREC, a 24-hr emergency response service for people who respond to emergencies involving chemicals; CHEMNET, a mutual aid agreement between chemical producers and emergency response contractors; and workshops and videotape training programs for first responders and other emergency response personnel.

Additional sources of assistance and information include other professional associations, such as the American Institute of Chemical Engineers, and publications, seminars, workshops, and videotapes offered by educational organizations. Considerable information is available on the Internet through web sites such as www.fema.gov which is maintained by the Federal Emergency Management Agency.

A bibliography; a list of laws, regulations, and standards; and a compilation of suggested reading material follow.

6.9 Select Bibliography

American Chemistry Council (Formerly Chemical Manufacturers Association) CAER: The Next Phase Program Handbook, Chemical Manufacturers Association (now American Chemistry Council), Washington, DC with assistance from HMM Associates, Concord, MA, 1989.

American Chemistry Council (Formerly Chemical Manufacturers Association) Community Emergency Response Exercise Handbook, Chemical Manufacturers Association (now American Chemistry Council), Washington, DC, 1986.

American Chemistry Council (Formerly Chemical Manufacturers Association) Emergency Warning Systems Guidebook, Chemical Manufacturers Association (American Chemistry Council), Washington, DC, 1987.

American Chemistry Council (Formerly Chemical Manufacturers Association) Site Emergency Response Planning Handbook, Chemical Manufacturers Association (now American Chemistry Council), Washington, DC, 1986.

The Conservation Foundation, Risk Communication: Proceedings of the National Conference on Risk Communication, The Conservation Foundation, Washington, DC, 1986.

Covello, V. T., Sandman, P. M., and Slovic, P., Risk Communication, Risk Statistics, and Risk Comparisons, Chemical Manufacturers Association (American Chemistry Council), Washington, DC, 1988.

Federal Emergency Management Agency, FEMA Handbook, Federal Emergency Management Agency, Washington, DC, 1985.

National Response Team, Hazardous Materials Emergency Planning Guide (NRT 1), National Response Team of the National Oil and Hazardous Substances Contingency Plan, Washington, DC, 1987.

National Response Team, Criteria for Review of Hazardous Materials Emergency Plans, National Response Team of the National Oil and Hazardous Substances Contingency Plan. Washington, DC, 1988.

Occupational Safety and Health Administration, How to Prepare for Workplace Emergencies, OSHA 3088 (revised), U.S. Department of Labor, Washington, DC, 1988.

Prugh, R. W., and Johnson, R. W., Guidelines for Vapor Release Mitigation, Center for Chemical Process Safety, The American Institute of Chemical Engineers, New York, 1988.

SARA Title III (Superfund Amendment and Reauthorization Act, Emergency Planning and Community Right-to- Know) 4OCFR 355.3, Section 302(c), Emergency planning notification Section 303(d), Appointment of emergency coordinator, provision of information; Section 304, Emergency release notification requirements.

EPA 40 CFR Part 68, Risk Management Plan.

U.S. Coast Guard 33CFR 126.9 Reporting requirements for discharge of petroleum products or dangerous liquid commodities into navigable waters of USA 33CFR 153.023 Reporting requirements for discharge of reportable quantity (RQ) of oil or hazardous substance into navigable waters.

Standards

ANSI/National Fire Protection Association Standards

1. # 72 National Fire Alarm Code
2. # 110 Emergency Power and Standby Systems
3. # 600 Private Fire Brigades
4. # 1561 Emergency Services Incident Management System

6.10 Suggested Reading

Regulations

Bureau of National Affairs, Environment Reporter, Bureau of National Affairs, Inc., Washington, DC.

ENSR Consulting and Engineering, Air Quality Handbook: A Guide to Permitting and Compliance under the Clean Air Act and Air Toxics Programs, 10th ed., ENSR Consulting and Engineering, Acton, MA, 1988.

ERT, Inc., and Sidley & Austin, Superfund Handbook, 2nd ed., ERT, Inc. and Chicago, IL: Sidley & Austin, Concord, MA, 1987.

Lowry, G. G., and Robert, C., Lowry's Handbook of Right-to-Know and Emergency Planning, Lewis Publishers, Inc., Chelsea, MI, 1988.

Office of Solid Waste, U.S. Environmental Protection Agency, Solving the Hazardous Waste Problem: EPA's RCRA Program (EPA/530-SW-86-037), U.S. Environmental Protection Agency, Washington, DC, 1986.

Prevention and Planning

Abrams, M. J., and Lewis, J., "Preplanning, the Key Emergency Response," Papers Presented at the Spring 1987 National Meeting of the American Institute of Chemical Engineers, American Institute of Chemical Engineers, New York, 1987.

Adler, V., Sorenson, J. H., and Rogers, G. O., "Chemical and Nuclear Emergencies: Interchanging Lessons Learned from Planning and Accident Experience," Proceedings of a Joint NEA/CEC Workshop on Technical Aspects of Emergency Planning, Brussels, Belgium (CONF-89906137-1), U.S. Department of Energy, Washington, DC, 1989.

U.S. Environmental Protection Agency, Guide to Exercises in Chemical Emergency Preparedness Programs, U.S. Environmental Protection Agency, Washington, DC, 1988.

6.11 Laws, Regulations, and Standards

Laws and Regulations

CAA (Clean Air Act)

1. 4OCFR 51 SIP Emergency air pollution episode plan

CAA Reauthorization Section 304, Prevention of sudden, accidental releases

CERCLA (Comprehensive Environmental Response, Compensation, and Liability Act)

1. 4OCFR 300-306 Section 103(a) Spill reporting requirements
2. 4OCFR 355 Emergency planning

CWA (Clean Water Act)

1. 4OCFR 112.3-7 Spill prevention, control, and countermeasure plan (SPCC) requirements, modifications, preparation, and implementation Energy Reorganization Act (was Atomic Energy Act)
2. 1OCFR 50.47, also Appendix E NRC (Nuclear Regulatory Commission) Standards for on-site and off-site emergency plans

HMTA (Hazardous Materials Transportation Act)

1. 49CFR Parts 171-177 DOT requirements for providing information and advice on meeting emergencies; FEMA requirements for evaluation of training programs for incident prevention and response

OSHA (Occupational Safety and Health Act)

1. 29CFR 1910.1200 Hazard communication standard
2. 29CFR 1910-210 and Appendices, Hazardous waste operations and emergency response
3. 29CFR 1910.35-38 Requirements for evacuation route and exit posting, emergency lighting, accessibility of exits, and emergency action plan
4. 29CFR 1910.156 Fire brigades
5. 29CFR 1910.157-164 Fire extinguishing and detection systems
6. 29CFR 1910.165 Employee alarm systems
7. 29CFR 1910.119 Process Safety Management of Highly Hazardous Chemicals

RCRA (Resource Conservation and Recovery Act)

1. 4OCFR 263.30,31 Transporter responsibilities in hazardous waste transportation incidents
2. 4OCFR 264.30-37 and 4OCFR 265.30-37 Preparedness and prevention
3. 4OCFR 264.50-56 and 4OCFR 265.50-56 Contingency Plan

American Nuclear Society. Emergency ResponsePlanning, Technologies, and Implementation; Proceedings of the ANS Topical Meeting (CONF-880913. UC-610), Savannah River Laboratory, E. I. du Pont de Nemours and Company, Aiken, SC, 1988.

Bell, D. W., and Burns, C. C., "Offsite Emergency Plan Development and Maintenance Considerations," Papers Presented at the Spring 1987 National Meeting of the American Institute of Chemical Engineers, American Institute of Chemical Engineers, New York, 1987.

Davis, D. S., et al., Prevention Reference Manual: Overviews on Preventing and Controlling Accidental Releases of Selected Toxic Chemicals (EPA/600/8-88-074), U.S. Environmental Protection Agency, Air and Energy Engineering Research Lab., Research Triangle Park, NC, 1988.

Dickerson, M. H., "Emergency Planning, Response, and Assessment: A Concept for a Center of Excellence," Proceedings from an International Seminar on Nuclear War International Cooperation: The Alternatives (CONF-8608149-1), U.S. Department of Energy, Washington, DC, 1986.

Federal Emergency Management Agency, Hazardous Materials Exercise Evaluation Methodology (HM-EEM) and Manual, Federal Emergency Management Agency, Washington, DC, 1989.

Fingleton, D. J., Tanzman, E. Z., and Bertram, K. M., "Development of a Model Emergency Response Plan for Catastrophic Releases of Toxic Gases," Proceedings of the Air Pollution Control Association Annual Meeting (CONF-860606-19), U.S. Department of Energy, Washington, DC, 1986.

Gudiksen, P. et al., Emergency Response Planning for Potential Accidental Liquid Chlorine Releases, U.S. Department of Energy, Washington, DC, 1986.

International Association of Fire Chiefs, Fire Service Emergency Management Handbook, Federal Emergency Management Agency, Washington, DC, 1985.

Jones, E., "Contingency Planning and Emergency Response in Construction Activities: Training the Construction Worker," Proceedings of an Oak Ridge Model Conference on Waste Problems (CONF-871075-5), U.S. Department of Energy, Washington, DC, 1987.

Kalnins, R. V., "Emergency Preparedness and Response," Symposium on the Characterization of Thermodynamic and Transport Properties of Polymer Systems, American Institute of Chemical Engineers, New York, 1986.

Krimsky, S., and Plough, A., Environmental HazardsCommunicating Risks as a Social Process, Auburn House Publishing Company, Medford, MA, 1988.

Michael, E. J. et al., "Emergency Planning Considerations for Specialty Chemical Plants," American Institute of Chemical Engineers Summer National Meeting, American Institute of Chemical Engineers, New York, 1986.

Morentz, J. W., and Griffith, D., "Using Computers for Chemical Emergency Planning and Response," Papers Presented at the Spring 1987 National Meeting of the American Institute of Chemical Engineers, American Institute of Chemical Engineers, New York, 1987.

National Response Team, Criteria for Review of Hazardous Materials Emergency Plans (NRT-1A), National Response Team of the National Oil and Hazardous Substances Contingency Plan, Washington, DC, 1988.

National Response Team, HAZMAT Planning Guide (WH-562A), National Response Team, Washington, DC, 1987.

National Response Team, Technical Guidance for Hazards Analysis: Emergency Planning for Extremely Hazardous Substances (Supplement to Hazardous Materials Emergency Planning Guide NRT 1), National Response Team, Washington, DC, 1987.

Philley, J. O., "Emergency Preparedness Training Tips," American Institute of Chemical Engineers National MeetingSummer '87, American Institute of Chemical Engineers, New York, 1987.

U.S. Environmental Protection Agency, Bibliography on Chemical Emergency Preparedness and Prevention, U.S. Environmental Protection Agency, Washington, DC, 1986.

U.S. Environmental Protection Agency, It's Not Over in October: A Guide for Local Emergency Planning Committees, U.S. Environmental Protection Agency, Washington, DC, 1988.

U.S. Environmental Protection Agency, Seven Cardinal Rules of Risk Communication (EPA 87-020), U.S. Environmental Protection Agency, Washington, DC, 1988.

U.S. Environmental Protection Agency, "Why Accidents Occur: Insights from the Accidental Release Information Program," Chemical Accident Prevention Bulletin (OSWER-89-008.1), Series 8, No. 1 (July 1989), U.S. Environmental Protection Agency, Washington, DC, 1989.

Waldo, A. B. (Ed.), The Community Right-to-Know Handbook, Thompson Publishing Group, Washington, DC, 1986.

Transportation

Office of Hazardous Materials Transportation, U.S. Department of Transportation, Emergency Response Guidebook. Available through the U.S. Government Printing Office (GPO) U.S. Government Bookstores. Updated every three years.

7.1 Introduction

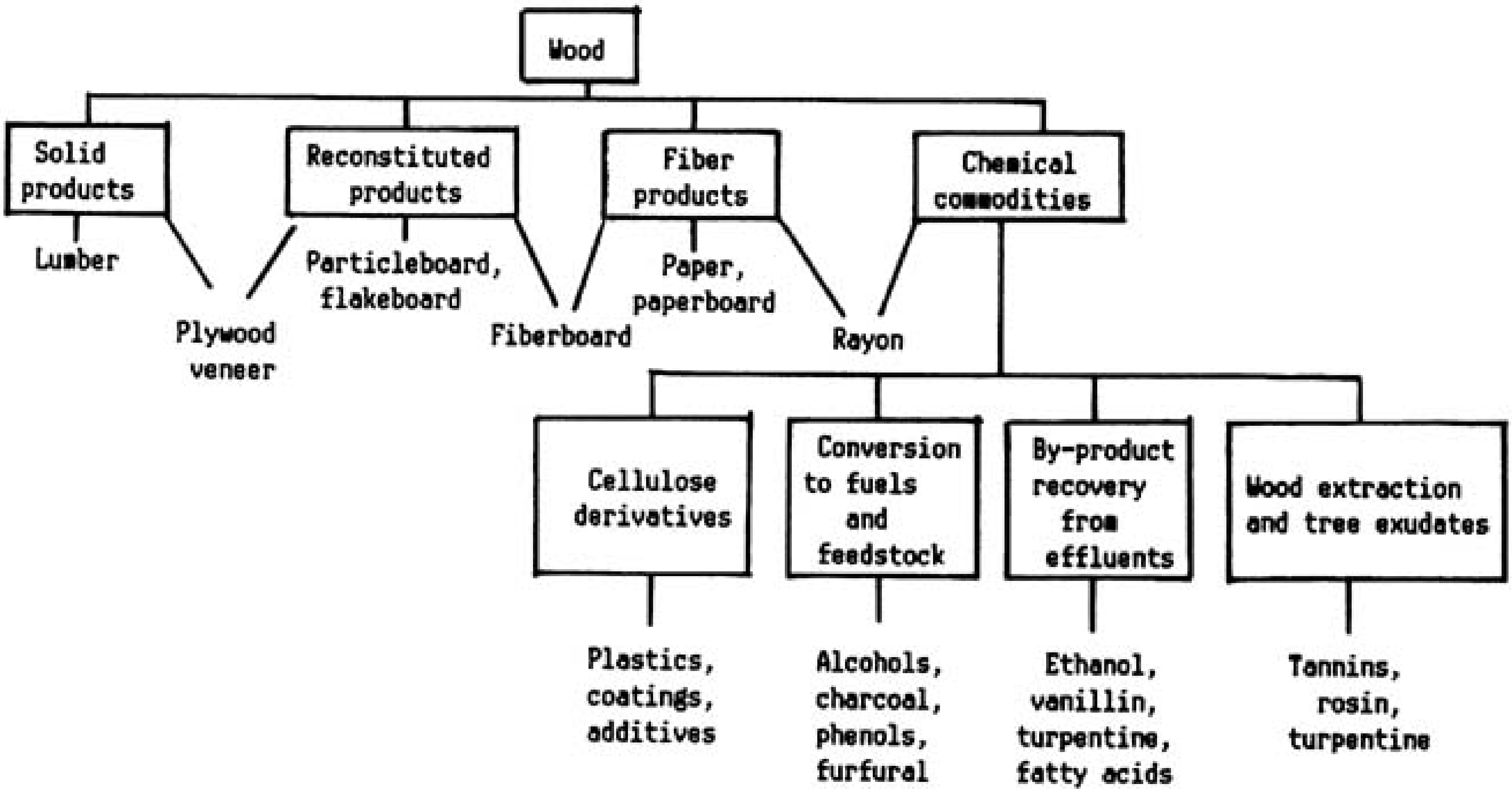
Wood has been utilized by humans since antiquity. Trees provided a source of many products required by early humans such as food, medicine, fuel, and tools. For example, the bark of the willow tree, when chewed, was used as a painkiller in early Greece and was the precursor of the present-day aspirin. Wood served as the primary fuel in the United States until about the turn of the century, and even today over one-half of the wood now harvested in the world is used for heating fuel.

Throughout history different types of wood have served many purposes. The tough, strong, and durable white oak, for example, was a well-proven raw material for ships, bridges, cooperage, barn timbers, farm implements, railroad ties, fence posts, flooring, paneling, and other products. In contrast, woods such as black walnut and cherry became primarily cabinet woods. Hickory was manufactured into tough, hard, resilient tool handles. Black locust was used for barn timbers and treenails. What the early artisan learned by trial and error became the basis for intelligent decisions concerning which species was best suited to a given purpose, as well as what characteristics should be looked for in selecting trees for different applications. It was known that wood from trees grown in certain locations was stronger, more durable, and more easily worked with tools than wood from the same species grown in other locations. Modern wood quality research has substantiated that location and growth conditions significantly affect wood properties.[1]

In the United States more than 100 kinds of wood are available to the prospective user, but it is very unlikely that all are available in any one locality. About 60 native woods are of major commercial importance. Another 30 wood types are commonly imported in the form of logs, cants, lumber, and veneer for industrial uses, the building trades, and crafts.[2,3]

The importance of wood as a raw material supplying fiber, energy, and chemicals is similar in magnitude to its use as a solid material. Lumber, plywood, and reconstituted boards consume about one-half of the timber harvest; usage for fiber, chemicals, and fuels accounts for the remaining half. Fiber production consumes about 25 percent of the total timber use. Although the relative value of wood as a source of energy and chemicals has varied considerably through the decades, wood continues to be an important source of specialty chemicals and renewable energy, and may be even more important in the future. The many uses for wood are summarized in Fig. 7.1. In this chapter we discuss the chemical nature of wood and provide a description of the technology for conversion of wood to pulp fibers for paper-making, to fuels for energy, and to chemicals for industry and consumers.

Fig. 7.1 Summary of uses for wood. (*From* Introduction to Forest Ecosystem Science and Management, *3rd ed., R. A. Young and R. L. Giese (Eds.), Copyright © 2004, John Wiley & Sons, New York; with permission of John Wiley & Sons Publisher*.)



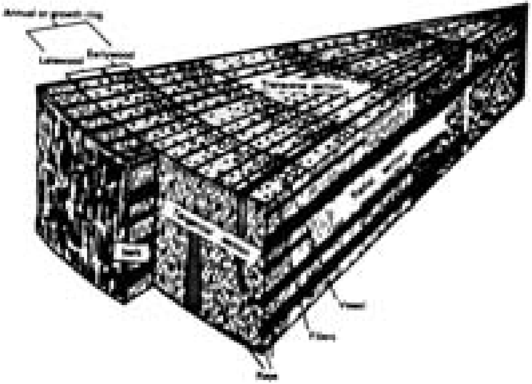
7.2 Wood Structure

The wood from the many different species of trees is divided into two major categories according to the botanical classification of the trees as seed plants. The hardwoods are derived from the angiosperms (encased seeds), which are generally the deciduous broad-leaved trees; the softwoods are from the gymnosperms (naked seeds), generally characterized as cone-bearing trees with persistent needlelike foliage. The majority of the trees in the latter classification fall into the division Coniferophyta and thus are often referred to as conifers. It is important to note that the terms "hardwood" and "softwood" have no direct relation to the hardness or softness of the wood. In fact, hardwood trees such as cottonwood, aspen, and balsa have softer wood than the western white pines and true firs; and certain softwoods, such as longleaf pine and Douglas fir, produce wood that is much harder than that of basswood or yellow poplar.

Wood can be considered as a biological composite of hollow tubes of cellulose fibers held together by a lignin matrix gluing material. Liquids are transported up and down the trees through the tubular plumbing system. About 90 percent of the wood tissue of softwood trees consists of fiber tracheid cells for liquid conduction and support. The hardwood trees evolved after the softwoods and have specialized water conduction cells called vessels.

Each vessel consists of numerous drum-shaped vessel members with open ends, stacked one above the other to form a continuous vessel pipeline from root tips to the leaves. The vessels in oaks and some other broad-leaved trees are large enough to be seen with the unaided eye, that is, as round pores at the end surface and as needlelike grooves at the lateral surface (Fig. 7.2).4

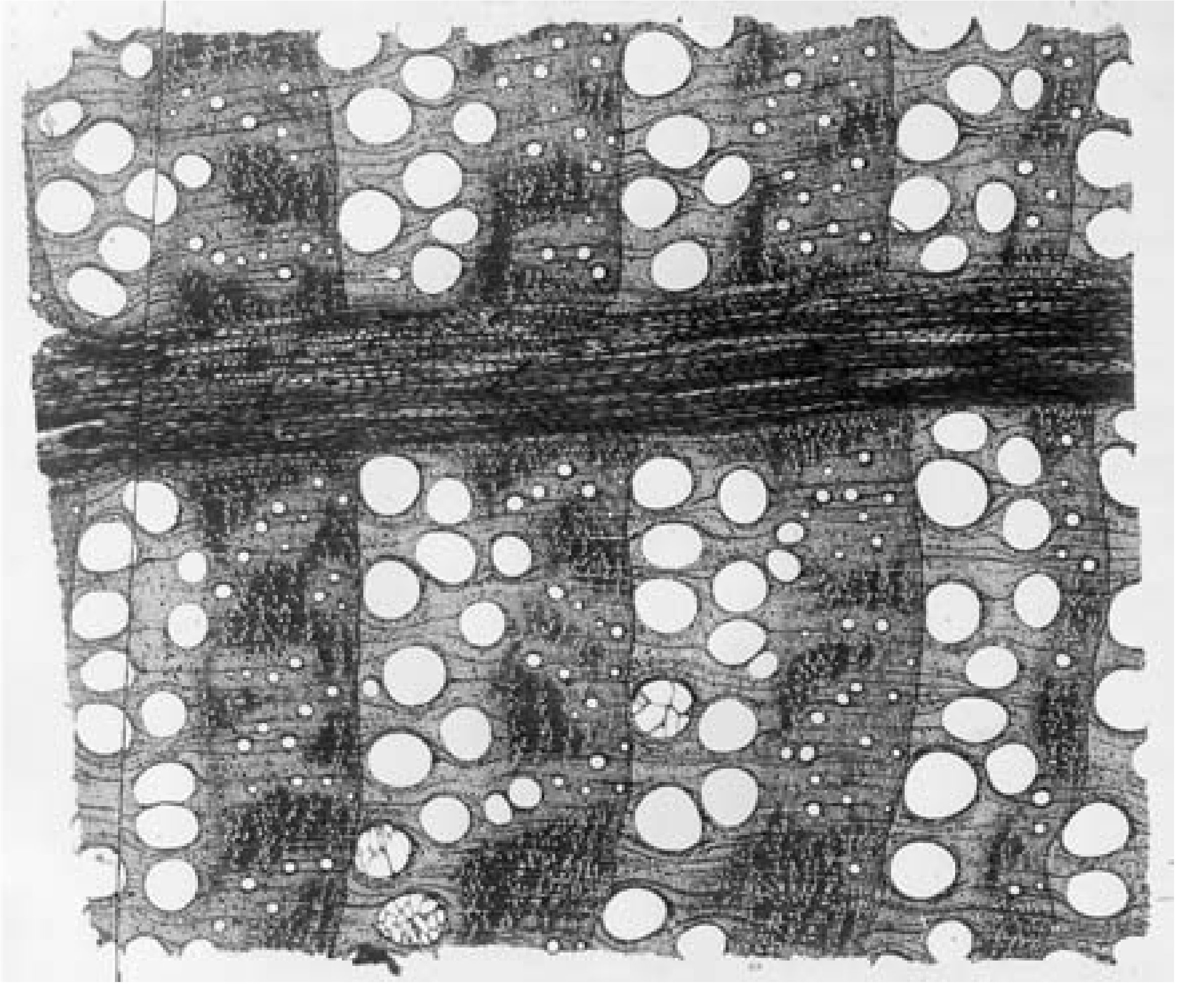
Fig. 7.2 Diagrammatic wedge section of a five-year-old oak trunk. The term "transverse section" is synonymous with end surface and cross-section. Both the radial section and tangential section are lateral surfaces. (*Courtesy U.S.D.A. Forest Products Laboratory*)



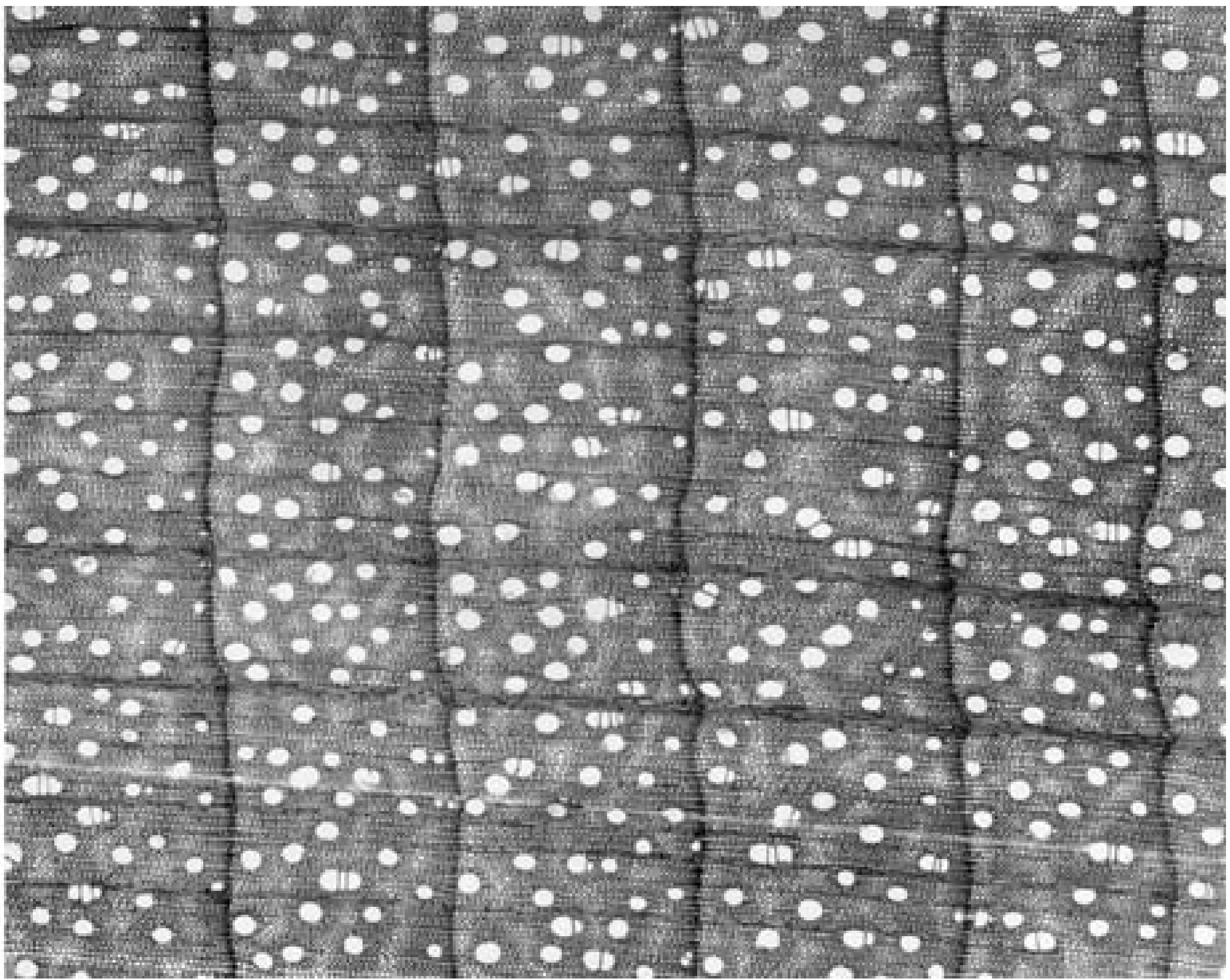
In some woods such as oak, the vessels in the earlywood (produced early in the growing season) are much wider than those in latewood and form the distinct growth rings of these ring-porous woods (Fig. 7.3a). In the diffuse-porous maple, birch, and many other genera, the vessels are of the same size throughout the growth ring, the boundary of which is marked only by extremely thin cells (Fig. 7.3b). Because coniferous woods lack vessels, they sometimes are designated non porous, in contrast to the porous broad-leaved trees, particularly those with relatively wide pores. Coniferous wood has growth rings because the earlywood tracheids have wider cavities and thinner cells than do the latewood tracheids; the earlywoodlatewood contrast is especially sharp in southern pines.

Fig. 7.3 Photomicrograph of hardwood cross-sections depicting (a) ring-porous and (b) diffuse porous wood. (*Courtesy of U.S.D.A. Forest Products Laboratory*)

(a)



(b)



On tree cross-sections, rays of parenchyma cells resemble spokes of a wheel as they extend from the bark into the wood, in the radial direction of the circular growth rings and perpendicular to the tangent of the rings or to the tangential direction (Fig. 7.2). Radial and tangential are both transverse directions, that is, horizontal to the trunk of the standing tree.

Fibers and vessels die at an age of a few weeks, after attaining their final shape and chemical composition. Parenchyma cells live much longer, some remaining alive until sapwood becomes heartwood. One of the tasks of parenchyma cells is to convert sugars into storage starch to serve as food reserves for the tree, and to reconvert starch to sugars when they are needed for growth. The starch and sugar content of sapwood is one reason why most wood-damaging organisms prefer dead sapwood over heartwood, and why sapwood is never quite as durable as heart-wood, except in living trees whose sapwood cells fight attacking organisms.

Virtually all cavities of wood cells are interconnected, as mentioned above for the vessel members of hardwoods. Tiny openings in the cell walls of overlapping ends of tracheids provide vertical passageways for water rising from cell to cell in conifers. Thin walls of ray cells are perforated for the radial transport of sugars, and tangential connections provide for growing cells. Hence, wood is permeable in all three of its cardinal directions. The longitudinal orientation of most cells, however, causes liquids and gases to flow and diffuse many times faster in the longitudinal or fiber direction than in the two transverse directions. This explains why stacked firewood dries predominantly at the exposed ends. The radial orientation of rays allows liquids and gases to move radially up to twice as rapidly as tangentially.4

As a porous material, wood contains air-filled or water-filled cell cavities surrounded by cell walls (Fig. 7.3). The cell wall substance is remarkably uniform; it differs very little even between conifers and broad-leaves trees and has a constant density of around 1.5 g/cm3. The density of wood thus depends on the proportion of the cell wall volume to the volume of the bulk wood.

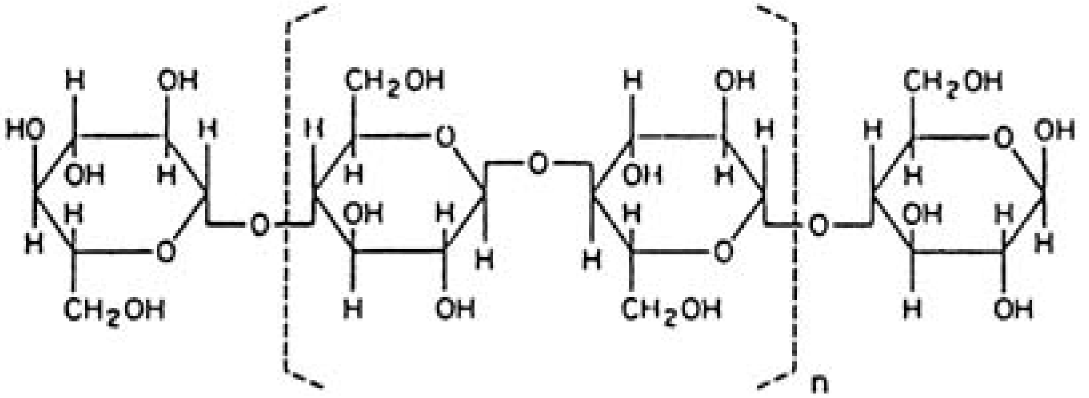
Wood density is *the* indicator of nearly all wood properties and of wood quality. Relatively dense species are strong, as one might expect from their large cell wall mass. Wood species also vary in features not related to density, such as color, percentage and arrangement of cell types, and minor chemical constituents. Therefore, even woods of the same density are likely to differ, with the properties varying more than the densities.

The widespread use of wood is due to its many advantages, including its very high strength-to-mass ratio, good toughness, low cost and availability, workability, attractive appearance, and good thermal and electrical insulating properties. However, there also are a number of disadvantages of wood, which have limited its usage for many applications; these include its lack of plasticity and fusibility, shrinkage and swelling properties, deterioration due to fungi and insects, and variability. However, most of its negative properties can be corrected by proper treatment and modification, as described later in this chapter.

7.3 Chemical Composition and Fiber Structure

Wood is composed of three major polymeric materials: cellulose, hemicelluloses, and lignin. Cellulose is the common name used for the glucan present in wood, which constitutes about 42 percent of wood's dry weight. Cellulose is the primary component of the walls of cells making up wood fibers and is the main structural material of wood and other plants. Paper, paperboard, and other wood fiber products thus also are composed mostly of cellulose. The chemical structure of the cellulose macromolecule is shown in Fig. 7.4. In the plant the degree of polymerization (DP) of cellulose is approximately 14,000.5

Fig. 7.4 The chemical structure of cellulose with the cellulose repeat unit in brackets.



Closely associated with cellulose in the wood structure and paper products are other polysaccharides called hemicelluloses, which often have been labeled as the matrix material of wood. In hardwoods the primary hemicellulose is a xylan (polymer of xylose), whereas in softwoods the primary hemicellulose is a glucomannan, although both of these polysaccharides occur to some extent in both types of wood. The DP of the hemicelluloses is much less than that of cellulose, in the range of 100200.

Table 7.1 compares the chemical composition of extractive-free hardwoods and softwoods. As cellulose and the hemicelluloses are both polysaccharides, it is obvious that the polysaccharide of wood is by far the dominant component, making up approximately 70 percent of both hardwoods and softwoods. Additional polysaccharides may occur as extraneous components of wood, which are not part of the cell wall; for example, the heartwood of species of larch can contain up to 25 percent (dry weight) of arabinogalactan, a water-soluble polysaccharide that occurs only in trace quantities in other wood species.5

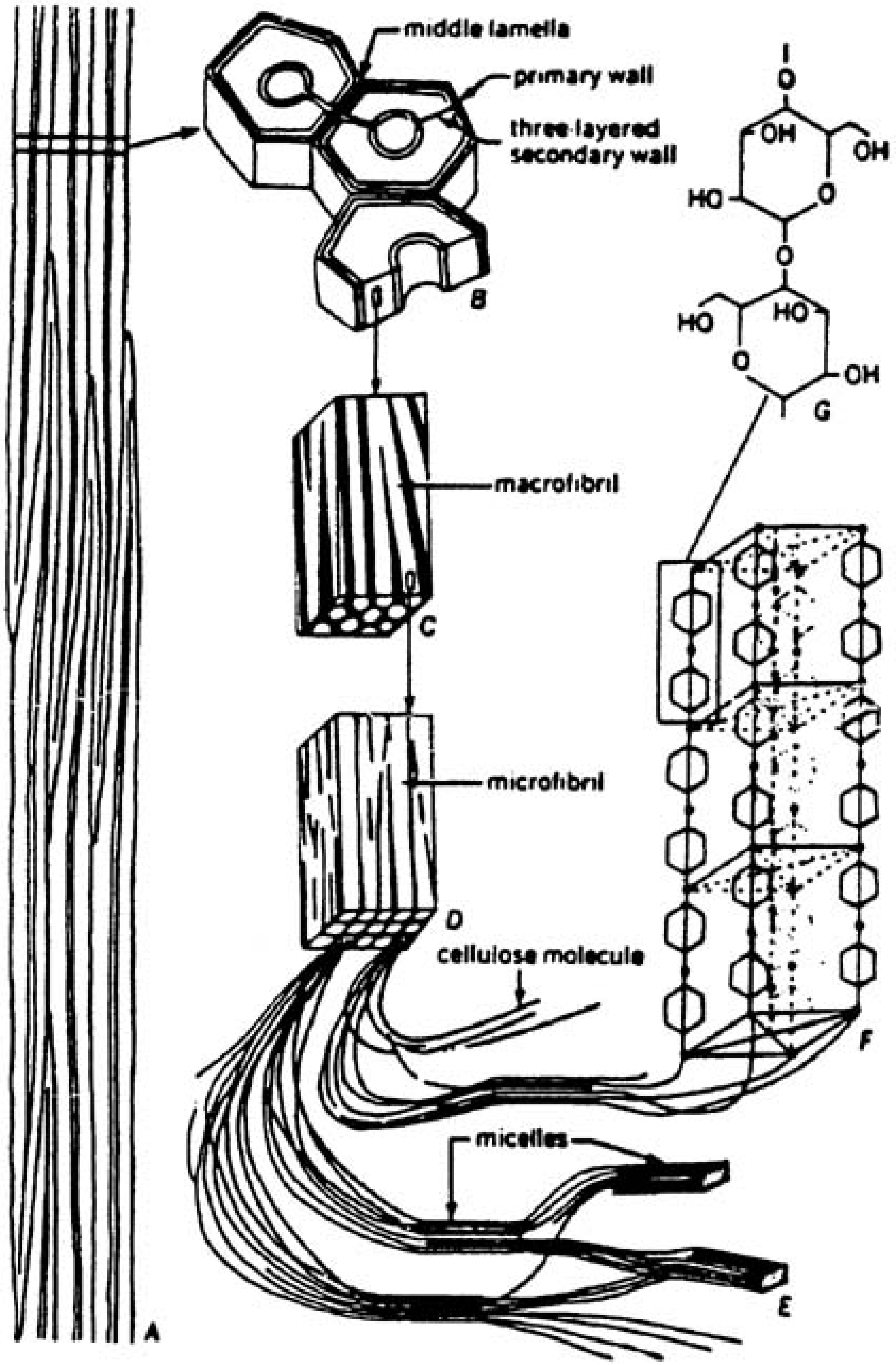
TABLE 7.1 Chemical Composition and Fiber Length of Extractive-Free Wood

	Hardwood (Red Maple)	Softwood (Balsam Fir)
Component	(%)	(%)
Cellulose	44	42
Hemicelluloses		
Xylan	25	9
Glucomannan	4	18
Lignin	25	29
Pectin, starch	2	2
Average fiber length (mm)	0.81.5	2.56.0

The third major component of wood shown in Table 7.1 is lignin. Although lignin also is a polymer, it has a different chemical structure from that of the polysaccharides. The monomeric units in lignin are phenolictype compounds, but the exact chemical structure of lignin still is not known after 100 years of intensive research. The spaces between fibers in wood are almost pure lignin and are termed the middle lamella. Lignin is considered the gluing or encrusting substance of wood and adds mechanical strength or stiffness to the tree and to wood. Higher plants commonly are referred to as lignocellulosic because of the typical joint occurrence in them of lignin and cellulose.

In the tree the cellulose polymers are laid down uniformly, the chains paralleling one another, and the long-chain molecules associate strongly through hydrogen bonds that develop between hydroxyl groups. These bonds create very strong associations between the cellulose macromolecules, producing very uniform crystalline structures known as micelles or microcrystallites (see Fig. 7.5).

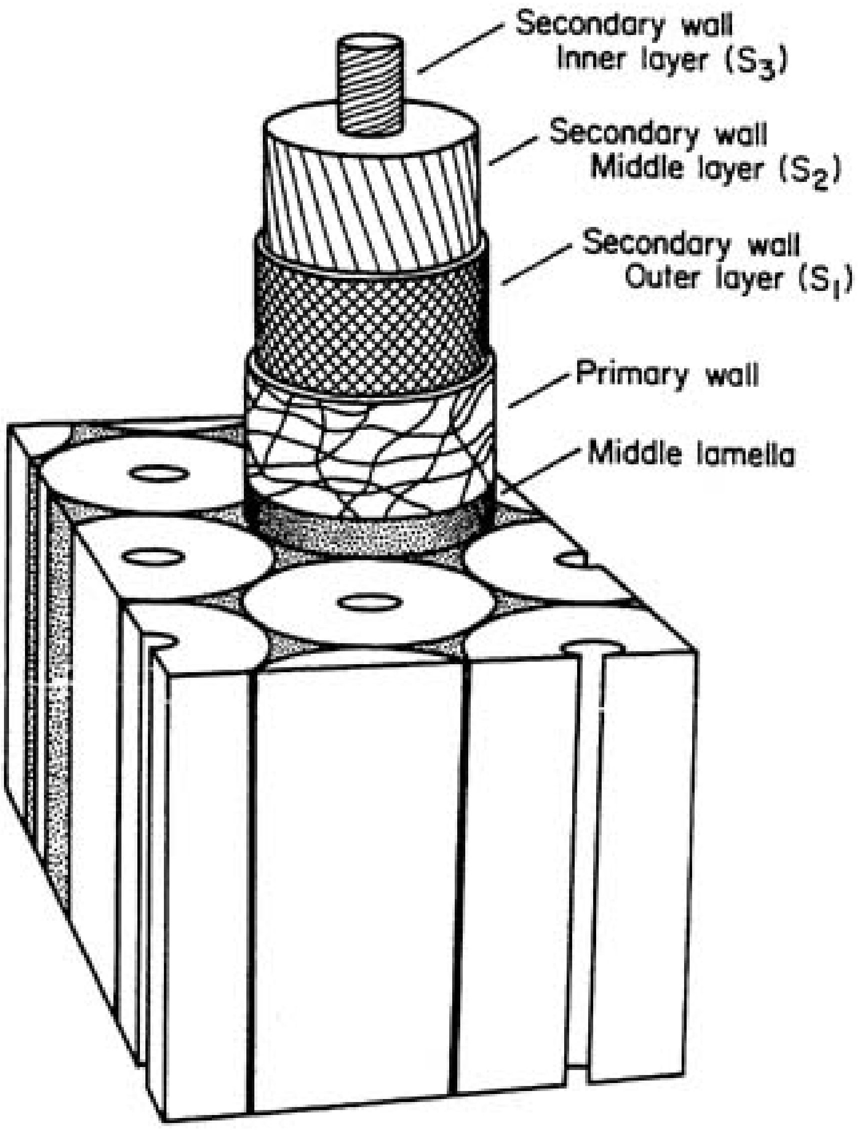
Fig. 7.5 Detailed structure of cell walls. (A) Strand of fiber cells. (B) Cross-section of fiber cells showing gross layering: a layer of primary wall and three layers of secondary wall. (C) Fragment from the middle layer of a secondary wall showing macrofibrils (white) of cellulose and interfibrillar spaces (black), which are filled with noncellulosic materials. (D) Fragment of a macrofibril showing microfibrils (white), which may be seen in electron micrographs (Fig. 7.6). The spaces among microfibrils (black) are filled with noncellulosic materials. (E) Structure of microfibrils: chainlike molecules of cellulose, which in some parts of the microfibrils are orderly in arrangement. These parts are the micelles. (F) Fragment of a micelle showing parts of chainlike cellulose molecules arranged in a space lattice. (G) Two glucose residues connected by an oxygen atoma fragment of a cellulose molecule. (*From* Anatomy of Seed Plants, Second Edition, 1977, *K. Esau* (Ed.), *Copyright © by John Wiley & Sons, 1977, reprinted by permission of the copyright owner*.)



The micelles also are associated in the tree to give long threadlike structures termed microfibrils (Fig. 7.5). However, the structure of the microfibrils is not completely uniform in terms of the alignment of the cellulose macromolecules, and the regions of nonuniformity between the micelles in the microfibrils are called amorphous regions. Thus the cellulose microfibril has a crystallineamorphous character. Water molecules enter the amorphous regions and swell the microfibrils; ultimately this is the mechanism by which fibers and wood swell in moist or wet environments.

The final fiber cell wall structure is essentially layers of the microfibrils or macrofibrils aligned in several different directions, as shown in Fig. 7.6. The microfibrils that make up the wood fiber are visible under the scanning electron microscope, which has a greater magnification than the light microscope.

Fig. 7.6 A section of wood made up of fibers and middle lamella. The structure of a fiber is given, showing the micro-fibrillar orientations in the different layers of the fiber cell wall. (*From R. A. Young and R. L. Giese (Eds.),* Introduction to Forest Science, *2nd ed., Copyright © 1990 by John Wiley & Sons, Inc. and reproduced by permission of the copyright owner*.)



The entity holding the fibers together, the middle lamella, is almost pure lignin (90%), as mentioned earlier. For the cellulose fibers to be separated, the middle lamella lignin must be chemically removed, a process that also removes most of the hemicelluloses, or must be mechanically degraded to free the fibers for papermaking. A paper sheet then can be formed from the separated cellulose fibers by depositing them from a water slurry onto a wire screen. The water drains away and the fibers collapse, leaving a fiber mat that derives its main strength from reassociation of the fibers through many hydrogen bondsthe same type of bond that gives mechanical integrity to the fibers.5

The long fibers from softwoods (Table 7.1) usually are preferred in papermaking for products that must resist tearing, such as grocery bags, whereas the shorter hardwood fibers give improved opacity, or covering power, and printability to the final paper sheet. The type of pulping process also affects the pulp properties, as described in a later section.

7.4 Pulp and Paper

The concept of making paper from the fibers from lignocellulosic materialsan integrated system of fiber separation (pulping) and re-forming of the fibers into a mat (papermaking)is attributed to T'sai Lun, a court official in southeast China in A.D. 105. The first fibers were obtained from old hemp rags and ramie fishnets, but shortly thereafter, the inner bark fibers from paper mulberry trees also were utilized for papermaking. Bamboo was used as a source of fiber several centuries later. The rags were macerated into a pulp in water with a mortar and pestle; then, after dilution in a vat, the pulp was formed into a wet mat on a bamboo frame equipped with a cloth screen to drain the free water. The mat was dried in the sun. The invention was based on the need for a writing material to replace the expensive silk and inconvenient bamboo strips. The invention was a closely guarded secret for many centuries but eventually filtered west, reaching Europe in 1151 and the United States (Pennsylvania) in 1690.5

The importance of the pulp and paper industry to the American economy is exemplified by the growth rate in the use of paper and paper products. New uses continually are being found for paper, and these developments together with a rising standard of living have resulted in a constant increase in the per capita consumption of paper. The industry has shown constant growth since about the middle of the nineteenth century, when processes were invented for the production of fiber from wood, replacing reliance on cotton rags as a source of fiber (Fig. 7.7). The relative paper consumption by grade is shown in Fig. 7.8, and differences in the economies of developed and less developed countries are demonstrated in Fig. 7.9 by the wide disparity in per capita consumption of paper and paperboard.

Fig. 7.7 Consumption of paper in the United States since the year 1860.

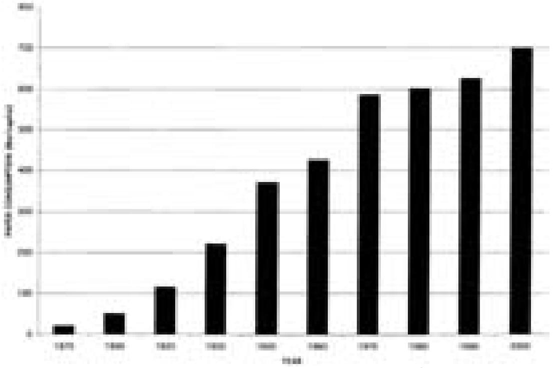


Fig. 7.8 Relative consumption of paper by grade.

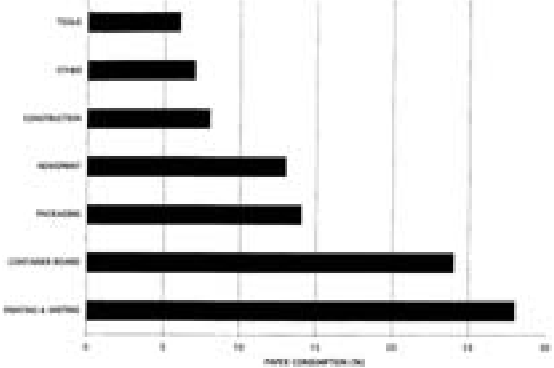
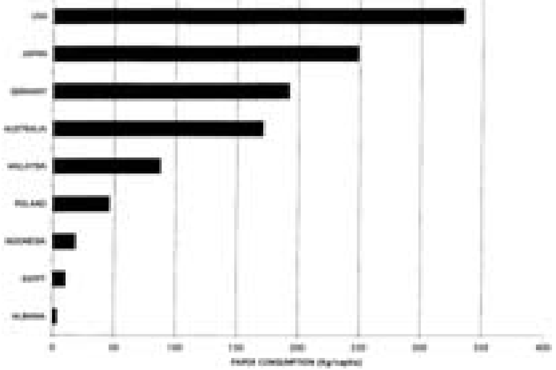


Fig. 7.9 Comparison of paper consumption in developed and developing countries.



In 1996, a total of 100,250,000 tons of paper and paperboard was produced in hundreds of separate mills. This paper and board was made from 82,580,000 tons of pulp, produced primarily from wood in over 200 pulp mills, supplemented by recycled paper.6

The five leading states for final paper production are Wisconsin, Maine, Washington, Alabama, and Louisiana. Only a small amount of paper is made from rags or other fiber sources (agricultural residues); so most wood pulp is produced from those areas of the country that are heavily forested. With the exception of Maine and Wisconsin, pulp production is concentrated in the southern and northwestern sections of the United States. Georgia and Washington are the top pulpproducing states. Although a high percentage of pulp is converted into paper or board at the same plant site, a significant portion (classified as "market pulp") is sold in bales to mills in other sections of the country for subsequent manufacture into paper. Thus, New York and Michigan are relatively important in papermaking, but they rank quite low as pulp producers. However, both New York and Michigan are large users of waste paper that is reprocessed into usable fiber. With the current emphasis upon recycling, it is probable that the use of waste paper will be greatly accelerated in the near future.

Raw materials for the pulp and paper industry can be classified as fibrous and nonfibrous. Wood accounts for over 95 percent of the fibrous raw material (other than waste paper) in the United States. Cotton and linen rags, cotton linters, cereal straws, esparto, hemp, jute, flax, bagasse, and bamboo also are used and in some countries are the major source of papermaking fiber.

Wood is converted into pulp by mechanical, chemical, or semichemical processes. Sulfite and kraft (sulfate) are the common chemical processes, and neutral sulfite is the principal semichemical process (NSSC). Coniferous wood species (softwoods) are the most desirable, but the deciduous, broad-leaved species (hardwoods) have gained rapidly in their usage and constitute about 25 percent of pulpwood. Table 7.2 summarizes the conditions utilized with the various pulping processes, and Fig. 7.10 shows the relative production of pulp by the major processes.5

Fig. 7.10 Relative wood pulp production by grade.

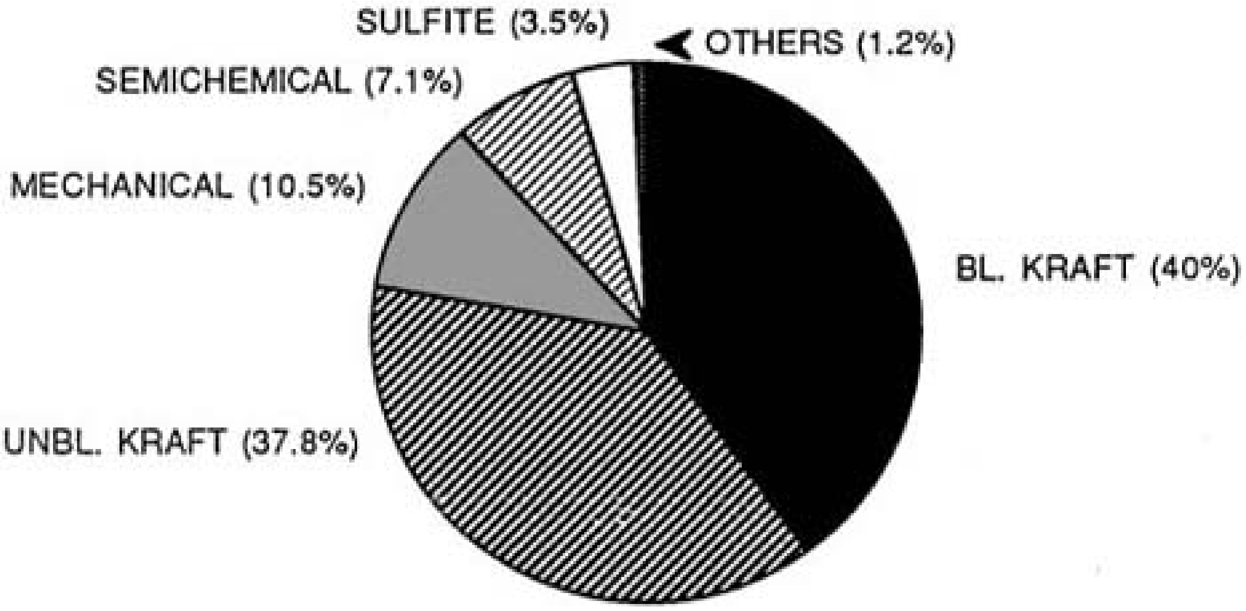


TABLE 7.2 Wood Pulping by Process and Yield

Process	Acronym	Treatment		Pulp Yield(%)
		Chemical	Mechanical	
Mechanical Processes				
Stone groundwood	SGW	None	Grinder	9395
Pressure groundwood	PGW	None	Grinder	9395
Refiner mechanical	RMP	None	Disk refiner (pressure)	9395
Thermomechanical	TMP	Steam	Disk refiner (pressure)	8090
Chemithermomechanical	CTMP	Sodium sulfite or sodium hydroxidea	Disk refiner (pressure)	8090
Chemimechanicalb	CMP	Sodium sulfite or sodium hydroxide	Disk refiner	8090
Semichemical Processes				
Neutral sulfite	NSSC	Sodium sulfite + sodium carbonate	Disk refiner	7085
Green liquor	GLSC	Sodium hydroxide + sodium carbonate	Disk refiner	7085
Nonsulfur		Sodium carbonate + sodium hydroxide	Disk refiner	7085
Chemical Processes				
Kraft		Sodium hydroxide + sodium sulfide	None	4555
Sulfite		Calcium bisulfite in sulfurous acidc, d	None	4050
Magnetite		Magnesium bisulfite in sulfurous acide	None	4555
Soda		Sodium hydroxide	None	4050
Soda-oxygen		Sodium hydroxide + oxygen	None	4555
Soda-anthraquinone	SAq	Sodium hydroxide + anthraquinone	None	4555
Dissolving Processes				
Prehydrolysis kraft		Steaming and kraft (two-step process)	None	35
Acid sulfite		Acid sulfite (Ca, Na)	None	35

aSodium sulfite or sodium hydroxide, 27% of wood; balso chemical treatment after fiberizing; calso sodium, magnesium, ammonia; dpH 2; epH 5.

Source: Data taken from references 1 through 5.

Nonfibrous raw materials include the chemicals used for the preparation of pulping liquors and bleaching solutions and the various additions to the fiber during the papermaking process. For pulping and bleaching, these raw materials include sulfur, lime, limestone, caustic soda, salt cake, soda ash, hydrogen peroxide, chlorine, sodium chlorate, and magnesium hydroxide. For papermaking, they include rosin, starch, alum, kaolin clay, titanium dioxide, dyestuffs, and numerous other specialty chemicals.

Wood Preparation

The bark of trees contains relatively little fiber and much strongly colored nonfibrous material; what fiber there is, is of poor quality. The nonfibrous material usually will appear as dark-colored dirt specks in the finished paper. Therefore, for all but low-grade pulps, bark should be removed as thoroughly as possible, and this removal must be very thorough in the case of groundwood and sulfite pulps if the finished paper is to appear clean.

Debarking usually is done in a drum barker, where the bark is removed by the rubbing action of logs against each other in a large rotating drum. Hydraulic barkers using highpressure water jets are excellent for large logs and are common on the West Coast. Mechanical knife barkers are becoming more common and are used extensively in smaller operations because of their relatively low capital cost. Also they have found widespread use in sawmills to debark logs prior to sawing so that the wood wastes can be used to produce pulp.

Wood cut in the spring of the year during the active growing season is very easy to peel. Much of the spruce and fir cut in the North is still hand-peeled during this season and usually represents the optimum in bark removal.

The standard log length used in the Northeast is 48 in., whereas 63 in. is common in the South. Wood generally is measured by log volume, a standard cord being considered to contain 128 ft3. Large timber on the West Coast generally is measured in board feet of solid volume. Measurement and purchase of wood on a weight basis are practiced, and are desirable because weight is directly related to fiber content.1

The growth in the use of wood residues has been substantial. By barking the sawlogs, the slabs, edgings, and other trimmings that formerly were burned now can be used to make pulp. Almost 28 percent of the wood used by the pulp industry is classified as waste wood. Several mills have been built that use no logs whatsoever but depend on residuals from satellite sawmill operations. Special sawmilling equipment has been developed to produce sawdust of a proper size so that it too can be used.

Wood used in producing groundwood or mechanical pulp requires no further preparation after debarking, but that used in the other chemical processes must first be chipped into small pieces averaging $\frac{1}{2} - 1$ in. in length and about $\frac{1}{8} - \frac{1}{4}$ in. in thickness. Chipping is accomplished with a machine consisting of a rotating disc with knives mounted radially in slots in the face of the disc. Modern chippers have up to 12 knives; the ends of the logs are fed against the disc at about a 45° angle.

Mechanical Pulping

Mechanical pulping, as the name implies, involves mainly a grinding and abrading action on wood to produce a mechanical pulpthe primary component of newsprint. Modern mechanical pulping includes stone ground-wood pulping (SGW), in which bolts of wood are pressed against a revolving grindstone, and refiner mechanical pulping (RMP), in which chips are passed between single- or double-rotating plates of a vertical-disc attrition mill. Recent developments in stone grinding include application of pressure to the grinder (PGW) and control of temperature.

At one time, natural sandstone was used for the grindstone, but modern stones are either silicon carbide or aluminum oxide grits in a vitrified clay binder. Thus the characteristics of the stone can be varied to produce pulps "tailor-made" to fit their desired end use. Groundwood pulps for roofing or flooring felts must be extremely coarse and free-draining, so they require a stone with large grits, whereas newsprint pulps are very fine and require the use of small grits. Pulp characteristics also can be varied by changing the stone surface pattern, the stone speed, the pressure of the logs against the stone, and the temperature of the ground-pulp slurry. Generally a coarser and more freely draining pulp is obtained with a coarse surface pattern and high speed, pressure, and temperature. The type and the condition of the wood also are factors, but groundwood pulps usually are made from the coniferous or long-fibered species because the deciduous or short-fibered species give very weak pulps.

Basic changes in mechanical pulping technology include pretreating chips with chemicals, steam, or both. These developments started 40 years ago when chips were pretreated with caustic soda, in a process called chemimechanical pulping (CMP). Presteaming and pressure refining of chips gives a thermomechanical pulp (TMP). The chips are steamed at 40 psi for 24 min and sent to a refiner modified to operate under pressurized conditions. When chemical pretreatment and pressure steaming are combined, the pulp is referred to as chemithermomechanical pulp (CTMP). Sulfite solutions commonly are used to produce CTMP, but there are many variations of these processes. These treatments are employed to improve pulp quality. The steam and chemicals aid fiberizing by giving a less-damaged fiber, which makes the final paper stronger. The pulps are sufficiently improved over stone groundwood that they can be used to partially substitute for chemical pulps. Newsprint consisting of 100 percent TMP or CTMP is produced commercially.

Groundwood or mechanical pulp is low in strength compared with the chemical pulps. It is composed of a mixture of individual fibers, broken fibers, fines, and bundles of fibers. Papers made from groundwood also lose strength and turn yellow with time. Thus, groundwood pulps are used only in relatively impermanent papers such as for newsprint, catalogs, magazines, and paperboard. Groundwood papers have excellent printing qualities because of high bulk, smoothness, resiliency, and good ink absorption. Newsprint contains about 80 percent groundwood, and the other publication paper grades contain about 3070 percent, with the remainder a chemical pulp to provide greater strength. Groundwood is the cheapest pulp made and also utilizes the entire wood, giving close to a 100 percent yield.

Chemical Pulping

The objective of chemical pulping is to solubilize and remove the lignin portion of wood so that very little mechanical treatment is necessary to fiberize the wood. Many variations are used throughout the world, the simplest classification of pulping methods depending on whether they are acidic or alkaline. Each has its own specific advantages and disadvantages, but the kraft (alkaline) process accounts for over 90 percent of all chemical pulp produced in the United States. All present commercial processes use aqueous systems under heat and pressure although several organosolv systems are at the demonstration-plant stage.

The lignin-removal and fiber-separation mechanisms involved in chemical pulping are alkaline hydrolysis (lignin bond cleavage) and the formation of soluble sulfur-containing lignin (thiolignins) in kraft pulping. The chemical mechanisms in sulfite pulping also involve lignin bond hydrolysis and the formation of soluble sulfur-containing lignin derivatives, termed lignosulfonates.5 Polysaccharide reactions are secondary to fiber separation in chemical pulping although polysaccharide retention, solution, or modification profoundly affect pulp yield and properties. Neutral sodium sulfite reactions in neutral sulfite semichemical (NSSC) pulping also are based on sulfonation for partial dissolution of lignin and a partial modification of the fiber bond to permit a clean separation of the fibers by mechanical action in disk refiner mills. The pulping chemicals are applied in water solutions (white liquor) at various concentrations and ratios of liquids to solids under different conditions of temperature and pressure to achieve the desired degree of delignification.

Kraft Process

The kraft (sulfate) process uses a mixture of sodium hydroxide and sodium sulfide as the active chemical. Although the designation sulfate process is misleading, it sometimes is used because sodium sulfate is the makeup chemical in the process. The word "kraft" now is mainly used to describe this process, derived from the Swedish or German word for strength, as this process produces the strongest pulp. Historically, sodium hydroxide alone (soda process) first was used as the alkaline pulping agent, but all these mills have converted to the kraft process because the other pulp is weak and inferior to kraft pulp.

The pulping (cooking) process traditionally was performed on a batch basis in a large pressure vessel called a digester. Conditions vary depending upon the type of wood being pulped and the quality of end product desired. Typical conditions for kraft cooking are listed in Table 7.3.

TABLE 7.3 Typical Kraft Pulping Conditions

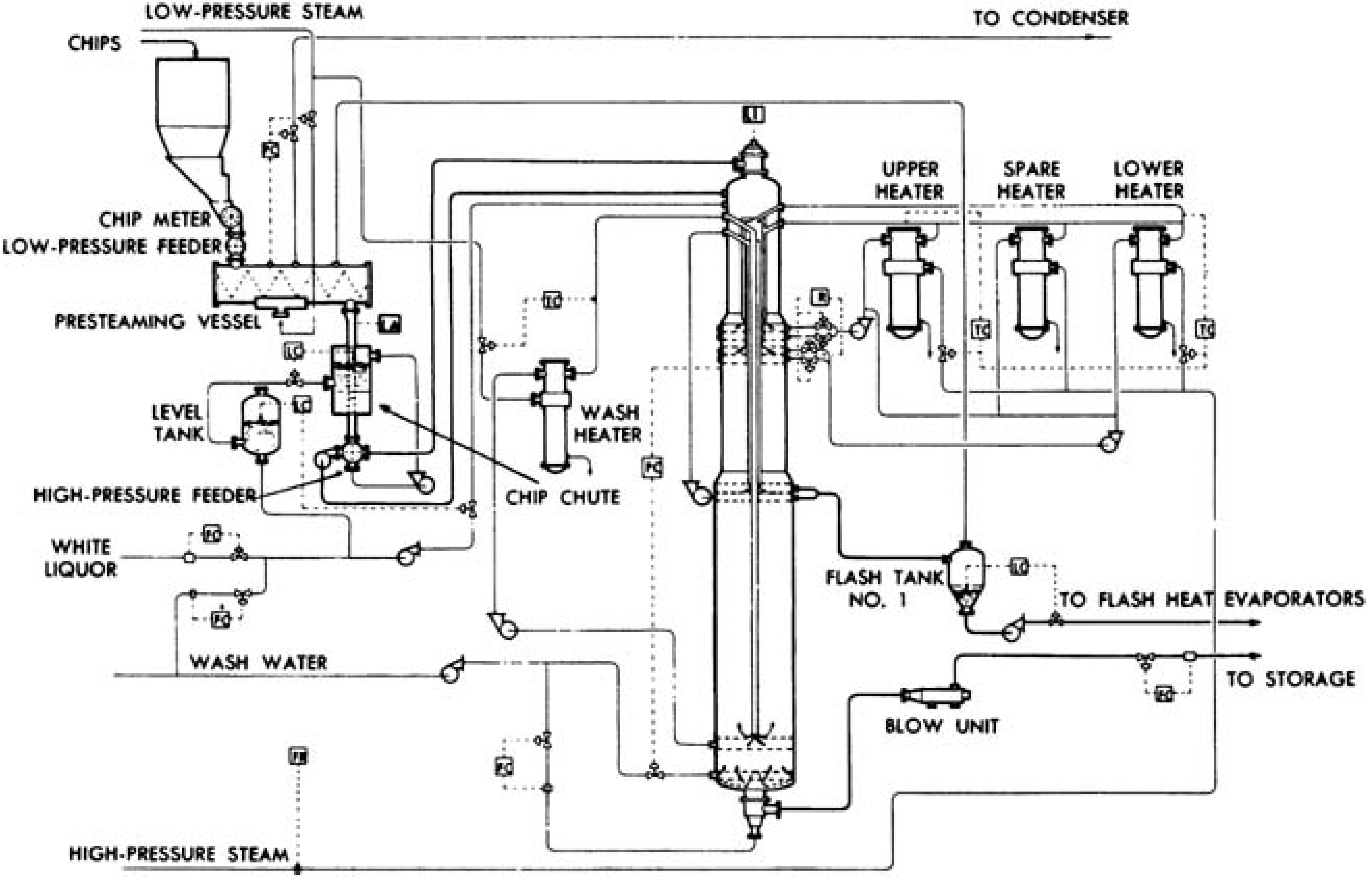
Pressure	100110 psig
Temperature	170175°C
Time	23 hr
Alkali charge	1525% of weight of wood (calculated as Na_2O but consisting of approximately $5NaOH + 2Na_2S$)

Liquor to wood ratio is 4:1 (by weight).

Digesters are cylindrical in shape with a dome at the top and a cone at the bottom. Ranging in size up to 40 ft high and 20 ft in diameter, the largest will hold about 7000 ft3 of wood chips (about 35 tons) for each charge. The chips are admitted through a large valve at the top, and at the end of the cook they are blown from the bottom through a valve to a large blow tank. During the cook the liquor is heated by circulation through a steam heat exchanger, which also avoids the dilution of the cooling liquor that would occur from heating by direct injection of steam.

The development of the continuous digester (Fig. 7.11) significantly facilitated the use of kraft pulping. Chips are admitted continuously at the top through a special high-pressure feeder, and the cooked pulp is withdrawn continuously from the bottom through a special blow unit. Recent installations range in size up to 150 ft high and are capable of producing about 1,000 tons of pulp per day in one unit. Cooking liquors and conditions are approximately the same as for the batch digesters. These units offer both good economics in the production of pulp and a quality advantage compared to the batch digester. However, because the capital investment for these units is somewhat higher, both systems are installed for new mills.

Fig. 7.11 Typical Kaymr continuous cooking system for kraft pulp. (*Courtesy Kaymr, Inc.*)



Because of the high alkali charge, the chemicals must be recovered and reused. This process also alleviates pollution problems because the yield of pulp is only about 45 percent of the original wood weight, and the organic residues must be eliminated. After being cooked in the digester, the pulp is washed in a countercurrent rotary vacuum washer system using three or four stages. The pulp then is ready for bleaching or for use in papers such as grocery bags where the brown color is not objectionable.

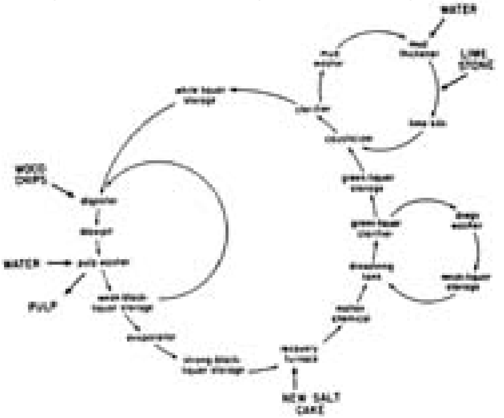
The separated liquor, which is very dark, is known as "black liquor." It is concentrated in multi-effect evaporators to 6065 percent solids. At this concentration the quantity of dissolved organic compounds from the wood (lignin and carbohydrate degradation products) is sufficient to allow the liquor to be burned in the recovery furnace.

By controlling the amount of excess air admitted to the furnace and the temperatures, the organics in the liquor can be burned. The inorganics collect on the bottom of the furnace as a molten smelt of Na2CO3 and Na2S. Sodium sulfate is added to the liquor as makeup and is reduced to Na2S by carbon. After dissolving in water, this mixture (called "green liquor") is reacted with slaked lime:

$$Na_2CO_3 + Ca(OH)_2 \longrightarrow 2NaOH + CaCO_3$$

The Na2S does not react with the lime, so the resultant mixture of NaOH and Na2S (called "white liquor") can be reused to pulp more wood. The CaCO3 sludge is filtered off, burned in a lime kiln, and reused. Thus, the chemical system is a closed one, as shown in Fig. 7.12, and this minimizes costs and pollution.

Fig. 7.12 Diagram showing cyclic nature of kraft recovery process. (*From* Pulp and Paper, *3rd ed., J. P. Casey (Ed.), WileyInterscience, New York. Copyright © 1980 by John Wiley & Sons, Inc. and reproduced by permission of the copyright owner*.)



The kraft process has had a serious problem with air pollution due to the production of hydrogen sulfide, mercaptans, and other vile-smelling sulfur compounds. In recent years, the use of various techniques such as black-liquor oxidation, improved evaporators and furnaces, and control of emissions has greatly improved this situation. However, older mills are being forced to expend large amounts of money to bring their operations up to the environmental standards.

Virtually any kind of wood can be pulped by the kraft process; and because it produces the strongest pulps with good economies, it has grown to be the dominant process in the world. Although the odor problem is very severe, it does not appear that this process will be supplanted in the near future; instead, it will be improved and modified. When pulping resin-rich woods such as southern pine, the kraft process yields turpentine and tall oil as valuable by-products. The steam generated in the recovery furnace is almost enough to make the pulp mill energy self-sufficient.

A recent discovery that small amounts of anthraquinone (0.050.25% on wood) added to alkali liquors can enhance delignification and produce kraft quality pulp has been of great interest. Although the price of anthraquinone (AQ) is high and the chemical is not recovered, only small quantities are required, and the benefits outweigh the costs sufficiently that a number of mills around the world now routinely add AQ to the kraft pulpwood digester.

Sulfite Process

The sulfite process uses a cooking liquor of sulfurous acid and a salt of the acid. Although calcium was the most widely used base at one time, it has been supplanted by sodium, magnesium, and ammonia. Lignin will react with the bisulfite ion (HSO-3) under acidic conditions to form lignosulfonates that are soluble in water. For many years this was the preferred process because it produced pulps of light color that could easily be bleached, it used cheap chemicals in fairly limited amounts so that no recovery was necessary, and it was a relatively simple process to operate.

Although the production of sulfite pulp has remained relatively constant for the last 30 years, the rapid growth of kraft pulping has reduced sulfite's share to less than 10 percent of the chemical pulp produced. There are several reasons for this, but the primary ones are the inability to cook resinous woods such as pine, problems in producing strong pulps from hardwoods, and of the greatest importance, the lack of a cheap and simple recov-ery system to reduce water pollution problems. However, this process produces pulps with special qualities such as high alpha-cellulose grades for rayon, so it will continue to be used.

Initially calcium was the preferred base because it was cheap and convenient to use. However, no recovery system is available for this base, so most calcium-base mills either have ceased operation or have converted to sodium, magnesium, or ammonia, for which recovery systems are available.

Regardless of the base used, the initial step is the burning of sulfur to produce sulfur dioxide (SO2). The air supply to the burner must be carefully controlled, as too much air will enhance the formation of sulfur trioxide (SO3) and subsequent production of sulfuric acid (H2SO4), which is very undesirable. The gas also must be cooled quickly from 1000°C, leaving the burner at below 400°C, to minimize formation of SO3. After cooling to 2030°C, the SO2 gas must be absorbed in water and reacted with the proper base to form the cooking liquor.

For calcium-base liquor, the gas is passed through towers packed with limestone with water flowing down through the tower. Because of the limited solubility of calcium bisulfite [Ca(HSO3)2], the pH of the liquor is very low (about 2), and free sulfurous acid is present. This usually is called the acid sulfite process. As mentioned before, calcium-base mills have essentially disappeared in the United States.

The so-called soluble bases now are used, with each having certain advantages. Because solutions of sodium, magnesium, and ammonium bisulfite are all soluble at pH 4.5, the current practice is to pulp at the higher pH, which is usually called bisulfite pulping. Extremely long cooking times (710 hr) are necessary with acid sulfite, whereas 45 hr is sufficient with bisulfite.

Sodium base is the easiest to prepare (Na2CO3 or NaOH usually is used as the makeup chemical) and gives the highest-quality pulp; however, recovery processes, though available, are complicated and expensive. Magnesium base [from Mg(OH)2] is somewhat more difficult to handle, but recovery systems are available, and the majority of the sulfite pulp now is produced from this base. Ammonium base (from NH4OH) was used in the past; the ammonia cannot be recovered, but the liquor can be evaporated and burned without leaving any solid residue, thus reducing water pollution. As long as aqueous ammonia remains low in price, this process will be attractive because the SO2 can be recovered from the waste gases by passing them through a wet scrubber flooded with fresh ammonium hydroxide. The various sulfite-based processes are summarized in Fig. 7.13.

Fig. 7.13 Summary of sulfite pulping processes.

Batch digesters usually are used in the sulfite process. Cooking temperatures are lower (140150°C), and times are longer than in the kraft process, with pulp yields about the same as in the kraft process. Spruce and fir are the preferred species for cooking by the sulfite process because they produce relatively strong, light-colored pulps. About 20 percent of newsprint consists of this type of pulp that has not been bleached. Thus, the sulfite industry is concentrated in Canada, the northern United States, and the Pacific Coast, where the supplies of spruce and fir are greatest and the largest quantities of newsprint are produced.

A large amount of research has been done on developing products from the waste sulfite liquor, and some success has been achieved. Vanillin, alcohol, and torula yeast can be produced as by-products, and the lignosulfonates are used as viscosity modifiers in drilling muds and for similar purposes. However, the majority of waste liquor is burned to recover the cooking chemicals and the heat values.

Semichemical Pulping

Various combinations of chemical and mechanical treatments have been used to produce pulps with specific properties. Mild chemical treatments to give partial delignification and softening are followed by mechanical means to complete fiber separation.

The NSSC process is one in which wood chips, usually from hardwoods, are cooked with Na2SO3 liquor buffered with either NaHCO3, Na2CO3, or NaOH to maintain a slightly alkaline pH during the cook. Unbleached pulp from hardwoods cooked to a yield of about 75 percent is widely used for the corrugating medium. Although bleachable pulps can be produced by this process, they require large quantities of bleaching chemicals, and the waste liquors are difficult to recover. Currently many NSSC mills are located adjacent to kraft mills, and the liquors can be treated in the same furnace. Thus, the waste liquor from the NSSC mill becomes the makeup chemical for the kraft mill, solving the waste problem. NSSC hardwood pulp is the premier pulp for the corrugating medium and cannot be matched by any other process.

Chemimechanical pulps usually are produced by soaking the chips in solutions of NaOH or Na_2SO_3, and then refining in disc refiners to produce a groundwood-type pulp. Chemical consumption is very low, and yields are usually 8595 percent.

Solvent Pulping

The concept of using solvents to dissolve lignin was first explored by Peter Klason in 1893. In the 1930s, Kleinert developed and patented the first solvent (organosolv) pulping process based on 50 percent aqueous ethanol.[7] Until recently there was little incentive to pursue alternatives to the kraft process because energy and chemicals were reasonably cheap, there were few environmental regulations, and there was little interest in by-products. However, new, strict environmental regulations, rising energy and chemical costs, and the very large capital investments needed to construct new kraft mills have stimulated research and investment in alternate solvent pulping processes.[8]

With organosolv pulping, wood chips are contacted with an aqueous solvent mixture and cooked (digested) for a period of time at elevated temperatures. Digestion pressures range from atmospheric to 500 psig. Cooking initially releases acetic and formic acids from ester functionalities in the wood, and these acids further promote hydrolysis of lignin and hemicelluloses to low molecular weight species. Catalysts such as mineral acids (HCl), organic acids (acetic, formic), sulfonic acids, and Lewis acids and bases ($AlCl_3$, $Fe_2(SiO_4)_3$, $Mg(SO_4)$, $CaCl_2$) can be employed to promote delignification. Cellulose pulp is filtered from the liquor, and the solvent is recovered, usually by distillation. Lignin becomes insoluble at this point and precipitates from the liquor, leaving an aqueous solution of primarily hemicellulosic sugars. Conditions for a variety of organosolv processes are compared in Table 7.4.[9]

TABLE 7.4 Solvent Pulping Process Characteristics

Process Name Developer	Solvent Used	Additional Reactants	Digestion Temperature	Digestion Time	Reference
Acetic acid (Herdle, Young)	90% acetic acid[a]	0.2% H_2SO_4	150°C	4 hr	(10)
Acetosolv (Nimz)	95% acetic acid	0.1% HCl	110°C	35 hr	
Ester (Young)	5075% acetic acid[a]	Ethyl acetate[d]	170°C	1 hr	(11)
Formic acid (Jordan)	80% formic acid	Unspecified organic catalyst	105°C	45 min	(12)
Peroxoic acid (Poppins, Sundquist)	80100% formic acid and hydrogen peroxide	None	7080°C	60 min	
Hydrotropic pulping (Springer)	2.0 M xylene-sulfonic acid[a]	None	100°C	60 min	(13)
BattelleGeneva process (Sachetto)	Aqueous phenol[b]	Acid catalyst	100°C	3060 min	(14)
Sulfolane (Clemont)	50% sulfolane	None	175°C	2 hr	(15)
Acetone (DeHaas and Lang)	Aqueous acetone[b]	NH_3/Na_2S	160190°C	12 hr	(16)
Glycol (Burkart, Quigley)	20100% glycol	Mineral, salicylic, sulfonic acids	100205°C	13 hr	
Phenol (Sweers)	2050% phenol	HCl. sulfonic acids	80205°C	14 hr	
Cresol (Sakakibara)	2080% cresol	Cuprous oxide, acetic acid, AQ	160190°C	14 hr	
Amine (Wallis)	2590% mono- or triethanol amine	None, AQ	165200°C	12 hr	
Kleinert process	2075% ethanol[a, c] usually 50%	Uncatalyzed or metal salt catalysts	160195°C	3060 min	(7, 17, 18)
Alcell (Pye)	50% aq. ethanol	Uncatalyzed	195°C	12 hr	
AEM (Paszner)	Aqueous methanol[b]	Alkali earth salt catalysts	195210°C	2545 min	(19)
MD-Organosolv (Edel)	Aqueous methanol	NaOH	170195°C	45 min	(20)
ASAM (Patt)	3050% aq. methanol	NaOH, Na_2SO_3, and AQ (all)	180°C	23 hr	

[a]The other compound is water; [b]Concentration not reported; [c]*n*-Butanol also has been employed; [d]A three-component (ethyl acetate/acetic acid/water) pulping process.

Two processes that reached the demonstration-plant scale are the Alcell and the MD-Organosolv processes, both of which are basic modifications on the original process patented by Kleinert.[7] In the Alcell process preheated chips (80°C) are packed in the digester using steam, and the steaming is continued to displace air from the chips. Preheated solvent that has been previously used as a wash liquor for two earlier batches is pumped in and rapidly brought up to the pulping temperatures of 190200°C, corresponding to an operational pressure of 500 psi. The liquor is continuously circulated in the digester during the pulping period. At the end of this period the liquor in the digester is displaced by wash liquor that was used in the second washing stage of the previous batch. The displaced liquor from the digester flows to the lignin and sugar recovery system. The third-stage wash liquor is displaced by fresh liquor and flows into an intermediate storage tank for use in the second stage. The second-stage wash liquor is drained into another storage tank for use in the digester. At this point, the digester contains soaked, delignified chips and alcoholwater vapor at pulping temperature and pressure. It is now depressurized, and the departing vapor is condensed for reuse as fresh pulping liquor. Finally, the alcohol remaining in the digester is driven off by steaming. The pulp is diluted with water and pumped out of the digester for cleaning and bleaching treatments. The first-stage spent liquor entering the by-product recovery area is first flashed and then diluted with process water to precipitate the dissolved lignin. After settling, the solid lignin is separated by centrifugation, washed, and dried. The filtrate enters a distillation tower, where alcohols and some acetic acid and furfural are recovered. Finally, the remaining aqueous liquor is further evaporated to a sugar syrup.[21]

A demonstration plant, producing 15 metric tons of organosolv pulp a day, was constructed in New Brunswick, Ontario. A comparison of the strength properties of Alcell and kraft pulps is given in Table 7.5. The process as described above is only suitable for hard-woods.[8,22]

TABLE 7.5 Comparison of Properties of Allcell and Kraft Hardwood Pulps[a]

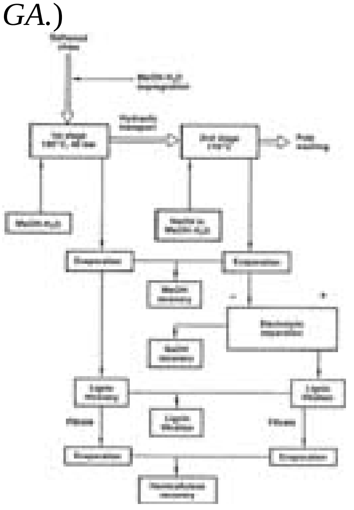
Property	Alcell Pulp	Commercial Kraft Pulp
Tensile (km)	7.47	7.40
Tear(mN · m^2/g)	7.20	6.75
Burst (kPa · m^2/g)	5.08	5.18
Brightness (ISO)	88.7	89.6

[a]Pulps 50% maple, 35% birch, 15% poplar; beaten to 400 Canadian Standard Freeness.

Source: McCready, M. (Ed.), *Paper Age* (Mar. 1991).

The principles of the MD-Organosolv process, developed in Germany, are illustrated in Fig. 7.14. The process consists of first impregnating softwood chips with 50 percent aqueous methanol before they enter the first-stage continuous reactor. The first pulping stage is operated under 40-bar pressure at 195°C and removes approximately 20 percent of the lignin and a major part of the hemicelluloses from the chips. The chips then, are transported hydraulically to the second-stage digester where alkali is charged at 1520 percent based on wood, and the methanol concentration is reduced to about 30 percent. The second stage is operated at 170°C, and the combined residence time in both reactors is only 45 min. The strength properties of the pulps produced are only slightly lower than those of corresponding kraft pulps. The recovery of organics and NaOH from the second-stage liquor is performed in a novel manner. After methanol evaporation, the liquor enters an electrolytic cell that not only separates NaOH but also precipitates lignin in a filterable form.[21]

Fig. 7.14 Flowsheet of MD-Organosolv pulping system. (*From A. Aziz, and K. V. Sarkanen,* Tappi, 72(3), 169175 (1989); *with permission of Tappi, Atlanta, GA.*)



Biomechanical Pulping

Another pulping method that has reached the pilot plant stage, and will probably be commercialized in the near future, is biomechanical pulping. This approach involves the use of white rot fungi to first breakdown the wood structure, and apparently modify the lignin, with the result that the wood chips can then be mechanically disintegrated with much less energy input. After screening of hundreds of different species of white rot fungi at the U.S. Forest Products Laboratory, it was found that *Ceriporiopsis subvermispora* produced the greatest reduction in energy consumption and also substantially improved the strength properties of the pulp in comparison with conventional refiner mechanical pulp. The approach is also suitable for the production of pulp from agro-based materials as well.8,23

Screening and Cleaning of Wood Pulp

The desired pulp fibers are usually between 1 and 3 mm in length with a diameter about one-hundredth as large. Any bundles of fibers or other impurities would show up as defects in the finished paper and must be screened out. Wood knots usually are difficult to pulp and must be removed.

Screening is usually a two-stage process with the coarse material being removed by screens with relatively large perforations ($\frac{1}{4} - \frac{3}{8}$ in.). Additional fine screening is done with screens using very small (0.0080.014 in.) slots to ensure the removal of oversized impurities. Screen size openings will depend on the species of wood being processed and the desired quality of the end product. Because of the tendency of the fibers to agglomerate when suspended in water, it is customary to screen at very low solids (consistencies) of about 0.5 percent fiber and 99.5 percent water.

To meet the ever-increasing demands for cleaner pulps, centrifugal cyclone cleaners have come into almost universal use. The screened pulp is pumped through these units at low consistencies and high velocities. The fiber slurry enters the cone tangentially at the top, and a free vortex is formed with the velocity of the flow greatly increased as the diameter of the conical section is reduced. Heavier particles of sand, scale, or other dirt are forced to the outside of the cleaner and are discharged from the bottom tip through a small orifice. Because of the velocity gradients existing in the cone, the longer fibers (7595%) are carried into the ascending center column and are discharged through the larger accept nozzle at the top. In a properly designed and operated unit, a shape separation also is made so that round particles, even though of the same specific gravity as the good fibers, will be discharged as rejects through the bottom orifice. In this way, small pieces of bark also are removed. To reduce the quantity of rejects to an acceptable level, they in turn are processed through a second, a third, or even a fourth stage of cleaners, thus holding the final loss of pulp to about 0.251 percent of the feed, depending upon quality demands and the dirt level of the incoming pulp.

Bleaching of Wood Pulp

The color of unbleached pulp ranges from cream or tan for the sulfite process to dark brown for the kraft process. Although about 7590 percent of the lignin has been removed by the pulping process, the remainder, along with other colored degradation products, must be removed by bleaching.

Although it is possible to improve the brightness (whiteness) of the pulp in one stage, the economical achievement of high brightness requires the use of several stages. Current practice uses combinations of chlorination with elemental chlorine (C), alkaline extraction with sodium hydroxide (E), and various oxidative stages using sodium or calcium hypochlorite (H), chlorine dioxide (D), or hydrogen peroxide (P). The pulp is washed between each stage to remove solubilized impurities. Many combinations are possible, and each mill selects the sequence that fits its requirements the best. The sequences most commonly utilized are:

CEH
CEHD
CEHDP
CED
CEDED
CEHDED

The greater the number of stages, the higher the quality of the final pulp but at increased cost. Sulfite pulps are much easier to bleach and usually use only the three- or four-stage sequences, whereas kraft pulps require additional stages. Typical conditions for bleaching a kraft pulp are given in Table 7.6.

TABLE 7.6 Typical Conditions for Bleaching Kraft Pulp

Stage	% Chemical on Pulp	Time (min)	Temperature (°F)
Chlorination	56	30	70
Extraction	21/23	60	140
Hypochlorite	12	90	105
Chlorine dioxide	1/21	240	160

Chlorine and caustic are purchased, but chlorine dioxide must be generated at the site by using sodium chlorate as the basic chemical. Effluents from bleach plants are a source of great concern in regard to pollution. Large quantities of water are discharged (typically 10,00020,000 gal/ton of pulp) that are high in color, especially the water coming from the chlorination and extraction stages.

New developments involve the replacement of the chlorination and extraction stages with a single stage involving gaseous oxygen and sodium hydroxide. The pulp is dewatered to about 1525 percent solids, treated with about 46 percent NaOH, and passed through a reactor in a fluffed condition using oxygen gas at about 150 psi. The pulp consumes about 12 percent oxygen in about 15 min and leaves in a semibleached condition. After washing, the use of conventional chlorine dioxide stages will produce the desired level of whiteness. The advantage of this process is that the effluent contains no chlorides and can be used as process water in the pulp mill. Any sodium and organics in the effluent then will enter the kraft recovery cycle and serve as makeup, thus reducing the pollution load.

With the removal of chlorine from the bleaching sequence, the process is termed Elemental Chlorine Free (ECF) bleaching and usually an Oxygen (O) stage is now substituted for the Chlorine (C) stage. Regulatory agencies in Europe, and particularly in Scandinavia, have imposed even greater restrictions on emissions from pulp mill bleach plants and another new approach has been developed, namely Totally Chlorine Free (TCF) bleaching of pulps. For TCF more radical changes are necessary with substitution of both (C) and (D) stages with ozone (O), peroxide (P), and enzyme (X) stages in a sequence such as OXZP.8

The use of enzymes is the newest development in bleaching technology. At least one enzyme-based process developed in Finland has been applied commercially. The process uses xylanase to make lignin more vulnerable to oxidation by attacking the surrounding polysaccharides that protect the lignin. Another exciting application would be to use these and other enzymes for removal of lignin pollutants from waste effluents. Biotechnology should lead to safer and cleaner methods for pulping and bleaching.8

These bleaching sequences are designed to remove lignin, yielding a highly purified fiber consisting only of carbohydrate material. For producing high-yield pulps such as groundwood, where all of the lignin is retained in the pulp, it is not possible to use these systems. However, extremely high brightness is not required in this case; thus, some improvement is attained by using one stage with either peroxide or hydrosulfite (dithionite). No yield loss is encountered as the action of both of these is merely to decolorize the pulp rather than to remove any impurities. Usually about 0.5 percent of either of these chemicals will give a noticeable increase in brightness, and they are widely used to upgrade the quality of groundwood.

BiotechnologyBiopulping and Biobleaching

The pulping of wood is at present based on either mechanical or chemical methods or combinations thereof, as previously described in this chapter. The interfiber lignin bond is broken down by the mechanical and chemical treatments to free the cellulose fibers for papermaking. In the forest, white rot fungi perform a similar task on wood left behind. The enzymes of the fungi do the work of lignin degradation as referred to in the section on "Biomechanical Pulping."

It should be possible to isolate these specific enzymes from the fungus for use in biological pulping or biopulping. Indeed, researchers at several laboratories are attempting to isolate lignin-degrading enzymes (lignases) from the white rot fungi. They are quick to point out, however, that there probably is a lignase complex in which several enzymes work together to degrade lignin. Once specific enzymes are isolated, work can commence on producing the enzymes in greater quantities through genetic engineering. Another drawback is the slow rate of reaction as already discussed in the section on "Bleaching of Wood Pulp."

The initial applications of lignases would be to degrade the lignin partially before mechanical treatment so that the process would require less energy, and the pulp would have more strength after this biomechanical pulping. Biobleaching also would be possible for brightening or whitening pulp fibers, in lieu of the toxic chlorine compounds utilized at present by the industry. A particularly exciting application would be to use these types of enzymes for removal of lignin pollutants from waste effluents. Biotechnology should lead to safer and cleaner methods for pulping and bleaching.

Recycling

Worldwide, approximately 38 percent of wastepaper is recycled; however, the rate of recycling varies between countries with the United States at a 45 percent recycle rate. The United States is a major exporter of wastepaper, with the majority going to "fiber-poor" countries such as Korea and Japan that have much less virgin fiber.

A variety of problems are associated with paper recycling, such as collection, distribution, and wild cyclic swings in the market. However, with landfill sites at a premium and paper representing 3040 percent of the municipal solid waste, it makes good sense in the long run to promote paper recycling, which reduces landfill needs and the consumption of virgin timber. The paper industry in the United States has now set a goal of 50 percent paper recycling rate.

There are a wide range of different grades of wastepaper available, depending on the source and the extent of separation. As the name implies, wastepaper designated as "direct pulp substitutes" is utilized with little treatment before reslushing in a hydrapulper. The direct pulp substitutes are the highest grade of wastepaper.

The majority of recycled paper (about 75%) is used with no attempt to remove inks, dyes, or pigments from the paper. The resultant pulp is of rather poor color and quality, and is used primarily as filler stock in paperboard. The bulk grades are the largest-volume wastepaper source.

Deinked grades of wastepaper require special techniques and equipment to remove inks, coatings, adhesives, and so on. The deinking process is complicated and time-consuming; depending on the quality of deinked pulp required, the process may involve a number of combinations of washing, flotation, dispersion, screening, and the use of cyclone cleaners. The newer noncontact inks present special removal problems, as do wet-strength agents, adhesives, and the so-called stickies. Stickies are made from hot-melt adhesives (vinyl acetate polymers and copolymers), pressure-sensitive adhesives (styrene-butadiene), and lattices (natural and synthetic rubber). Tackifiers and waxes also are usually included in these adhesive formulations. A variety of additives are used to help remove stickies and other contaminants from wastepaper, including solvents, nonionic and cationic surfactants, zirconium compounds (to reduce tackiness), and talc. Of course, all of the additional steps and additives add to the expense of recycling. There is also the problem of waste disposal from the deinking process, which must be properly handled.

Stock Preparation

Stock preparation in a paper mill includes all intermediate operations between preparation of the pulp and the final papermaking process. It can be subdivided into (1) preparation of the "furnish" and (2) "beating" or "refining." Furnish is the name for the water slurry of fiber and other chemicals that goes to the paper machine. Beating or refining refers to the mechanical treatment given to the furnish to develop the strength properties of the pulp and impart the proper characteristics to the finished paper.

Cellulosic fibers are unique in that, when suspended in water, they will bond to each other very strongly as the water is removed by filtration and drying, with no need for an additional adhesive. This is so because of the large number of hydrogen bonds that form between the surfaces of fibers that are in close contact as the water is removed. This bonding is reversible and accounts for the well-known fact that paper loses most of its strength when wet. If paper is suspended in water and agitated, it will separate into the individual fibers, a behavior that allows the easy reuse of wastepaper or the processing waste from the paper mill.

In order to enhance the bonding capability of the fibers, it is necessary to mechanically beat or refine them in equipment such as beaters, jordans, or disc refiners. This treatment of the pulp slurry at about 36 percent consistency is done by passing the pulp between the two rotating surfaces of the refiner. These surfaces contain metal bars and operate at very close clearances. As the fibers pass between the bars, they are made more flexible, and a larger surface for bonding is developed by the mechanical action.

This refining brings about fundamental changes in the pulp fibers and increases the degree of interfiber bonding in the final sheet of paper. Thus, the final properties of the paper can be significantly changed by varying the degree and type of refining. As additional refining is performed, properties such as tensile strength, fold, and density are increased while tear resistance, opacity, thickness, and dimensional stability are decreased. Thus, the proper refining conditions must be selected to bring out the desired properties without detracting too much from other properties.

The furnish of a paper machine varies widely, depending on the grade of paper being made. Newsprint usually consists of about 80 percent groundwood and 20 percent chemical fiber (sulfite or semibleached kraft). Bag papers and linerboard are usually 100 percent unbleached softwood kraft. Printing papers are made from bleached pulps and contain both hardwoods and softwoods. By selecting the proper pulps and refining conditions, a wide variety of paper qualities can be achieved.

The paper industry is a large user of chemicals because relatively small quantities of additives can materially change the properties of paper. The use of 12 percent rosin size and 23 percent alum [Al2(SO4)3] will greatly increase the resistance of paper to penetration by water or ink. Pigments such as kaolin clay, calcium carbonate, and titanium dioxide are added in amounts up to 15 percent to increase opacity and give a better printing surface. Organic dyes and colored pigments are added to produce the highly colored papers used for business and printing papers. Other additives such as wet-strength resins, retention aids, and starch can be used to give particular properties that are needed. Thus, in order to produce the wide variety of grades of paper now available, the papermaker selects the proper pulps, refining conditions, and additives, and then combines the pulp and additives before sending them on to the paper machine for the final step in the process.

Papermaking Process

Some paper mills are not integrated with pulp mills, and it is necessary for these mills to use dried, baled pulp manufactured at a separate location. Many mills making limited quantities of highly specialized papers fall into this category, as it allows maximum flexibility in selecting the optimum pulps for a particular paper grade. However, the papermaking process is the same regardless of the source of pulp.

After the furnish has been prepared with the proper refining treatment and additives, it is stored in the machine chest and then fed continuously into the paper machine system. A refiner or jordan is placed in this line to give the paper machine operators the opportunity to make small adjustments in the quality of the furnish as needed to give the desired paper properties. Screens and centrifugal cleaners also are included to ensure a high-quality paper.

The papermaking process is essentially a system whereby the pulp is diluted to a very low consistency (about 0.5%) and continuously formed into a sheet of paper at high speeds, and then the water is removed by filtration, pressing, and drying. The basic units of the fourdrinier paper machine are diagrammed in Fig. 7.15; a picture is shown in Fig. 7.16.

Fig. 7.15 Diagram of a fourdrinier paper machine. (*Courtesy Hammermill Paper Co.*)

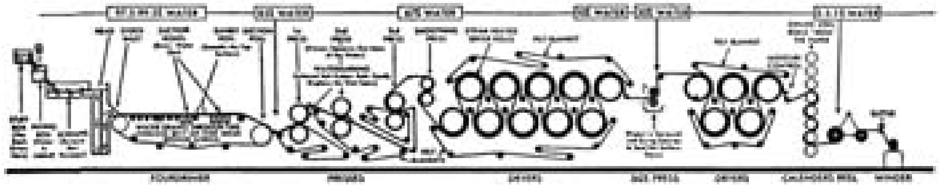
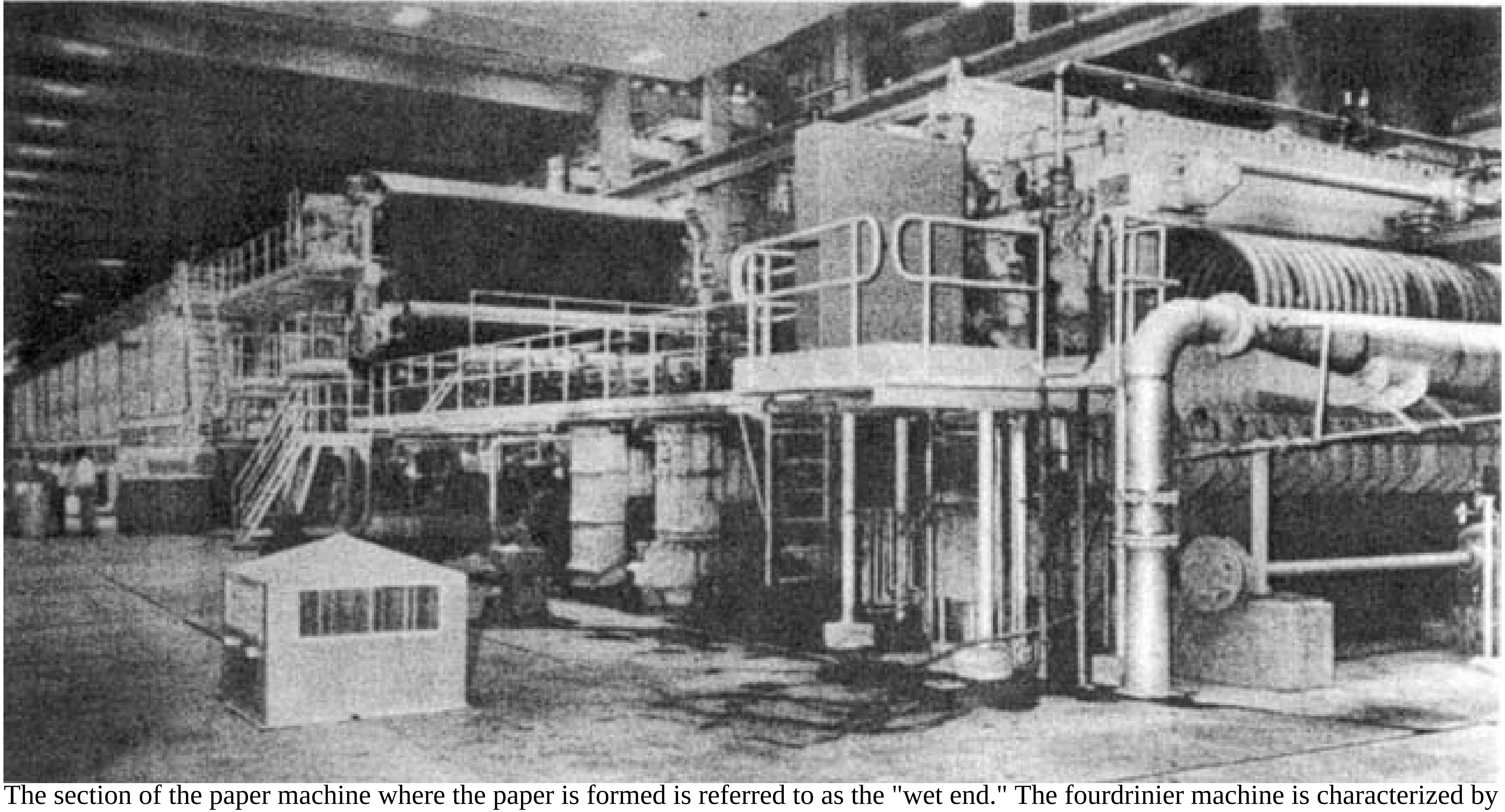


Fig. 7.16 Modern high-speed paper machine. (*Courtesy Beloit Corp.*)



The section of the paper machine where the paper is formed is referred to as the "wet end." The fourdrinier machine is characterized by a headbox that allows the diluted stock to flow through an orifice (slice) onto the flat moving wire. This is actually an endless wire belt that returns on the underside of the machine, thus allowing the process to proceed continuously. Because a low headbox consistency (about 0.5%) is necessary for good formation, the volumes of water handled are very large (about 10,000 gal/min for a machine producing 300 tons/day). Much of the water is removed through the wire by the action of the table rolls and foils that support the wires in the forming area immediately following the headbox. At this point the stock consistency has been raised to about 2 percent, and the stock no longer drains freely. By passing it over suction boxes operating at fairly high vacuum (68 in. Hg), the consistency is raised to about 1520 percent. A dandy roll (covered with woven wire) rotating on top of the wet paper is used to improve the formation and can impact a watermark if it contains the proper pattern. The suction roll after the suction boxes has several functions: it removes additional water; it serves as the driving roll for the wire; and it serves as the point at which the paper separates from the wire and passes into the press section while the wire returns to pick up additional pulp at the headbox.

Modern fourdrinier machines are available up to a width of 360 in. and can operate at speeds up to 3000 ft/min. Newsprint machines are usually the widest and fastest, whereas those making heavier grades such as linerboard run somewhat more slowly. In the United States, there are several machines capable of producing over 1000 tons/day of linerboard. More specialized grades such as bond and printing papers are usually produced at a lower speed on narrower machines, and 150250 tons/day would be considered a high output. Many specialized grades such as filter paper and tracing paper are produced on very small, slow machines producing only a few tons per day. Machines making tissue paper for sanitary purposes use modifications of the standard fourdrinier to produce tissue at speeds of up to 5000 ft/min. Because of the light weight of this paper, it is necessary to make many modifications in the equipment shown in Fig. 7.15.

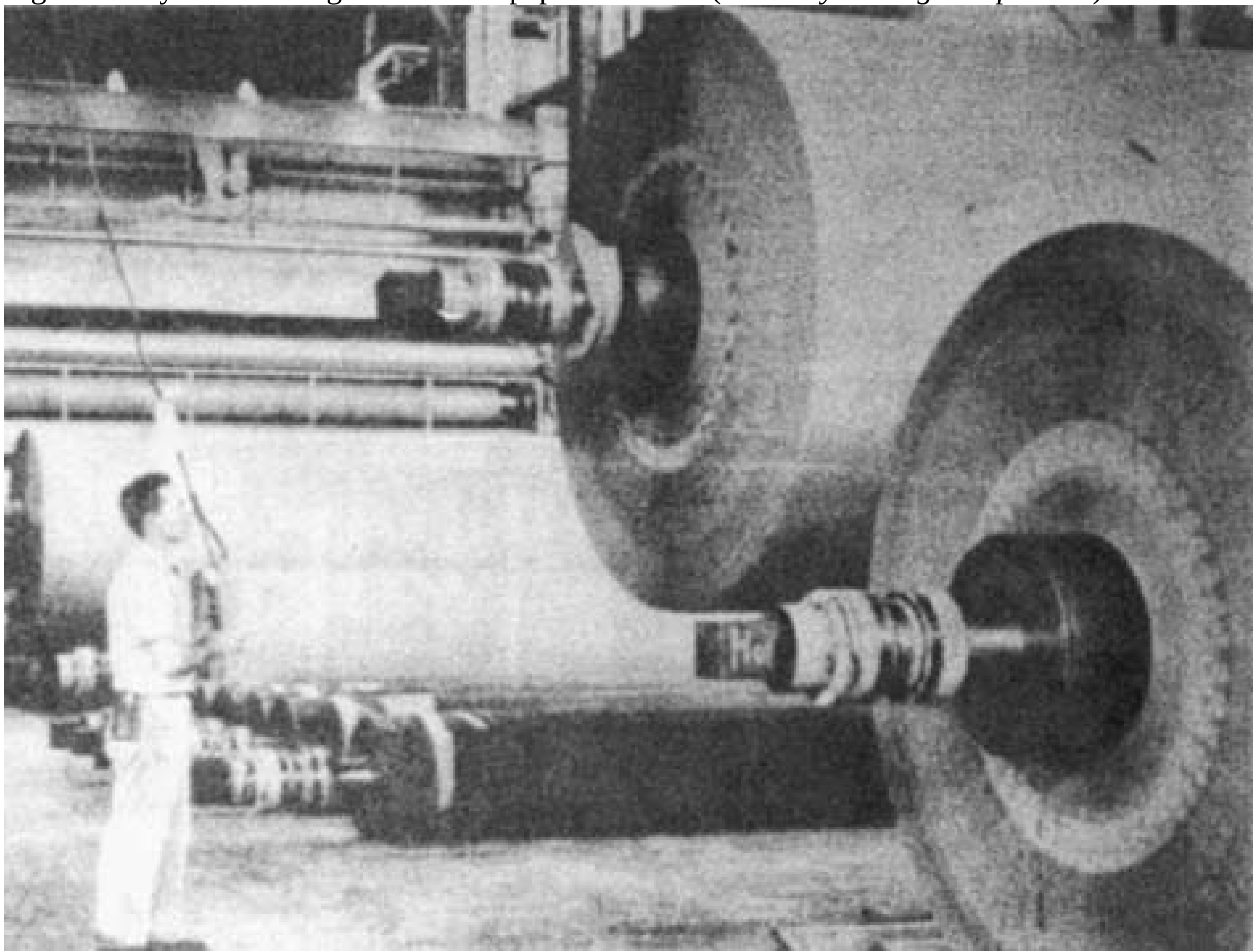
Many new methods of forming a sheet of paper are being developed. The most common concept is to introduce the pulp slurry between two converging wires and remove the water from both sides. This is termed "twin-wire" forming, and many variations are now in operation. This technique has many advantages and may eventually replace the fourdrinier.

A second method of forming paper uses the cylinder machine. Actually these machines are used primarily to form the multi-ply board used in packaging such as cereal boxes. The cylinder wet end consists of one or more (up to eight) cylinder vats, each of which forms a separate wet web of fibers. Each vat contains a woven wire-covered cylinder rotating in the diluted pulp slurry. The liquid head on the outside of the cylinder is greater than that on the inside, resulting in a flow of water through the wire and out of the vat. The pulp mat that is formed on the face of the cylinder is removed by an endless woolen felt that moves in contact with the cylinder by means of a rubber roll riding on top of the felt. With machines having more than one cylinder, the same felt moves from one cylinder to the next, and the wet sheets from each cylinder are laminated to each other on the bottom side of the felt. Thus, very heavy papers or boards can be fabricated by multi-vat machines. Because each vat can be supplied with a different type of pulp, it is possible to make boards with a white surface of high-quality pulp and a center of low-cost pulp from waste newspaper or other cheap grades. Because of hydrodynamic problems, cylinder-machine speeds are limited to about 250 ft/min and widths of about 150 in. However, because of the heavyweight board produced and the widespread use of cheap waste paper for most of the furnish, the cylinder machine is widely used. New forming units looking like miniature fourdrinier units (Ultra-former, Inverformer) have been developed and are rapidly replacing the old-fashioned cylinder vat because their speed is not so limited.

From the wet end of the machine, the wet sheet is conveyed by woolen felts through a series of roll-type presses for further water removal, increasing the consistency to about 35 percent. The sheet then is threaded through the dryer section, consisting of a long series of steam-heated cast iron cylinders that reduce the moisture content to approximately 5 percent, which is about the equilibrium moisture content for cellulosic fibers at 4050 percent relative humidity. Tissue machines use one large dryer (called a Yankee dryer), ranging from 8 to 18 ft in diameter. Because of the light weight of the tissue paper, it can be dried at high speeds on a single dryer.

After drying, the paper is compacted and smoothed by passing through a calender stack consisting of a vertical row of highly polished cast iron rolls. Then the paper is wound into rolls on the reel, as shown in Fig. 7.17.

Fig. 7.17 Dry end of a large fourdrinier paper machine. (*Courtesy St. Regis Paper Co.*)



Finishing and Converting

The objectives in the final stage of the total papermaking system are to improve the paper surface, to reduce rolls and sheets in size, to modify paper for special properties, such as coat or emboss, to convert to finished products, such as bags and corrugated boxes, and to package for shipping.

Corrugated boxboard is made by gluing sheets of linerboard to each side of a fluted sheet of the corrugating medium. Papers for packing may be laminated to polyethylene film or aluminum foil, or coated with waxes and hot-melt resins. The printing and bag- and box-making industries depend on the production of many mills that produce the several hundred grades of paper used in the United States, and each user may require special paper characteristics to match its process.

The quality of many papers is improved by a surface treatment. A size press about two-thirds of the way along the dryer section can apply a solution of starch to improve surface bonding. Also, paper generally is coated to improve printing properties. A surface coating of a pigment (usually kaolin or china clay, calcium carbonate, or titanium dioxide) and an adhesive (such as starch, casein, and others) is applied to the partially dried web by brush, blade, spray, or other method, and dried during the papermaking (on-machine) operation or in a separate operation. The paper surface is brought to a high finish by passage through the calenders referred to above or through supercalenders. Supercalenders are stacks of alternate steel and densified fiber rolls that create a rubbing action on the sheet, imparting an extra-high gloss to the sheet surface.

Environmental Protection

The manufacture of pulp and paper is a chemical-process industry and produces air emissions, effluents, and solid and toxic wastes that are potential hazards. The paper industry uses large volumes of water as a fiber carrier and as a solvent. An increasing volume of water is being recycled, but makeup water still is required to cover losses. A bleached-pulp and paper mill may use 100 m3 (26,400 gal) of fresh water per metric ton of product and 50,000 m3 (13.2 million gal) daily for a plant producing 500 metric tons of products. In addition to this aqueous effluent that the mill must clean up, it must contend with polluted air and solid and toxic wastes. The nature of these emanations, their sources, and their treatments are summarized in Table 7.7. In this connection the paper industry generally has been in good compliance with government environmental regulations although at considerable nonreimbursed capital expense, which amounts to about 10 percent of the cost of the mill.

TABLE 7.7 Summary of Pollution from Pulp and Paper Mill Operations

Type of Pollutant	Mill Operation	Treatment
Effluents		
Suspended solids (SS): fiber fragments, inerts, clay	Papermaking	Primaryclarification

Pulping residuals	Pulping	Secondarybiological treatment and clarification
Air emissions		
Total reduced sulfur (TRS)	Kraft liquor recovery	Oxidation, precipitation, scrubbing, incineration
Sulfur dioxide	Sulfite pulping	
Particulates	Steam generation	Precipitation, scrubbing
Solid wastes	Effluent treatment	Landfill, utilization, incineration
Toxic wastes: chlorinated compounds	Bleaching	Lime pretreatment, oxidation, biological

Using a revolving cylinder or other equipment (save-alls), in the 1930s mills recovered for reuse fibers and clay from the paper machine water (white water) system. In the 1970s, procedures for removal of the fibers and clay from the paper mill effluent were incorporated through settling or clarification or primary effluent treatment. About the same time, secondary effluent treatment (biochemical treatment) of the pulp mill effluent was necessary to remove pulping residuals. The purpose of this treatment is to reduce the biological oxygen demand (BOD) of the effluent, which, if untreated, reduces the oxygen content of the stream to a level incapable of supporting aquatic life. The most common method uses microorganisms that react with the wood sugars and other oxygen-consuming compounds in the spent liquors; this is called the activated-sludge method. The products of primary and secondary treatments are sludges, the handling of which is discussed later under solid wastes.

Two objectionable air emissions have characterized pulp mills for years: the sulfur dioxide of the sulfite pulping mill and the |malodorous reduced sulfur compounds (TRS) (mercaptans and hydrogen sulfide) of the kraft mill. Still another less noxious air emission is the particulate matter from steam boilers. Coal-burning boilers also emit sulfur dioxide, as is well known. The treatments for the particulate emissions are shown in Table 7.7.

Solid wastes represent the ultimate in mill residues and include the accumulated refuse of the mill and the sludges from primary and secondary effluent treatment. There is difficulty in removing water from the secondary sludge; the primary and secondary sludges often are mixed to aid in water removal, which is important if the sludge is to be incinerated for disposal. The sludges from pulp and paper mills are handled mostly as landfill, and sometimes, if not toxic, they are spread for agricultural purposes.

Most mill solids are slightly toxic, predominantly from chlorination compounds in the wash waters from bleaching. This toxicity can be reduced with lime pretreatment and biological treatment. Toxicity has been the main concern of governmental regulating bodies in recent years.

7.5 Board and Structural Materials

Board, sheets, panels, and other structural materials are manufactured from wood fibers and various other vegetable fibers, from wood particles, and from paper. The industries making these products are not generally classified as chemical industries; nevertheless, they are closely related to the chemical industry. Fiberboard manufacture is similar to paper-making; particleboards and paper laminates involve the use of synthetic resins and, therefore, chemical technology.

Fiberboard

Fiberboard is the name for rigid or semi-rigid sheet materials of widely varying densities and thicknesses manufactured from wood or other vegetable fibers. The board is formed by the felting of the fibers from a water slurry or an air suspension to produce a mat. Bonding agents may be incorporated to increase the strength, and other materials may be added to give special properties, such as resistance to moisture, fire, or decay.

Fiberboards are manufactured primarily for panels, insulation, and cover materials in buildings and other structures where flat sheets of moderate strength and/or insulating capacity are required. They also are used as components in doors, cupboards, cabinets, furniture, and millwork.

The classification of fiberboards is best done on the basis of density, as there is a great deal of overlap when classifying by use only. Table 7.8 shows the density classification of fiberboards as well as some of their major uses. Fiberboard factories equalize the mat of loose fibers between screens and rollers, whereas hardboard is press-bonded between hot plates to make it dense and strong. The dividing line between the two types of panels lies at a density of 0.5 g/cm3. Both are manufactured for many specific uses, and they vary accordingly.

TABLE 7.8 Classification and Uses of Fiberboards[a]

	Density Classification		
Fiberboards	(g/cm3)	(lb/ft3)	Major Uses
Noncompressed (insulation board) semi-rigid insulation	0.020.15	1.259.5	Heat insulation as blankets and batts; industrial cushioning
Rigid insulation board (includes wallboard and softboard)	0.150.40	9.525.0	Heat and sound insulation as sheathing, interior panelling, base for plaster or siding, thick laminated sheets for structural decking, cores for doors and partitions, acoustical ceilings
Compressed	0.400.80	2550	Structural use and heat insulation as sheathing base for plaster and siding, interior panelling, containers, underflooring
Intermediate or medium density fiberboard (includes laminated paperboards and homogeneous boards)			
Hardboard	0.801.20	5075	Panelling, counter tops, components in doors, cabinets, cupboards, furniture, containers, and millwork, concrete forms, flooring
Densified hardboard (superhardboard)	1.201.45	7590	Electrical instrument panels, templets, jigs, die stock

[a] From information in *Fiberboard and Particleboard*, Food and Agriculture Organization of the United Nations, Rome, 1958.

The production of fiberboards goes back to 1898 when the first plant was built in Great Britain. However, large-scale production, mainly of insulation board, developed in the United States between the two world wars. The United States is still the largest-producing country and accounts for about one-third of the world output. There has been a much more rapid increase in the production of compressed fiberboards (hardboards) than noncompressed fiberboards (insulation board) during recent years. In 1950 consumption of hardboard was, by weight, roughly 50 percent of insulation board; by 1976 the percentages were almost exactly reversed. Although the total use of insulation board has increased over the past 25 years by 73 percent, per capita consumption has remained constant or even declined, but for hardboard it has doubled. These figures indicate the faster growth rate of the hardboard industry, brought about because favorable prices have stimulated substitution of hardboard for plywood and lumber in construction and furniture. About 85 percent of all insulation board is used in construction, but because it is a nonstructural material, it is not readily substituted for either plywood or hardboard.24

Wood is the principal raw material for the manufacture of fiberboards. The species used are numerous, including both softwoods and hardwoods, and vary from region to region. The wood may come from the harvesting of commercial timber, from species not commonly used for lumber or pulp, or from cull timber, logging and forest management residues, and industrial wood residues. Other fiber raw materials for fiberboard manufacture are bagasse (sugarcane residue after sugar extraction) and waste paper. Only minor amounts of other plant fibers are used.

Wood handling and preparation for fiberboard manufacture is much the same as described for pulp and paper. Wood is debarked and chipped with the same type of equipment. If the chips are to be extracted first for rosin or tannin, then cylinder or drum-type chippers may be used instead of disc chippers.

Fibers for fiberboard are coarser and less refined chemically than those used for paper. Processes are used that bring about fiber separation with a minimum loss of chemical components and a maximum yield. The pulping processes used are generally the following: mechanical, thermomechanical, semichemical, and explosion methods. The pulping processes are essentially the same as those described for making paper pulp.

A unique explosion process for defibering wood was developed by W. H. Mason. Wood chips, about $\frac{3}{4}$ in. long, prepared in conventional chippers and screened, are subjected to high pressure, in a cylinder, commonly called a gun, about 2 × 6 ft in size, and ejected through a quick-opening valve. The elevated temperature softens the chips, and, upon ejection, they explode into a fluffy mass of fibers and fiber bundles. The process involves thermal plasticization of the lignin, partial hydrolysis, and disintegration by the sudden expansion of the steam within the chip.

About 260 lb of wood chips are fed into the cylinder and steamed to 600 psi for 3060 sec. Then the pressure is quickly raised to 1000 psi (about 285°C) and held only about 5 sec before sudden release of the charge into a cyclone. The time of treatment at this high pressure and temperature is critical and depends upon the species of wood, chip size, moisture content, and quality of product desired. The steam is condensed in the cyclone, and the exploded fiber falls into a stock chest where it is mixed with water and pumped through washers, refiners, and screens.

The high temperature to which the chips have been subjected causes appreciable hydrolysis of the hemicelluloses, resulting in a somewhat lower yield of pulp than is obtained by mechanical or thermomechanical pulping. The hydrolysis results in a final board product with an enriched lignin content, about 38 percent compared with about 26 percent in the original softwood. The lignin content of the pulp can be varied by controlling the steaming process.

Board Formation

Pulp prepared by any of the above processes may be used for making insulation board. Mechanical (groundwood) pulp was the first type of pulp used in the large-scale production of insulation boards, and is still being used in many plants. Pulps from other sources, such as disc mills, may be admixed with it. Groundwood pulp is not considered satisfactory for hardboard, and most hardboard is made from pulp prepared by the explosion process or by defibering with disc refiners.

Board-making is basically similar to papermaking and involves refining, screening, mixing of additives, sheet forming, and drying operations. Pressing also is required for hardboards. The pulp is refined and screened prior to sheet formation.

Sizing agents in amounts up to 1 percent of the fiber are added to the pulp in mixing chests. Paraffin wax emulsion is commonly used for all types of boards. For insulation boards, rosin, cumarone resin, and asphalt also are used. Often a mixture of rosin and paraffin emulsion is used, with 1025 percent rosin in the mixture. For hardboards, paraffin wax is the most common sizing agent, although tall oil derivatives and phenolformaldehyde resins also are used. The sizing agent is precipitated on the fibers by alum, with careful control of pH; the latter may be between 4.0 and 6.5, according to the conditions.

The strength properties of a fiberboard depend mainly upon the felting characteristics of the individual fibers and, to a lesser degree, upon their interfiber bonding. The felting or forming process usually is done from a water suspension of the fiber at a consistency of around 1 percent. This is the wet-felting process. An air-felting or dry-forming process is used in a few U.S. plants for hardboard manufacture.

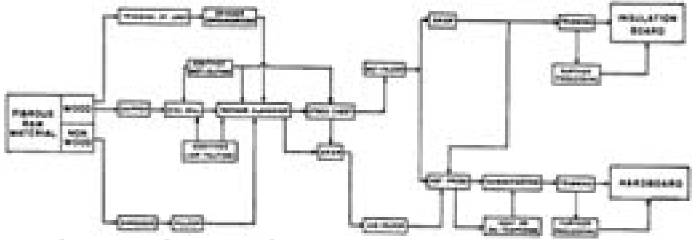
The wet-felting process generally is carried out in a manner similar to papermaking, that is, in a continuous operation on a fourdrinier machine or on a cylinder machine. The machines move more slowly than in the case of papermaking (545 ft/min on the fourdrinier), and a coarser mesh wire is employed. In the cylinder-machine method, a single large vacuum cylinder, 814 ft in diameter, or two cylinders counter-rotating and forming a two-ply sheet, are most commonly used. Further water removal is effected by section boxes and press rolls. The wet sheet is cut to length and conveyed on rollers from the press sections to a tunnel-type dryer. Then the dried sheets are cut into desired sizes.

A third type of wet-felting is a discontinuous method, known as the deckle-box method. The deckle box consists of a bottomless frame that can be raised or lowered onto a wire screen. A measured quantity of stock sufficient to form one sheet is pumped into the deckle box, and vacuum is applied to the lower side of the screen. After most of the water has drained off, pressure is applied from the top to express more water and compact the sheet, reducing its thickness. The deckleframe then is raised and the sheet conveyed to the dryers.

For the recent air-felting process, defibering usually is done in disc mills with control of the moisture content to give the minimum amount possible, consistent with good defibering conditions. Additional moisture may be removed by preheating the air that conveys the fiber from the refiners to the cyclone. The fiber may be further dried in a tunnel or other type of dryer. Fines are removed by either air classification or screens after the dryer. Wax for sizing is introduced either with the chips or added as a spray before or after passing the disc mills (about 2.5% of the weight of the fiber). Sometimes 0.55 percent of phenolic resin is added, depending upon the quality of board desired. The fiber-blend is fed to a moving screen by a metering unit through a combined air and mechanical action. The fibers felt as they fall on the screen, and the fiber mat thus formed is precompressed between belts and/or rollers. If the board is to be wet-pressed, water is added by spraying.

After the felting or sheet-forming operation, the subsequent operations differ for insulation board and hardboard (see Fig. 7.18). For insulation board, the sheets are dried without further compression, whereas for hardboard, the sheets are either pressed and dried simultaneously (wet-pressing) or are first dried and then pressed (dry-pressing). Air-felted sheets are pressed directly after forming.

Fig. 7.18 Schematic outline for insulation board and hardboard manufacture. (*From* "*Fiberboard and Particleboard*," *Food and Agricultural Organization of the United Nations, Rome, 1958*.)



Drying and Pressing

The wet-felted sheets for insulation board or for dry-pressed hardboard, containing 5080 percent water, may be dried by any of three methods: (1) tunnel kilns using racks or carts to support the sheets; (2) steam-platen dryers; or (3) continuous roller dryers of single or multideck arrangement. Most widely used is the continuous roller-type multideck dryer, which has an average length of 150300 feet but may be more than 600 ft long. An average dryer will have eight decks and be 12 ft wide.

The pressing conditions greatly affect the board properties. The conditions of time, temperature, pressure, and moisture content will depend upon the fiber in the board and the product desired. In wet pressing, a typical cycle has three phases and lasts from 6 to 15 min. The first is a short high-pressure stage (up to 710 psi) to remove most of the free water and bring the board to the desired thickness; the second step serves to remove water vapor and requires most of the time; the third stage is a final short period at high pressure to effect a final "cure" or bonding by plastic flow of the lignin. To secure this fiber-to-fiber bond, a temperature of 185°C must be attained, and temperatures up to 210°C may be used to increase the production rate. In the dry-pressing process the cycles are shorter (1.53.5 min), and the temperatures and pressures usually are higher.

Conditioning

After the hardboard has been hot-pressed or has been heat-treated or oil-tempered, the moisture content is well below what will be attained at equilibrium with the atmosphere in normal use. Such very dry boards will change dimensions upon picking up moisture and may warp. It is important, therefore, to humidify the boards under controlled conditions before packaging. The desired equilibrium moisture content (EMC) reached will vary from 5 to 12 percent, depending upon the nature of the board and the general humidity conditions in the region of use.

Most humidifying is done in the chambers or tunnels, kept at 8085 percent R.H. and 3850°C. A lesser-used system of conditioning utilizes water spraying and dipping followed by standing to allow uniform absorption.

Special Treatments

Hardboards often are given a special treatment to improve their strength and resistance to moisture. The heat treatment method has come into wide use; the boards, which are kept apart to permit hot air circulation, are heated in chambers by either batch, continuous, or progressive systems. Typical heating conditions are 5 hr at 155160°C. The heating increases the strength (except impact), sometimes as much as 25 percent, and the water resistance is improved. This operation may replace sizing wholly or in part. Some exothermic reaction takes place in the board, and the heat developed must be removed by the hot circulating air to prevent burning. Probably some chemical condensations occur in the wood fiber, producing an internal resin system, and there is the possibility of some crosslinking of large molecules.

Some hardboard is oil-tempered. A drying oil, such as linseed, tung, perilla, soya, or tall oil, or an alkyd resin, is impregnated into the board, by passing the hot-pressed board through a hot oil bath. About 48 percent of the oil is absorbed. The board then is heated in a kiln with circulating air at 160170°C for 69 hr. This treatment hardens the oil as well as bringing about chemical reactions in the fiber, and results in greater strength and moisture resistance.

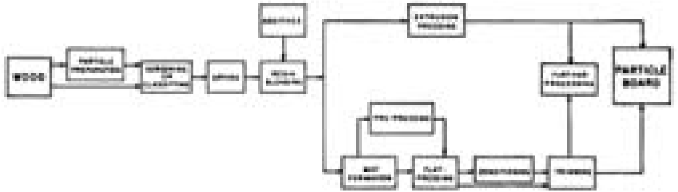
Various additives may be incorporated into insulation boards and hardboards, or they may be surface-treated to bring about resistance to decay, insects, and fire. Pentachlorophenol and copper pentachlorophenate are commonly used for preservative treatments. The sodium salt of pentachlorophenol is added before sizing and is precipitated onto the fiber along with the size by the alum. Special fire-retarding-paint coatings sometimes are used to give resistance to the spread of flame.

Particleboard and Oriented Strandboard (OSB)

Although wood particleboard was patented in the United States as early as 1905, commercial development of the product usually is credited to either Germany or Czechoslovakia in 1941. Production was restricted by a shortage of gluing resins during most of World War II, but in 1945 the first U.S. plant began operation. Particleboard did not really begin to gain market acceptance until the 1960s. From 1964 to 1979 consumption increased at an average annual rate of 9.5 percent. Total consumption of particleboard was about 2.5 million m3 in 2000.

Particleboards are composed of discrete particles of wood bonded together by a synthetic resin adhesive, most commonly urea-formaldehyde or phenol-formaldehyde. The material is consolidated and the resin cured under heat and pressure. The strength of the product depends mainly upon the adhesive and not upon fiber felting as in the case of fiberboards, although the size and shape of the particles influence strength properties. They may be fine slivers, coarse slivers, planar shavings, shreds, or flakes, and they are divided into two main groups: (1) hammer-mill-produced particles (slivers and splinters from solid wood residues, feather-like wisps to block-shaped pieces from planar shavings) and (2) cutter-type particles, sometimes termed "engineered" particles (flakes and shreds). The various steps in particleboard manufacture are illustrated in Fig. 7.19.

Fig. 7.19 Schematic outline for particleboard manufacture. (*From* "*Fiberboard and Particleboard*," *Food and Agricultural Organization of the United Nations, Rome, 1958*.)



Hammer-milled particles usually vary appreciably in size. Dry raw material produces greater amounts of fines than green wood. Cutting machines (either cylinder-type or rotating-disc-type) give more uniform particles, with the length dimension in the direction of the grain of the wood. The thickness, size, and shape of particles influence the strength of the board. Boards made from sawdust have the lowest strength properties, hammer-milled particles give boards of intermediate strength, and solid wood, cut to flakes, gives boards of highest strength. The latter often are referred to as waferboards.

In another relatively new type of particleboard, the manufacturers align long strands of wood at the surface for increased strength in the direction of panel length. This oriented-strandboard (OSB) and waferboard replaced a great deal of plywood sheathing in the early 1980s.

Particleboards may be made in a wide range of densities. Low-density or insulating types are a comparatively recent development in Central Europe, whereas the high-density-hardboard types are a U.S. development. Most particleboard production is in the middle-density range.

Particleboards most commonly are used as core stock for veneer in furniture and in doors, as interior panels for walls and ceilings, as subflooring, as sheathing and siding, and as components in interior millwork. The dense types are used in the same way as fiberboard hardboard, described above. Both dense particleboards and hardboards, after receiving a surface coating, may be printed with decorative designs.

Particleboard production has increased rapidly, both in the United States and worldwide. Until recently the most significant limitation on the market for particleboard was the availability of inexpensive plywood. Now plywood is no longer inexpensive relative to particleboard, and the cost factor would seem to favor continued growth of the particleboard industry.

OSB is a newer product, being less than 20 years old. However, because small trees and crooked stems can be effectively utilized the raw material cost, compared to product cost, is about 38 percent. Plywood requires more than double the raw material cost. Lumber requires even more. As a result, OSB has rapidly gained acceptance in the construction market. OSB plant capacities exceed 20 million cubic meters annually; more new plants are being constructed each year.1

The typical material flow in an OSB plant provides very high conversion efficiencies. Any "waste" is used for fuel for the manufacturing process. Logs are debarked, heated in hot water, cut into short lengths, flaked into wafers or strands using knives. The strands are dried in large drum driers, coated with adhesive, and assembled into a thick mat. The strands in the mat are oriented to provide parallel strands in the various layers in the mat. The mat is compressed and heated, and cut to size. By controlling the panel density, the amount of adhesive, and the strand orientation, a panel can be designed and produced with properties that meet the engineering needs of the building designer and engineer.1

If lumber were used to do to the same job as OSB, an excess of 8 billion more board feet of lumber would be needed annually, and this lumber would have to come from larger, higher quality, more expensive trees. It is easy to understand why OSB is considered as the wood product of future in the construction business. Veneer lumber, parallel strand lumber, and laminated strand lumber are also rapidly moving into the market as engineered wood products.25,26

Paper-Base Laminates

Paper-base laminates are panels or other laminated assemblies composed of many plies of resin-treated paper molded together under high temperatures and pressure to produce rigid structures that no longer have the characteristics of paper. These products are widely used in the electrical and machine industries for insulators, gears, pulleys, and a multitude of machine parts. They possess high impact strength and toughness, good electrical insulation, and high dimensional stability, and are not subject to corrosion; they also have a dampening effect on sound, eliminating rattling and drumming in steel cars and machinery. Furthermore, they easily can be manipulated into complex shapes and can be drilled, turned, and sawed. With these properties, these products are of great industrial value. They also are used in making trays, light flooring panels, table and counter tops, and many other products employing panels.

Paper-base laminates fall into four major classes: (1) mechanical or structural, (2) electrical, (3) punching, and (4) decorative. Phenolic resins are especially suitable and mostly are used where mechanical strength and resistance to heat, water, and electricity are required. For punching-grade laminates, phenolic resins are specially modified with plasticizers or drying oils to yield laminates having good plasticity and elasticity. For electrical grades the phenolics generally are catalyzed with ammonia, amines, or less conductive catalysts in place of the stronger alkaline catalysts otherwise used.

For decorative laminates, urea and melamine resins are the principal resins used. The melamines are used where translucent, light-colored products with good heat and water resistance are required. Polyesters and some melamines are used for low-pressure laminating, enabling the continuous production of counter and table tops by passing the assembly through a set of rolls and then through a heating chamber.

There are two broad general methods for introducing resins into papers for making paper-base laminates: (1) the beater addition process and (2) impregnation of the paper sheet. In the beater addition process the resin is added to the pulp in the beater and then precipitated on the fiber by alum or an acid. The resin adheres to the pulp fibers, which then are formed into a sheet. This type of paperplastic combination commonly is called resin-filled paper. These resin-filled papers may be made into flat or shaped preforms. The former are flat sheets; the latter are vacuum-felted to a shape closely conforming to that of the final molded product. Shaped preforms are used for deep forms requiring high-strength contours. Little flow is required on molding; hence the paper sheet does not break, and good strength properties are obtained. The use of stretchable cross-creped paper for postformable laminates is a new development.

The more common method of application of the resin is by impregnation of the wet or dry paper sheet (web) with a solution of resin. Such papers are called resin-impregnated papers. Both water-soluble and alcoholsoluble types of phenolic resins are used for paper-base laminates. The former tend to give more brittle but more dimensionally stable products than the latter. Phenolic resins are applied mostly by impregnation of the wet or dry paper sheet, producing resin-impregnated papers, although some progress has been made in the slush stock addition procedures to produce phenolic resin-filled papers.

In the wet-web process for impregnating paper for laminates, the wet sheet of paper on the paper machine is carried through a resin bath while the sheet is supported by a wire. The sheet contains up to 65 percent water, and the amount of resin taken up depends upon the moisture content of the paper, solids content of the resin solution, viscosity of the resin solution, temperature of the bath, machine speed, and pressure of the squeeze-rolls. Only water-soluble or water-dispersible resins can be used in the wet-web process.

Generally, however, phenolic impregnations of paper for laminates and other purposes demanding a high resin content are done with dry paper on off-machine equipment. The dry paper sheet is passed continuously through a resin solution, generally an alcoholic solution, and then moves under and between two metering rolls, after which the paper is dried to remove solvent and to complete the condensation of the thermosetting phenolic resin.

The addition of phenolic resins in the form of emulsions to the slush stock by the beater-addition process also has received considerable study. Papers containing 4555 percent resin, and even as high as 65 percent, have been prepared. The resin is precipitated by adding alum to the beater. Papers prepared in this way are highly plastic and are suitable for low-pressure molding. The phenolic resins can also be used in combination with elastomers, such as GR-S types, neoprene, hycar, and vinyl polymers, to produce laminates of high impact strength and greater elongation under tension.

Lignin also is commercially used as a resin for paper laminates. Its low cost and availability as a by-product in pulp manufacture make it attractive. One of the most successful products is made with lignin recovered from the spent liquor of the soda pulping process. The lignin may be added to the pulp suspension in solution and precipitated directly among the fibers, or added in a preprecipitated form.

Properties of paper-base laminates depend upon both the resin and the paper. In general, the final product has the characteristics of the resin used, provided that over 30 percent is resin. The paper acts as a structural reinforcer, greatly enhancing toughness and tensile, flexural, and impact strength.

Electrical properties of the laminate depend upon both the paper and the resin, although the amount of resin absorbed is the most important factor affecting electrical insulating properties. Papers of low power factor, good dielectric strength, and high dielectric constant are necessary for electrical grades of laminates. Other desired paper properties for good paper-base laminates are uniformity, cleanliness (freedom from slime, dirt, and fiber bundles), low finish, neutral pH, freedom from chemicals (bleach residues, etc.), low and uniform moisture content (below 4%), and uniform absorbency. High absorbency is desired when high electrical resistance and minimum water absorption are required, as greater amounts of resin are taken up by high-absorbency papers than by those with a low absorbency. For high-impact-strength punching grades of laminates, a low-absorbency paper is necessary to reduce the amount of resin absorbed.

Polymer-Modified Papers

In addition to paper-base laminates, polymers are combined in many ways with paper to develop new properties or to modify or enhance certain properties. Paper commonly is combined with polymers, such as synthetic resins, elastomers, and plastics in general, to produce products that may be classified as follows: (1) wet-strength paper; (2) special-purpose papers containing large amounts of resins or elastomers (chemical-resistant papers, sandpaper backings, gaskets, imitation leather, shoe parts, wood overlay and honeycomb core papers, air and oil filters, battery separators, etc.); (3) plastic-coated papers; and (4) paper-base laminates. The use of plastics in combination with paper has grown enormously during the past three decades. Large quantities of urea and melamine resins as well as polyethylene are used for coating and heating paper and paperboard.

Many other polymers are used in treating and coating papers, including polyesters, styrene, polyvinyl chloride, their copolymers, and others. Polymer-modified papers thus are very important in the present economy.

Overlaid lumber is a composite of lumber and phenolic resin-treated kraft paper. Similar paper overlays are applied to veneer, plywood, fiberboard, and particleboard. Paper honeycomb cores also are made from phenolic resin-treated kraft paper, which is formed into a honeycomb of different geometrical designs, such as figure of eight or hexagonal, in special machines. Simpler types are made from resin-impregnated corrugated sheets, which can be assembled in several ways. This material permits construction of sandwich panels of light weight and high strength. Kraft papers for wood overlays or for honeycomb cores are treated with phenolic resin in the same ways as described above for paper-base laminates.

Paper overlays have three basic uses: masking, decoration, and structures. Masking overlays are used to cover minor defects and provide a more uniform paintable surface. Such overlays contain 2025 percent phenolic resin based on the weight of the paper. For structural purposes, one or more sheets of paper may be used in the laminate. High-density and medium-density types are produced for plywood. The high-density type contains not less than 40 percent of a thermosetting resin, phenolic, or melamine; it has a hard, smooth surface not requiring further finishing and may be used for exterior service. The curing of the resin is completed at the same time that the paper is bonded to the wood material in a hot press or, in the case of plywood, at the same time that the veneer is assembled into plywood. The overlay sheet swells and shrinks much less than wood and thus exerts a resistance to the dimensional changes of the wood and may reduce lateral swelling by as much as 40 percent. Overlay papers also upgrade the appearance of low-grade lumber, increase strength properties, improve finishability, and increase resistance to weathering.

Decorative surfaces are obtained by applying a top sheet of white paper on which there is a printed design. This is covered with a clear coating of melamine or vinyl resin. Or, a thin transparent (when cured) paper impregnated with melamine resin may be applied to a decorative veneer, providing a permanent protective finish.

For honeycomb cores, either water- or alcohol-soluble phenolic resins may be used. Many types of facings may be glued to the honeycomb cores: veneers, plywood, hardboards, asbestos board, aluminum, stainless steel, and paper-plastic laminates. Thin sheet material may be used because of the almost continuous support of the core.

The honeycomb sandwich possesses great strength in relation to its weight. It may carry loads as much as 25 tons/ft2. Its strength and weight vary with the weight of the paper, quantity of resin impregnated, and honeycomb design.

Honeycomb sandwich construction has many uses, such as in airplanes, cargo containers, truck and trailer bodies, railway passenger cars, cabins, barns, airplane hangars, house floors, walls, roofs, doors, and a variety of other products. Besides combining strength with lightness, honeycomb sandwich material has high rigidity, good insulation properties, resistance to fungi and pests, and durability to temperature extremes.

Modified Wood and Wood Composites

Wood has a number of disadvantageous properties, as described earlier in this chapter. It is biodegradable, flammable, dimensionally unstable to water, and degradable by ultraviolet light, acids, and bases. There has been a considerable amount of research expended over the years to correct these defective properties for both solid wood and reconstituted wood products.

A wood product of improved dimensional stability can be produced through heat treatment of wood at temperatures varying from 150 to 320°C. For good temperature control and to avoid strength loss the heat treatment preferably is performed under the surface of molten metal or a fused salt for exclusion of oxygen. The modified wood product is referred to as staybwood.

Treatment of wood with polyethylene glycol (PEG) is the most commonly used method to impart dimensional stability to wood. PEG in a water solution can be used to permeate the lumens and penetrate the cell wall. A 3050 percent solution of PEG in water generally is used on green wood; no pressure is required. Treatment times are adjusted so that a uniform uptake of 2530 percent of PEG is achieved, based on the dry weight of the wood; temperatures range from 21 to 60°C. The wood may require weeks of soaking, depending on species, density, and thickness. After treatment, the wood is dried in a well-ventilated space, preferably heated.

The PEG is not "fixed" or cured in the wood, and it leaches out if the wood contacts water. For this reason, PEG-treated wood usually is finished with a penetrating oil or polyurethane varnish. PEG is used primarily to treat cross-sections of trees for plaques, clock faces, and tabletops. It also is used to stabilize bowls and other turnings, carvings, and rifle stocks, and to treat archaeological water-logged wood.

Resin impregnation can be used to add bulking agents to the cell wall. These resins have a low degree of polymerization, or the monomers are polymerized after impregnation. The best fiber-penetrating thermosetting resins are highly water-soluble, phenol-formaldehyde, resin-forming systems with low initial molecular weights. Green or dry veneers usually are treated by soaking or pressure treatment with a water-soluble resin in a 30 percent aqueous solution. Following impregnation, the wood is dried slowly and then heated at about 150°C for 20 min to set the resin. Laminates of treated wood are constructed by gluing the individual sheets together. This product, called Impreg, contains 2535 percent resin and has a density about 20 percent higher than the original wood and an anti-shrink efficiency (ASE) of about 65 percent. As with PEG, the watersoluble resin penetrates the cell wall and keeps the wood in a swollen state. Unlike PEG, the phenol-formaldehyde resin is cured by heating to form a high-molecular-weight, water-insoluble resin in the cell wall. Thus, Impreg can be used in contact with water. Its compressive strength is higher than that of untreated wood, but it has much less toughness. It shows considerable resistance to decay and termite and marine-borer attack. Treating with phenolic resins imparts high acid resistance, greatly increases the temperature to which wood can be heated without appreciable change in properties, and increases electrical resistance. The largest industrial application of Impreg is in die molds for automobile body parts and other uses. Despite its high cost, the dimensional stability and plasticity of Impreg make it commercially viable.

Compreg is a highly dense product obtained by applying pressure to dry, resin-treated veneers during heat cure. It offers most of the advantages of Impreg. Its mechanical properties are appreciably better than those of the original uncompressed wood because of a two- to three-fold increase in density (1350 kg/m3). The strength properties of Compreg are increased in proportion to the degree of compression. Only its toughness is lower than that of untreated wood, although greater than that of Impreg. Because of the plasticizing action of the resin-forming materials, treated wood can be compressed under considerably lower pressures than dry, untreated wood. For example, when subjected to a pressure of only 1.7 MPa (244 psi) at 149°C, treated spruce, cottonwood, and aspen wood, dried to a moisture content of about 6 percent but not cured, are compressed to about half the original thickness and a specific gravity of about 1.0.

In a 24-hr water-soaking test, Compreg has an ASE value of 95 percent. The rate of water pickup is so slow that complete swelling equilibrium of a 13-mm specimen is not reached after one year at room temperature. The product is brown and can be buffed to a high polish. Compreg is produced commercially in small quantities and is used for knife handles, gears, certain musical instruments, and decorative articles.

Composites of wood with vinyl polymers have been developed. Because the vinyl polymers are clear, colorless, thermoplastic materials, they do not significantly discolor the wood; thus its natural beauty is retained, whereas the phenolic resins darken the wood. However, the hygroscopic characteristics of the wood substance are unaffected because little, if any, resin penetrates the cell walls; the ASE is only about 1015 percent. The high resin content (70100%) greatly reduces the normally high void volume of wood, however. Thus, the elimination of this important pathway for vapor diffusion slows down the response of the wood substance to changes in relative humidity, and its moisture resistance is greatly improved compared with the original wood. The hardness of the woodplastic composite also is greatly improved.

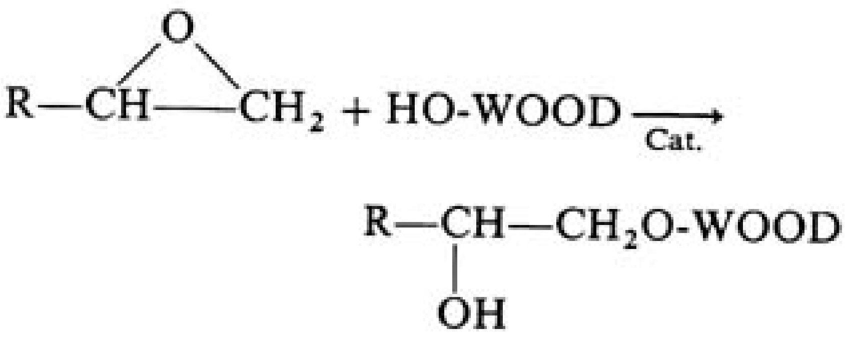
A variety of vinyl monomers, such as methyl methacrylate and styrene, may be used. Complete filling of the cell lumens and other voids (the "full-cell process") is easily accomplished by first subjecting the wood to a partial vacuum (about 0.3 in. of Hg) and then covering it with the monomer and soaking it for 26 hr, depending upon the species of wood and its dimensions. Some penetration of the monomer into the cell walls also may be obtained by using a diffusion process, such as a solvent-exchange method.

Polymerization of the vinyl monomer in the wood may be done with either radiation or free radical catalysts. The polymerization of the vinyl monomers in both processes depends upon the same mechanism, that is, initiation by free radicals. In the radiation process, the gamma rays passing through the monomer and the woody tissue create a large number of excited and ionized molecules, many of which break into fragments, namely organic free radicals (R·). These act as the initiator for the polymerization of an unsaturated monomer. Alternatively, the free radicals may be formed by thermal decomposition of compounds involving a weak bond. Commercially, the catalyst 2,2′-azobisisobutyronitrile now is most widely used, as it forms free radicals at a lower temperature than benzoyl peroxide.27

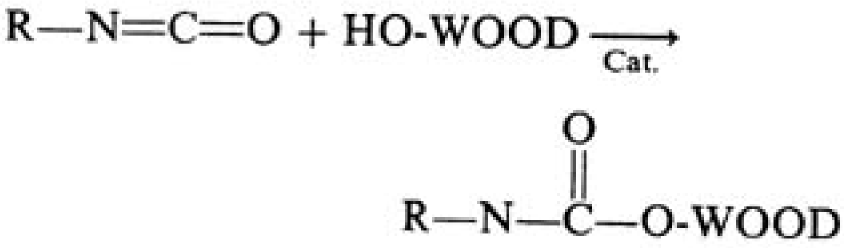
If the end use of the woodpolymer composite requires an abrasive (sanding) or cutting process that brings about high temperatures, the thermoplastic polymer will melt, causing machining difficulties. To prevent such melting, a crosslinking substance such as diethylene glycol dimethacrylate is added to the monomer before impregnation into the wood (about 5% of the volume of the monomer). Woodplastic materials are used in parquet flooring, certain sporting equipment, musical instruments, and decorative articles.

Wood also can be modified by chemical reaction directly with the hydroxyl groups present in the polymer constituents in the cell wall of the wood. This treatment confers bulk to the cell wall with the help of a permanently bonded chemical. Many reactive chemicals have been used to modify wood chemically. For best results, the chemicals should be capable of reacting with the wood hydroxyl groups under neutral or mildly alkaline conditions at temperatures below 120°C. The chemical system should be simple and must be capable of swelling the wood structure to facilitate penetration. The complete molecule must react quickly with wood components, yielding stable chemical bonds; the treated wood still must possess the desirable properties of untreated wood. Anhydrides, epoxides, and isocyanates best conform to these conditions, conferring ASE values of 6075 percent at chemical weight gains of 2030 percent. Reactions of the chemicals with wood, as shown below, are fast and complete, and stable chemical bonds are formed. The systems swell the wood and penetrate well.28

Epoxides:

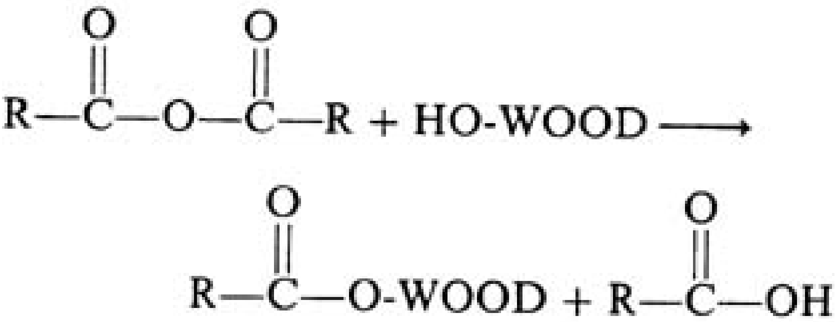


Isocyanates:



$$R{-}N{=}C{=}O + HO\text{-}WOOD \xrightarrow[\text{Cat.}]{} R{-}N{-}\overset{\overset{\displaystyle O}{\|}}{C}{-}O\text{-}WOOD$$

Anhydrides:



$$R{-}\overset{\overset{\displaystyle O}{\|}}{C}{-}O{-}\overset{\overset{\displaystyle O}{\|}}{C}{-}R + HO\text{-}WOOD \longrightarrow R{-}\overset{\overset{\displaystyle O}{\|}}{C}{-}O\text{-}WOOD + R{-}\overset{\overset{\displaystyle O}{\|}}{C}{-}OH$$

In the epoxide reaction, a new hydroxyl group originating from the epoxide is formed; and from this new hydroxyl, a polymer can begin to form. Given the ionic nature of the reaction and the availability of alkoxyl ions in the wood components, the chain length probably is short because of chain transfer.

The optimum reaction conditions for reaction with epoxides are a 95/5 (v/v) mix of epoxide with triethylamine at 120°C and 150 psi nitrogen. Depending on the isocyanate used, reaction conditions of 120°C and 150 psi with no catalyst and 5 percent triethylamine or 535 percent dimethylformamide are optimum.

Of several anhydrides studied, acetic anhydride reacted the most readily. Reactions were carried out by refluxing the wood in a xylene/acetic anhydride solution or with acetic anhydride vapors alone at 120°C. With this system, for each mole of acetate bonded onto the wood a mole of acetic acid is generated as a by-product. Although this by-product generation is a disadvantage of the process, the chemical system does penetrate and react quickly with wood, without a catalyst. It is not so sensitive to moisture as are the epoxide and isocyanate systems.

A very facile dip acetylation procedure was developed for wood chips and fibers. The wood is dried at 105°C and then dipped in acetic anhydride for 1 min and drained. Then the sample is placed in a preheated (120°C) reactor for different lengths of time, depending on the degree of substitution of acetyl groups desired in the sample.29 The swelling of aspen flakeboard made from such chemically modified flakes is dramatically reduced. For example, while unmodified flakeboard swelled by over 60 percent after immersion in water for 5 days, the modified material swelled by only about 822 percent, depending on the extent of chemical modification.

Molded and Formed Wood

Historically wood has been bent and shaped first by steaming, which plasticizes the wood polymers. The new shape then is set in place by cooling and drying in the desired conformation. However, the curvature possible by this method is limited, and the process is time-consuming. Some chair backs and solid hickory and ash skis still are steam-bent.

Anhydrous ammonia also is known to cause temporary platicization of wood. The ammonia swells and plasicizes both the lignin and the cellulose, and the crystalline structure of the cellulose is converted to a different form in the process. To shape the wood, it is immersed in liquid ammonia or treated with gaseous ammonia under pressure until the cell walls have been penetrated and the wood becomes pliable and flexible. In this condition it is easily shaped and formed by hand or mechanically. The ammonia readily vaporizes and evaporates from the wood, so that the wood regains its normal stiffness but retains the new form into which it has been shaped. With this process the wood can be distorted into quite complex shapes without springing back to its original form. Treating plants have been developed on a pilot-plant scale, but the process has not been widely adopted.

There has been a concerted effort over the years to develop molded wood-based materials. The Japanese emphasized chemical modification approaches and made considerable progress towards the development of a thermoplastic wood-based material.30,31 Their approach was mainly based on esterification of wood meals with subsequent thermomolding. As the size of the aliphatic group was increased, the melting temperature of the modified wood under pressure was decreased. It also was found that a very small degree of additional grafting of polystyrene to the esterified wood resulted in a dramatic improvement in the thermoplasticity of the product. Matsuda31 developed a very facile method for the production of moldable wood by simply heating the wood meal with an acid anhydride for 3 hr to temperatures greater than 60°C. The wood meal was readily moldable at high temperatures (180°C) under pressure (570 kg/cm2) for 10 min. The moldability of the esterified wood decreased in the following order: succinic anhydride > maleic anhydride > phthalic anhydride.

Although it is possible to make wood thermoplastic through chemical modification, the handling of the finely divided wood (wood meal) becomes a problem when more complex shapes or deep draws are desired. Therefore, a totally new approach for the creation of molded products based on wood was developed by Hunter Brooks, formerly of the Triad Corp., Detroit, Michigan. The raw material is a dry attrition-milled wood fiber (80%), a polyester staple fiber 12 in. long (10%), and a phenolic resin (10%). The materials are formed into a nonwoven web on a Randowebber machine. The longer polyester fibers are incorporated so that the web can be needle-punched to form a nonwoven fabric with good mechanical integrity. The phenolic resin is applied to the web either as a dry powder or as an aqueous solution. This fiber composite then is easily compression-molded into any of a variety of shapes. For deep-draw compression molding, patterns may be cut into the nonwoven web to allow the fiber mat to drop deep into the mold. Any type of fibrous material, synthetic or natural lignocellulosic, can be utilized in the web; so the process offers considerable flexibility for production of a wide variety of products engineered to specific applications.32

Major advances have been made in recent years in the development of woodthermoplastic composites formed by injection molding. In this approach attrition-milled wood fibers are mixed with, typically, polypropylene in a thermoplastic mixing device such as a Brabender mixer. The material is then injection molded to produce a variety of useful products. Often times, compatibilizers are included in the blend to improve the interaction of the inert polypropylene with the more hydrophilic lignocellulosic materials. The advantage offered by these composites are lower cost and weight and improved strength properties, although impact strength is usually reduced. Many new companies have started up in recent years to produce these useful composite materials.32

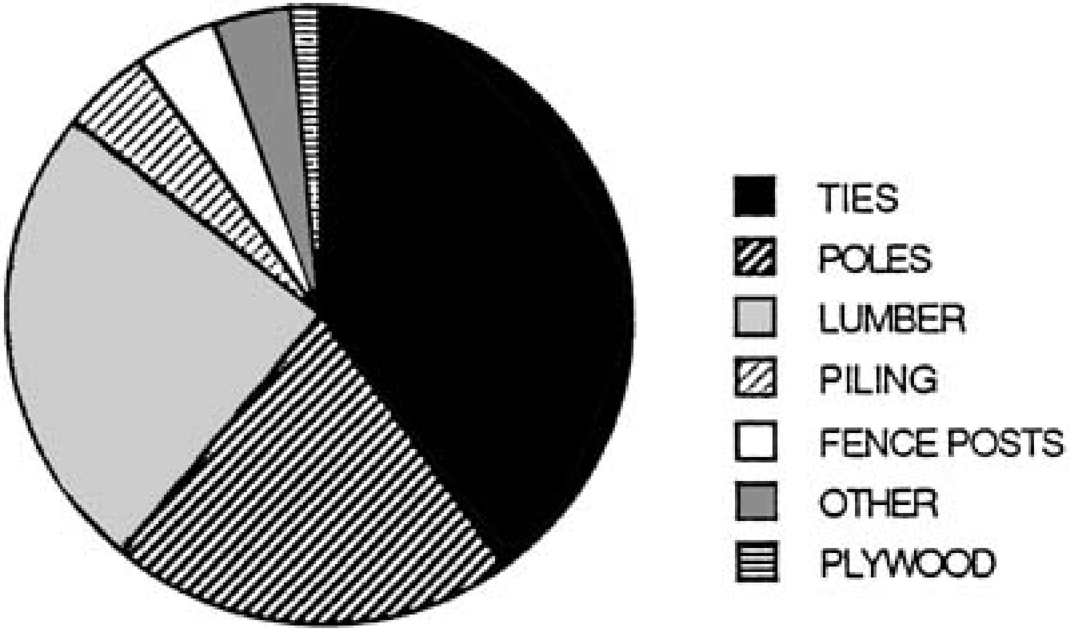
A major application of these types of molded products would be for interior uses in automobiles, such as head liners, door panels, and dashboards. Although this is a low-cost, low-performance application, it represents a very large-volume market. Indeed, wood is already utilized in applications of this type, but as a finely ground flour that serves as a filler (up to 40%) in extrusion-molded polyolefin products. The use of recycled fiber in this process and the one described above offers the potential of even greater cost reductions, combined with alleviation of solid waste disposable problems.

7.6 Preservative Treatment of Wood

Wood, as a natural plant tissue, is subject to attack by fungi, insects, and marine borers. Some species of wood are more resistant to decay than others (e.g. the heartwood of cedars, cypress, and redwood) because of the presence of natural toxic substances among the extractable components. Most woods, however, are rapidly attacked when used in contact with soil or water, or when exposed to high relative humidities without adequate air circulation. Wood for such service conditions requires chemical treatment with toxic chemicals, collectively termed wood preservatives. The service life of wood may be increased 5- to 15-fold, depending upon the conditions of preservative treatment and the nature of the service.

The preservative treatment of wood is the second largest chemical wood-processing industry; pulp and paper manufacture is the most important. The more important types of wood products treated are shown in Fig. 7.20.

Fig. 7.20 Relative distribution of products treated with preservatives.



Preservative Chemicals

Owing to their toxic nature, all the commercial wood preservatives presently used in the United States are effective in preventing attack by microorganisms. However, because of concern that these chemicals will have a deleterious effect on the environment, alternative methods based on nontoxic procedures also are being investigated. These nonconventional approaches are based on chemical modification, and are described briefly at the end of this section.

Toxic chemicals used for the preservation of wood may be classified as follows:

1. *Organic liquids of low volatility and limited water solubility*:
    1. coal-tar creosote
    2. creosotecoal tar solutions
    3. creosotepetroleum solutions
    4. other creosotes
2. *Chemicals dissolved in organic solvents, usually hydrocarbons*:
    1. chlorinated phenols (principally pentachlorophenol)
    2. copper naphthenate
    3. solubilized copper 8-quinolinolate
3. *Water-soluble inorganic salts*:
    1. acid copper chromate
    2. ammoniacal copper arsenite
    3. chromated copper arsenate
    4. chromated zinc chloride
    5. fluor chrome arsenate phenol

Creosote from coal tar is the most widely used wood preservative for several reasons: (1) it is highly toxic to wood-destroying organisms; (2) it has a high degree of permanence due to its relative insolubility in water and its low volatility; (3) it is easily applied, with deep penetration easily obtained; and (4) it is relatively cheap and widely available. For general outdoor service in structural timbers, poles, posts, piling, and mine props, and for marine uses, coal-tar creosote is the best and most important preservative. Because of its odor, dark color, and the fact that creosote-treated wood usually cannot be painted, creosote is unsuitable for finished lumber and for interior use.

Coal-tar creosote is a mixture of aromatic hydrocarbons containing appreciable amounts of tar acids and bases (up to about 5% of each), and has a boiling range between 200 and 355°C. The important hydrocarbons present include fluorene, anthracene, phenanthrene, and some naphthalene. The tar acids are mainly phenols, creosols, xylenols, and naphthols; the tar bases consist of pyridines, quinolines, and acridines.

Often coal tar or petroleum oil is mixed with coal-tar creosote, in amounts up to 50 percent, as a means of lowering preservative costs. Because coal tar and petroleum have a low toxicity, their mixtures with creosote are less toxic than is creosote alone.

A number of phenols, especially chlorinated phenols and certain metalorganic compounds, such as copper naphthenate and phenyl mercury oleate, are effective preservatives. Pentachlorophenol and copper naphthenate are most commonly used, and are carried into the wood in 15 percent solutions in petroleum oil. Pentachlorophenol is colorless, and can be applied in clear volatile mineral oils to millwork and window sash requiring a clean, nonswelling, and paintable treatment.

Inorganic salts are employed in preservative treatment where the wood will not be in contact with the ground or water, such as for indoor use or where the treated wood requires painting. They are also satisfactory for outdoor use in relatively dry regions.

Preservation Process

The methods for applying preservatives to wood are classified as follows:

1. *Nonpressure processes*:
    1. surface (superficial) applications by brushing, spraying, or dipping
    2. soaking, steeping, and diffusion processes
    3. thermal process
    4. vacuum processes
    5. miscellaneous processes
2. *Pressure processes*:
    1. full-cell process (Bethell)
    2. empty-cell processes (Rueping and Lowry)

Brush and spray treatments usually give only limited protection because the penetration or depth of capillary absorption is slight. Dip treatments give slightly better protection. Organic chemicals dissolved in clear petroleum solvents often are applied to window sash and similar products by a dip treatment of 13 min.

Cold soaking of seasoned wood in low-viscosity preservative oil for several hours or days and the steeping of green or seasoned wood in waterborne preservatives for several days are methods sometimes employed for posts, lumber, and timbers on a limited basis. The diffusion process employs water-borne preservatives that will diffuse out of the treating solution into the water in green or wet wood.

The most effective of the nonpressure processes is the thermal method of applying coal-tar creosote or other oil-soluble preservatives, such as pentachlorophenol solution. The wood is heated in the preservative liquid in an open tank for several hours, after which it is quickly submerged in cold preservative in which it is allowed to remain for several hours. This is accomplished either by transferring the wood at the proper time from the hot tank to the cold tank, or by draining the hot preservative and quickly refilling the tank with cooler preservative. During the hot treatment, the air in the wood expands, and some is expelled. Heating also lowers the viscosity of the preservative so that there is better penetration. When the cooling takes place, the remaining air in the wood contracts, creating a partial vacuum that draws the preservative into the wood. For coal-tar creosote, the hot bath is at 210235°F, and the cold bath at about 100°F. This temperature is required to keep the preservative fluid.

The hot- and cold-bath process is widely used for treating poles and, to a lesser extent, for fence posts, lumber, and timbers. This process gives the most effective results of the common nonpressure processes, which most nearly approach those obtained by the pressure processes.

The vacuum processes involve putting the wood under a vacuum to draw out part of the air. The wood may be subjected to a vacuum alone or to steaming and a vacuum before being submerged in a cold preservative. These methods are used to a limited extent in the treatment of lumber, timber, and millwork.

The commercial treatment of wood is most commonly done by one of the pressure processes, as they give deeper penetrations and more positive results than any of the nonpressure methods. The wood, on steel cars, is run into a long horizontal cylinder, which is closed and filled with preservative. Pressure is applied, forcing the preservative into the wood.

There are two types of pressure treatment, the full-cell and the empty-cell. The full-cell process seeks to fill the cell lumens of the wood with the preservative liquid, giving retention of a maximum quantity of preservative. The empty-cell process seeks deep penetration with a relatively low net retention of preservative by forcing out the bulk liquid in the wood cells, leaving the internal capillary structure coated with preservative.

In the full-cell process, the wood in the cylinder first is subjected to a vacuum of not less than 22 in. Hg for 1560 min, to remove as much air as possible from the wood. The cylinder then is filled with hot treating liquid without admitting air. The maximum temperature for creosote and its solutions is 210°F, and for water-borne preservatives it is 120150°F, depending upon the preservative. Then the liquid is placed under a pressure of 125200 psi, and the temperature and pressure are maintained for the desired length of time, usually several hours. After the liquid is drawn from the cylinder, a short vacuum is applied to free the charge of surface-dripping preservative.

In the empty-cell process, the preservative liquid is forced under pressure into the wood, containing either its normal air content (Lowry process) or an excess of air, by first subjecting the wood to air pressure before applying the preservative under pressure (Rueping process). In the former case, the preservative is put in the cylinder containing the wood at atmospheric pressure, and, in the latter case, under air pressure of 25100 psi. After the wood has been subjected to the hot preservative (about 190200°F) under pressure (100200 psi in the Lowry process and 150200 psi in the Rueping process) and the pressure has been released, the back pressure of the compressed air in the wood forces out the free liquid from the wood. As much as 2060 percent of the injected preservative may be recovered, yet good depth of penetration of the preservative is achieved.

Preservative Retention

Retention of preservative generally is specified in terms of the weight of preservative per cubic foot of wood, based on the total weight of preservative retained and the total volume of wood treated in a charge. Penetration and retention vary widely between different species of wood, as well as with woods of the same species grown in different areas. In most species, heartwood is much more difficult to penetrate than sapwood. Also, within each annual growth ring there is variability in penetration, the latewood generally being more easily treated than the earlywood.

The American Wood-Preservers' Association Standards specify methods of analysis to determine penetration and retention. They also specify minimum retention amounts for different preservatives according to the commodity, the species, the pretreatment of the wood, such as kiln drying, and the end use of the commodity. Heavier retention is required for products in contact with the ground (poles, timbers, etc.) or with marine waters (piles, timbers, etc.).

Unprotected wood in contact with the ground is subject to severe attack by fungi and insects, and, in contact with seawater, it is quickly destroyed by marine borers. For wood products to be used in contact with the ground or marine waters, creosote is the major preservative employed because it can be readily impregnated to give high retention and good protection, and it is not leached out by water.

Increasing environmental regulations are causing greater restrictions on the use of traditional wood preservatives and alternate, nonpolluting methods are urgently needed. The use of chromated copper arsenate (CCA) to treat wood for residential structures will be phased out by the end of 2003. Currently wood pressure-treated with CCA is the most common material used to build decks, play equipment, picnic tables and residential fences. Arsenic, a carcinogen, imparts resistance to rot and termite damage, but there are concerns that it leaches from wood and exposes children. Substitutes for CCA include a copper boron azole compound and ammoniacal copper quat.

Nonconventional Wood Preservation

Chemical modification as a possible preservative treatment for wood is based on the theory that enzymes (cellulose) must directly contact the substrate (cellulose), and this substrate must have a specific configuration. If the cellulosic substrate is chemically changed, this highly selective reaction cannot take place. Chemical modification also can change the hydrophilic nature of wood. In some cases water, a necessity for decay organisms, is excluded from biological sites. The chemicals used for modification need not be toxic to the organism because their action renders the substrate unrecognizable as a food source to support microbial growth. For wood preservation, this means that it is possible to treat wood in such a manner that attack by wood-destroying fungi will be prevented, and the material will be safe for humans to handle. For applications of wood in which human contact is essential, nontoxic preservatives may well be specified or required in the future. An added benefit to wood of most chemical modification treatments is that the resulting bulking action gives the treated wood very good dimensional stability. The chemical modification of wood involves a chemical reaction between the hydroxyl groups in wood and a simple single chemical reagent, with or without a catalyst, to form a covalent bond between the two, as previously described for dimensional stability treatments.33 The wood component may be cellulose, hemicellulose, or lignin.

The treated wood must continue to possess the desirable properties of untreated wood: the strength must remain high, there must be little or no color change (unless a color change is desirable), and it must be a good electrical insulator, safe to handle, gluable, paintable, and so on. These chemicals, once reacted, are effective in preventing attack by microorganisms, but they are not toxic to the decay organisms. The important factor in preventing attack is to attain a treatment level that inhibits the growth of the organisms. A recent review on this subject shows that treatments with acetic anhydride, dimethyl sulfate, acrylonitrile, butylene oxide, phenyl isocyanate, and β-propiolactone all give good rot resistance at a 1725 weight percent gain (WPG).33 The decay resistance of acetylated wood is proportional to the WPG, and the degree of dimensional stability also is proportional to the WPG; so the exclusion of cell wall or biological water may be a very important factor in the decay resistance mechanism.34

In preliminary tests, alkylene oxide-treated southern pine was found to be resistant to termite attack and attack from the marine borers *Teredo* (shipworm) and *Limnoria*.

In conclusion, chemical modification of wood results in a treatment that is nontoxic, effective, and nonleachable. The high chemical treatment level required for effectiveness, however, results in a rather expensive treatment. Dimensional stability also is obtained at these high (1725 WPG) substitution levels; so, for those products where both not resistance and dimensional stability are important, the present state of the technology is close to a viable industrial process. The use of toxic chemicals that are permanently bound to the wood components by this treatment may be an environmentally acceptable preservation method.

< previous page

next page >

7.7 Fire-Retardant Treatment of Wood

The Romans first treated wood for fire retardancy in the first century A.D. They used solutions of alum and vinegar to protect their boats against fire. In 1820, Gay-Lussac advocated the use of ammonium phosphates and borax for treating cellulosic material. Many of the promising inorganic chemicals used today were identified between 1800 and 1870. Since then, the development of fire retardants for wood has accelerated. Commercially treated wood became available after the U.S. Navy (1895) specified its use in ship construction, and New York City (1899) required its use in buildings over 12 storeys tall. Production reached over 65 million board feet in 1943, but by 1964 only 32 million board feet was treated annually.35

However, starting in 1979 changes were made in some building codes that allowed structural changes if fire-retardant (FR) treated wood was used in place of untreated plywood for roof sheathing. The structural changes resulted in sufficient savings such that FR-treated wood was utilized more extensively. Also the use of FR-treated plywood was mandated at other specific sites such as prisons.

Fire-retardant treatments for wood can be grouped into two general classes: (1) those impregnated into the wood or incorporated into wood composite products and (2) those applied as paint or surface coatings. Chemical impregnation has the greater use, primarily for new materials, whereas coatings have been limited primarily to materials in existing constructions. There are advantages and disadvantages to each class. Coatings are applied easily, and they are economical. Chemical impregnation usually involves full-cell pressure treatment and can be costly. A coating is subject to abrasion or wear that can destroy the effectiveness of the fire retardant. Chemical impregnations deposit the fire retardant within the wood, so that if the surface is abraded, chemicals are still present. On-site application of surface coatings requires strict control of the amount applied to ensure correct loading levels for a particular flame-spread rating. Both coating and impregnation systems are based on the same chemical compounds although the formulations for each vary. Among the most commonly used chemicals for impregnation treatments are diammonium phosphate, ammonium sulfate, borax, boric acid, zinc chloride, and most recently the leach-resistant amino-resin systems. These compounds have different characteristics with respect to fire resistance. Ammonium phosphate, for example, is effective in checking both flaming and glowing; borax is good in checking flaming but is not a satisfactory glow retardant. Boric acid is excellent in stopping glow but not so effective in retarding flaming. Because of these different characteristics, mixtures of chemicals usually are employed in treating formulations.

Fire-Retardant Formulations

Many chemicals have been evaluated for their effectiveness as FRs. Today most FRs for wood are based on phosphorus, nitrogen, boron, aluminum trihydrate, and a few other compounds. Phosphorus and nitrogen frequently are used together because they behave synergistically; amino-resins are an example of such a combination.

Most FR formulations are not resistant to leaching by water. Therefore, there have been increased efforts to develop leach-resistant chemicals that can be impregnated into wood products for use in exterior or high-humidity applications. Some of the proposed leach-resistant systems include chemical combinations that form insoluble complexes, amino-resin systems, and monomers that polymerize in the wood. A common amino-resin system for exterior use is dicyandiamide phosphoric acid formaldehyde. Guanylurea phosphate-boric acid also is a commonly used organic phosphate salt for modern commercial FR wood.

The American Wood-Preservers' Association Standards specified the four types of FR formulations given below (A, B, C, D). Many newer formulations have been developed by commercial enterprises and are proprietary.

Type A

Chromated zinc chloridea mixture of sodium dichromate and zinc chloride having the composition: hexavalent chromium as CrO_3, 20% and zinc as ZnO, 80%.

Type B	
Chromated zinc chloride (as above)	80%
Ammonium sulfate	10%
Boric acid	10%
Type C	
Diammonium phosphate	10%
Ammonium sulfate	60%
Sodium tetraborate, anhydrous	10%
Boric acid	20%
Type D	
Zinc chloride	35%
Ammonium sulfate	35%
Boric acid	25%
Sodium dichromate	5%

Minimum and maximum limits of variation in the percentage of each component in the above formulations are specified in the standards.

The impregnation methods are similar to those employed for the preservative treatment of wood by water-borne salts using pressure processes. The maximum temperature of the solution must not exceed 140°F for formulation Types A, B, and D, and must not be above 160°F for Type C. After treatment, the wood must be dried to remove the water solvent to a moisture content of 19 percent or less. For most uses, the wood is kiln-dried to a moisture content of below 10 percent.

For effective fire protection larger amounts of chemicals must be deposited in the wood than the amounts necessary when water-borne chemicals are used for decay prevention. Whereas for the water-soluble toxic salts retentions from 0.22 to 1.00 lb/ft3 of wood are specified according to commodity standards in order to give good protection against decay and insects, as much as 56 lb of some fire retardants may be required for a high degree of effectiveness against fire. Usually, however, smaller amounts will give a good degree of protection. For example, formulation Type B when impregnated in amounts of 1.53 lb/ft3 of wood provides combined protection against fire, decay, and insects. Because the amino-resin systems are leach-resistant, high application levels are not necessary in practice.

Fire retardancy of wood involves a complex series of simultaneous chemical reactions, the products of which take part in subsequent reactions. Most FRs used for wood increase the dehydration reactions that occur during thermal degradation so that more char and fewer combustible volatiles are produced. The mechanism by which this happens depends on the particular FR and the thermal-physical environment. The effectiveness of a FR treatment depends upon the performance rating of the treated material when tested in accordance with ASTM E84 (no greater flame spread than 25).

7.8 Conversion of Wood to Energy, Fuels, and Chemicals

Wood has been a source of energy and chemicals for hundreds of years and continues to be an important raw material for specific chemicals. The use of wood as a primary source of industrial chemicals decreased dramatically in the 1940s when oil became the preferred raw material. The term "silvi-chemicals" sometimes is used to refer to wood-derived chemicals analogous to petrochemicals.

The use of wood for energy, fuels, and chemicals can be conveniently divided into four major categories: direct combustion, saccharification-fermentation (SF), thermal decomposition, and thermochemical liquefaction. These methods are discussed in more detail in the following sections.5

Direct Combustion

The concept of using wood as a source of energy through direct combustion dates back to the very beginning of human existence. As soon as early people learned to use fire, wood became the major source of energy. It is important to note that fuel is still the major use of wood on a worldwide basis, amounting to an estimated 1.2 × 109 m3/year compared with 0.8 × 109 m3/year for lumber and other solid wood products, 0.34 × 109 m3/year for pulp, and 0.2 × 109 m3/year for other uses. Thus, on a global basis, nearly 50 percent of the wood harvest is used directly for fuel. In many nonindustrialized countries, fuel amounts to 8090 percent of total wood consumption, compared with an average of 7 percent for industrialized regions.

The average stored energy content of wood is 8600 BTU/lb. Variations in heat values for wood as a function of species occur within a relatively narrow range of 800010,000 BTU/lb. Bark values are slightly higher than wood values, about 10,000 BTU/lb.

The greater the oxygen content of a carbon compound or carbonaceous substance, the lower the heat of combustion. Lignin, with an elemental analysis of $C_{10}H_{11}O_2$ compared with $C_6H_{10}O_5$ for cellulose, has a lower degree of oxidation and a higher heat of combustion than cellulose. The fossil fuels that have very low or no oxygen content have much higher energy contents, for example 12,00013,500 BTU/lb for coal, 18,00019,900 BTU/lb for oil, and 18,550 BTU/lb for natural gas.

Wood and other biomass materials are hygroscopic and retain substantial amounts of moisture (1560%). Hence, some of the thermal energy of the wood is lost in vaporizing this water to steam during combustion. In addition, wood has a lower density than coal, so that a greater volume of material must be gathered and transported to provide the same heating value as that of coal. Wood has a negligible sulfur content, less than 0.1 percent compared with much larger amounts in coal (avg. 2%) and petroleum. Sulfur from coal and petroleum creates serious environmental problems by causing acid rain. Also, ash from coal presents a disposal problem and creates unhealthy fly ash, whereas the ash content of wood is less than 0.5 percent. The only environmental hazard produced from burning wood as a fuel is the production of particulates.

Most wood and wood-derived materials (spent pulping liquors) that are used for energy are consumed by the forest products industry itself. The source of this fuel is almost entirely in the wood-processing and manufacturing operations, termed "manufacturing residuals" or "mill residues." Forest residues from logging operations and the noncommercial trees in the forest also are utilized for fuels.

Wood wastes supply the forest products industry with roughly 4050 percent of the energy requirements in the form of spent liquors and wood and bark residues. The spent pulping liquor contains 4050 percent of the wood as dissolved lignin and a large variety of other organic compounds. The spent liquor is used almost totally, after evaporation to about 65 percent solids, to produce steam and to recover the inorganic chemicals used in the pulping operation. The Tomlinson furnace is used almost universally and is the most efficient system for combined energy and chemical recovery yet devised.

The mixed wood and bark residues burned directly are collectively termed "hog fuel." Over the past decade, there has been a trend to channel more and more of the cleaner, drier, and larger-sized mill residues into raw materials for products. The coarse wood residues (slabs, trim, edgings, etc.) are chipped and sent to pulp mills. Also, considerable sawdust now is being pulped instead of being burned for fuel. Some of the residues, including shavings, also are used for particleboard. This means that hog fuel contains the less desirable, dirtier, and wetter forms of waste.

Combustion Technology

The hog fuel boiler represents the conventional technology of using wood for energy to produce process steam. Modern wood-fuel boilers have traveling grates and use mechanical draft fans. Many are of gigantic size, handling as much as 500,000 lb of wood per hour. Spreader stokes distribute the fuel evenly over the large grate areas to ensure efficient combustion. The ash can be removed continuously with traveling grates. Modern instrumentation meters air flow, controls fuel-to-air ratios, and meters combustion efficiency and stack emissions. With these developments, the energy recovery and environmental acceptability of wood-burning systems have been improved in the forest products industry.

The deterioration of hog fuel quality referred to above makes it more difficult to obtain efficient combustion, high reliability, and low stack emissions simultaneously. To overcome this difficulty, the fuel can be processed before combustion to remove dirt and moisture and produce clean-burning, efficient fuel. This can be done by drying, screening, grinding, washing, and pelletizing or densification of the hog fuel in order to remove noncombustible dirt and moisture, reduce the size of large and slow-burning material, and agglomerate small and fine material. Size uniformity is important for ease in transportation, storage, and fuel feeding. Dirt and moisture reduction reduces transportation costs, promotes efficient combustion, and minimizes air emissions. Another approach for using low-quality hog fuel is the development of improved combustors. Examples are: (1) the fluidized-bed combustors, which obtain excellent wood combustion at relatively high efficiencies with lowquality, nonuniform fuels; and (2) the pyrolytic burner, which has very low stack emissions with relatively high thermal efficiencies and wood combustion rates.

Present-day wood-fired boiler systems are complicated and cost considerably more than a comparable petroleum-fired installation. As much as 25 percent of the capital cost is in the fuel-handling equipment, and another 20 percent is in the air pollution control system. Because of the high capital costs and the lower thermal efficiencies of burning wood compared with oil (68% vs. 82%), the success of the wood-fired systems depends on the low cost of the wood fuel supplies.

The North American forest products industry has been successful in developing useful technologies for recovering energy from wood residues as process heat or steam, which are cost-effective. New technologies promise even greater economic benefits.

Cogeneration Technology

Cogeneration is the concurrent generation of electricity and the use of exhaust heat, usually in the form of process steam, for manufacturing operations. This is done by burning fuel (in this case, wood) to make high-pressure steam, 6001200 lb/in.2, passing this steam through a back-pressure or extraction turbine to drive a generator, and then using the steam exhausted from the turbine at lower pressures, 50300 lb/in.2, for process heat. This technology gets full use of the energy contained in the fuel. Wood at 55 percent moisture will generate power at about 60 percent efficiency.

The forest products industry is a major user of cogeneration technology, as it requires large quantities of process steam as well as electricity. It produces about 50 percent of its electricity needs in this way. The pulp and paper industry alone is the largest producer of energy by cogeneration of any U.S. industry. Electricity self-sufficiency is likely to increase to 8090 percent for forest industries in the future through cogeneration. However, there is a minimum plant size for economical power generation. Steam usage should be more than 70,000120,000 lb/hr, equivalent to 35 MW of back pressure, for a plant economically to employ cogeneration systems.

SaccharificationFermentation

The saccharificationfermentation (SF) method for the derivation of fuels and chemicals from wood is based on the breakdown or hydrolysis of the polysaccharides in wood to the constituent monomeric sugars. The six carbon or hexose sugars (glucose, galactose, and mannose) then are fermentable to ethyl alcohol (ethanol or grain alcohol, C_2H_5OH) by yeast fermentation in the same way that ethanol is produced from grains or fruits. Obviously the concept is not a new one; the polysaccharide character of wood has been known for over 100 years. The limitations on the use of wood for ethanol production primarily have been the difficulties in separating and hydrolyzing the crystalline cellulose component in wood. Both acids and enzymes can be used to hydrolyze the cellulose to glucose, but only acids have been utilized commercially for wood hydrolysis to sugars, and only in foreign countries.36

In contrast to the Western economy, the Soviet Union continued expansion of wood hydrolysis facilities, and about 40 such plants are presently in operation. All the Russian plants are based on dilute sulfuric acid in percolation towers. In the West interest in producing alcohols from wood was revitalized by the dramatic increase in the price of petroleum in the 1970s and the push to decrease oil imports by substituting gasohol, which is one part alcohol in nine parts gasoline, for 100 percent gasoline at gas pumps. Both ethanol and methanol can be used in gasohol blends.

Because of high oil prices, Brazil (most of whose sugars are produced from sugarcane) took the dramatic step of shifting to a much greater use of fuel alcohol. One wood hydrolysis plant was constructed, but it was uneconomical to operate and was shut down. However, Brazilian experience has demonstrated that fermentation ethanol (95 percent ethanol and 5 percent water) is a perfectly satisfactory motor fuel. At least 500,000 Brazilian automobiles operate on undried alcohol continuously, and most of the rest of their fleet operates on this fuel on weekends when only alcohol is available at the gas stations.36 A number of methods can be used for the production of ethanol from wood, as described below.

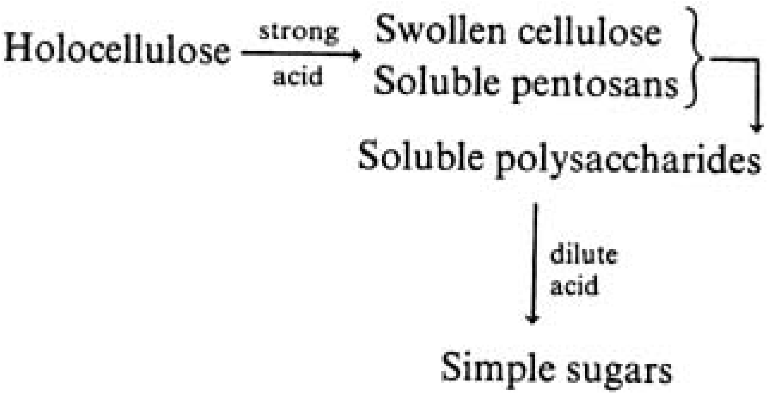
Cellulose, the major component of wood, gives about a 90 percent yield of pure glucose under laboratory conditions of hydrolysis, according to the following equation:

$$(C_6H_{10}O_5)_n + nH_2O \xrightarrow{\text{acid}} nC_6H_{12}O_6$$

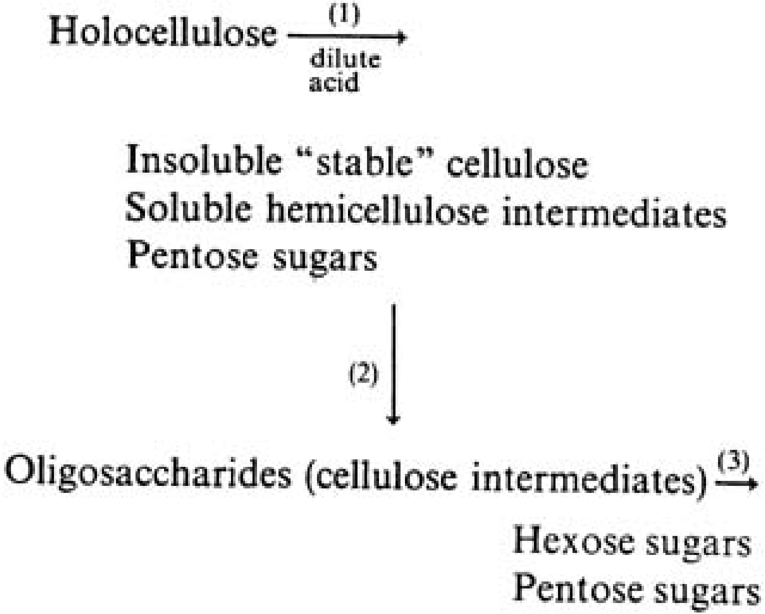
where *n* is in the range 10,00015,000. The hemicellulose fraction gives a mixture of sugarsxylose, arabinose, mannose, galactose, and glucose. Glucose, galactose, and mannose are yeast-fermentable sugars, whereas the pentoses (xylose and arabinose) are non yeast-fermentable. The potential total reducing sugar yield from wood averages 6570 percent, whereas the fermentable sugar yield is about 50 percent for hardwoods and 58 percent for softwoods. The lower quantity of fermentable sugar from hardwoods is due to their higher content of pentosans, compared with the coniferous woods.

Hemicelluloses hydrolyze much more easily and rapidly than cellulose. Temperatures and acid concentrations that hydrolyze the cellulose to glucose in a matter of a few hours readily convert much of the hemicellulose into simple sugars in minutes or even seconds. Under industrial conditions of hydrolysis, the sugars formed undergo decomposition, with the pentoses decomposing more rapidly than the hexoses. Thus, the conditions of hydrolysis cause variations in the ratio and yields of the various sugars due to (1) their different rates of formation by hydrolysis and (2) their different rates of decomposition.

The polysaccharides of wood (holocellulose) may be hydrolyzed by two general methods: (1) by strong acids, such as 7072 percent sulfuric acid or 4045 percent hydrochloric acid; or (2) by dilute acids, such as 0.52.0 percent sulfuric acid. The hydrolysis by strong acids is constant, proceeds as a first-order reaction, and is independent of the degree of polymerization. The reaction may be represented as follows:



In dilute-acid hydrolysis, the reactions are heterogeneous and more complex than this because no swelling and solubilizing of the cellulose occurs. Cleavage of the insoluble cellulose results directly in low-molecular-weight oligosaccharides (intermediate products), which are rapidly converted to simple sugars, as indicated below:



Reaction (1) is rapid and occurs under mild conditions, hydrolyzing mainly the hemicelluloses. Reaction (2) is slow, proceeds as a first-order reaction, and is the limiting reaction in this process. Reaction (3) is rapid.

On the basis of the above methods, two classes of industrial processes have been developed, namely the BergiusRheinau process, based on the use of concentrated hydrochloric acid at ordinary temperatures, and the SchollerTornesch process, in which very dilute sulfuric acid is used at temperatures of 170180°C (338356°F). The latter method in an improved form is known as the Madison process, based on work done at the U.S. Forest Products Laboratory in Madison, Wisconsin. A number of modifications have been developed, including four in Japan.

BergiusRheinau Process

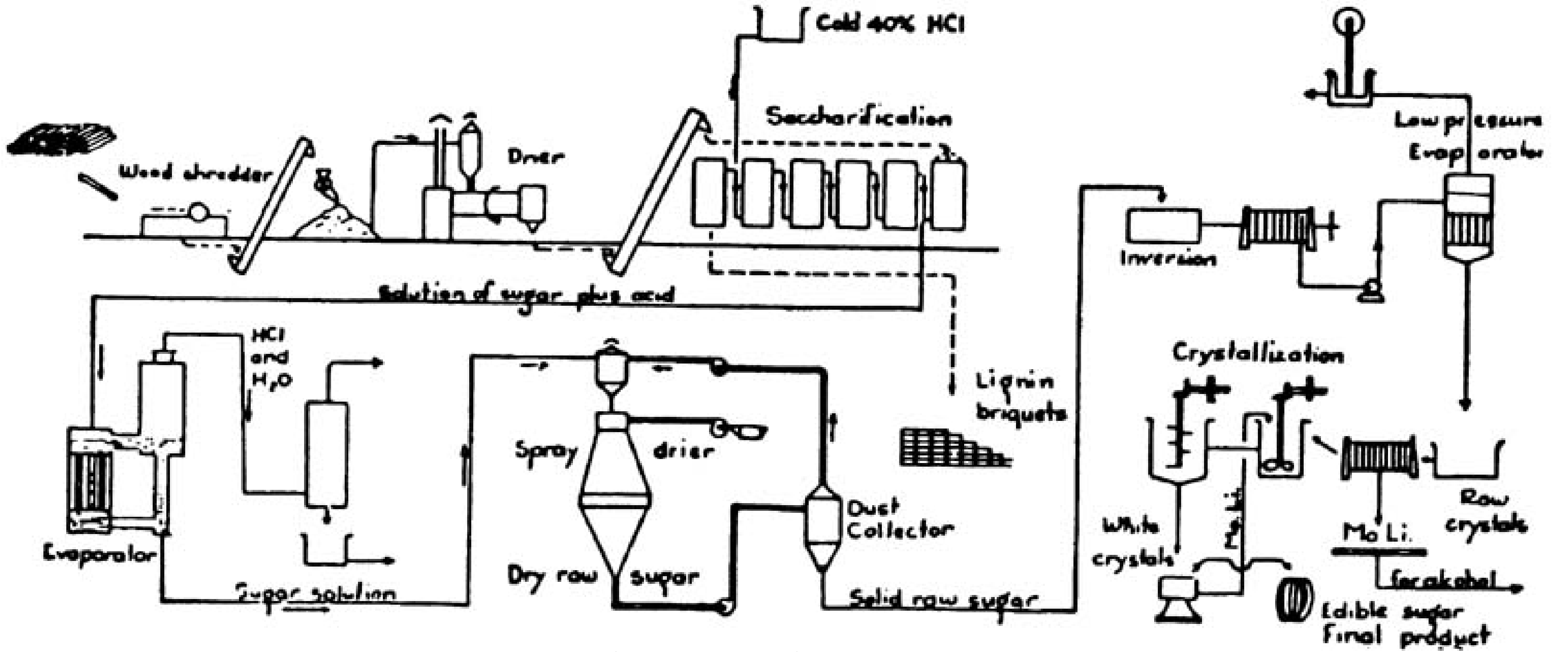
Hydrochloric acid of about 4045 percent (by weight) is produced by reinforcing recovered, weaker acid with hydrogen chloride from salt-sulfuric acid reactors, or by burning chlorine with illuminating gas.

Wood chips are air-dried, then charged into a tile-lined reactor, and extracted countercurrently by the acid. The fresh strong acid enters that part of the battery of diffusers or reactors which contains the most nearly exhausted wood, and is pumped through the following containers until it is nearly saturated with the carbohydrates dissolved from the wood. Part of this solution is mixed, under slight cooling, with fresh wood, and the mixture is charged into the head container of the battery. After the container is filled and a few hours are allowed for reaction time, an amount of solution is forced from this container, through the pressure of the incoming acid, that corresponds to the yield from one charge.

The drawn-off solution contains hydrochloric acid and carbohydrates in about equal parts, at a concentration of about 25 percent (by weight) each. It then is concentrated in stoneware tubes under vacuum. The distillate, containing about 80 percent hydrochloric acid with minor proportions of acetic acid and furfural, is reused after fortification. The concentrated sugar solution is dried to a powder in a spray-dryer, where it also loses most of the remaining acid.

The dry, somewhat acid, powder contains the carbohydrates in the form of intermediate polymers (oligosaccharides), which are water-soluble and must undergo further hydrolysis to yield simple sugars, either for fermentation or for crystallization. This is done by dissolving the oligosaccharides, diluting the solution to approximately 20 percent sugar concentration, and heating it for 2 hr in the presence of 2 percent acid at 125°C. Part of the glucose can be crystallized from the neutralized and reconcentrated solution, while the mother liquors are fermented to alcohol or used for growing yeast. A diagrammatic flow chart for the Bergius process is shown in Fig. 7.21.

Fig. 7.21 Diagrammatic flowsheet of Bergius wood hydrolysis process.



When hardwoods are to be used, it is necessary to remove a part of the hemicellulose first by prehydrolysis. This has been done on a large scale with straw, a substance chemically similar to hardwood, by heating it in an 8-to-1 liquid-to-solid ratio, with 0.5 percent sulfuric acid for 23 hr at 130°C (266°F). Without prehydrolysis, hardwoods and straws form slimy materials, probably because of their high hemicellulose content, which prevent the flow of the hydrolyzing acid.

In the BergiusRheinau process the concentrated hydrochloric acid employed requires dried wood, and recovery of the acid is essential. The process gives high yields of sugars (to 65%) at high concentrations. The intermediate sugars first obtained, however, call for an extra processing step to reduce them to monomers, before fermentation or crystallization.

Two Japanese processes have been developed using concentrated (above 62%) sulfuric acid. In the NihonMokuzaiKagaku process, the acid is neutralized by lime, and the gypsum may be used for making gypsum board and for other purposes. In the Hokkaido process, the acid is recovered by dialysis. Two other Japanese developments are the Udic Rheinau process, using 41 percent hydrochloric acid, and the NoguchiChisso process, using anhydrous hydrochloric acid. All four of these Japanese processes operate at atmospheric pressure and at temperatures below 100°C. All involve a prehydrolysis step to remove the hemicellulose because the main product is crystalline glucose.

Madison Process

In the Madison process, dilute sulfuric acid with an average concentration of 0.5 percent is pressed through wood in the form of sawdust and shavings. Regular flow of the acid and of the resulting sugar solution is one of the two principal requirements; the other is a lignin residue that can be discharged from the pressure vessels without manual labor. Both depend upon careful charging of the wood, which should not contain too many very fine particles, and upon maintaining a pressure differential of not more than 56 lb/in.2

between the top and the bottom of the digester.

The digesters or percolators are pressure hydrolyzing vessels, commonly employing a pressure of 150 lb/in.2, and having a capacity of 2000 ft3 each. In the original Scholler plants in Germany, the digesters were lined with lead and acid-proof brick; in the Madison process a lining of "Everdur" metal was found to give sufficient protection.

The wood, about 15 tons, is pressed down with steam and then is heated by direct steam, after which the acid is introduced. The practice in Germany was to bring the dilute acid into the digester in several batches, with rest periods of about 30 min, heating the wood to temperatures of 130°C (266°F) at first, then to 180°C (356°F), while keeping the temperature of the entering acid 1020°C lower. A total of about 14 hr was required to exhaust the wood, yielding about 50 lb of carbohydrates for 100 lb of dry wood substance.

In the Madison process, continuous flow of the acid, and, correspondingly, of the sugar solution, is providedin other words, continuous percolation. The cycle is thereby reduced to 6 hr, and the yields are increased somewhat. The lignin is blown out of the digester by opening the specially constructed bottom valves while the vessel is still under pressure.

The sugar solutions usually contain about 5 percent sulfuric acid. The solutions, still under pressure (150 lb/in.2), are flash-evaporated to 35 lb/in.2, neutralized with lime at that pressure, and filtered. Calcium sulfate is much less soluble at the elevated temperature corresponding to the pressure than it is at 100°C. This is a fortunate circumstance, for it must be removed to an extent sufficient to avoid difficulties caused by the formation of incrustations in the subsequent alcohol distillation. The filtered solution is cooled by further flash evaporation and heat-exchanged with water to the fermentation temperature.

Sugar yields from coniferous woods (softwoods) are about 50 percent at an average concentration of 5 percent. When fermented, the average ethyl alcohol yield per ton of drywoods is 5060 gal and sometimes higher.

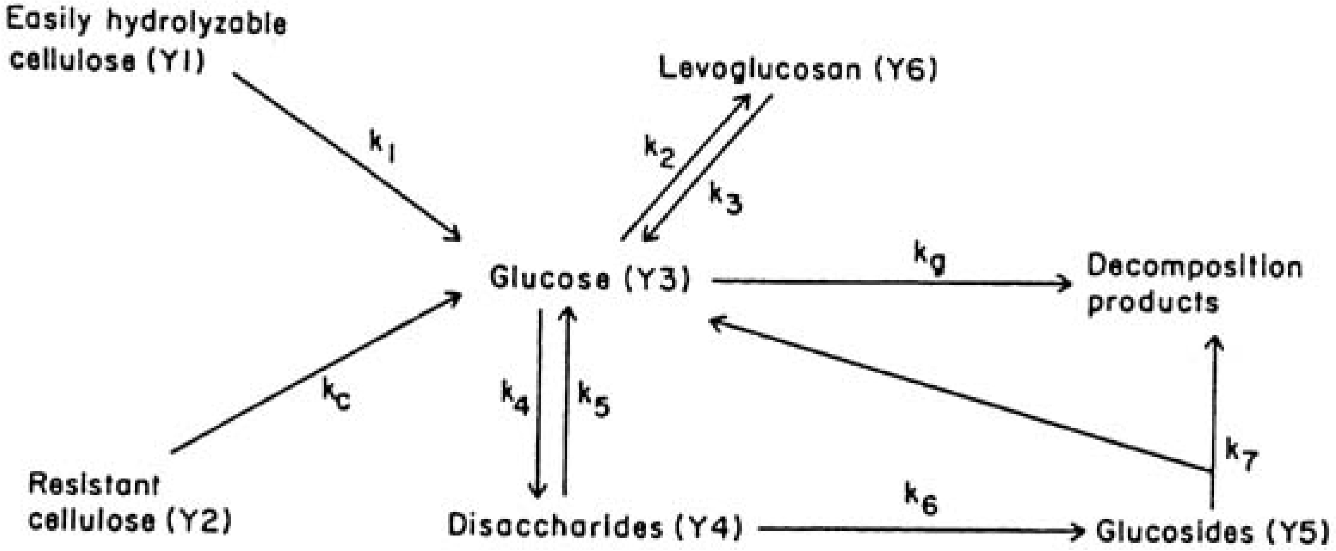
The dilute sulfuric acid employed in the Madison process gives lower yield (4955%) of sugars than the BergiusRheinau process, and only very dilute solutions are obtained directly. Recovery of heat is easier in the Madison process, and the acid need not be recovered.

The dilute-acid hydrolysis method presently is preferred for production on a commercial scale although the yield of glucose is only about 50 percent based on the cellulose. Nearly quantitative yields can be obtained by strong-acid hydrolysis, but are offset by higher capital costs for corrosion-resistant equipment and higher operating costs for acid recovery plus acid losses. More recent efforts to hydrolyze cellulosics with dilute acid have involved innovative reactor schemes. Rugg and Brenner described equipment for a screw-fed, continuous, single-stage hydrolyzer, and Thompson and Grethlein investigated a plug-flow reactor.36 Both processes take advantage of the improved kinetics of cellulose hydrolysis relative to glucose decomposition at higher temperatures and shorter retention times to improve sugar yields. Both also have the advantage of being simple one-stage processes.

The U.S. Forest Products Laboratory (FPL) in cooperation with the Tennessee Valley Authority (TVA) has been studying a two-stage dilute acid hydrolysis process based in part on studies of Cederquist in Sweden during the 1950s. The first stage (pre-hydrolysis) selectively removes the hemicellulosic sugars with dilute sulfuric acid at about 170°C prior to hydrolysis of the lignocellulosic residue to glucose in the higher-temperature (230°C) second stage. The two-stage dilute acid process has a number of important advantages: (1) the carbohydrates are fractionated into hemicellulosic sugars and glucose, so that the separate utilization of each fraction is facilitated; (2) glucose is isolated in moderately good yield (•50%); (3) the glucose solution from the second stage is moderately concentrated (•1012%); and (4) the consumption of acid and steam is relatively low.

Despite recent efforts such as those to produce glucose from agricultural residues and wood, little attention has been given to improving the kinetic modeling of cellulose saccharification with dilute acid. Kinetic modeling plays an important role in the design, development, and operation of processes to hydrolyze cellulosics to fermentable sugar. A new model for the dilute acid hydrolysis of cellulose was developed at FPL in connection with studies of the two-stage dilute sulfuric acid hydrolysis process (Fig. 7.22).37 The model incorporates the effect of the neutralizing capacity of the substrate, the presence of readily hydrolyzable cellulose, and the reversion reactions of glucose in acid solution. Although general in nature, the model was developed specifically for application to the dilute, sulfuric acid hydrolysis of prehydrolyzed wood. A computer program simulating the new model can be used to predict yields of free glucose, reducing sugars, reversion material, remaining cellulose, and glucose loss due to dehydration as a function of acid concentration, temperature, and reaction time.

Fig. 7.22 Model for dilute acid hydrolysis of cellulosics. (*From A. H. Conner et al., in* Cellulose: Structure, Modification and Hydrolysis, *R. A. Young and R. M. Rowell (Eds.), Copyright © John Wiley & Sons, New York, 1986; with permission of the publisher*.)



Enzymatic Hydrolysis

Saccharification of wood polysaccharides to sugars can be accomplished by enzymatic techniques instead of acid hydrolysis. The U.S. Army Natick Laboratories developed a method for conversion of cellulose to glucose with a cellulose enzyme from an active strain of the fungus *Trichoderma viride*. However, extensive pre-treatment of wood is necessary before sufficient enzymatic hydrolysis will take place.

Enzymatic methods show the biggest promise for conversion of waste paper from municipal waste into glucose for ethanol production. Because paper is composed primarily of wood cellulose fibers, the enzyme inhibition due to lack of accessibility with whole wood is partially alleviated. As mentioned previously, waste paper can represent up to 50 percent of typical municipal waste. Currently the separated paper from the waste is burned for its fuel value.

The Gulf Oil Company developed a method called simultaneous saccharification and fermentation (SSF) for the enzymatic conversion of waste paper to ethanol.38 In this process the cellulose is enzymatically hydrolyzed and the glucose yeast-fermented in one operation. This modification, along with improved enzyme production and performance, has made the enzymatic technique more economically viable for the conversion of waste paper to ethanol. The process was donated to the University of Arkansas for further development.

Fermentation of Sulfite Waste Liquor

The sulfurous acid used in the sulfite pulping liquor causes hydrolysis of the more easily hydrolyzable components of wood, especially the pentosans in the hemicellulose. About 35 percent of the potentially fermentable sugars in the wood are hydrolyzed. However, most of them are decomposed during the long pulping procedure so that only one-fourth to one-third of the sugars, including the more resistant hexoses, remain in the waste sulfite liquor. If these sugars are fermented by yeast, about 12.5 gal of 95 percent alcohol per ton of wood may be produced.

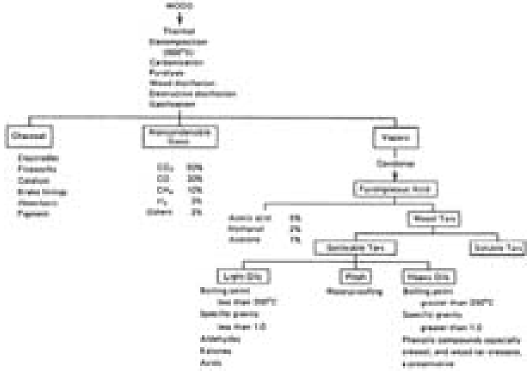
A large number of plants in Europe and a few in North America have been constructed to utilize the sugar hydrolyzate in sulfite waste liquors. The procedure as carried out in one American operation is as follows.

The liquor is recovered from the digester by discharging it in such a manner that as much sugar as possible is removed with a minimum of dilution by washing. Free sulfur dioxide is removed and recovered by blowing steam through the solution, which decreases the acidity from a pH of 2.2 to a pH of about 3.9. The liquors are cooled by a vacuum flash and neutralized by lime to a pH of about 4.2. A small amount of inorganic nitrogen is added for yeast growth, and about 1 percent by volume of yeast is added continuously. Fermentation is carried out in a series of tanks, the solution flowing from one to the other with agitation to keep the yeast in suspension. The yeast is recovered by centrifuging and mixed with the new sugar solution entering the fermenter. About 30 hr is required for fermentation. The alcohol content of the fermented liquor is about 1 percent by volume. The ethyl alcohol is recovered in stainless steel stills. Methanol and other alcohols are obtained in small amounts as by-products.

Thermal Decomposition

When wood is heated in the absence of air or with only limited amounts of air, thermal degradation takes place. This begins at about 100°C and increases with rising temperature. At about 270°C exothermic reactions set in, causing a rise in temperature (usually held at 400500°C) bringing about complete carbonization. The products are charcoal, condensable liquids, and noncondensable gases.39 The condensable liquids separate into aqueous (pyroligneous acid) and oil and tar fractions (Fig. 7.23). The charcoal, the gases (low BTU gas), and the oil fractions all may be used as fuels.

Fig. 7.23 Products obtained from the thermal decomposition of wood. (*From R. A. Young and R. L. Giese (Eds.),* Introduction to Forest Ecosystem Science and Management, *3rd ed., Copyright © 2004 John Wiley & Sons, New York; with permission of the publisher*.)



A number of terms are used for the thermal decomposition of wood and generally refer to similar processing methods: *carbonization*, *pyrolysis*, *gasification*, *wood distillation*, *destructive distillation*, and *dry distillation*. All result in the thermal breakdown of the wood polymers to smaller molecules in quantities dependent on reaction conditions. The hydrogen content of the gas increases with increasing temperature of pyrolysis. The wood gas has a fuel value of 300 BTU/ft3.

The yields of the different materials obtained by pyrolysis vary with the species of woods used and the type of equipment and system employed. Manipulation of three variables, mainly the temperature, heating rate, and gas residence time, can greatly alter the relative proportions of the gas, liquid, and char produced. For example, the oil and tar yields can be varied between 1 and 40 percent and the char between 40 and 10 percent or less. Low temperatures favor liquids and char, low heating rates favor gas and char, and short gas residence favors liquids. Conversely, high temperatures favor gas, high heating rates favor liquids, and long gas residence times favor gas. Thus, the various product fractions can be preferentially manipulated by proper combinations of these variables.

During World War II in Germany, automobiles were fueled by the gases produced from thermal decomposition of wood; and research is ongoing today on the more efficient gasification of wood. Destructive distillation has been used throughout most recorded history to obtain turpentine from pinewood, as discussed later in the chapter.

Low BTU Gas and Oil

There has been considerable research activity to produce low BTU gas and oil from wood for energy. The Georgia Institute of Technology and the Tech-Air Corporation have intensively studied pyrolysis of biomass materials and have built several pilot plants. Their pyrolytic process (GT/T-A) has been applied to forestry residues. Their plant differs from the older wood-distillation plants, such as the BadgerStafford retorts used in the 1930s, in several respects, namely the use of smaller-sized equipment to process 7 dry tons/hr compared with 34 tons/hr, the use of wood dried to 7 percent moisture instead of to less than 0.5 percent, and the use of a small quantity of air inside this report (approx. 0.25 lb/lb feed) to sustain the reaction instead of using gases in order to heat it to 1000°F. The on-line time of the rated capacity for the GT/T-A system is claimed to be better than 90 percent, as compared with 67 percent for the BadgerStafford units.

For the GT/T-A system the wood material is hogged so that the maximum size is not more than about 1 in. in any dimension. Sawdust may be used directly. The relatively uniform material is conveyed to a dryer where it is dried to about 7 percent water content. The dryer is heated by a portion of the wood gas from the reactor, and should this not be sufficient, the oil produced during pyrolysis can be used as a backup fuel.

The dried wood is fed into the top of the reactor though an air-lock and moves downward by gravity. A sensing device measures the bed height and controls the input of the wood. The temperature increases from 350 to 500°F at the top of the bed to 10001700°F in the pyrolysis zone. The gases move upward through the bed of descending wood and leave the top at 350500°F. The reactor is operated at slightly below atmospheric pressure by an induced draft fan. The charcoal passes through an outfeed device at the bottom of the reactor into a sealed chamber, where it is cooled by a water spray and then conveyed to a storage bin. The rate of charcoal discharge controls the rate of passage of material though the unit.

The gases from the reactor pass into a scrubber where they are sprayed with cooled pyrolytic oil, which removes particulates and cools the gas stream to 180200°F, causing condensation of the condensable organic substances into an oil mixture. The oil is filtered, the filter cake returned to the reactor, and the filtered oil pumped to a tank, cooled, and then recirculated through the scrubber. Excess oil is pumped to a storage tank.

The gases from the scrubbercondenser contain the noncondensable gases, lowboiling organic vapors, and water vapor. A portion of the gases is used to heat the wood dryer, and the remainder is available as a fuel for other purposes, but it is desirable that it be used near the pyrolysis plant.

Operating conditions can be controlled to vary the distribution of the energy among the products: gas, oil, and charcoal. Thus, high yields of charcoal can be obtained with correspondingly reduced yields of oil and gas and vice versa, or the unit may be operated as a gasifier with only a 3.8 percent yield of charcoal. The off-gases in the latter case can be burned directly in a gas-fired boiler, or the condensable oils can be removed first and used separately as fuel. The total energy recovery in the form of wood gas, oil, and charcoal is about 95 percent of the energy in the dry wood used in the process.

Gasification

Gasification is the thermal degradation of wood or other carbonaceous material in the presence of controlled amounts of oxidizing agents, such as air or pure oxygen. It is carried out at higher temperatures than those used in the pyrolysis process, up to around 1000°C. Hence, the reaction rates are very fast, making equipment design critical. The thermal efficiency for conversion of wood to gases is 6080 percent, which compares favorably with that of coal. Wood gasification offers several advantages over coal: (1) much lower oxygen requirements, (2) virtually no steam requirements, (3) lower costs for changing H2/CO ratios, which are already higher in wood gas, and (4) no or very little desulfurization costs. Coal has an advantage in that larger plants can be built for coal gasification than is normally the case for wood, because of its procurement advantages.

As noted above, the GT/T-A pyrolysis system also can be used as a gasifier by increasing the amount of air to the reactor and screening the charcoal and recycling the coarser particles to the incoming feed material. In this way, most of the wood is converted to gas and oil. If all the off-gases from the converter are conducted directly into a boiler as a hot gaseous fuel, then the GT/T-A process is essentially a gasifier. This requires that the gases be utilized as a fuel close to the pyrolysis plant. The use of air in the reactor results in the presence of nitrogen in the gas, causing a reduction in fuel value from 360 to 420 BTU/ft3 (without air) to 140200 BTU/ft3 (with air).

Several gasifiers have been designed to handle municipal refuse, wood wastes, and other biomass materials. Basically, gasifiers fall into two types: (1) those that use air and (2) those using oxygen. An example of the air system is the MooreCanada gasifier. The wood residues are carried on a moving bed through stages of drying, reduction, and char oxidation, and the ash is discharged in granular form. The maximum temperature in the reaction zone is about 1222°C. The hydrogen content of the crude gas is increased from 810 percent up to 1822 percent by adding steam to the air intake.

The Union Carbide Company Purox gasifier is an example of the oxygen system. This unit also uses a moving-bed reactor. Pure oxygen is the oxidizing agent used to convert the char into CO and CO2. Molten ash leaves the bottom at about 1670°C.

If air is used, the crude gases contain about 46 percent nitrogen, which must be removed by cryogenic means; but if oxygen is used, it first must be separated from air, with oxygen and nitrogen separated by a cryogenic system (see Chapter 13). The crude gases also contain an oil and tar fraction, about 2 percent of the wood (dry basis). Gasifiers designed for wood operate at atmospheric pressure, in contrast to coal gasifiers that operate at pressures up to 400 psi.

The composition of the wood gas varies according to the technology used. If the limited oxygen required is supplied by air, a typical gas might contain 1018 percent H2, 2230 percent CO, 69 percent CO2, 4550 percent N2, and 35 percent hydrocarbons and have a heating value of about 1700 kcal/m3 (180 BTU/ft3). If pure oxygen is supplied to the process, the gas might contain 2426 percent H2, 40 percent CO, 10 percent hydrocarbons, and 2325 percent CO2 and have a heating value of 2900 kcal/m3 (350 BTU/ft3). The gas may be used directly as a fuel in a variety of ways; for example, supplying boiler energy in an industrial plant at the rate of 2.5 × 108 kcal/hr from about 136 kg/day of dry wood.

The technology of gasification is under active development in both equipment and process design to reduce capital costs, and in the chemistry of the process to improve yields and obtain more favorable gas ratios. The latter includes the use of catalysts to enhance the process and to promote the formation of specific products, such as methane or ethylene for increased thermal efficiency or for feed-stocks for chemical synthesis.

Methanol

As is the case with ethanol, the concept of producing methanol from wood is not new. Methanol obtained from the destructive distillation of wood represented the only commercial source until the 1920s. The yield of methanol from wood by this method is low, only about 12 percent or 20 L/metric ton (6 gal/ton) for hardwoods and about one-half that for softwoods. With the introduction of natural gas technology, the industry gradually switched to a synthetic methanol formed from a synthesis gas (syngas) produced from reformed natural gas. Two volumes of H2 and one volume of CO are reacted in a catalytic converter at pressures of 15004000 psi to produce methanol. Presently, 99 percent of the methanol produced in the United States is derived from natural gas or petroleum.

Methanol also may be produced from wood gas; so wood could be a future raw material for making methanol, especially for use as an additive to gasoline for internal combustion engines. Thus, reforming the gasification products obtained at high temperatures is a second method for the production of methanol from wood. This is in contrast to the older method (destructive distillation), which directly yields small quantities of methanol at lower temperatures as mentioned above.

If wood gas is produced in a gasifier using air as the oxidizing agent, the nitrogen (about 40%), as well as the CO2 must be removed. The CO2 is removed by passing the gas through hot potassium carbonate followed by scrubbing with monoethanolamine. The nitrogen, hydrocarbons, residual CO2, and water are removed by freezing them out in a cryogenic system. By means of an iron catalyst, a "shift conversion" of a portion of the CO is carried out with steam in order to reach the required two parts of H2 to one part of CO, as follows:

$$\mathrm{CO} + \mathrm{H_2O} \xrightarrow{\text{catal.}} \mathrm{H_2} + \mathrm{CO_2}$$

$$\Delta H = -590\ \mathrm{BTU/lb\ CO}$$

The pressurized gas then is led to the methanol reactor. Two different catalyst systems may be used: (1) a zincchromium catalyst requiring gas pressures of 2000 4000 psi or (2) a copper catalyst system at 10002000 psi. About 95 percent of the gas is converted to methanol by this exothermic reaction:

$$2\mathrm{H_2} + \mathrm{CO} \xrightarrow{\text{catal.}} \mathrm{CH_3OH}$$

$$\Delta H = -1200\ \mathrm{BTU/lb\ CO}$$

The yield of methanol with present technology is about 390 L/ton of oven dry wood. Conversion of wood to methanol is less efficient than producing methanol from natural gas or coalabout 38 percent compared with about 60 percent, based on the heat value of methanol as a percent of the total energy input into the plant.

The raw materials cost is the most significant operating cost. The production cost for methanol from wood is estimated to be almost twice that produced from natural gas. Improving the efficiency of the gasifier reactor to increase the quantities of CO and H2 produced from wood would enhance the process.

Charcoal and Other Chemicals

Production of charcoal and tars by destructive distillation is the oldest of all chemical wood-processing methods. Charcoal probably was first discovered when the black material left over from a previous fire burned with intense heat and little smoke and flame. For centuries, charcoal has been used in braziers for heating purposes. Destructive distillation of hardwoods has been carried out with charcoal the product sought and volatiles as by-products; with softwoods (pines), volatiles were the principal products (naval stores), with charcoal considered a by-product.

In the United States, charcoal production began in early colonial days. During this period, principal uses of charcoal were as a fuel in blast furnaces for the production of pig iron and as an ingredient of gunpowder. Charcoal needed by the iron industry had to have a high crushing strength, and thus was made from dense hardwoods such as maple, birch, oak, and hickory. A softer charcoal was preferred for making gunpowder and was produced from willow and basswood. The first successful blast furnace was built in Saugus, Massachusetts, in 1645. Other furnaces were started in close proximity to iron ore deposits. These early furnaces were small, usually producing only 13 metric tons of pig iron per day. In the late 1880s, wood charcoal lost its metallurgical market to coke, which was better suited to the demands of the larger furnaces than being built.

In 1812, the additional collection, by condensation, of the volatile substances from hardwood carbonization began; the products then were charcoal, crude pyroligneous acid, and noncondensable gases. The pyroligneous acid was refined to produce methanol, acetate of lime, which in turn was used to make either acetic acid or acetone and tar. The noncondensable gases in a normal wood distillation consisted of about 50 percent carbon dioxide, 30 percent carbon monoxide, 10 percent methane, 3 percent heavier hydrocarbons, and 3 percent hydrogen.5 Table 7.9 gives the yields of products from an industrial operation. The tars and noncondensable gases usually were used as fuel. From the late 1800s until the 1920s, destructive distillation of hardwoods was an important source of industrial acetic acid, methanol, and acetone. This market was lost when these materials were made synthetically from petroleum. In 1920, there were approximately 100 plants recovering these products from hardwood distillation; the last of these plants ceased operation in 1969. The reader interested in details of the wood distillation processes is referred to the 1962 edition of *Riegel's Industrial Chemistry*, Chapter 15.

TABLE 7.9 Yield of Products from Destructive Distillation of One Ton of Dry Hardwood

Product	Quantity
Charcoal	270 kg (600 lb)
Noncondensable gasa	150 m3 (5000 ft3)
Soluble tar	83.6 L (22 gal)
Pitch	29.7 kg (66 lb)
Creosote oil	12.5 L (3.3 gal)
Methanol	11.8 L (3.1 gal)
Acetone	2.7 L (0.7 gal)
Allyl alcohol	0.2 L (0.5 gal)
Ketones	0.8 L (0.2 gal)
Methyl acetate	3.8 L (1.0 gal)
Acetic acid	45.5 kg (101 lb)

aComposition in order of highest to lowest: CO2, CO, CH4, H2, others.

Source: U.S.D.A. Forest Service Data.

In the early 1900s, charcoal from by-product recovery plants usually was used for cooking and heating in low-income areas and was known as a "poor man's fuel." Beginning in about 1950, there was an upturn in demand for charcoal for recreation use. In this era of suburban living, the use of charcoal briquettes for cookouts represents a significant market. The charcoal briquette now can be considered a luxury fuel, as it is too expensive for heating.

Basic techniques for producing charcoal have not changed over the years although the equipment has. Charcoal is produced when wood is burned under conditions in which the supply of oxygen is severely limited. "Carbonization" is a term that aptly describes the thermal decomposition of wood for this application. Decomposition of carbon compounds takes place as the temperature rises, leading to a solid residue that is richer in carbon than the original material. Wood has a carbon content of about 50 percent, whereas charcoal of a quality suitable for general market acceptance will be analyzed as follows: fixed carbon 7481 percent, volatiles 1823 percent, moisture 24 percent, and ash 14 percent. Charcoal with a volatile content over 24 percent will cause smoking and is undesirable for recreational uses.

Earthen "pit kilns" originally were used to produce charcoal. A circular mound-shaped pile of wood (1545 cords) was built up with an open core 3060 cm (12 ft) in diameter to serve as a flue. The entire surface of the pile, except for the top flue opening and several small openings around the bottom periphery, then was covered with dirt or sod sufficiently thick to exclude air. The mound was allowed to "coal" for 2030 days to give the final product.

In the second half of the nineteenth century, brick or masonry "beehive" kilns came into widespread use. The capacity of many of these kilns was from 50 to 90 cords, and operation was essentially the same as for the pit kilns. Many other types of kilns have been used from time to time. Small portable sheet-metal kilns of 12 cords have been widely used, as have rectangular masonry block kilns of various designs. These were predominantly used by farmers and small woodlot owners.

Large-scale production of charcoal was done by distilling the wood in steel buggies in long horizontal ovens. The buggies rode in steel rails that carried the cars in line from predryers to the ovens and then to coolers. Charcoal was produced by this method in a matter of 24 hr. The latest types of charcoal-producing equipment are designed for continuous operation and make use of residues instead of roundwood. An example is the Herreshoff multiple hearth furnace, in which several hearths or burning chambers are stacked on top of one another, the number depending on capacity. Production in this type of furnace is from 1 to 2.5 metric tons of charcoal per hour.

Thermochemical Liquefaction

Although a reasonable amount of research effort has been expended on thermochemical liquefaction of wood, extensive commercialization of this process is not anticipated in the near future. The basis of the method is a high-pressure and high-temperature treatment of wood chips in the presence of hydrogen gas or syngas to produce an oil instead of a gas. The low-grade oil produced potentially could be substituted for some present petroleum uses.

An oil of a heating value of about 35,200 BTU/kg can be obtained by reaction of wood waste for 1 hr with syngas, a catalyst, a temperature of 750°F, and a pressure of 5000 psi. The feasibility of the process has been tested in a pilot plant in Albany, Oregon, based on laboratory work conducted at the U.S. Bureau of Mines. So far, it has been determined that a barrel of oil equivalent to No. 6 bunker fuel can be produced from about 405 kg of wood chips.40

A similar process was developed in Japan in the 1950s specifically to degrade lignin. Called the Noguchi process, it was thought to hold promise for the production of phenols from lignin. By 1955, the Japanese investigators had discovered superior catalysts that converted a substantial portion of the lignin into a relatively few phenols. The CrowrZellerbach Corporation (now defunct) subsequently obtained an option on the process and initiated its own trials. Despite several improvements researchers were able to make, the process did not prove profitable at the time. However, the company was routinely able to obtain a yield of 55 percent (and up to 65%) of distillable products. The major drawback was the inability to separate cleanly even the few different phenols remaining after the reactions.

7.9 Naval Stores

The U.S. naval stores industry began in the very early colonial days, when wooden vessels used tar and pitch from the crude gum or oleoresin collected from the wounds of living pine trees. The demand for tar and pitch from crude gum is now of minor importance.

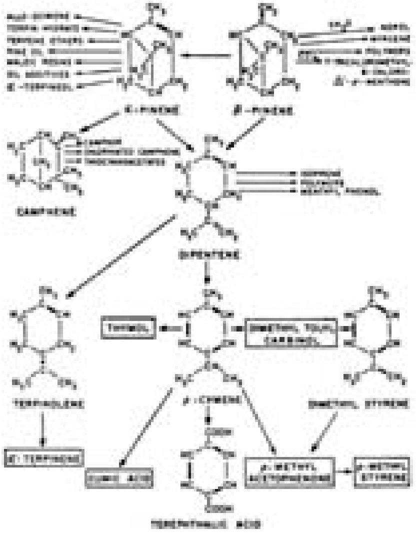
The industry is centered in the southeastern United States and is confined to the longleaf and slash pine areas. There is also a small, but locally important, naval-stores-producing area in the Landes region of southwestern France, based on the maritime pine.

There are three routes by which naval stores are produced. The oldest method is the tapping of living trees to cause a flow of oleoresin. The second method is removal of naval stores by solvent extraction. The latter process now has replaced steam distillation as a means of recovering turpentine. In the United States, the latest and now the most important route is kraft (sulphate) pulping of pine, during which turpentine and tall oil are recovered as by-products of kraft pulp manufacture. A fourth process, no longer used in the United States, is recovery of turpentine and pine oils by the destructive distillation of pine wood.41

Worldwide, about 60 percent of naval stores are produced by tapping living trees, whereas gum naval stores account for only 4 percent of U.S. production. The sulfate process is now the major U.S. process.

Turpentine is a volatile oil consisting primarily of terpene hydrocarbons, having the empirical formula $C_{10}H_{16}$. These 26 atoms can have many different arrangements, only six of which are present in appreciable amounts in commercial turpentines: α-pinene (b.p. 156°C), β-pinene (b.p. 164°C), camphene (b.p. 159°C), Δ3-carene (b.p. 170°C), dipentene (b.p. 176°C), and terpinoline (b.p. 188°C). The molecular configurations of some of these are shown in Fig. 7.24.

Fig. 7.24 Some reactions of alpha-pinene and beta-pinene. (*From L. A. Goldblatt*, Yearbook of Agriculture, *U.S.D.A., 195051*.)



Gum and sulfate turpentines have similar compositions. Gum turpentine contains 6065 percent α-pinene, 2535 percent β-pinene, and 58 percent other terpenes, compared with 6070 percent α-pinene, 2025 percent β-pinene, and 612 percent other terpenes for sulfate turpentine. Wood turpentine, which has 7580 percent α-pinene, has no or very little β-pinene (02%), and also contains 48 percent camphene and 1520 percent other terpenes.

Rosin, the other major naval stores product, is a brittle solid that softens at 80°C. Chemically it is composed of about 90 percent resin acids and 10 percent neutral matter. The resin acids are mainly *l*-abietic acid and its isomers, $C_{20}H_{30}O_2$. These are tricyclic monocarboxylic acids and are diterpenes.

Rosin is graded and sold on the basis of color, the color grades ranging from pale yellow to dark red (almost black). The color is due almost entirely to iron contamination and oxidation products. Fresh oleoresin, as it exudes from the tree, will yield a rosin that is nearly colorless. Color-bodies are removed by selective solvents and selective absorption from a 1015 percent gasoline solution passed through beds of diatomaceous earth. About 70 percent of the world's rosin is produced in the United States. Naval stores also include important fatty acids, as well as steroids and other products.

Gum Naval Stores

The crude gum or oleoresin is obtained from healthy trees by exposing the sapwood. The lower part of the tree is faced; that is, a section of bark is removed, giving a flat wood surface for the gutters, which are inserted into a slanting cut made by a special ax. The gutters conduct the gum to a container that can hold 12 quarts of gum. At the top of the exposed face, a new V-shaped strip of bark is removed about every two weeks.

The operations of inserting gutters, hanging cups, and cutting the first bark preferably are done in December or January, as early facing stimulates early season gum flow. The gum continues to flow until November, with the height of the season being from March to September. The collected gum is distilled from a copper still; turpentine and water pass over, and the rosin is left in the still. The remaining molten rosin, plus impurities, is passed through a series of strainers and cotton batting to remove dirt particles. Then the liquid rosin is run into tank cars, drums, or multiwall paper bags for shipment.

Increases in yield of naval stores are brought about by chemical treatment of the exposed wood, especially with paraquat herbicides (dipyridyl compounds). This treatment stimulates extensive oleoresin formation and diffusion into the wood, extending to the pith of the tree and several feet above the treatment level. As much as 40 percent oleoresin content in the wood has been produced. Such treatment could double naval stores production, for both gum and sulfate processes. It also has the potential of providing a new type of wood naval stores by solvent extraction prior to kraft pulping, or a combination of both methods.

Wood Naval Stores

"Wood" naval stores are produced by solvent extraction of resin-rich wood from old southern pine stumps and roots. The depletion of these stumps from the large trees of virgin forests, combined with high labor costs, has brought about a major decline in the production of wood naval stores by this process.

In modern practice, all the resin products are removed from the shredded wood by solvent extraction; the solvent retained by the extracted wood chips is recovered by steaming. Extraction is carried out with naphtha (b.p. 90115°C fraction). Multiple extraction is carried out in a series of vertical extractors in a countercurrent manner, whereby fresh solvent is used for the final extraction of a charge.

The solution from the extractors is vacuum-distilled and the solvent recovered. The remaining terpene oils are fractionally re-distilled under vacuum and recovered as turpentine, dipentene, and pine oil. The nonvolatile rosin is of dark color and is upgraded by clarification methods, such as selective absorption of its solution (bed-filtering).

Sulfate Naval Stores

Sulfate turpentine is obtained as a by-product during the kraft pulping of pine woods. Vapors periodically released from the top of the digesters are condensed, and the oily turpentine layer is separated and purified by fractional distillation and treatment with chemicals to remove traces of sulfur compounds. Sulfate turpentine is very similar to gum turpentine obtained from the oleoresin of the tree and contains 6070 percent α-pinene, 2025 percent β-pinene, and 612 percent other pinenes. Sulfate turpentine from western North America woods contains appreciable amounts of Δ3-carene, which is used as a solvent. In the United States, about 80 percent of the annual production of turpentine is from the kraft pulping of southern pines (approximately 125 million liters).

The spent black liquor from the kraft pulping of pines contains the less volatile products of the wood resin in the form of sodium salts or soaps. The liquor first is concentrated in multiple-effect evaporators, and then the concentrate is sent to settling tanks. The soaps rise to the surface, are skimmed off, and then are acidified with sulfurous or sulfuric acid. The crude tall oil rises to the top and is mechanically separated. Crude tall oil from southern pines contains 4060 percent resin acids and 4055 percent fatty acids with 510 percent neutral substances. These components are separated by fractional distillation under vacuum. One metric ton of crude tall oil yields about 350 kg of rosin, 300 kg of fatty acids, and 300 kg of head and pitch fractions. For each metric ton of pulp produced, northern pines yield about 50 kg of tall oil, and the southern pines yield about 125 kg. The U.S. capacity for fractional distillation of tall oil is nearly one million metric tons per year.

The rosin component of tall oil is mostly made up of resin acids, which are diterpene derivatives. The major compounds (figures indicate averages) are as follow:

1. Abietic typeabietic (32%), neoabietic (4%), palustric (10%), and dehydroabietic(30%) acids.
2. Pimaric typepimaric (4%), isopimaric (10%), and sandara copimaric acids (small amounts).

The fatty acids from tall oil have the following components: oleic (50%), linoleic (35%), conjugated linoleic (8%), stearic (2%), palmitic (1%), and others (4%). From tall oil heads, a fraction is produced that is composed predominantly of saturated acids, containing 55 percent palmitic acid. Oleic acid is distilled in grades that are 99.5 percent pure.

The neutral or "unsaponifiable" materials present in tall oil include anhydrides, phenolics, diterpene aldehydes and alcohols, stilbenes, and steroids. In the neutral fraction of southern pine tall oil soap, 80 compounds have been identified. They include 25.1 percent sistosterol and a total of 32.4 percent steroids. The sistosterol content of crude tall oil is 23 percent and is the main component of the neutral fraction.

Uses of Naval Stores Products

Naval stores products have a wide range of uses from ordinary household commodities to complex industrial applications, as outlines below.42

Turpentine

Turpentine is used directly as a solvent, thinner, or additive for paints, varnishes, enamels, waxes, polishes, disinfectants, soaps, pharmaceuticals, wood stains, sealing wax, inks, and crayons, and as a general solvent. The chemistry of its monoterpenes offers many possibilities for conversion to other substances, as illustrated in Fig. 7.24. There is increasing use of turpentine to produce fine chemicals for flavors and fragrances. An important use of turpentine is in conversion by mineral acids to synthetic pine oil. It also is a raw material for making terpin hydrate, resins, camphene, insecticides, and other useful commodities. These uses are included in the following summary of its applications:

1. solvents for paints, etc. (11%)
2. synthetic pine oil (48%), used for mineral flotation, textile processing, solvents, odorants, bactericides, and conversion to their chemicals such as terpin hydrate, etc.
3. polyterpene resins (6001500 mol. wt.) (16%), used for adhesives, pressure-sensitive sizes (dry cleaning, paper, chewing gum)
4. camphene → toxophene insecticides (16%)
5. flavor and fragrance essential oils (9%)

Dipentene

Dipentene is present in the higher-boiling fractions of wood turpentine. It is used in paints and varnishes and as a penetrating and softening agent in rubber reclamation.

Pine Oil

Pine oil obtained from wood naval stores has similar uses to those of the synthetic pine oil made from turpentine, given above.

Rosin

Rosin is used mainly in some modified form. Because the abietic-type acids in rosin each contain a carboxyl group and double bonds, they are reactive and can be used to produce salts, soaps, esters, amines, amides, nitriles, and DielsAlder adducts; and they can be isomerized, disproportionated, hydrogenated, dimerized, and polymerized. When destructively distilled, rosin produces a viscous liquid, termed rosin oil, used in lubricating greases.

The paper industry uses large amounts of the sodium salt of rosin as paper size, which accounts for the greatest single use of rosin. The synthetic rubber industry is the second most important user of rosin. In making styrene-butadiene rubber, disproportionated rosin soaps are used alone or in combination with fatty acid soaps as emulsifiers in the polymerization process. Disproportionation decreases the number of double bonds in the abietic acid of the rosin, making a more stable material.

The adhesives industry is the third most important market for rosin. Rosin, modified rosins, and rosin derivatives are used in several types of adhesives, including the pressure-sensitive, hot-melt, and elastomer-based latices, and solvent rubber cements.

Protective coatings are the fourth major user of rosin, either directly or in a modified or derivative form. Varnishes and alkyds are the most common types of protective coatings using rosin. Rosin is combined with a heat-reactive phenol-formaldehyde resin to produce a widely used varnish. Printing inks also use substantial amounts of rosin.

The above rosin uses are summarized as follows:

1. paper size (33%)
2. chemical intermediates and rubber (42%)
3. resins and ester gums (18%)
4. coatings (3.9%)
5. other uses (3.1%)

Fatty Acids

Of the total fatty acids produced annually in the United States, amounting to more than 450,000 tons, 35 percent come from tall oil. The solvent extraction of pine wood yields only 1 percent fatty acids and their esters. The yield is not increased, however, by paraquat (dipyridyl herbicides) treatment. Hence, the kraft-pulping industry will continue to be the major source of fatty acids from wood. The approximate distribution of uses of the fatty acids are:

1. intermediate chemicals (43%)
2. protective coatings (28%)
3. soaps and detergents (11%)
4. flotation (3%)
5. other uses (15%)

A future product from the neutral fraction of tall oil may be sistosterol. This chemical has potential use in the synthesis of cortisones and other steroids and hormones by fermentation processes.

7.10 Additional Chemicals from Wood

A variety of additional chemicals are derived directly from wood or as by-products from the pulp and paper industry.

Tannins and Other Extractives

The tissues of wood, bark, and the leaves of trees contain a great variety of chemical substances of considerable scientific interest and some of practical value. Turpentine, pine oil, and rosin from the resins of pines are the most important commercial extractives from American woods.

Tannin is a commercially important substance that can be extracted from the wood, bark, or leaves of certain trees and other plants. Tannins are complex dark-colored polyhydroxy phenolic compounds, related to catechol or pyrogallol, and vary in composition from species to species. They have the important property of combining with the proteins of animal skins to produce leather.

For many years, most of the leather in the United States was tanned with domestic tannins from hemlock and oak bark and from chestnut wood. Today only a small amount of tannin comes from these and other domestic sources. The most important source of vegetable tannin today is the wood of the quebracho tree, which grows mainly in Paraguay and Argentina. The tannin content of this tree and a few other sources of vegetable tannin are shown in Table 7.10.

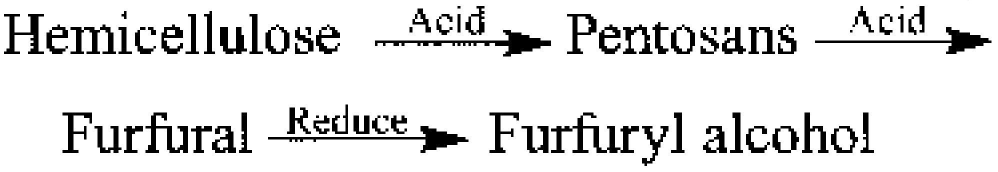
TABLE 7.10 Tannin Content of Some Plant Material

Plant Material	Percent Tannin
Domestic sources	
Eastern hemlock bark	913
Western hemlock bark	1020
Tanbark oak	1516
Chestnut oak	1014
Black oak	812
Chestnut wood	415
Sumac leaves	2532
Foreign sources	
Quebracho heartwood	2030
Mangrove bark	1542
Wattle (*acacia* bark)	1550
Myrobalan nuts	3040
Sicilian suma leaves	2530

The wood or bark for tannin production is reduced to chips and shreds by passing the material through hoggers or hammer mills. Then it is extracted with warm water in diffusion batteries. The dilute solutions are evaporated to the desired concentration. Loss of solubility of the tannin can be counteracted by treatment of the concentrate with sodium sulfite.

Furfural

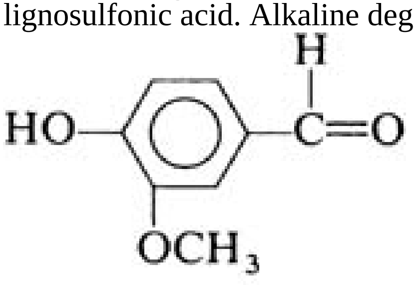
An additional, potentially important chemical derivable from wood as a result of hydrolytic (acid) treatment is furfural. Furfural is derived, from the hemicellulose fraction of woodspecifically from the five-carbon or pentose sugars (primarily xylose). The pentose sugars are not yeast-fermentable by standard methods to ethanol but can be treated with acid, which causes dehydration and yields furfural.



Most furfural is produced from corncobs and oat and rice hulls, primarily by the Quaker Oats Company. The product is used in the chemical industry as a solvent and in wood rosin refining. A large amount of furfural is treated further to give furfuryl alcohol. The furfuryl alcohol is added to urea-formaldehyde resins in applications for adhesives and foundry core binders.

Vanillin

Vanillin is not a product of hydrolysis or of fermentation; in fact, it does not originate from the holocellulose, but rather than the lignin portion of the wood. The major organic material in sulfite waste liquor (spent liquor, after pulping wood by the sulfite process) is the lignin dissolved from the wood as lignosulfonic acid. Alkaline degradation of this lignin product produces vanillin, the same substance that occurs naturally in the vanilla bean:



Vanillin

With the Howard process, 510 percent vanillin is produced, based on the lignin in the waste sulfite liquor. Some vanillin is produced from sulfite waste liquor in both the United States and Canada.

Dimethyl Sulfide and DMSO

Dimethyl sulfide (DMS) is recovered from the waste black liquor in the kraft pulping process by flash drying of the liquor after the DMS content has been increased by additions of sulfur. DMS is an odorant for natural gas and a solvent. DMS is oxidized to give dimethylsulfoxide (DMSO), an important solvent and a medicinal of interest in the treatment of arthritis.

Medicinals

The bark of the cascara tree of the northwestern region of the United States yields cascara, a laxative used in medicine. Several hundred tons of bark are harvested annually. One of the most promising anti-carcinogenic compounds to be discovered in the last 25 years was found in the bark and wood of the Pacific Yew tree from the Pacific Northwest of the United States. This compound, taxol, was found to be quite effective by the National Cancer Institute for the treatment of breast and ovarian cancers. The compound is now produced by partial synthetic routes and is available commercially under the generic name, paclitaxel.

The red gum tree of the southern United States exudes a yellowish balsamic liquid or gum from wounds, which is known as storax. It is produced by removing a section of bark and incising the wood in much the same manner as that used for the production of naval stores gum described above. Storax is used in medicinal and pharmaceutical preparations, such as adhesives and salves, and as an incense, in perfuming powders and soaps, and for flavoring tobacco.

Biotechnology Chemicals

As with pulping and bleaching, biotechnology too could have a considerable impact on the production of chemicals from wood and other forms of plant biomass. The effects of biotechnology probably will be noticed first in areas of enzymatic hydrolysis of polysaccharides and fermentation technology.

It should be possible to improve the efficiency of the cellulose enzyme complex for hydrolyzing cellulose to glucose. The enzyme complex apparently contains decrystallizing and hydrolysis enzymes that work together to convert cellulose to glucose. Isolation of the specific enzymes and genetic engineering could provide a more efficient complex.

As discussed earlier in this chapter, enzymes are the basis for yeast conversion of hexose (six-carbon) sugars, such as glucose and mannose, to ethanol. These enzymes also could be genetically engineered to improve the efficiency of alcohol production; several biotechnology firms are exploring this possibility.

The yeast (*Saccharomyces cerevisiae*) enzymes are specific to six-carbon sugars, but wood and other forms of biomass also contain large quantities of pentose sugars, especially xylose in hardwoods. The pentoses are not fermentable to ethyl alcohol with conventional yeasts. However, researchers at the U.S. D.A. Forest Products Laboratory have discovered a xylose-fermenting yeast (*Candida tropicalis*). Thus it now is possible to convert all wood sugars to ethyl alcohol with a combination of yeasts. Isolation of the specific enzymes and genetic engineering of the enzymes could dramatically improve the efficiency of this conversion.

Many other chemicals can be obtained from both yeast and bacteria fermentation of sugars and pulp mill effluents. Potential fermentation products from wood hydrolysates include acetone, organic acids (acetic, butyric, lactic), glycerol, butanediol, and others.42

< previous page

next page >

7.11 References

Young, R. A., and Giese, R. L. (Eds.), Introduction to Forest Science, 2nd ed., John Wiley & Sons, New York, 1990; see also 3rd ed., Introduction to Forest Ecosystem Science and Management, 2004.
Anon., Wood Handbook: Wood as an Engineering Material, Gen. Tech. Rept., FPL-GTR-113, U.S.D.A. For. Serv., Forest Products Laboratory, Madison, WI, 1999.
Anon., Wood Colors and Kinds, U.S.D.A. For. Serv., Forest Products Laboratory, U.S. Govt. Printing Office, Washington, DC, 1956.
Kubler, H., "Properties and Utilization of Wood," in Introduction to Forest Science, 2nd ed., R. A. Young and G. L. Giese (Eds.), pp. 451477, John Wiley & Sons, New York, 1990.
Young, R. A., McGovern, J. N., and Rowell, R. M., "Wood for Fiber, Energy and Chemicals," in Introduction to Forest Science, 2nd ed., R. A. Young and R. L. Giese (Eds.), pp. 478507, John Wiley & Sons, New York, 1990.
M. Kwok, http://member.netease.com/~mitchell/pulp/96/wpdp.html:wpp.html
Kleinert, T. N., and Tayenthal, U.S. Patent 1,856,567, 1931.
Young, R. A., and Akhtar, M. (Eds.), Environmentally Friendly Technologies for the Pulp and Paper Industry, John Wiley & Sons, New York, 1998.
Klausmeier, W. H., "Economic Analysis of an Innovative Process for Cellulose Acetate Production," in Cellulose: Structure, Modification and Hydrolysis, R. A. Young and R. M. Rowell (Eds.), pp. 187221, WileyInterscience, New York, 1986.
Herdle, C. E., Pancoast, L. H., and MacClaren, R. H., Tappi, 47(10), 617 (1964).
Young, R. A., Tappi, 72, 195 (1989).
Young, R. A., University of Wisconsin-Madison, unpublished results.
Springer, E. L., and Zoch, L. C., Tappi, 54(12), 2059 (1979).
Johansson, A., Sachetto, J. P., and Roman, A., "A New Process for the Fractionation of Biomass," Battelle-Geneva Labs, Geneva, Switzerland, 1982.
Clemont, L. P., Tappi 53(12), 2243 (1970).
Deltaes, G. G., and Lang, C. J., Tappi, 57(5), 127 (1974).
Kleinert, T. N., Tappi, 57(8), 99 (1974).
Katzen, R., Frederickson, R., and Brush, B. F., Chem. Eng. Progr., 76, 62 (1980).
Pazner, L., and Chang, P. C., International Symposium on Wood and Pulping Chemistry, Tsakuba Science City, Japan, 1983.
Edel, V. E., Dtsch Papier Wirtschaft, 1, 39 (1984).
Aziz, S., and Sarkanen, K. V., Tappi, 72(3), 167 (1989).
McCready, M. (Ed.), "Allcell Pulping; It Works!," Paper Age, 14 (Mar. 1991).
Young, R. A., "Processing of Agro-Based Resources into Pulp and Paper," in Paper and Composites from Agro-Based Resources, Rowell, R. and Young, R. A. (Eds.), Lewis Pub., CRC Press, New York, 1997.
Stier, J. C., and Buongiorno, J., "The Forest Products Economy," in Introduction to Forest Science, R. A. Young and R. L. Giese (Eds.), pp. 508531, John Wiley & Sons, New York, 1990.
Smulski, S. (Ed.), Engineered Wood Products: A Guide for Specifiers, Designers and Users, PFS Research Foundation, Madison, WI, 1997.
Haygreen, J., and Bowyer, J., Forest Products and Wood Science, 3rd ed., Iowa State Univ. Press, Ames, IA, 1996.
Meyer, J. A., "WoodPolymer Materials," in The Chemistry of Solid Wood, R. M. Rowell (Ed.), pp. 257290, Am. Chem. Soc., Washington, DC, 1984.
Rowell, R. M., "Penetration and Reactivity of Cell Wall Components," in The Chemistry of Solid Wood, R. M. Rowell (Ed.), pp. 175210, Am. Chem. Soc., Washington, DC, 1984.
Rowell, R. M., Tillman, A. M., and Simonson, R., J. Wood Chem. Technol., 6(3), 427 (1986).
Shiraishi, N., Aoki, T., Norimoto, M., and Okumuso, M., "Thermoplasticization of Cellulose and Wood," in Graft Copolymerization of Lignocellulosic Fibers, D. N.-S. Hon (Ed.), pp. 321348, Am. Chem. Soc., Washington, DC, 1932.
Matsuda, H., Wood. Sci. Technol., 21, 75 (1987).
Rowell, R., and Young, R. A. (Eds.), Paper and Composites from Agro-Based Resources, Lewis Pub., CRC Press, New York, 1997.
Rowell, R. M., "Chemical Modification of Wood," Commonwealth Forestry Bureau, Oxford, England, 6(12), 363382 (1982).
Rowell, R. M., "Nonconventional Wood Preservation Methods," in Wood Technology: Chemical Aspects, Am. Chem. Soc., Symp. Ser. No. 43, 4756, Washington, DC, 1977.
LeVan, S. L., "The Chemistry of Fire Retardancy," in The Chemistry of Solid Wood, R. M. Rowell (Ed.), pp. 531574, Am. Chem. Soc., Washington, DC, 1984.
Wayman, M., "Comparative Effectiveness of Various Acids for Hydrolysis of Cellulose," in Cellulose: Structure, Modification and Hydrolysis, R. A. Young and R. M. Rowell (Eds.), pp. 265279, WileyInterscience, New York, 1986.
Connor, A. H., Wood, B. V., Hill, C. G., and Harris, V. F., "Kinetic Modeling of the Saccharification of Prehydrolyzed Southern Red Oak," in Cellulose: Structure, Modification and Hydrolysis, R. A. Young and R. M. Rowell (Eds.), pp. 281296, WileyInterscience, New York, 1986.
Evert, R. G. H., Katzen, R., Frederickson, R. E., and Kapisch, K. F., Chem. Eng. Progr., 76(9), 47 (1980).
Shafizadeh, F., "The Chemistry of Pyrolysis and Combustion," in The Chemistry of Solid Wood, R. M. Rowell (Ed.), pp. 489530, Am. Chem. Soc., Washington, DC, 1984.
Appell, H. R., "The Production of Oil from Wood Waste," in Fuels from Waste, Academic Press, New York, 1977.
Zinkel, D. F., J. Appl. Polym. Sci., 28, 309 (1975).
Hajny, G. J., Biological Utilization of Wood for Production of Chemicals and Foodstuffs, U.S.D.A. For. Serv. Res. Rept. FPL 385, Forest Products Laboratory, Madison, WI, 1981

7.12 Selected References

Books

American Wood Preservers' Association, American Wood Preservers' Standards, Washington, DC, 1971.
Bierman, C. J., Essentials of Pulping and Papermaking, Academic Press, New York, 1993.
Browning, B. L., Methods of Wood Chemistry, Vols. I and II, John Wiley & Sons, New York, 1967.
Casey, J. P., Pulp and Paper, 2nd ed., Vols. I, II and III, John Wiley & Sons, New York, 1980.
Fengel, D., and Wegener, G., Wood: Chemistry, Ultrastructure, Reactions, Walter de Gruyter, New York, 1984.
Glasser, W. G., and Sarkanen, S. (Eds.), Lignin: Properties and Materials, Am. Chem. Soc. Symp. Ser. 397, Am. Chem. Soc., Washington, DC, 1989.
Glasser, W., Northey, R., and Schultz, T. (Eds.), Lignin: Historical, Biological and Material Perspectives, ACS Symp. Ser. 742, Am. Chem. Soc., Washington, DC, 2000.
Goldstein, I. (Ed.), Organic Chemicals from Biomass, CRC Press, Boca Raton, FL, 1981.
Haygreen, J., and Bowyer, J., Forest Products and Wood Science, 3rd ed., Iowa State Univ. Press, Ames, IA, 1996.
Hillis, W. E. (Ed.), Wood Extractives, Academic Press, New York, 1962.
Hoadley, R., Identifying Wood: Accurate Results with Simple Tools, Taunton Press, Newtown, CT, 1990.
Hon, D. N.-S., and Shiraiski, N., Wood and Cellulosic Chemistry, Marcel Dekker, New York, 1991.
Hunt, G. M., and Garratt, G. A., Wood Preservation, McGraw-Hill, New York, 1953.
MacDonald, R. G. (Ed.), Pulp and Paper Manufacture, 2nd ed., Vols. I, II and III, McGraw-Hill, New York, 1969.
Rowell, R. M. (Ed.), The Chemistry of Solid Wood, Am. Chem. Soc., Washington, DC, 1984.
Rowell, R., and Young, R. A. (Eds.), Paper and Composites from Agro-Based Resources, Lewis Pub., CRC Press, New York, 1997.
Rydholm, S. A., Pulping Processes, John Wiley & Sons, New York, 1965.
Sarkanen, K. V., and Ludwig, C. H. (Eds.), Lignins, John Wiley & Sons, New York, 1971.
Sjostrom, E., Wood Chemistry: Fundamentals and Applications, Academic Press, New York, 1981.
Smulski, S. (Ed.), Engineered Wood Products, A Guide for Specifiers, Designers and Users, PFS Research Foundation, Madison, WI, 1997.
Young, R. A., and Giese, R. L. (Eds.), Introduction to Forest Science, 2nd ed., John Wiley & Sons, New York, 1990.
Young, R. A., and Giese, R. L. (Eds.), Introduction to Forest Ecosystem Science and Management, 3rd ed., John Wiley & Sons, New York, 2004.
Young, R. A., and Rowell, R. M., Cellulose: Structure, Modification and Hydrolysis, WileyInterscience, New York, 1986.
Young, R. A., and Akhtar, M. (Eds.), Environmentally Friendly Technologies for the Pulp and Paper Industry, John Wiley & Sons, New York, 1998.

Articles and Chapters

Akhtar, M., Blanchette, R., Myers, G., and Kirk, T., "An overview of Biomechanical Pulping," in Environmentally Friendly Technologies for the Pulp and Paper Industry, R. A. Young and M. Akhtar (Eds.), John Wiley & Sons, New York, 1998.
Chum, H. L., "Structural Materials for the Automobile of the FutureComposite Materials," in Assessment of Biobased Materials, H. L. Chum (Ed.), pp. 1.11.19, Solar Energy Res. Inst., Rpt. No. SERI/TR-234-3650, Dept. of Energy, Golden, CO, 1989.
Hegert, H., "Developments in Organosolv PulpingAn Overview," in Environmentally Friendly Technologies for the Pulp and Paper Industry, R. A. Young and M. Akhtar (Eds.), John Wiley and Sons, New York, 1998.
Jahn, E. C., and Stannett, V., "Polymer Modified Papers," in Modern Materials, Vol. II. H. H. Hauser (Ed.), Academic Press, New York, 1960.
Johansson, A., Aaltonen, O., and Ylinen, P., "Organosolv PulpingMethods and Pulp Properties," Biomass, 13, 45 (1987).
Levan, S. L., Ross, R. J., and Winandy, J. E., Effects of Fire Retardant Chemicals on the Bending Properties of Wood at Elevated Temperatures, U.S.D.A. For. Serv. Res. Rept. FRL-RP-498, Forest Products Laboratory, 1990.
Nelson, P., "Elemental Chlorine Free (ECF) and Totally Chlorine Free (TCF) Bleaching of Pulps," in Environmentally Friendly Technologies for the Pulp and Paper Industry, R. A. Young and M. Akhtar (Eds.), John Wiley & Sons, New York, 1998.
Sanadi, A., Caulfield, D., and Jacobsen, R., "Agro-Fiber/Thermoplastic Composites," in Paper and Paper Composites from Agro-Based Resources, R. Rowell and R. A. Young (Eds.), Lewis Pub., CRC Press, New York, 1997.
Shiraishi, N., Aoki, T., Norimoto, M., and Okumura, M., "Make Cellulosics Thermoplastic," Chemtech, 336 (June, 1983).
Young, R. A., "Structure, Swelling and Bonding of Cellulose Fibers," in Cellulose: Structure, Modification and Hydrolysis, R. A. Young and R. M. Rowell (Eds.), pp. 91128, WileyInterscience, New York, 1986.

8.1 Introduction

Fats and oils predominantly are triesters (triacylglycerols (TAG), triglycerides) of glyecrol and aliphatic fatty acids containing up to 22 carbon atoms. Waxes are esters of long-chain fatty acids, usually containing 2428 carbon atoms, with long-chain primary alcohols (1636 carbon atoms) or with alcohols of the steroid group.1

Fats and oils are members of a broader group of chemical substances called *lipids*, which has been classified by the National Research Council into: (1) *nonpolar lipids*including esters of fatty acids (triacylglycerols and cholesteryl esters) that are virtually insoluble in water, but soluble in most organic solvents and enter metabolic pathways only after hydrolysis; and (2) *polar* or *amphipathic lipids*including fatty acids, cholesterol, sphingolipids, and glycerolphospholipids (mainly lecithins). The term *phospholipids* (phosphatides) includes lecithins and sphingomyelins.2 Other minor natural compounds, also extracted by low-polarity organic solvents, include fat-soluble vitamins, colors, and flavors.

Fats and oils have major roles in human nutrition. They are concentrated dietary sources of energy, providing approximately 9 kcal/g when metabolized compared with 4 kcal/g for carbohydrates and proteins, and account for about 36 percent of domestic caloric intake per capita.2 Dietary lipids also can provide essential molecular structures that are synthesized by the body into compounds required for selective functioning of cell membranes and regulation of life processes.

Fats and oils modify product texture in the preparation of foods; act as heat transfer media in food frying; carry flavors, colors, and oil-soluble vitamins; improve mouthfeel; provide a sensation of product richness; and induce satiety. They are used as energy sources in feeds for domesticated animals, and as components of many industrial products, including soaps and detergents, lubricants, plastics and protective coatings, and printing inks, and as carriers of pesticides for aerial spraying, for controlling grain dust, and as feed stocks for manufacturing chemicals. Considerable public interest has developed in the last decade in replenishable biodegradable carbon sources and in liquid fuels such as biodiesel.

Although many chemistry and processing principles have long been established, the industry has undergone major changes in the last three decades, with most starting in the mid-1980s. Rising costs of energy have led to more efficient equipment designs and to the installation of heat recapture systems throughout modern extraction plants and refineries. In the United States, Occupational Safety and Health Environmental Protection Agency (OSHA) regulations to prevent injuries and protect the health of workers have led to increased use of safety guards, dust collection systems with shrouding of equipment, and improved ventilation, oversight of workers entering dangerous areas such as bins, and periodic checks for hearing loss. Redesigning and retrofitting the equipment was expensive. Process control computers became available concurrently in the early 1990s, and many companies chose instead to install robots in health- and safety-risk areas, and to automate processes for operation from control rooms. The few people now seen on extraction plant refinery floors are mainly repair and cleaning personnel, with hardly any "operators." With computers making process adjustments, *product quality*, defined as *uniformity*, typically has improved.

Furthermore, Environmental Protection Agency (EPA) air emissions regulations have led to increased dust controls and to reduced solvent losses in extraction plants. Regulations on the discharge of polluting process streams into public waterways have led to containment, and treatment where required, of rain runoff from grain storage and processing propertiesand even from employee parking lots. Silica gel adsorption (*modified caustic refining*) processes have been developed to avoid production of wash waters in refining oils, and problems associated with their disposal. Similar changes are being adopted at various rates throughout the world.

The effects of economic development loans and private investments in developing countries in the 19601990 era have matured and are changing the global fats and oils industry. The growing world population requires 1.52.0 million metric tons more fats and oils each year, in addition to the accompanying oilseed feed and food proteins. World production of soybeans has increased by 6.8 times since 1960, with approximately 45 percent of the world's crop now entering global trade as soybeans, meal, and oil.3 The growth of palm oil production has been even more spectacular. Hardly known as a crop after World War II, its production increased by nearly 14 times since 1970, and is projected to surpass soybean oil within the next several years; approximately 70 percent enters world trade. This has brought equatorial countries such as Malaysia and Indonesia into the group of leading edible oil producing and exporting nations. The establishment of trading rules, product definitions and standards, and analytical procedures for the erupting world market also has been part of the technical challenge.

Strong competition in the world oil market, and concerns about the environment, have refocused interest in the United States on nonfood, nonfeed uses of crops, including biodegradable applications, renewable hydrocarbon sourcesincluding liquid fuels, and *chemurgy*the use of agricultural crops as chemicals feed stocks. Henry Ford's Edison Institute, and General Electric, DuPont were among the better known of this early movement, which began in the early 1930s and lasted until the start of World War II.4 Nor is the United States alone in the current movement. Germany, Hungary, France, and other European countries, whose climates are too cold for raising soybeans or oil palm but can raise rapeseed, recently have installed biodiesel production facilities.

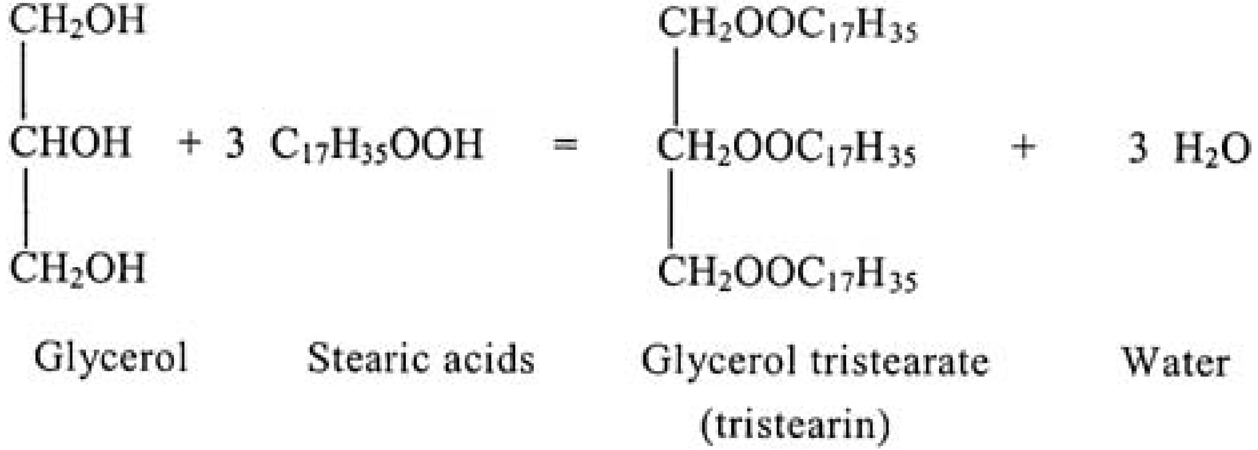
Anyone who says "nothing is new" in the oils and fats industry, doesn't live in it. This chapter cannot hope to summarize advances in all related technologies. It concentrates on major current practices, and leaves the reader to realize that the speeds of implementation differ within specific countries, and to browse the Internet for commercial information and computer-based abstract services for research reports. Much of the world's research on fats and oils extraction, processing, and utilization is reported in the *Journal of the American Oil Chemists' Society (JAOCS)* published by the AOCS Press, an activity of AOCS (An International Society for the Science and Technology of Fats, Oils and Related Materials), Champaign, Illinois. The AOCS Press also publishes proceedings of the Society's international conferences, a broad variety of related books, and the following journals: *INFORM*a monthly business, news, and scientific publication addressed to professionals interested in the science and technology of fats and oils, surfactants, detergents, proteins, oleochemicals, and related substances; *Lipids*a monthly journal on basic chemistry and nutrition of lipids; *Journal of Surfactants and Detergents (JDS)*a quarterly science and news journal on the practical and theoretical aspects of oleochemical and petrochemical surfactants, soaps, and detergents. The Technical Services function of AOCS establishes, revises, and annually updates the *Official Methods and Recommenced Practices of the American Oil Chemists Society* (AOCS Methods)5 for fats, oils, and soap technology; *Spanish AOCS Methods*a Spanish translation of the more commonly used AOCS Methods; and *Physical and Chemical Characteristics of Oils, Fats and Waxes*. Leaders of the methods development committees coordinate closely with AOAC International (formerly the Association of Official Analytical Chemists); AOCS Methods are recognized as "official methods" in U.S. Food and Drug Administration (FDA) activities and when litigation becomes necessary in industry trade. Additionally, the Technical Services function operates a Laboratory Proficiency Program (formerly the *Smalley Check Sample Program*) and oversees the distribution and statistical analysis of 30 different series of basic laboratory quality assurance/quality control test samples. Certification as AOCS Approved Chemists or AOCS Certified Laboratories (for referee analysis) requires successful participation in the Laboratory Proficiency Program.

Comments on fats and oils in nutrition and human health are limited in this chapter. Discoveries in fats nutrition are frequent, and interpretations often change. Even less is said about "nutraceuticals," a class of products often with relatively little documentation of efficacy in human trials, whose marketing was allowed by the government. The products were expected to be occasional food supplements, but have become the leading growth market of the decade. The reader should be aware that many of the extraction, concentration and purification techniques, and equipment used in the preparation of nutraceuticals originated in the vegetable oils and proteins processing industry.

8.2 Nomenclature and Molecular Structures

General

The choice of the term "oil" or "fat" usually is based on tradition and the physical state of the material. Generally, oils are liquid at ambient temperatures, and fats are semisolid mixtures of crystals in oil. Fats usually are of animal origin (beef tallow, pork lard, and butter fat) or hardened (hydrogenated, interesterified, or thermally fractionated) vegetable oils, while "oils" are extracted from fish or plant tissue or seeds. Nutritionists generally use the term "fat." Over 95 percent of the weight of most extracted/separated food fats are TAG (triacylglycerols or triglycerides) formed by the enzymatic combination of glycerol (a trihydric alcohol) with three fatty acids also yielding one molecule of water for each ester linkage:



$$\begin{array}{l} CH_2OH \\ | \\ CHOH \\ | \\ CH_2OH \end{array} + 3\ C_{17}H_{35}OOH = \begin{array}{l} CH_2OOC_{17}H_{35} \\ | \\ CH_2OOC_{17}H_{35} \\ | \\ CH_2OOC_{17}H_{35} \end{array} + 3\ H_2O$$

The reaction is reversible and favored by the presence of moisture and catalysts including lipases, alkalis, and alkaline metals. In the oleochemicals industry, TAG are split by high pressure steam. Unassociated fatty acids are called *free fatty acids or FFA*.

Fatty Acids

Fatty acids are the building blocks of TAG. More than 90 percent of fatty acids have an even number of carbon atoms, and are in aliphatic chains ranging from 4 to 24 carbons in length. Trace amounts of odd-number carbon fatty acids are found in most fats, and also have been synthesized for research purposes. Microorganisms frequently produce odd-number carbon fatty acids, with heptadecenoic (17 carbon) acid a major component of *Candida tropicalis* yeast fat. Up to 8 percent C17 fatty acids have been found in milk and meat fats of ruminants (cattle, sheep, goats) and are of rumen microbe origin.

The names of common fatty acids under several conventions, carbon numbers, and selected properties are shown in Table 8.1.

TABLE 8.1 Names and Characteristics of Some Important Fatty Acids

Carbon Atoms & Abbreviations	Common Name	Symbol	Systematic Name	Melting Point (° C)	Iodine Value	Common Sources
Saturated Fatty Acids						
3:0	Propionic		Propanoic	-20.8		Bacterial fermentation
4:0	Butyric	B	Butanoic	-7.9		Milk fats
5:0	Valeric		Pentanoic	-33.8		Bacterial fermentation
5:0	Isovaleric		3-Methylbutanoic	-51.0		Dolphin and porpoise fats
6:0	Caproic	H	Hexanoic	-3.4		Milk fats, some seed oils
8:0	Caprylic	OC	Octanoic	16.7		Milk fats, *Palmae* seed oils
10:0	Capric	D	Decanoic	31.6		Sheep and goat milk, palm seed oils, sperm head oil
12:0	Lauric	La	Dodecanoic	44.2		Coconut-oil
14:0	Myristic	M	Tetradecanoic	54.4		Palm and coconut oils
16:0	Palmitic	P	Hexadecanoic	62.9		Palm oil, most oilseeds and animal fats
18:0	Stearic	St	Octadecanoic	69.6		Animal fats
19:0	Tuberculostearic		10-Methylstearic	11.0		Tubercle bacillus lipids
20:0	Arachidic	Ad	Eicosanoic	75.4		Some animal fats
22:0	Behenic		Docosanoic	81.0		Peanut and various other seed oils
24:0	Lignoceric		Tetracosanoic	84.2		Minor amounts in some seed oils
26:0	Cerotic		Hexacosanoic	87.8		Plant waxes
28:0	Montanic		Octacosanoic	90.9		Beeswax and other waxes
30:0	Mellisic		Triacontanoic	93.6		Beeswax and other waxes
Unsaturated Fatty Acids						
10:1	Caproleic		9-Decenoic		149.1	Milk fats
10:2	Stillingic		2,4-Decadienoic			*Stillingia* oil
12:1	Lauroleic		2-Dodecenoic		128.0	Butterfat
14:1	Myristoleic		9-Tetradecenoic	18.5	112.1	Some feed fats, milk fats
16:1 (*n* 7)	Palmitoleic		9-Hexadecenoic	0.5	99.8	Many fats and marine oils
16:3	Hiragonic		6,10,14-Hexadecatrienoic			Sardine oil
17:1			9-Heptadecenoic	14.0		*Candida tropicallis* yeast
18:1 (t-oleic)	Elaidic		9-Octadecenoic	43.7		Butterfat
18:1 (*n* 9)	Oleic		9-Octadecenoic	16.3	89.9	Almost all fats and oils
18:1	Petroselinic		6-Octadecenoic	3033		Parsley seed oil
18:1 (*n* 7)	Vaccenic		11-Octadecenoic	44.0		Butterfat, seed oils
18:2 (*n* 6)	Linoleic	Lo	9,12-Octadecadienoic	-6.5	181.0	Most vegetable oils
18:3 (*n* 6)	(gamma) Linolenic		6,9,12-Octadecatrienoic			(Omega-6); Evening primrose, borage, vegetable oils
18:3(*t*)	Eleostearic		9,11,13-Octadecatrienoic			Tung oil
18:3 (*n* 3)	(alpha)Linolenic	Ln	9,12,15-Octadecatrienoic	-12.8	273.5	(Omega-3); Linseed, soybean, canola, other vegetable oils
20:1	Gadoleic		11-Eicosenoic	2324	81.8	Some fish oils
20:3 (*n* 9)	Eicosatrienoic		5,8,11-Eicosatrienoic			Brain phospholipids
20:3 (*n* 6)	Dihomo-γ linolenic		8,11,14-Eicosatrienoic			Shark liver oil
20:5 (*n* 3)	EPA		5,8,11,14,17-Eicosapentanoic	-33.5	75.0	Fish, plants
22:1	Erucic	E	13-Docosenoic	34		Rapeseed oil

22:2		13,16-Docosadienoic			
22:5 (*n* 3)	DPA	7,10,13,16,19-Docosapentaneoic			Fish oils
22:6 (*n* 3)	DHA	4,7,10,13,16,19-Docosahexaenoic			Cooler climate fish oils
Fatty Acids of More Unusual Structure					
18	Chaulmoogric	13, (2-Cyclopentenyl) tridecanoic	68.5	90.5	*Chaulmoogra* oil
18	Malvalic	8,9-Methylene-8-heptadecenoic			*Malvaceae* seeds, cottonseed
18	Ricinoleic	12-Hydroxy-9-octadecenoic	5.5	86.0	Castor oil
18	Vernolic	12,13-Epoxy-9-octadecenoic	3031		Some *Compositae* seeds
19	Sterculic	9,10-Methylene-9-octadecenoic			*Sterculiaceae* seeds, cottonseed
20	Arachidonic	5,8,11,14-Eicosatetraenoic	-49.5	333.5	Lard
20		5,8,11,14,17-Eicosapentaenoic			Some fish oil
20	Lesquerolic	14-Hydroxy-11-Eicosenoic			*Lesquerella* seed oil
2:22		4,7,10,13,16,19-Docosahexaenoic			Some fish oil

The common (trivial) names of some fatty acids are of long standing, and often indicate the initial source studied. As examples: butyric acid is a major component of butter flavor; the 6, 8, and 10 saturated fatty acids have been called the *goaty acids* because they impart the characteristic flavors of goat and sheep milk and cheese; and the terms *olein* and *stearin* were applied to the liquid and solid fractions, respectively, of tallow separated by pressing in early manufacture of oleomargarine and compounded shortenings. The iodine value (IV) is an indicator of the unsaturation of a fatty acid or fat/oil. It is determined by the AOCS Method5 Tg 1a-64 or Cd 1-25, respectively; the higher the IV, the more unsaturated the product.

Fatty acids sometimes are designated by the number of carbon atoms in the chain, followed by a colon with additional numbers indicating the number of double bonds. In the 18-carbon series, C18:0, C18:1, C18:2, and C18:3 represent *stearic, oleic, linoleic,* and *linolenic* acids, respectively; one or two letter abbreviations sometimes are used, with these four acids designated by St, O, L, and Ln, respectively.

Under the most common convention, fatty acids are named on the basis of the number of carbon atoms, starting with the terminal carboxyl (COOH) carbon as number "1." The terminal letter *e* of the respective alkane hydrocarbon is replaced with *oic* to indicate an acid; thus:

$CH_3CH_2CH_2CH_2CH_2CH_2CH_2CH_3$

Octane

8 7 6 5 4 3 2 1

$CH_3CH_2CH_2CH_2CH_2CH_2CH_2COOH$

Octanoic acid

The suffix *dioic* is used if the acid contains two carboxyl groups.

Occasionally in the literature, the carboxyl unit is regarded as a group substituted for hydrogen. In this case, the number 1 location is moved one position away from the reactive end, and the suffix *carboxylic* acid is added:

5 4 3 2 1

$CH_3CH_2CH_2CH_2CH_2COOH$

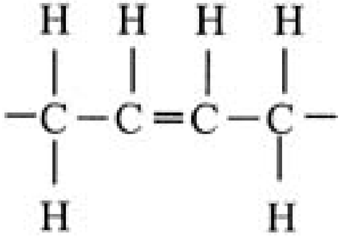
1-Pentanecarboxylic acid

(hexanoic acid)

(caproic acid)

A double bond between two carbon atoms indicates the site, and possibly type, of hydrogen unsaturation. When double bonds are present, the suffix *anoic* is changed to *enoic, dienoic,* or *trienoic* to indicate the number of bonds. The location of the first carbon in the double bond is indicated by a number preceding the systemic name. Under the International Union of Pure and Applied Chemistry (IUPAC) convention, stearic, oleic, linoleic, and linolenic acids are called octadecanoic, 9-octadecenoic, 9,12-octadecadienoic, and 9,12,15-octadecatrienoic acids, respectively.

The real-world, three-dimensional geometric configuration of hydrogens at double bonds is indicated on paper by the Latin prefixes *cis* (both hydrogens on one side) and *trans* (hydrogens across from each other). Linoleic acid, with the *cis* configuration in both double bonds, is called *cis*-9, *cis*-12-octadecadienoic acid. Most fatty acids occur in nature in the *cis* form. Oleic acid is in the *cis* configuration and the corresponding *trans* form is called *elaidic* acid.



$-\overset{H}{\underset{H}{C}}-\overset{H}{C}=\overset{H}{C}-\overset{H}{\underset{H}{C}}-$

cis

Increasing the number of double bonds lowers the melting point of the fatty acid from its fully or partially saturated form. Double bonds also are the sites of oxidation initiation on FFA and within TAG, and breakdown into aldehydes, ketones, other volatile compounds, and polymer-forming compounds. The presence of *trans* bonds has less effect on reducing fatty acid melting points than *cis* bonds.

The locations of the double bonds, and especially the last double bond in long-chain polyunsaturated fatty acids, are of special interest. Whereas chemists traditionally count with the carboxyl carbon (COOH) assigned "number 1;" biochemists and nutritionists assign "number 1" to the methyl carbon (CH3). Thus, linoleic acid (9,12-octadecadienoic acid), known as C18:2 to a chemist, carries the same trivial name for biochemists but is known as C18:2ω6 or C18:2*n*6 with ω (omega) or *n* signifying "count from the methyl carbon." From a nutrition viewpoint, four families of fatty acids (*n*2 7, *n*2 9, *n*2 6, and *n*2 3) exist. Members of the *n*2 7 and *n*2 9 families generally are synthesized by each species as needed, but members of the *n*2 6 and *n*2 3 families may be dietarily essential either because the species is unable to synthesize the fatty acid, or metabolic mechanisms are impaired in specific individuals. Plants, including plankton, are the ultimate source of dietary essential fatty acids. Humans and most animals are considered to require linoleic and linolenic acids (C18:2*n* 6 and C18:3*n* 3, respectively). Fish vary by species, with carnivorous members like the salmonoids (salmon and trout) requiring EPA (C20:5*n* 3; 5,8,11,14,17-eicosapentaenoic acid) and DHA (C22:6*n* 3; 4,7,10,13,16,19-docosahexaenoic acid) when raised in captivity.6

The contents of individual fatty acids can vary by over 100 percent in the fat of a species. Table 8.2 summarizes the general fatty acids contents of the major edible fats and oils, and Table 8.3 does likewise for industrial fats and oils.

TABLE 8.2 Fatty Acid Composition of Some Edible Oils and Fatsa

Source	<14:0	14:0	16:0	16:1	18:0	18:1	18:2	18:3	20:0	20:1	22:0	22:1	24:0	24:1
Almond oil		0.0	6.5	0.6	1.7	69.4	17.4							
Avocado oil			11.0	3.4	0.7	71.5	12.0	1.5						
Barley bran oil		0.5	10.8	0.2	1.0	17.8	55.3	4.4						
Borage oil			11.3		3.7	16.3	38.1	23.0	0.2	3.9		2.4		1.4
Buffalo gourd seed oil			11.8		3.5	21.9	60.6	0.0	0.0					

Butter fat	23.8	8.2	21.3	1.8	9.8	20.4	1.8	1.2						
Canola oilb			4.8	0.5	1.6	53.8	22.1	11.1	1.1	1.5	0.3	0.1	0.1	
Cherry pit oil			7.8	0.4	2.4	43.9	44.8	0.5	0.7					
Cocoa butter		0.1	25.4	0.2	33.2	32.6	2.8	0.1		0.0				
Coconut oil	58.7	16.8	8.2		2.8	5.8	1.8							
Corn oil	0.0	0.0	10.9		1.8	24.2	58.0	0.7						
Cottonseed oil		0.8	22.7	0.8	2.3	17.0	51.5	0.2						
Evening primrose oil			8.5		2.5	8.5	72.5	11.0						
Fish (manhaden) oil		9.6	20.5	12.6	3.3	11.0	0.7	1.6	0.3			0.8		
Grapeseed oil		0.1	6.7	0.3	2.7	15.8	69.6	0.1						
Illipe butter			23.7		19.3	43.3	13.7							
Lard	0.5	1.3	23.8	2.7	13.5	41.2	10.2	1.0		1.0				
Lupine oil			8.3		2.5	55.0	17.7	9.3						
Macademia nut oil		0.6	8.5	21.7	3.7	56.0	1.7			1.4				
Mango kernel oil			7.6		36.0	49.4	5.0	0.5	1.4					
Mustard seed oil		0.1	1.9	0.3	0.1	17.7	9.1	0.5	0.6	3.91	1.8	55.1	0.2	1.9
Okra seed oil		0.2	33.7	0.6	3.3	17.9	42.2	0.2	0.1		0.2			
Olive oil		0.0	11.0	0.8	2.2	72.5	7.9	0.6						
Palm oil	0.1	1.0	43.5	0.3	4.3	36.6	9.1	0.2		0.1				
Palm kernel oil	54.2	16.4	8.1		2.8	11.4	1.6							
Peanut oil		0.1	9.5	0.1	2.2	44.8	32.0			1.3			1.8	
Rapeseed oilc			1.7		0.9	12.3	12.7	7.6	1.2	5.8	0.9	59.4	0.5	1.6
Rice bran oil		0.7	16.9	0.2	1.6	39.1	33.4	1.6						
Safflower oil		0.1	6.2	0.4	2.2	11.7	74.1	0.4						
Safflower oild			4.8		1.3	75.3	14.2							
Sal seed oil			5.3		34.0	49.1	3.8	3.3	4.0					
Sesame oil			8.9	0.2	4.8	39.3	41.3	0.3		0.2				
Shea butter	1.7	0.1	4.4	0.1	38.8	43.5	4.9	0.3		0.0				
Soybean oil		0.1	10.3	0.2	3.8	22.3	51.0	6.8						
Sunflower oile			7.0		5.0	19.0	68.0	1.0						
Sunflower oilf			4.0		5.0	65.0	26.0							
Tallow	0.9	3.7	24.9	4.2	18.9	36.0	3.1	0.6		0.3				
Teaseed oil	0.1	0.1	17.5	0.5	3.1	49.9	22.2	0.7		1.0				
Tomato seed oil		0.2	15.0	0.5	4.4	21.9	50.8	2.3						
Walnut oil			7.0	0.1	2.0	22.2	0.4	52.9	10.4					
Wild cucurbit oil			19.0			34.0	47.0	15.0						

aThese are average values from recent years' crops;
bLow-erucic-acid variety rapeseed;
cHigh-erucic-acid variety;
dHigh-oleic variety;
eMaturing in cooler climates;
fMid-oleic sunflower flower oil.

TABLE 8.3 Fatty Acid Composition of Some Industrial Oils and Fatsa

Source	<14:0	14:0	16:0	16:1	18:0	18:1	18:2	18:3	20:0	20:1	22:0	22:1	22:2	24:1
Caster oilb			1.1	0.2	1.0	3.3	3.6	0.32	0.4					
Chinese tallow	1.3	2.1	65.0		4.4	22.5	0.8							
Crambe oil			2.0	0.4	0.4	16.9	8.6	6.4	0.5	3.2	2.0	57.2		1.4
Crepsis foetida oilc		0.1	4.8		2.9	4.3	27.8			0.3				
Croton oil	2.5	5.4	6.2	0.2	3.2	15.8	49.4	3.0	2.9	8.9	0.2	0.6		
Cuphea oil		76.4	7.8	2.4		0.7	5.9	6.9	0.1					
Jojoba oil			1.0			9.0				70.7		16.3		3.0
Lesquerella seed oil		1.4	1.4	2.0	17.6	7.9	13.2		52.5				2.7	
Linseed oil			5.3		4.1	20.2	12.7	53.3						
Meadowfoam oil										64.5		18.5	13.5	
Neatsfoot oil		0.7	16.9		2.7	64.4	2.3	0.7	0.1					
Oitica oild			7.0		5.0	6.0								
Rapeseede		0.1	2.6	0.3	0.9	11.2	12.8	8.6		7.5		48.1		
Rubber seed oil		0.2	19.1		17.8	24.5	30.5	2.4		0.1	0.9	0.4		
Stokes aster oilf			2.8		0.9	7.0	16.5							
Tall oil						50.0	7.0	41.0						
Tung oil			3.1		2.1	11.2	14.6	69.0						
Veronia seed oilg			2.7		1.3	2.0	8.8	0.4						
Whale oil		3.3	8.1	26.9	1.1	33.3				10.9		2.2		

aThese are average values from recent years' crops;
bContains 89.2% ricinoleic and 1.4% dihydroxystearic acids;
cContains 59.8% crepenynic acid;
dContains 78.0% licanic acid and 4.0% hydroxy acids;
eHigh-erucic-acid variety;
fContains 71.3% venolic acid;
gContains 78.5% vernolic acid and 5.8% hydroxy fatty acids.

On a global basis for all species, oils produced in the tropics are more completely saturated and have the highest melting points, with these properties decreasing with distance from the equator (in north and south latitudes). Many dietary fatty acids are transposed from plant or plankton feed sources to body tissues; and fish oils from the Arctic and Antarctic generally melt at lower temperatures than those from the tropics. Many oilseed plants are daylight length determinate; that is, they require a specific number of hours of sunlight to blossom and produce seed. Soybean and sesame mainly are light determinate, although indeterminate varieties also exist now. Thus, a plant grown at one latitude from seed adapted to another latitude may not produce seed at the second location.

Each oil has its unique properties and history, as demonstrated by sunflower. Sunflowers generally are indeterminate, with the same seed plantable and productive in the prairie provinces of Canada, and in Mexico. When polyunsaturated oils were promoted in the 1960s and 1970s, the United States, producer of *table oils* (for salads and light cooking), purchased only sunflower seed grown north of the 39th parallel in the United States and Canada to maximize the polyunsaturated fatty acids (PUFA) content of their products. Fall seed maturation temperatures in the Northern climates are lower, resulting in higher PUFA and lower monounsaturates (oleic) contents than sunflower seed maturing at the same time in Texas. However, the higher oleic acid oil of Texas seed is less susceptible to oxidation and produces deep fried foods and snacks with longer shelflives. Also, if growers in areas of Texas, capable of raising two crops annually, wanted to produce high-polyunsaturated oil, they merely had to time their second planting for seed to mature in the cooler winter months.[7,8]

Several important changes occurred in the early 1980s. First, oil processors realized that a significant market existed for high-stability oils and began seeking higher oleic acid content oils. Second, the medical community and nutritionists realized that, although PUFA do not cause cancer, they promote the growth of existing cancer cells more than monounsaturated fatty acid (oleic). Emphasis was changed from encouraging consumption of polyunsaturated vegetable oils to reducing fat intake in general, with recommendations that no more than 30 percent of dietary calories come from fats, of which no more than 10 percent are saturated (animal, tropical, or hydrogenated) fats,[2] with some nutritionists believing that PUFA also be no more than 10 percent, essentially leaving at least 10 percent for monounsaturated fat.

The world's traditional monounsaturated fat is olive oil (•70% oleic acid), historically consumed in countries surrounding the Mediterranean Sea, thus the term "Mediterranean Diet." However, olive oil is expensive and too limited in supply to satisfy the growing popularity for high-oleic acid oil. Thus, U.S. table oil producers turned to importing high-oleic acid content canola oil from Canada. A "mid-oleic" sunflower oil (•65% oleic acid), grown in the Northern States, also was introduced to the fried foods industry in 2000. High oleic acid varieties of safflower and peanuts have been introduced, and a transgenic *genetically modified organism* (GMO) high-oleic acid soybean was patented in late 2001. Two non-transgenic lines of high-oleic acid sunflowers, containing 8082 percent and 9092 percent oleic acid, were developed in the early 1980s, and their oils were marketed as feed stock for oleochemicals production. The venture was not commercially successful, although the genetic lines still exist.

Interest in nontraditional fats/oils sources, including newly domesticated crops, forest oilseeds, bacteria, yeasts, molds, and algae has increased in recent years. The reader should consult the cited and later references for unusual fatty acids and their occurrence in various sources.[9,10]

Triacylglycerols

Glycerol esterified with one, two, or three fatty acids is found in nature and can be made commercially. The designations monoacylglycerol, diacylglycerol, and triacylglycerol (TAG) respectively, now are encouraged in the scientific literature although the older mono-, di-, and triglyceride terminology remains widely used in commerce. As the number of fatty esters on the glycerine "backbone" decreases, the compound becomes more polar and functionally effective as a surfactant. Mono- and diacylglycerols are described in more detail throughout this chapter.

The TAG are named in various ways. For example, unsaturated fatty acids sometimes are indicated as U and the saturated as S. If glycerol is completely esterified with stearic acid, the resulting monoacid TAG may be designated as SSS, or, more descriptively as StStSt, tristearin, tristearoylglycerol, or glycerol tristearate.

If more than one species of fatty acid is present, its relative location on the glycerol may be important to its functionality, enzyme susceptibility, and storage stability of the fat/oil. Several conventions have been developed to specify arrangements of fatty acids on the glycerol molecule (if known). To avoid confusion from inversion of the 1 and 3 carbon positions, hierarchies have been established to designate the number 1 carbon under the R/S system[11] and the *sn* (stereospecific numbering) system.[12] In the R/S system, the longest chain fatty acid is assigned to the 1 position, the second longest to 2, and the shortest to 3.

If the positions of fatty acids on the TAG molecule are known, the *sn* system is preferred for identifying their locations: (1) *sn* immediately before the word glycerol, as in 1-stearoyl-2-oleoyl-3-myristoyl-*sn*-glycerol, indicates that the respective fatty acids are in the 1, 2, and 3 positions; (2) the term *rac* (racemic mixture), as in *rac*-StOM, indicates that the middle acid is in the 2-glycerol position and the remaining fatty acids are equally divided between the *sn*-1 and *sn*-3 positions; (3) the term β, as in β-StOM, indicates that the middle acid is in the 2-glycerol position, but distribution of the other two acids is unknown.[13]

Oxidation

As shown in Table 8.4, oxidation and hydrogenation reactivity of a fatty acid, in free form or as part of a TAG, increases with the number of double bonds.[14] The *oxidation* of fatty acids and TAG (aldehyde formation, breakdown into shorter chains, and cross-linking to form polymers) is initiated at double-bond sites. However, linoleic acid, C18:2, *n*6, does not degrade into a mixture of C9, C3, and C6 compounds. Instead, as the molecule starts degrading, the positions of the double bonds migrate and provide many opportunities for splitting. Over 250 different breakdown compounds have been found in the end products.[1518] Part of the confusion about oxidation reactions is related to the type present, with light-induced photosensitized singlet oxygen oxidation the fastest.[19] For this reason, removal of photosensitizers like chlorophylls and pheophytins during the refining process of oil is extremely important.

TABLE 8.4 Relative Rates of Oxidation and Hydrogenation of Fatty Acid Chains[a]

Fatty Acid	Iodine Value	Relative Oxidation Rate	Relative Hydrogenation Rate
Stearic (18:0)	0	1	0
Oleic (9*cis*-18:1)	90	10	1
Linoleic (9*cis*, 12*cis*-19:2)	181	100	20
Linolenic (9*cis*, 12*cis*, 15*cis*-18:3)	274	150	40

[a] Modified from Beckman, H. J., "Hydrogenation Practice," *J. Am. Oil Chem. Soc.*, 60, 234A242A (1983).

The extent of prior oxidative activity in a fat sample may be estimated by the following: peroxide value (PV, a titrametric methodAOCS Method[5] Cd 8-53), thiobarbituric acid test (TBA) (which measures the presence of malonaldehyde), anisidine valueTotox, Kreis test, oxirane test, total and volatile carbonyl compounds, chromatographic analysis, ultraviolet spectroscopy, fluorescence, and especially organoleptic evaluation. However, the resulting peroxides are unstable and decompose. Peroxide values are not cumulative and are not always indicative of the extent of earlier oxidation. An oil sample may be starting to degrade, or may already have passed through a serious oxidation cycle; thus, age and history of the sample should be considered in forming a conclusion.

Predisposition (susceptibility) to oxidation can be estimated for oils and fats by the Active Oxygen Method (AOM) and Oil Stability Index (OSI), and for fat-containing products by the Schall oven test and the oxygen bomb method as a function of oxygen absorption. The stability of triglycerides formerly was determined by the *Active Oxygen Method* (AOM) (AOCS Method,[5] Cd 12-93), in which heated air was bubbled through a heated liquid sample of the oil or fat, and the number of hours for the sample to reach 100 milliequivalents (meq) of peroxide was recorded. The AOM procedure was put in "surplus status" (still legal but not preferred) in 1997 with focus then shifted to the *Oil Stability Index* (OSI) (AOCS Method,[5] Cd 12b-92). In the OSI procedure, heated air is bubbled through heated liquid triglycerides, and is scrubbed on exiting in a bath of deionized water whose conductivity is continuously monitored spectrophotometrically. Absorption of polar degradation produce is noticed immediately. Whereas the AOM method determined the period for the triglyceride to reach a specific level of oxidation, OSI determines the "induction period," or time required to exhaust the antioxidant properties in the oil. As such, OSI values always are lower (less time) than AOM values.

Biohydrogenation

Rumen microorganisms, in cattle, sheep, and other ruminants, biohydrogenate fatty acids to protect themselves against their toxicity. Being parsimonious about energy input, rather than saturating completely, they take the reaction only part way by changing the *cis* bond to a *trans* bond. The reaction is catalyzed by enzymes in the microbial cell wall, and prevents permeation of the fatty acid into the cell. As a result, ruminant and dairy fats, may contain up to 7 percent *trans* fatty acids. Additionally, the bonds shift to *conjugated* positions between adjacent carbons (C=CC=C), eliminating the methylene carbon in di- or tri-unsaturated fatty acids (C=CCC=C) found in fresh *cis*-type plant oils. Much interest has been shown in *conjugated linoleic acids* (CLA) in the last decade for treating cancer and controlling obesity.[20] Capsules containing CLA, produced by microbial fermentation, are sold as food supplements. Recent research reports that more plant oils, especially in spices, are conjugated than previously recognized. Conjugation also is the first step in chemical hydrogenation and initiation of oxidative degradation of fats and oils.

8.3 Melting Characteristics

The position of fatty acids on the triglyceride chain also is important for additional reasons. If unsaturated fatty acids are limited in availability, nature tends to place them in the number 2 position to obtain the lowest melting point for a plant's triglycerides. If more plentiful, the preferred positions will be 2 and 1 or 3 or both. In contrast, saturated fatty acids are more likely to be in the 2 position in fats of warm blooded animals, with palmitic acid in the 2 position in (pork) lard, a specific example.21 Many digestive tract or microbial lipases are either 2, or 1, 3 specific, and are used for controlled hydrolysis and structuring of modified or specialty fats and emulsifiers. Preparation of the unusual triglycerides of cocoa butter has been a popular subject for demonstration. Factors affecting the melting points of specific fat samples include the types of fatty acids present (lengths of the fatty acids chains, number and location of *cis* and *trans* double bonds in the chains), location of specific fatty acids on the glycerol, compatibility of the different TAG in the mixture, and types of crystals present. Melting points increase with chain length. *Trans* fatty acids always have higher melting points than their *cis* counterparts for any chain length. Where only one double bond exists in a fatty acid, as in C18:1, the melting point is lower if it is located after an odd-number carbon than an even-number carbon, and also if the double bond is located near the middle of the chain as compared with a location at either end.22 Compatibility of mixed fats from different sources can be a factor. For example, the melting points of multi-fatty acid TAG, consisting primarily of 1618 carbon fatty acids, generally rise smoothly with increased content of higher melting fats. However, when TAG with 1216 carbon fatty acid fats (from coconut and palm kernel oils) are added, concentration-related eutectic points and noncompatibility (miscibility gaps) may occur.2325

Polymorphism and Crystal Types

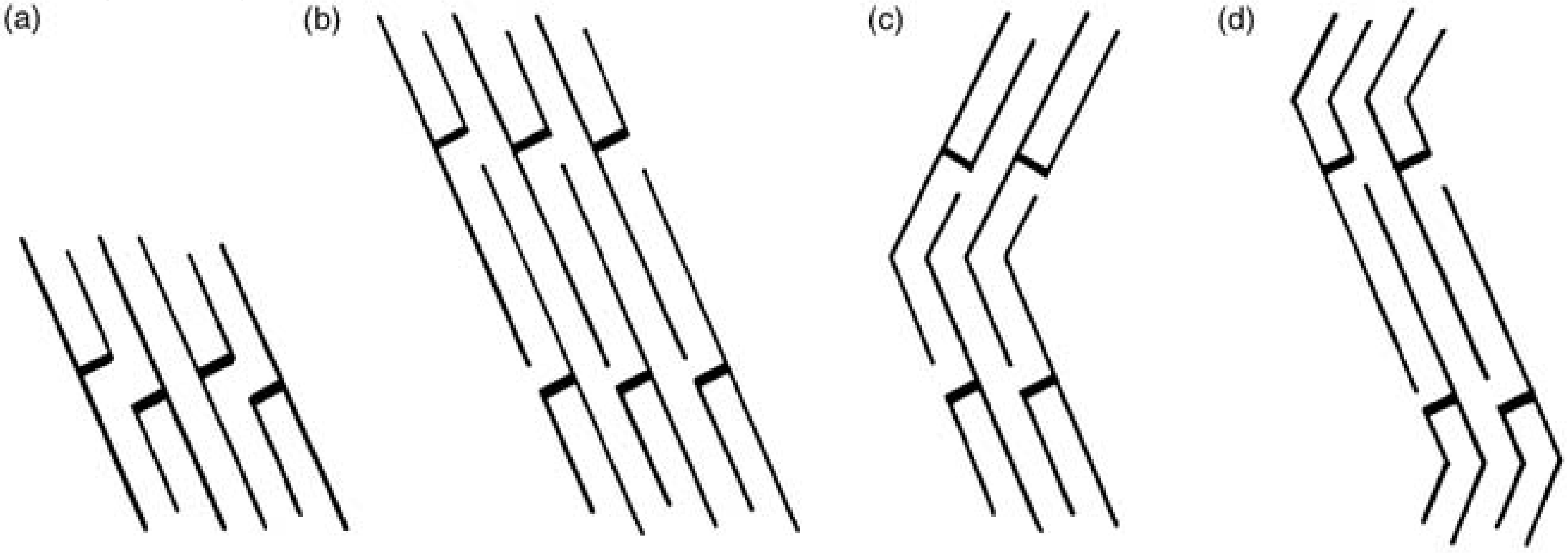
Polymorphism means "many bodies." Having determined the melting point of tristearin in 1849, Heintz continued to heat the capillary and observed resolidification and a second melting point at a higher temperature for the same sample. Duffy confirmed the principle and reported three melting points, at approximately 52°C, 64°C, and 70°C) in 1853. Later, five and even seven crystal forms were reported for some fatty acids and mono- and mixed-acid TAG, depending on the heating and cooling history of the sample.26

As fatty acids or TAG cool, Gibbs free energy ($G = H TS$) decreases by reduction of both enthalpy (H) and entropy (S, the degree of disorder). At decreased S, fatty acid chains assume pole-like structures that are less co-repulsive and pack more tightly into crystal lattices. It is widely accepted that, to participate in a crystal structure, TAG assume an "h" configuration (also called two-legged chair or tuning fork). This can be envisioned in our gravity-oriented macro world by assuming that each ester linkage acts as a hinge. If the glycerol number 2 carbon chain is held upright at the methyl end, the number 3 carbon chain hangs directly beneath it forming the back and one leg of the chair, and the number 1 carbon chain juts out at a right angle and curves downward to form the second leg.25,2731 In order to save space, half of the chairs are packed upside down to form a palisade-like structure. A stack of two such structures forms a bilayer whose outer surfaces, consisting of methyl groups, display low interattraction. This helps explain why TAG crystals are relatively flat, grow rapidly in length and less rapidly in width by adding parallel chair structures, and grow slowly in thickness by adding additional bilayers.

Within each layer, the chair backs and legs can be further envisioned to act like a vertical loose palisade of knobby-surfaced posts. (Although the hydrocarbon chains have stiffened into pole shape, the carbon atoms are not positioned as beads on a tight string, but rather in sawtooth-like knob fashion with carboncarbon bonds of 112°.) The poles can slide behind each other to obtain a tighter packing with lowered free energy, and can be tilted in two directions to allow the sawtooth configured carbon atoms on adjacent chains to slip by each other and pack more tightly. Furthermore, all the vertical units in one layer can be tilted at an opposing angle to the units in the other layer.

As shown in Fig. 8.1, if the three fatty acids in the TAG are saturated and approximately of the same length, each of the layers in the fat bilayer will be approximately two fatty acid chains in height. If the fatty acid on the number 2 glycerol carbon is appreciably shorter than those on the 1 and 3 carbons, each of the layers will be three chains thick. If the TAG is symmetrical, with either the number 2 glycerol carbon or both the 1 and 3 carbons unsaturated, the layer will be three chains long but with a zig-zag configuration to accommodate the *cis* configuration.

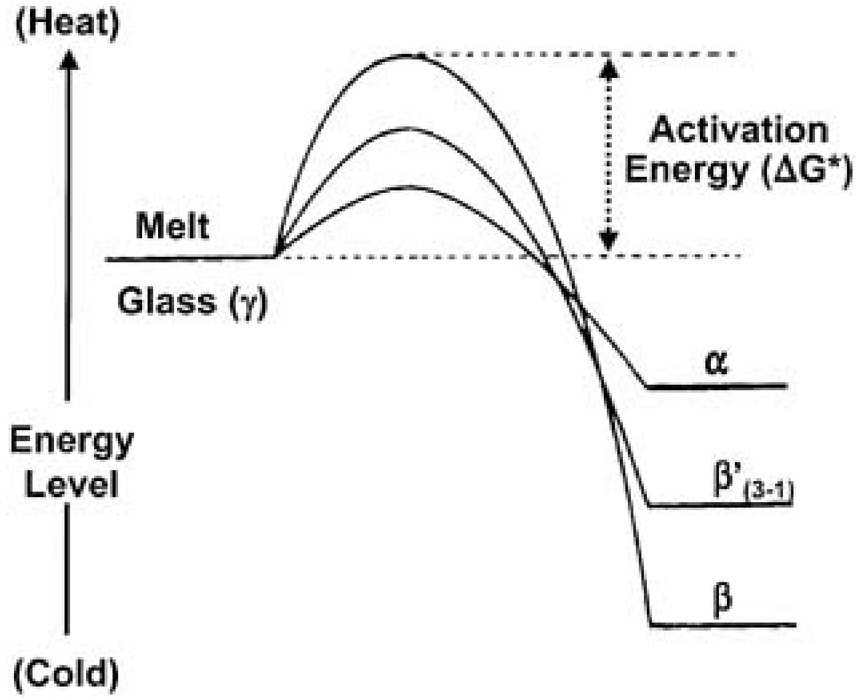
Fig. 8.1 Drawings of two- and three-chain triacylglycerol layers: (a) saturated monoacid SSS-type; (b) saturated symmetrical PSP-type where 2 chain differs from 1 and 3 chains in length; (c) symmetrical POP-type where 2 chain is unsaturated; and (d) symmetrical OPO-type where 1 and 3 chains are unsaturated. Crystals grow in bilayer units.



On rapid chilling, a glass (vitreous, γ) form occurs, which can change into an α or β′ form as activation energy becomes available. The form showing the least amount of crystalline order for a TAG, as determined by X-ray diffraction and infrared spectroscopy, is called the *alpha* (α) form. The most compact crystalline form, with the lowest free energy and the highest melting point, is called the *beta* (β) form. One or more intermediate *beta prime* (β′) forms also may exist, and are indicated as β3′, β2′, β1′ as the crystal progresses to tighter packing, a lower free energy state, and an increased melting point. When a TAG is cooled very slowly and without mixing, it preferentially assumes the β crystal form. But, the required tight packing of crystals requires time for alignment, and may be thwarted by increased viscosity as the fat/oil mixture cools.

The free energy relationships between the different crystal forms are depicted in Fig. 8.2. Although the β crystal form has the lowest free energy G, induction of its formation also has the highest activation energy G. Differential scanning calorimetry (DSC) often is used to follow free energy changes as fats melt and change between their polymorphic forms.

Fig. 8.2 Gibbs energy relationships of poly-morphs of a triglyceride. (*Modified from Sato, K., "Crystallization and Polymorphism of Fats and Fatty Acids", in* Crystallization and Polymorphism of Fats and Fatty Acids, *N. Garti and K. Sato (Eds.), pp. 227263, Marcel Dekker, Inc., New York, 1998.*)



The formation of β crystals may intentionally be encouraged or hindered, depending on the processing application. The α crystals are relatively unstable, and commercial interest is primarily placed on the differences between the β′ and β forms. Generally, β′ crystals are smaller (about 0.52.0 µm diameter in shortening, and 510 µm in margarine32), whereas β crystals can grow to as large as 2030 µm. When the objective is to thermally fractionate fats by crystallization, production of the β form is encouraged by carefully controlling temperatures (to not shock the fat into a lower melting β′ form), gentle stirring, and nuclei seeding. The smaller β′ crystals have smoother mouthfeel, minimize oiling off of margarine, and entrap more air in creaming cake batters. Their production is intentionally encouraged by formulating mixtures of natural or preprocessed fats, by the inclusion of emulsifiers to interfere with crystal growth, and by rapid agitation during plasticizing of the margarine or shortening.

Figure 8.2 also helps explain the mechanisms of defects appearing in fatty products. If a fat that has been conditioned into a stable β′, fine crystal form is suddenly exposed to thermal shock and then left unattended, the energy may activate it to settle into the coarse crystal, β, lower energy form. Thus, chocolate bars, left to melt and cool several times in an automobile during the summer, turn coarse in texture and mousy in color.

An awareness of crystal packing characteristics and polymorphism helps one to understand incompatibility problems of different fats. Crystal formation has specific demands, and individual crystals in mixed systems each consists of only one species of TAG. However, surfactants and other molecules can act as impurities and interrupt crystal growth. Different TAG are considered compatible when they co-crystallize as separate crystals under the same conditions without the formation of a eutectic.

Waxes

Waxes are fatty acid esters of alcohols and are formed by the general reaction:

$$\underset{\text{Alcohol}}{CH_3(CH_2)_nCH_2OH} + \underset{\text{Fatty acid}}{CH_3(CH_2)_nCOOH}$$

$$= \underset{\text{Wax ester}}{CH_3(CH_2)_nCH_2COCH_2(CH_2)_nCH_3} + \underset{\text{Water}}{H_2O}$$

A major role of waxes in nature is the protection of plant tissues. Examples include coating upper surfaces of leaves to reduce dehydration by the sun, and protecting seeds against moisture loss during storage. Minor quantities of waxes always are present in oils by solvents. Waxes extracted from seed hulls have been a problem in sunflower oil because their presence reduces the cold test of refrigerated oil, resulting in a cloudy oil. Sunflower oil may be dewaxed by first degumming or miscella-refining to remove the natural emulsifier lecithin, which limits the growth of wax crystals, and then winterizing (chilling and filtering) the oil or its miscella.33,34 With the improvement of dehulling equipment, some sunflower seed processors remove the hulls before extraction. Domestic interest only now is turning to commercial extraction of rice bran, whose oil contain waxes with potential commercial promise. In a sense, methyl and ethyl fatty acids esters, used as liquid fuels, solvents, and in other applications, are "waxes" of short-chain alcohols; these will be discussed later.

In some species, waxes are high-energy storage forms. Waxes are common in the oils of fish and other marine animals. The major lipids of commercial whale oil consist of approximately 65 percent waxes and 35 percent TAG. The lipids of Australian orange roughly (*Hoplostethus atlanticus*) and dory fish oils, are 97.1 and 90.9 percent wax esters, respectively.35 Essentially all the oil in jojoba (*Simmondsia chiensis*) seed is in wax form. Whale and jojoba oils have been valued for stability in cosmetics and heavy-duty lubrication applications.

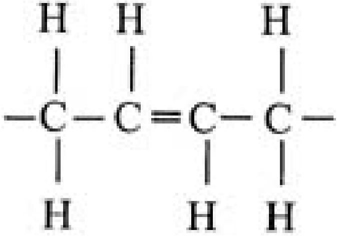
Extraction processes for waxes vary in sophistication, from boiling crushed leaves and berries and skimming, as in the (regulated) production of Candelilla wax (a substitute for hard carnauba wax) from *Euphorbia antisyphilitica* in the Big Bend area of Texas and Mexico, to screwpressing seeds such as those of jojoba. Cold-pressed jojoba oil is preferred for cosmetics because of concern about other lipids that might be extracted in solvent processes. Prepressed jojoba meal has been extracted by a secondary hexane process to recover oil for industrial uses. Waxes are susceptible to hydrolysis by nonspecific lipases, and are at least partially digestible. They may be hydrogenated or sulfurized into solid forms. The compositions of significant commercial waxes from natural sources are given in Table 8.5.

TABLE 8.5 Sources and Compositions of Natural Waxes

Type	Melting Point (°8C)	Main Components
Animal waxes		
Beeswax	64	Myricyl palmitate
Chinese	8284	Isoheptacosyl isoheptacosanoate, ceryl lignocerate
Shellac	8182	Ceryl lignocerate, ceryl cerotate
Spermaceti		Cetyl palmitate
Wool (anhydrous lanolin)	3642	Cholesteryl estolidic esters, alcohol esters of iso- and anteiso acids
Mineral waxes		
Montan	86	Tricontanyl esters of C2830 acids
Petroleum waxes		
Microcrystalline	7188	Hydrocarbons (490800 molecular weights)
Paraffin	5457	Hydrocarbons (350420 molecular weights)
Vegetable waxes		
Bayberry	4348	Trimyristin, tristearin
Candelilla	7080	C2933 hydrocarbons, simple esters and lactones
Carnauba	8085	Esters of C2630 alcohols and C2630 ω-hydroxy acids
Esparto	6981	Hydrocarbons, esters of C2632 acids and alcohols
Japan	5162	Tripalmitin
Jojoba (a liquid wax)	1112	Docosenyl eicosanoate
Ouricury	7985	Myricyl cerotate and hydroxycerotate
Sugarcane	7981	Myricyl palmitate stigmasteryl palmitate

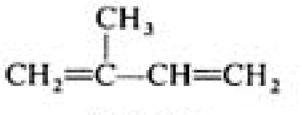
Terpenes, Carotenoids, Fat-Soluble Vitamins, and Steroids

Terpenes are condensation products of the five-carbon hydrocarbon *isoprene* (2-methyl-1,3-butadiene), and are extractable by nonpolar solvents. They are classified according to the number of isoprene units: two units, *monoterpenes*; three units, *sesquiterpenes*; four units, *diterpenes*; six units, *triterpenes*; eight units, *tetraterpenes*; and *polyterpenes*. Terpenes may be linear or cyclic. Taken together, this class of compounds includes major essential oils (used in perfume and flavorings), fat-soluble colors, fat-soluble vitamins, and steroids. A sawtooth short-hand form often is used to depict the longer chains. Examples of monoterpenes include the linear aldehyde *citral* which is found in many essential oils, and the (*cis*) alcohol *geraniol*a major component of oil of geranium. Cyclic monoterpenes include *limonene, menthol, pinene, camphor*, and *carvone*major components of lemon oil, mint oil, turpentine, camphor oil, and caraway oil, respectively. Sesquiterpenes include *farnesol*, a component of rose oil, and *bisabolene*, a component of Bisabol myrrh. The diterpenes include *phytol*, a component of chlorophyll, and vitamin A, which is one-half of the tetraterpene β-*carotene*. The triterpenes include *squalene*, a precursor of cholesterol. Examples of tetraterpenes are the oil-soluble *carotenoid* plant pigments: *xanthophylls*, including the yellow pigments *lutein* in plant leaves and *zeaxanthin* in corn (*Zea mays*); *capsanthin*, the red pigment in red peppers (*Capsicum annum*); *lycopene*, the red coloring of tomatoes (*Lycopersicum esculentum*); and β-*carotene*, a yellow-orange pigment that is the precursor of Vitamin A. *Bixin*, from the seed pods of *Bixa orellana*, is the yellow-orange pigment in annatto food color; it is considered to be a form of carotene oxidized to remove both six-membered end rings. Polyterpenes include *gutta*, a natural water repellant and electrical insulating material from *Palaquium gutta*, and natural rubbers with molecular weights of up to 1.2 million daltons obtained from the latex of the *Hevea brasiliensis* tree and from rubber-filled cells in the branches and roots of the guayule shrub (*Parthenium argentatum*). Important steroids derived from squalene include *cholesterol*, *stigmasterol*, *sitosterol*, and *campesterol*.

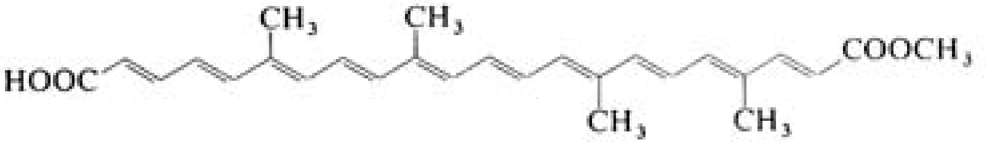


trans

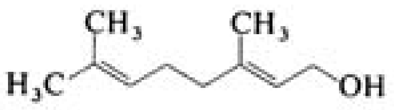
The fat-soluble vitamins include: vitamin A (*retinol*), a colorless compound resulting from cleavage of β-carotene; vitamin D3 (*cholecalciferol*), a steroid; and vitamins E and K (*a-tocopherol* and *menadione and its derivatives*, respectively), which consist of isoprene and phenolic-quinone components.



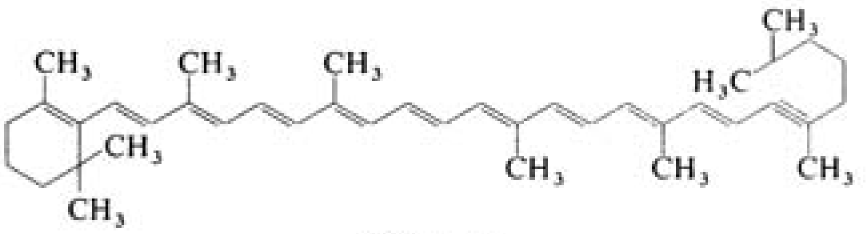
Isoprene



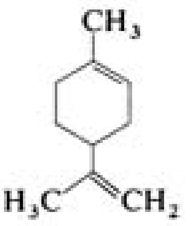
Bixin



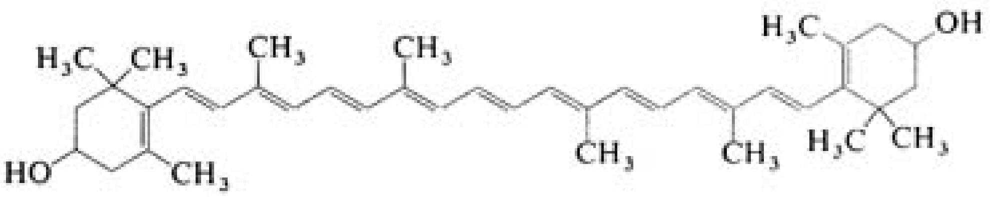
Geraniol



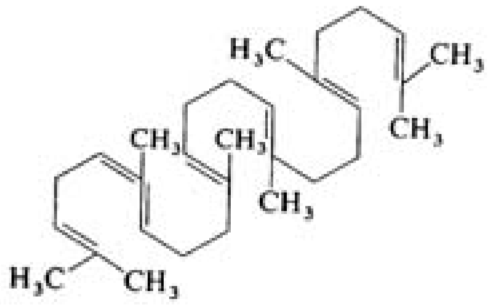
β-Carotene



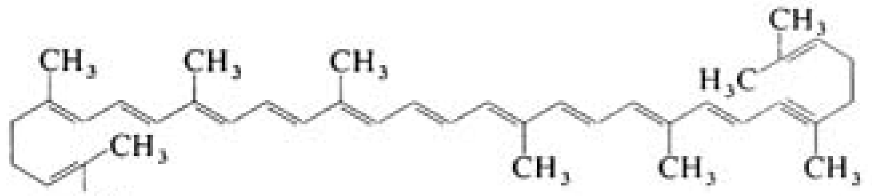
Limonene



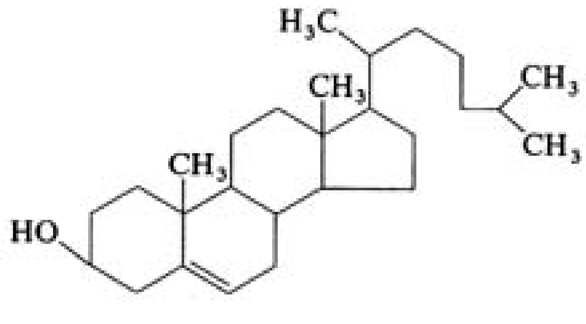
Zeaxanthin



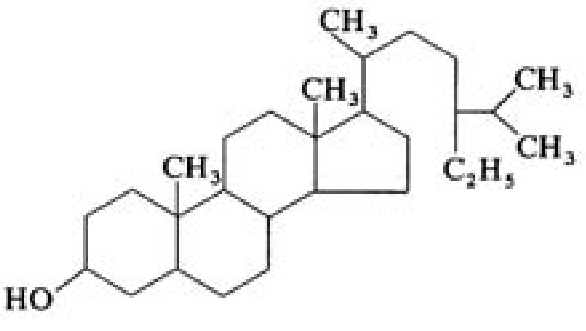
Squalene



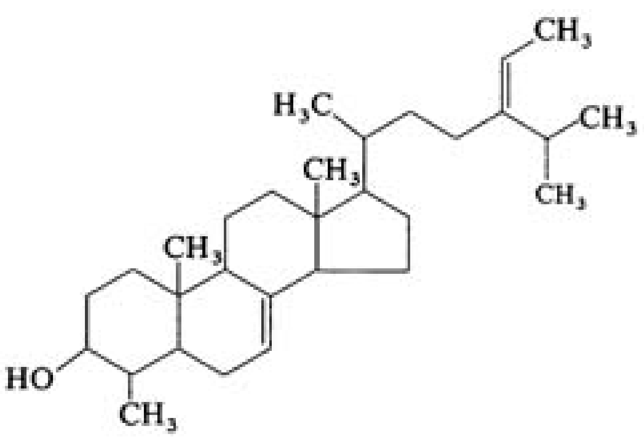
Lycopene

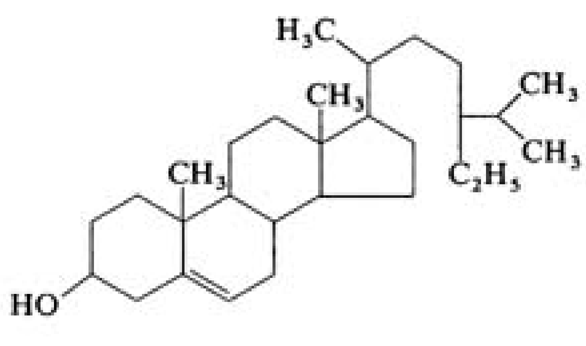


Cholesterol



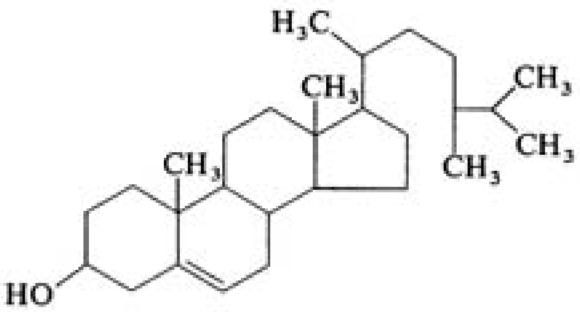
Stigmasterol





α1-Sitosterol

β-Sitosterol



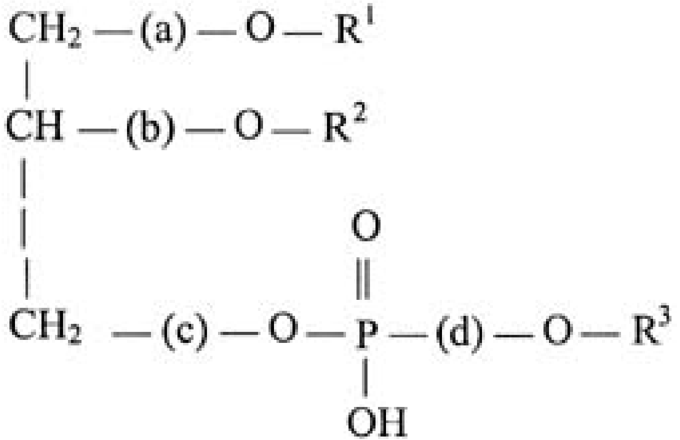
Campesterol

The carotenoids generally are highly susceptible to oxidation. Only selected members (xanthophylls, β-carotene, and bixin) are extracted for food coloring. They are used in products such as butter, margarine, and cheese, where exposure to oxygen and light is limited. The more stable yellow-orange pigments (lutein and xanthophylls) are extracted from alfalfa and marigold blossoms for use as poultry feed additives to increase the yellow color in egg yolks and in the skins of broilers. Much of the world's β-carotene is produced synthetically, but may be obtained from (red) palm oil in the future.

Oils from the livers of cod and other fish species were early sources of vitamin D3; also, *ergosterol* was irradiated by ultraviolet light to produce vitamin D2. Currently, most of the commercial vitamin D is produced synthetically. Commercial vitamin E is recovered from deodorizer condensates from plants refining palm and other oils, including soybeans. Wheat germ oil was a major earlier source.

Phosphatides

Phosphatides are natural emulsifiers with the following structures:36



$$\begin{array}{l} CH_2-(a)-O-R^1 \\ | \\ CH-(b)-O-R^2 \\ | \\ CH_2-(c)-O-\overset{\displaystyle O}{\overset{\|}{\underset{\underset{\displaystyle OH}{|}}{P}}}-(d)-O-R^3 \end{array}$$

R1 = fatty acid

R2 = fatty acid, or the phosphatidyl group in the 2 position.

When R3 is one of the following, the phosphatide is named:

$$R^3 = CH_2-CH_2-N^+(CH_3)_3 \quad \text{diacylphosphatidyl choline}$$
$$R^3 = CH_2-CH_2-NH_3{}^+ \quad \text{diacylphosphatidyl ethanolamine}$$
$$R^2 = CH_2-CH(NH_3{}^+)-CO_2H \quad \text{diacylphosphatidyl serine}$$
$$R^3 = C_6H_6-(OH)_6 \quad \text{diacylphosphatidyl inositol}$$
$$R^3 = H \quad \text{diacylphosphatidic acid}$$

and preceded by names of the two fatty acids if relevant.

In the pharmaceutical industry, the word *lecithin* is synonymous with phosphatidyl choline. However, the entire or modified mixture is sold as *lecithin* in food supplement capsules, and for food, feed, and industrial uses. Soy oil phosphatides consist of 2939 percent diacylphosphatidyl choline, 2026 percent diacylphosphatidyl ethanolamine, 1317 percent diacylphosphatidyl inositol, 5.96.3 percent diacylphosphatidyl serine, and 59 percent diacylphosphatidic acid.36

The phosphatides have polar and nonpolar sites, and generally act as water-in-oil emulsifiers. They are extracted by solvents with the oil, but preferentially will absorb available water, form "gums," and precipitate. Commercial lecithin is produced by *water degumming* (precipitation from oil with ion-exchange-treated water) as explained later.

Phosphatides precipitate on hydration during the storage of oils, foul bleaching earths, poison hydrogenation catalysts, and cause darkening of the oil during deodorization/physical refining and also if it is used in frying applications. Their removal is desirable, but requires close supervision to preserve oil yields.

Four lipases are able to hydrolyze phosphatides. Phospholipase A cleaves fatty acids at the 1 position; phospholipase B cleaves fatty acids at the 2 position; phospholipase C cleaves the phosphatidyl chain next to carbon 3; and phospholipase D cleaves between the phosphate structure and choline, ethanolamine, serine, or inositol. Cleaving by phospholipase D and dissociation of the phosphorous group exposes two negatively charged sites which can complex with divalent cations (mainly calcium and magnesium, but also including iron, copper, and others) present in soybean tissue during solvent extraction. As a result, the phosphatide becomes *nonhydratable*.

The production of nonhydratable phosphatides by phospholipase D can be minimized by heat inactivation of the enzyme by expanders or extruders while preparing seed for solvent extraction, by heating the seed to more than 85°C/185°F before flaking, and by rushing seed through preparation to minimize the exposure period to enzyme-favoring moisture and temperature conditions. However, some nonhydratable phosphatides are unavoidable, especially in wet falls when high moisture seed may begin to sprout. Unless the nonhydratable phosphatides content is extremely high, they can be changed to the hydrated form by treating the crude oil with chelating agents (*acid degumming*) to withdraw the divalent cations.

8.4 Sources and Consumption of Fats and Oils

Production

The global production and export estimates for fats and oils for 2000/2001 are shown in Table 8.6. Approximately 39 percent of the world's production of vegetable oils, and an additional 22 percent of oilseeds grown, enter international trade.37 Extensive records of production, imports, and exports are kept, country by country, by the Foreign Agriculture Service of the U.S. Department of Agriculture, and *Oilseeds World*, a publication in Hamburg, Germany.

TABLE 8.6 Estimated Area, Yield, Oilseeds, Products and Exports of World's Major Edible-Type Vegetable Oils, 2001/2002a, b

			Seed/Copra		Meal		Oil	
Oil/Fat Source	Area (MHad)	Seed Yield/ Hectarec (MTe)	Production (MMf)	Exports (MMT)	Production (MMT)	Exports (MMT)	Production (MMT)	Exports (MMT)
Edible fats								
Soybean	78.07	2.34	182.91	58.91	124.70	43.57	28.46	8.65
Palm							24.61	17.07
	24.64	1.48	36.41	7.64	20.78	4.04	12.75	2.48
Canola/rapeseed	19.24		21.16	2.09	8.55	2.31	7.60	2.69
	22.51	1.50	33.76	1.63	6.13	0.39	4.81	0.29
Sunflowerseed	32.99	1.11	36.67	1.37	12.05	0.60	3.83	0.14
Peanut			5.69	0.25	1.81	1.06	3.45	1.98
Cottonseed			7.20	0.06	3.77	3.33	3.03	1.33
							2.33	0.93
Copra/coconut					5.91	3.59		
Palm kernel	177.44	1.75	323.80	71.93	183.70	58.89	90.87	35.56
Olive								
Fish								
Total								

aFrom: Foreign Agriculture Service Circular Series FOP 02-02, U.S. Department of Agriculture, February 2002.
bSplit year includes Northern Hemisphere crops harvested in late months of the first year combined with Southern Hemisphere crops in the early months of the following year.
cYield per hectare (2.47 acres).
dMHa = million hectares.
eMT = metric ton.
fMMT = million metric tons.

The leading producers of the major oilseeds are: *soybeans*the United States, Brazil, Argentina, China, and Paraguay; *rapeseed/canola*China, the European Union, Canada, India, and Eastern Europe; *sunflower seed*the former Soviet Union 12, Argentina, European Union, Eastern Europe, China, and the United States; *groundnuts/peanuts*China, India, and the United States; and *cottonseed*China, the United States, India, Pakistan, the former Soviet Union 12, and Brazil.

Soybean oil is still the world's largest supply of visible fats, accounting for approximately 31.3 percent. It is followed by palm, rapeseed/canola, sunflower seed, and peanut oils in tonnages produced. The production of palm oil has been increasing, and now accounts for 27.1 percent of the world's supply; added to the 3.4 percent palm kernel oil produced, the palm crop provides 30.5 percent of total oil supply. With oil palm plantings still to mature in various tropical countries, palm oil production alone is expected to outpace soybean oil within the next several years.

Currently, the United States is the world's leader in soybean production (43%), followed by Brazil (23%), Argentina (16%), and China (8%). Approximately 35 percent of the world's soybean meal (for animal feed), 29 percent of the oil (mainly in degummed form), and 31 percent of the world's soybean seed (mainly for overseas extraction) enters global trade. The United States exports 36 percent of its soybean crop in seed form, 20 percent of its processed meal, and 8 percent of its soybean oil. In contrast, Brazil exports 36, 58, and 30 percent, and Argentina exports 23, 97, and 99 percent, respectively. The latter two nations each surpass the United States in tonnage of meal and oil exports. This results partially from increasingly more of the domestic crop retained in the United States to supply the local population's needs. Also, larger tariffs were imposed on seed exports than on meal or oil in Brazil and Argentina in the 1970s and 1980s to encourage the development of domestic oils extraction industries; these strategies appear to have been successful. Because of population growth, China now imports as much soybeans as produced internally.

Rendered beef, pork, poultry, and other fats are not well reported internationally, and global statistics are unreliable. Total production of fats in the United States by the rendering industry for 2000 is estimated at 4.18 million metric tons.38 Outputs of all rendering facilities captive to integrated broiler operations might not be included. Of the amount reported, 76 percent is inedible tallows and greases, 18 percent is (beef or mutton) edible tallows, and 6 percent is edible (pork) lard.

Fish oils are not well reported either, with annual estimates at about 1.3 million metric tons from sustainable (sea catch) fisheries. Production has ranged by as much as 50 percent between years, depending on availabilities of fish.

Changing Prominence of Fat and Oil Sources

The maturing of nations as raw materials suppliers follows a sequence. Centuries ago, bands of marauders attacked villages to steal crops after harvest. Later, countries conquered neighboring lands for "a place in the sun" for their growing populations. With improvements in transportation and discovery of the New World, followed by Central-South America, Africa and Australia, it no longer was necessary to relocate large populations. Colonies could be established to supply the "mother country" with raw agricultural materials in addition to minerals and fossil fuels. As local education improved, colonies typically declared their independence, but the new countries needed something to trade for goods they were unable to produce themselves.

The oils of Ancient Times were olive and sesame oil in the Mediterranean basin, rapeseed and animal fats in Europe, and coconut (copra) oils in the tropics. Cottonseed was the world's first new oil of the Industrial Revolution Age (early 1800s), and became the dominant oil after the U.S. Civil War (during the late 1860s until the mid 1930s). But, cottonseed oil is a by-product of cotton fiber growing, and oil and animal feed requirements of the world's growing populations soon exceeded supply capabilities of this crop. Solvent extractors, invented in Germany in the early 1920s, maximized oil recovery and produced animal feed protein meals with less heat damage. European demands grew for a closer and more reliable source of soybeans than had been available from the Manchuria area of China since 1910, and soybean export opportunities opened for the United States. Much of continental Europe's oil mills were demolished during World War II, placing the United States in the position of major soybean seed, oil, and meal supplier to the world. In 1960, the United States grew and traded about 60 percent of the world's soybeans. China grew about 32 percent of the world soybean crop and supplied about 19 percent of the trade, but soon dropped out as a world supplier because of its own population growth. Also, 1960 was a key year because the poultry broiler industry became global, requiring large amounts of feed proteinstill best-supplied by soybean meal. World production of soybeans has increased by 6.8 times since 1960.

Although slightly more than 4 tons of meal are produced for each ton of oil, the sheer volume of the meal business has co-produced enough soybean oil to keep it the world's major oil until now. Much of the production and processing technology was developed by the United States and European nations. In time, roads, canals and port systems, crop production, and processing infrastructures in South America were funded as economic development programs by the World Bank, Regional Banks, and by private investors. The position of "lowest cost producer of soybeans and soybean oil" has passed from the United States to South American countries. At the same time, economic assistance and private investment in Southeast Asia and tropical countries have led to production of palm oils at prices lower than soybean oil.

Many factors dictate which oilseed species will be grown and/or imported into a country. Climate and local demand for high-protein feed meals are leading factors. Cool weather and short growing seasons have essentially limited Canada, Northern Europe, and the former European Russian republics to growing rapeseed/canola or sunflower seed. Some European countries grow and export seed or oil of canola, a healthier type of rapeseed, but import soybeans for their animal feeding industries. Insect problems were important factors in selecting crops before modern insecticides and integrated pest management systems became available. Devastation of the south-eastern United States cotton crop by the boll weevil led to the introduction of large-scale peanut growing in the early 1900s.

As shown in Table 8.7, the oil content of row crop oilseeds varies from about 18 percent for soybeans to 43 percent for sunflower seed, and 4345 percent for rapeseed/canola. More feed co-products always are produced than oilsby a ratio of as much as 4:1 in the case of soybeans. Soybean meal is the major feed protein source for production of poultry, currently the leading domestic and global meat source, and in the rapidly developing aquaculture industries.

TABLE 8.7 Gross Composition of Major Undehulled Oilseeds

Crop/Source	Moisture (%)	Protein (%)	Fat (E.E.a)	Crude Fiber (%)	Ash (%)	Oil:Co-products Ratio
Soybean	8.5	36.5	19.5	5.8	4.9	1:4.1
Cottonseed	8.0	23.0	21.0	24.0	4.8	1:3.8
Peanut	6.5	25.7	49.2	4.9	2.3	1:1.0
Sunflower	6.0	21.1	42.0	17.4	3.3	1:1.4
Safflower	5.8	19.4	43.5	20.8	3.7	1:1.3
Coconut (copra)	4.0	7.5	67.3	5.0	1.9	1:0.5
Rapeseed/canola	8.0	22.0	41.2	11.5	5.1	1:1.4
Palm kernel	10.5	9.7	58.1			1:0.7
Sesame	8.0	24.2	47.6	11.2	6.1	1:10

aE.E. = Ether extraction method.

The relative availability of fat and oil-bearing by-products of other current agribusinesses is an additional consideration in types and amounts of oilseeds grown. Since only about 12 percent of the return to cotton farmers comes from the seed, the domestic supply of cottonseed for crushing is dictated by world demand and the price of cotton and is hardly affected by the price of cottonseed oil. Corn oil has become the second major oil in the United States due to large quantities of corn germ provided by rapid growth of the domestic corn sweetener and ethanol industries. There is sufficient processing of rice in the United States now to warrant two rice bran oil extraction plants, with quantities of stabilized rice bran also shipped to Japan. Beef, pork, and poultry packing operations always produce fatty tissues for rendering into inedible tallow, lard, and chicken fat used in animal feeds. Generally, fats/oils from co-products of other local processing industries must clear the market first, at whatever price they can get, before growing or importing of significant quantities of high-oil-content crops becomes economically attractive.

Consumption

Average per capita consumption of fats/oils is difficult to determine from gross disappearance figures because these materials also are used in animal feeds and industrial applications. Generally, consumption is related to personal income and local availability, but once fats are introduced into the diet, their priority in expenditures for food remains high. The consumption of oils is estimated at 10.4 kg per person for the world, and ranges from 27.2, 42.9, and 29.3 kg for the United States, Belgium, and West Germany, respectively, to 13.1, 4.7, and 5.9 kg, respectively, for Egypt, China, and India. Approximately 60 percent of the total fat consumed domestically is invisible in meat, poultry, fish, dairy products, and eggs, and the 40 percent visible fat is used primarily in the form of salad and cooking oils, shortenings, and margarine. The world needs about 2.0 million additional tons of processed fats/oils each year for its rapidly growing population. From 1965 to 1990, average domestic consumption of fat decreased from 50.7 to 32.5 kg per capita for men 1950 years of age and 30.3 kg/yr to 23.4 kg/yr for women. Since 1990, consumption increased to 36.9 kg/yr for men and 23.7 for women in 1995.39 Approximately two-thirds of visible fats available per capita in 1940 was from animal sources. The use of vegetable source fats has grown significantly, accounting for about two-thirds of the visible fats consumed in 1965, and 90 percent in 1985.40 More recent data have shown that fat consumption per capita has not deceased in grams per day intake, but has decreased as a percentage of total diet with average caloric intake increasing. Currently, the public acts indifferently regarding caloric intake. In the mid-1990s, various low-fat or nonfat snack foods were introduced, requiring different processing machinery and techniques and also flavor technology. Although promising at first, sales of these products showed a decline in 1997.41 The Procter & Gamble Company developed olestra (Olean™) during a 28-year period (19681996) before the U.S. FDA's approval of its use as a frying oil in savory snacks (chips, crackers, and tortilla chips), spending many several hundred million dollars in the process. Olean™ is a polyester of sucrose and 68 fatty acids, which is nondigestible by human lipases, and thus noncaloric. Many of these snack products have been withdrawn from the marketplace. Public interest has shifted from snacks promoted as "healthy" to indulgence in salty, traditional flavor, higher fat content products.31 Various newspaper articles relating diarrhea, intestinal cramps, and anal leakage (spotting of underwear and clothing by unabsorbed fat), published at the time of olestra introduction, and FDA requirements that products made with olestra carry a label stating: "may cause abdominal cramping and loose stools" also may have deterred public trials.42

Effects of Biotechnology

The development of new plant varieties using *biotechnology* includes a broad scope of traditional (selective crossing), tissue culture, and *genetic engineering* activities focused on the transgenetic transfer of plant or animal genes into unrelated species.

Currently, *genetically modified organisms* (GMOs) produced by transgenetic means are receiving much attention in the news media. Technical feasibility has been demonstrated by the release of various GMO food crops in the United States. However, some scientists from many countries, including the United States, have expressed concerns that: (1) the rigorous food safety tests of the earlier food additives review era were relaxed too far by the U.S. FDA's adoption of a policy of "substantial equivalence," wherein confirmatory multi-generation animal safety tests are no longer required if the altered plants appear similar to the parent; and (2) the effects of new GMO products on public health have not been specifically monitored after release. The U.S. government has not required labeling of foods containing GMO crops, but has proposed very strong restrictions and oversight regarding labeling foods as "non-GMO." This position has raised many objections from citizens in the United States and other countries where the "right to know" what is in foods is considered to be a hard-won achievement.

Round-Up™ is Monsanto Company's (St Louis, MO) trade name for glyphosate, an effective herbicide. Round-Up Ready Soybeans™ and Round-Up Ready Canola™ are GMOs resistant to glyphosate. Seeds are planted and after sprouting and initial growth, the growing area is sprayed with Round-Up™ to kill weeds without damage to the crop. Several trials at universities have shown little, if any, improvement in soybeans yields, with slight reductions in some cases. But, general agreement exists that, because of reduced field tillage expenses, per bushel costs of producing soybeans are reduced. First grown in 1996, approximately 68 percent of the domestic soybean crop was planted to Round-Up Ready Soybeans™ in 2001, with estimates of 74 percent for 2002. The European Union, China, Korea, and other countries have objected to GMO soybeans, but continue buying them because of limited world options. Thus far, U. S.-grown GMO soybeans have been moved in world trade. The GMO issue has rapidly developed Identity Preserved (IP) export markets and entrepreneurs for supplying non-GMO soybeans in the United States and other producing nations.

The bacterium *Bacillus thuringiensis* contains a protein that becomes highly toxic in the digestive tracts of certain insect larvae. *B. thuringiensis* cells have been grown, and dusted or sprayed onto flowers and vegetables in home gardens as a safe insecticides for many years. It is one of the few insecticides permitted in organic farming. The gene responsible for making the protein has been transferred to (BT) corn and (BT) cotton, which produce their own insecticides active against the European corn borer and the cotton bollworm, respectively. Secondary beneficial effects may include a reduction in aflatoxin contamination of corn and cottonseed because of a reduction in insects carrying *Aspergillus flavus*.

Thus far, only four GMO oil sources have been released: Round-Up Ready™ GMO soybeans in the United States and elsewhere, Round-Up Ready™ canola in Canada, corn germ from BT corn, and oil from BT cottonseed in the United States. Over the past several years, samples of oils from GMO soybean varieties with various traits, like high oleic acid or stearic acid contents, reduced linolenic acid, and low lipoxygenase have been shown and evaluated in the trade and among researchers. A new law, classifying GMO crops as services protectable by 20-year patents, was passed at the end of 2001. A stream of patents on these type of crops is being issued, and they will now become visible to industry and the general public.

One of the better examples of cell culture applications is the Palm Oil Research Institute of Malaysia (PORIM), which had an extensive laboratory operating in the mid-1980s that led to the development of improved varieties of oil palm trees.

The first major example of changing the fatty acid composition of oil seeds was the development of Canola, a low erucic acidlow glucosinolate seed, by selective crossing of rapeseed varieties by Canadian scientists in the 1960s and 1970s. The program was initiated after the medical community reported the development of fat deposits in heart muscle of traditional high-erucic rapeseed oil consumers. At the same time, researchers also worked on reducing thioglucosinolates because of adverse effects on thyroid function, and poisoning effects of sulfur on nickel hydrogenation catalysts. Later (the 1980s, 1990s), emphasis was placed on developing high-oleic, reduced-linolenic acid varieties of sunflower seed, soybeans, safflower seed, and peanuts to reduce oxidative deterioration in table oils and to prolong the service life of frying oils in fast-food shops and industrial frying operations.

8.5 Extraction of Fats and Oils

Basic Processes

By trial and error over the centuries, man has learned five basic skills in handling oilseeds:

1. *Preservation of seed by natural or artificial drying* and cooling to a dormant state before storage, with protection from insects and rodents. Regardless of species, once seed rises to about 75 percent Relative Humidity, it should be processed as soon as possible.
2. *Removal of trash and hulls*, by stamping, threshing, disc hullers, cracking rolls, and other devices, followed by winnowing, sieving or aspiration to separate kernels ("meats") from the chaff.
3. *Freeing the oil*, by pounding with rocks, mortar and pestle-type grinders (which later became human, beast, water, or electrical-driven ghanis); use of vertical stone/iron wheels known as "edge rollers" to crush the seed; cracking and flaking rolls; and, more recently, expanders/extruders.
4. *Heating the seed to increase oil recovery*. This was first interpreted as denaturation and making the protein matrix brittle to surrender the oil on pressing. Later, it was recognized that concurrent inactivation of enzymes also arrested development of various types of degradations.
5. *Separation of oil* from crushed seed, by draining, squeezing bagged cooked mash by wedge and hydraulic presses, followed by continuous screw presses, and currently solubilization in solvent extractors.

Various societies have used plant oils for medicinal purposes. Later, they were used for lighting and then for food as extraction and refining techniques improved. Some of the older techniques are still used in remote areas of the world.

Extraction of vegetable oils with high-pressure CO_2 in a critical state (at minimum pressure for liquid phase) was heavily researched in the 1980s.43 Thus far, loading and unloading systems that can operate continuously against the pressure differentials at the high rates required for the major oilseeds have not been developed. However, the process is used in batch mode for extracting high-value products like essential oils, hops, and decaffeinated coffee. A critical CO_2 extraction laboratory analysis instrument for rapid analysis of fat content in oil mills has been marketed since the mid-1990s.

Extraction Plants

Concerns about marketing the meal co-product (accounting for 6575% of returns per bushel of soybeans grown), and disposal of refinery by-products have reshaped the domestic soybean processing industry in the last 20 years. Soybean extraction plants now are located close to domestic meal feeding markets, and/or on major barge waterways and railways to reduce transportation costs to export facilities. By-products of refining are difficult to handle, and their production exceeds market demands. The major oil processors now generally locate extraction plants and oil refineries on the same property to enable spraying and drying surplus gums (crude lecithins) and soapstock on the meal for sale as animal feed. Independent edible oil refineries have almost disappeared, although refineries already associated with an extraction operation may purchase crude oils from other sources.

Economies of scale have led to minimum capacities of about 20003000 metric tons per day for new U.S. regional soybean extraction plants, and larger facilities (40006000 tons per day) for soybean plants that pool some of their output for export. At the U.S. average yield of 38 bushels per acre, each 1000 metric tons per day capacity solvent plant requires the output of 966.65 acres per day, or 338,328 acres per year (136,975 ha). A 2750 metric tons per day plant would use the soybean crop of 930,402 acres per year (376,681 ha). An extraction plant of this size will support a 500 tons per day refinery. Regional plants are smaller than export plants because accumulating this much soybeans, in competition with existing processing facilities, can be difficult in some areas of the United States.

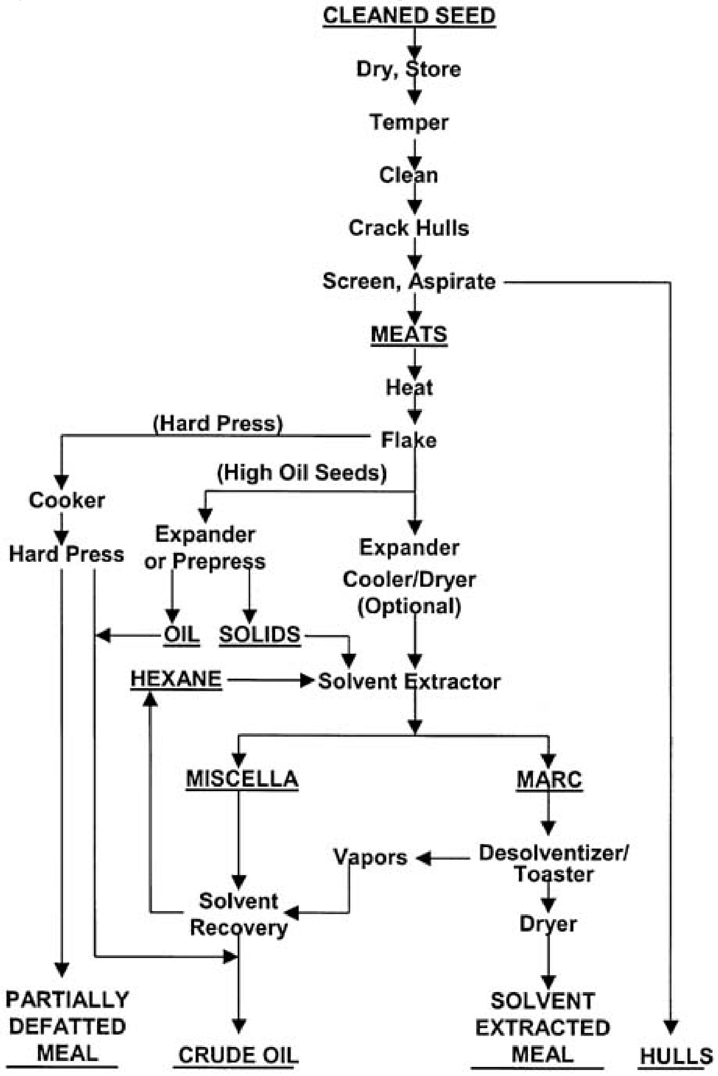
The largest solvent extractors can process 8000 and 10,000 metric tons of soybeans per day, and are used primarily for international trade. Two large extractors are used in the world's largest oilseed extraction plant in Argentina, reported to process 16,000 metric tons of soybeans per day. Installations typically include facilities for unloading railroad cars, barges, or ships; storage of at least several weeks' supply of seed; solvent, meal, and crude oil storage tanks or facilities; cleaning, extraction, and desolventizing facilities; an oil refinery; quality control laboratory; and day facilities for personnel and supervisors. A photo of an operating soybean extraction seaport plant is shown in Fig. 8.3.

Fig. 8.3 Photograph of an operating soybean extraction plant. (*Courtesy of Archer Daniels Midland Company, Decatur, IL*)

Seed Preparation for Extraction

A general flow sheet for direct solvent extraction of many row crop oilseeds is shown in Fig. 8.4. The initial seed quality, and its preparation for extraction, have the most effect on required subsequent refinery operations and yields of *neutral* (saleable) oil. Freshly harvested seed should be cleaned of trash, which harbors moisture that may accelerate the formation of Free Fatty acids (FFA), and/or become ignited during drying before storage. The oil of most good quality dry row crop seed contains 0.50.75 percent free fatty acids (AOCS Methods Aa 6-38, Ac 5-415) with seed containing twice this range typically accepted in trading without discounts. The maximum moisture content for holding seed for long periods without spoilage varies with storage temperature and ranges from 8 percent to 13 percent among species. It is inversely related to the oil content, since less protein and carbohydrate is available to compete for water in high-oil seeds; however, the optimum relative humidity for storage is constant at about 6570 percent for all seeds. It is determined by allowing a sample of seed to equilibrate in a closed container at about 21°C/70°F, and determining Relative Humidity (RH) or Water Activity (AW) in the headspace. If not adequately cleaned before storage, the seed should be completey cleaned before further processing to prevent clogging and damage of equipment. The processing system should be well equipped with magnets to arrest tramp metal that arrives with the seed or that is shed by equipment.

Fig. 8.4 General flow sheet for extraction of row, crop oilseeds.



Oilseeds do not have fat cells like those of animals for storing fats. Instead, oil is stored in microscopic globules throughout the cells of dicotyledonous oilseeds, or in the seed germ or bran in monocotyledonous plants such as maize (corn) and rice, respectively. Yields and processing costs are highly dependent on the effectiveness of preextraction operations to disrupt cells and free the oil for recovery. Operations differ among various oilseeds, mainly in techniques of dehulling. The traditional process has been to heat soybeans and let them stand (temper) for several days to enhance dehulling (Fig. 8.5).44 Several "hot dehulling systems" have been introduced in the last decade to loosen the hull and crack the seed in one operation immediately before flaking and extraction (Fig. 8.6). These processes save energy by avoiding a second heating step.

Fig. 8.5 Preparation system for conventional dehulling and flaking of soybeans. (*From Moore, H. N.*, J. Am. Oil Chem. Soc., 60, *141A144A. With permission*.)

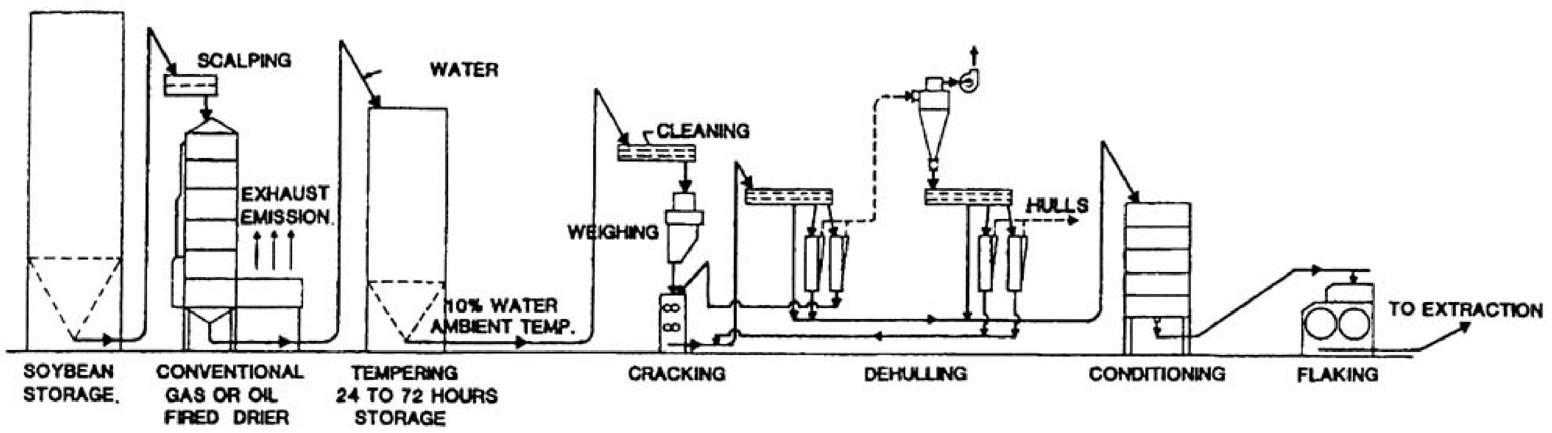
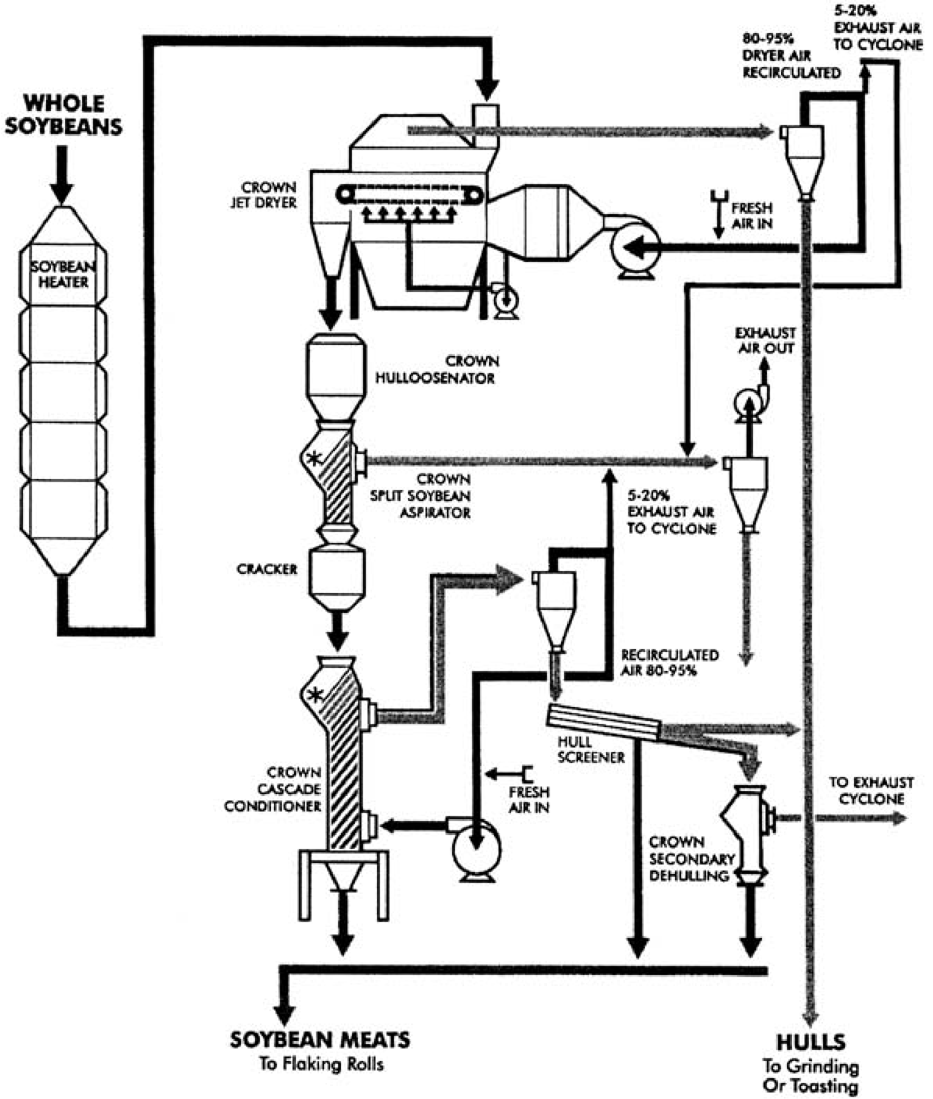


Fig. 8.6 Hot dehulling system for soybeans. (*Courtesy of Crown Iron Works, Minneapolis, MN*)



Hulls are removed in a two-step process, called *dehulling* or *decortication*, in which they are first cracked and then separated by screening and/or aspiration to produce a processing intermediate material sometimes called *meats*. The major objective of removing the hulls is the reduction of the fiber content in meals for feeding poultry and swine. As a result, protein content is increased, and the volume of material sent to the extractor is reduced. Partial retention of hulls in the meats was required earlier to improve handling in flaking and screw pressing operations, but hulls are no longer required with the advent of expanders equipped with oil drainage cages. It is common practice either to leave sufficient hulls with the meats to just surpass minimum protein guarantees of meals, or to adjust high protein meals to industry trading standards by adding back hulls after extraction. Currently, dehulled (low fiber) soybean meal trades at 48 percent protein, nondehulled soybean meal at 44 percent, cottonseed meal at 41 percent, and dehulled sunflower seed meal at 42 percent protein.

The "meats" are heated, with the addition of steam, to increase the moisture content for plasticizing if needed in flaking. Heaters (cookers), desolventizer-toasters (DT), dryers, and coolers (DC) used in the oilseed processing industry often are the shallow circular ring-cooker pan with sweep arm in a multi-stack design, shown in a later figure. In earlier techniques, seed was heated to about 74°C/165°F before flaking to about 0.3 mm (0.012 in.) thickness. However, now it is realized that phospholipase, an enzyme that makes the phospholipids nonhydratable and more difficult to remove from the oil by water degumming, is highly active at this temperature, and seeds preferably are heated to less than 57°C/135°F or to more than 85°C/185°F to avoid the range of maximum phospholipase activity. The seed next passes to the expander for rapid heating to 105121°C/220250°F, homogenization, and reshaping.

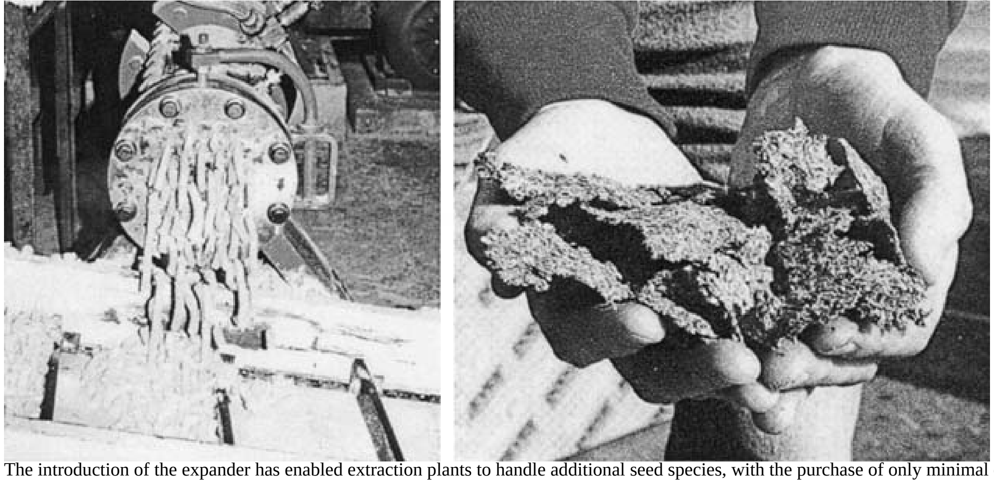
Before the mid-1980s, processing concepts were based on the classification of oilseeds into two groups. *High-oil* seeds were those containing over 30 percent oil on a dehulled or as-processed basis, and included rapeseed/ canola, oil-type sunflower seed, peanuts, safflower seed, and copra. Typically, these seeds were dehulled (shredded in the case of copra), heated, flaked, and *hard screw-pressed* leaving 56 percent residual oil in the meal. Processors desiring to recover additional oil would first prepress high-oil seeds to 1518 percent oil content using lighter-duty screw presses, and then solvent extract the press cake to less than 1 percent residual oil content; this process is called *prepress-solvent extraction*. Hard pressing was considered too inefficient for low-oil content seeds like soybeans unless nearby markets were available for the oil-rich meal. Soybeans typically have been *direct-solvent extracted* to less than 0.75 percent residual oil content. Cottonseed was processed originally by hard press, later by prepress-solvent extraction, and now mainly by direct solvent extraction techniques.

The introduction of the expander, a high-shear extruder with an interrupted-flight screw, in the latter half of the 1980s revolutionized oilseed extraction practices. Essentially all cottonseed, and approximately 70 percent of the soybean tonnage processed domestically, use this technique. The expander heats, homogenizes, and shapes seeds or flakes into porous collets (pellets) that are more dense (weigh more per unit volume) but are more rapidly extracted than flakes, approximately doubling the throughput of continuous solvent extractors. Even though the expander homogenizes the seed, prior flaking still enhances oil recovery, but can be done at 0.5 mm (0.020 in.) thickness instead of the typical 0.3 mm (0.012 in.). The flakes or collets are cooled, with some drying occurring, to about 6°C/10°F below the boiling point of the solvent before entering the extractor. Collets also drain more completely than flakes, greatly reducing steam costs for desolventizing the extracted meal.

The Anderson International Company of Cleveland, Ohio, has patented an expander with a drainage cage (Fig. 8.7) to reduce the oil content of high-oil seeds to less than 30 percent, thus enabling the production of intact collets for direct solvent extraction from completely dehulled seeds such as sunflower seed and peanuts. Replacement of the die plate (Fig. 8.7) with a hydraulically operated cone discharge head in the mid-1990s solved many of the problems first experienced in using oil-drainage-cage-equipped expanders to replace prepresses for high-oil content seeds Fig. 8.8.

Fig. 8.7 Anderson International Crop. Hivex-Series Expander™ model An-88-H with oil drainage section for preparing extraction collects from high-oil-content seeds. (*Courtesy of Anderson International Corp., Cleveland, OH*)

Fig. 8.8 Left: soybean collects for extraction exiting die plate head, Solvex Expander™; strands break into random lengths. Right: puff sheets made by hydraulically positioned cone choke head on expander. (*Courtesy of Anderson International Company, Cleveland, OH*)



The introduction of the expander has enabled extraction plants to handle additional seed species, with the purchase of only minimal cleaning and dehulling equipment where needed. Prepress-solvent extraction facilities are being replaced by expander-direct solvent extraction equipment, leaving two basic extraction processes in the modern oilseed industry: expander-direct solvent extraction, and hard press for applications where seed supplies are limited or other considerations do not warrant the construction of solvent extraction plants or the expense of skilled personnel and additional safety precautions for their operation. Hydraulic cage presses still are used in processing industrial crops such as castor seed, and for edible oils in developing countries.

Solvent Extractors

Hardly any batch-type oilseed extractors remain. Three of the more popular solvent extractors currently manufactured include: (1) *shallow bed-type* extractors, where a 0.51.0 m thick layer of collets or flakes is pulled across a linear screen and extracted by drenching with a counter-current flowing miscella consisting of solvent and solubilized oil (Fig. 8.9); (2) *diffusion belt-type*, where beds of collets or flakes are conveyed on a folding-pan belt while drenched in counter-current fashion with miscella (Fig. 8.10); and (3) *deep bed-type* which are constructed as revolving carousels with pie-shaped cells that are filled with collets or flakes (•3 m deep), extracted in counter-solvent flow fashion by drenching miscella, and unloaded. Marc (extracted flake) dropping doors and moving parts have been eliminated in the newer models and the bottoms of the cells are dragged across a fixed screen (Fig. 8.11). Shallow bed extractors are built in capacities of up to 8,000 metric tons per day, and a new carousel-type extractor at up to 10,000 metric tons per day.

Fig. 8.9 Rectangular loop-type continuous countercurrent solvent extractor. (*Courtesy of Crown Iron Works Company, Minneapolis, MN*)

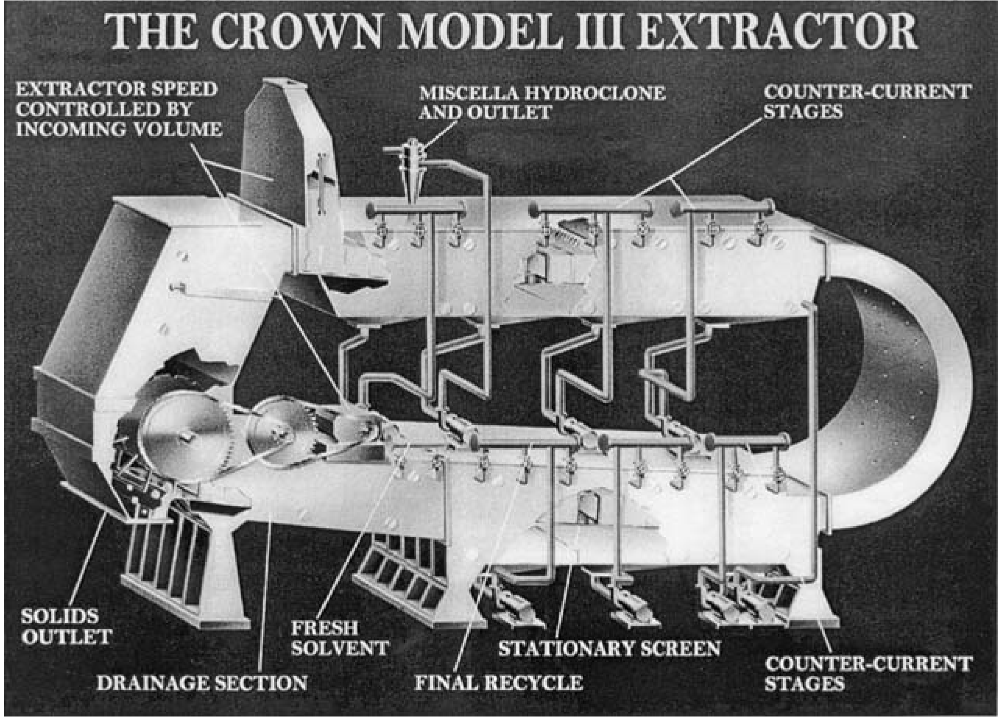


Fig. 8.10 Drawing of DeSmet LM perforated belt diffusion-type extractor. (*Courtesy of N.V. Extraction DeSmet S.A., Edegem, Belgium*)

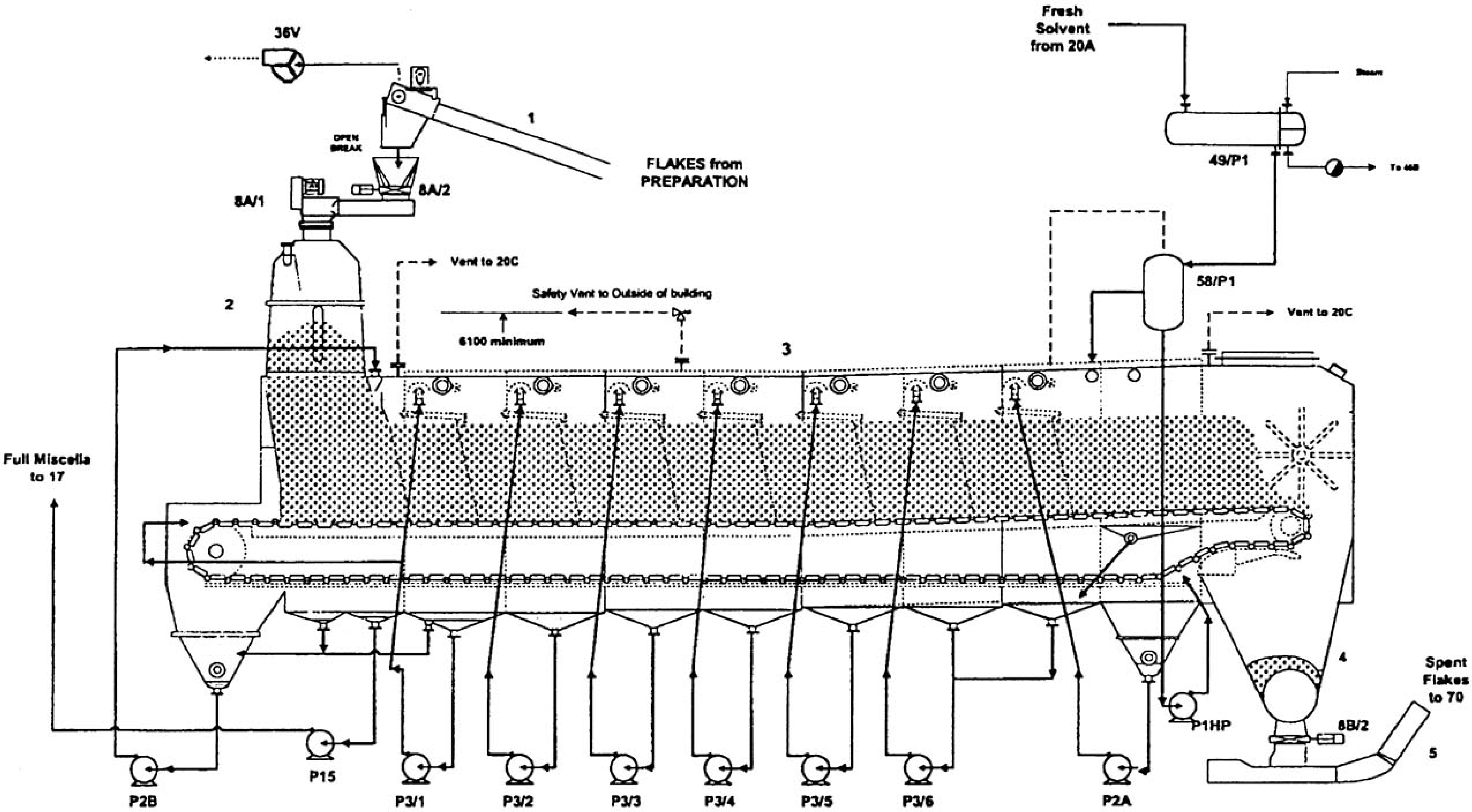
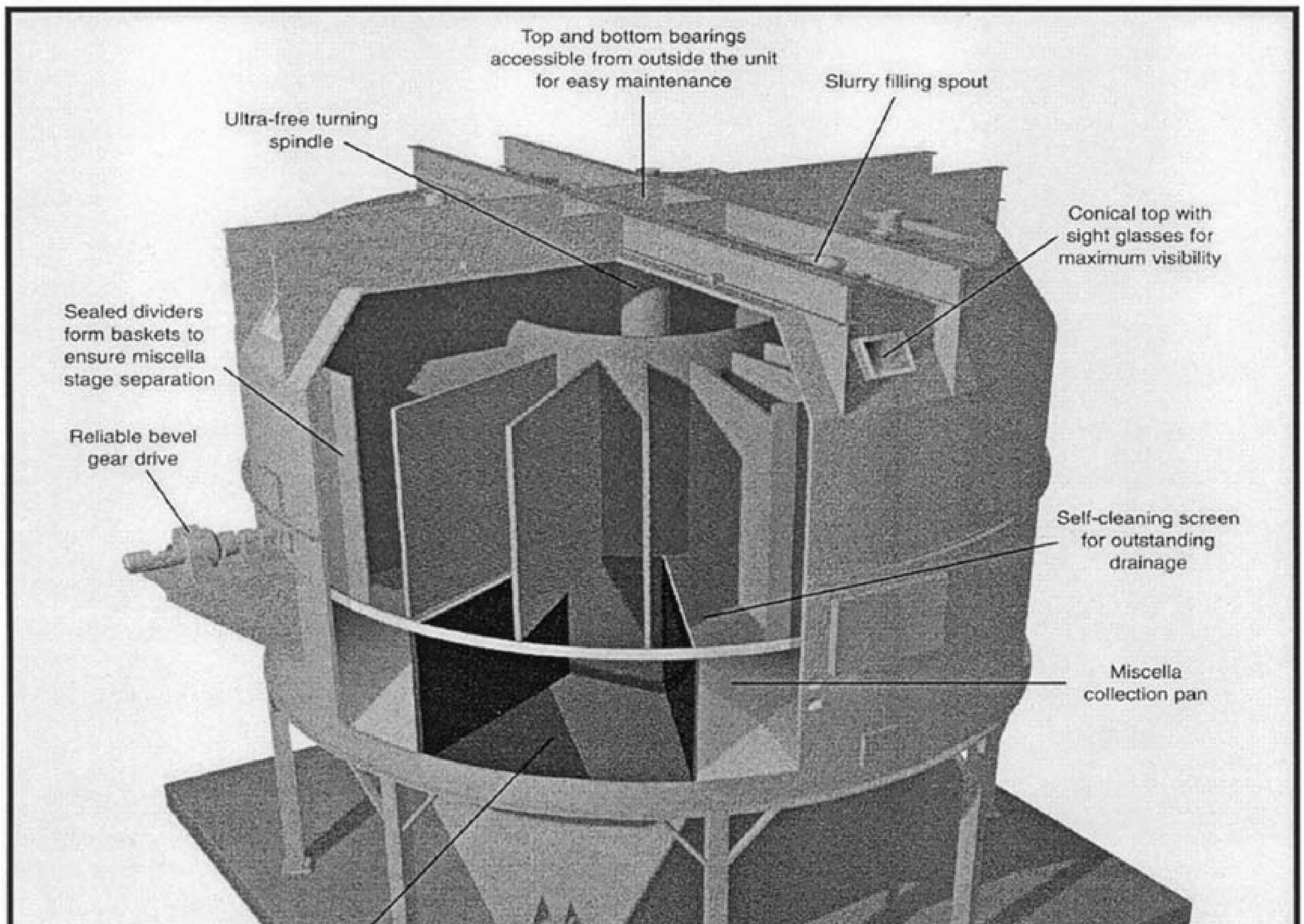
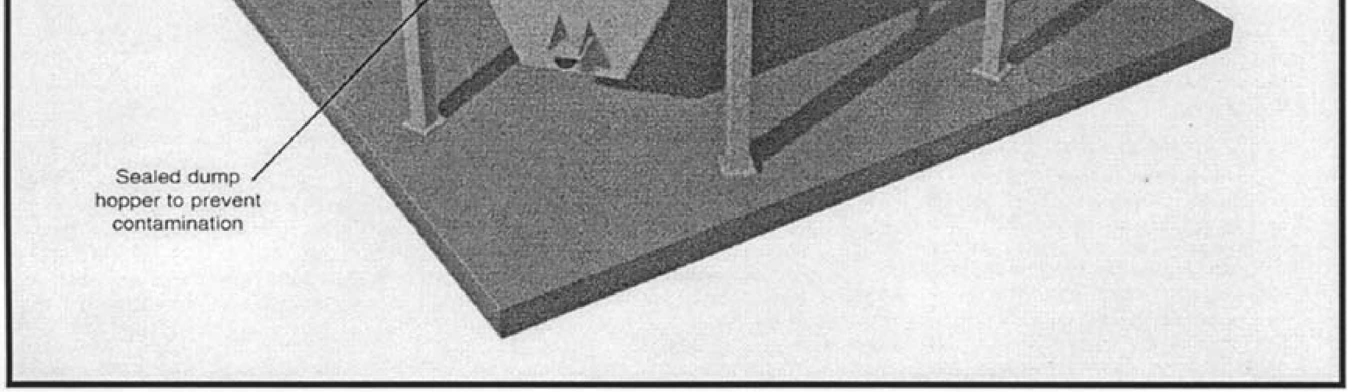


Fig. 8.11 Reflex™ "basket or circular type" 10,000 ton/day extractor. Note: basket revolves within shell. (*Courtesy of N.V. Extraction De Smet S.A., Edegem, Belgium*)





Solvents

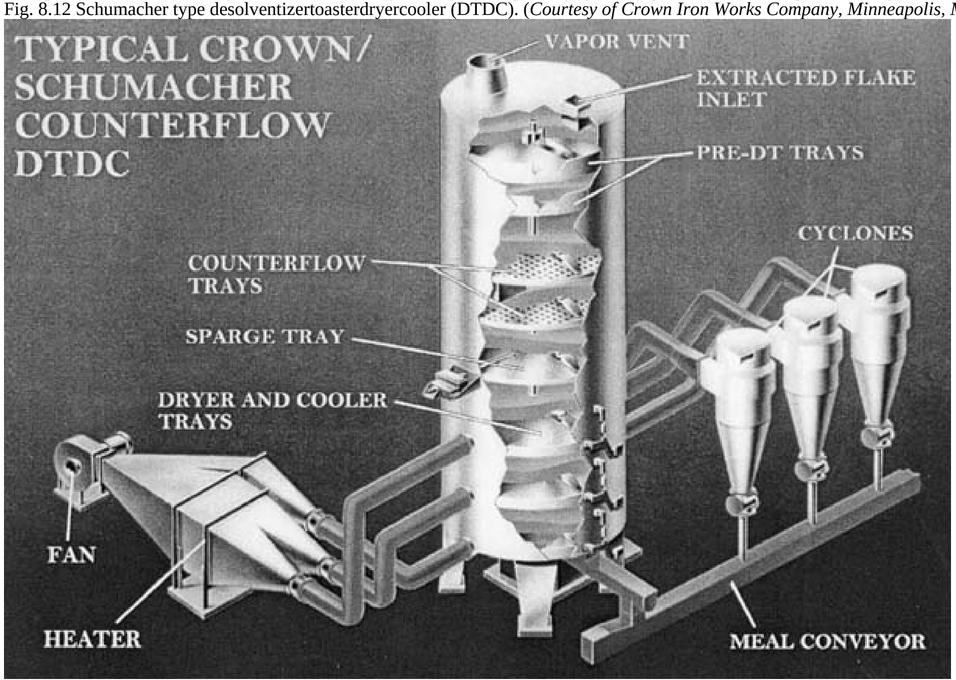
Many solvents have been proposed for extracting oilseeds, but later were found ineffective; others were used for a period, but disallowed because of health concerns about residues in food and feed products, and exposure of employees.45 Most commercial oil extraction now is done with hexane, a petroleum refinery fraction with a boiling point of 6568°C/149155°F that consists of at least 60 percent *n*-hexane with the balance being short-chain homologs and branched compounds. Some plants are just beginning to use isohexane, which boils at a lower temperature. The U.S. EPA raised many concerns about extraction solvents in the late 1980s, early 1990s. FDA's issue that *n*-hexane is a neurotoxin was put aside after industry-sponsored research showed that the problem does not exist in mixed solvent systems.46 The EPA raised concerns about the discharge of volatile organic compounds (VOC) as contributors to ozone production.47 A major containment program was instituted and resulted in domestic extraction plants reducing hexane losses from as high as more than 1 gallon per ton of seed extracted to 0.25 gallon, with some plants reducing losses to less than 0.16 gallon per ton.

A major project added to information about isopropyl alcohol (IPA) from earlier trials,48,49 and showed it can be as effective an extraction solvent as hexane.50 However, because of high retrofitting costs, it is not likely to be implemented while hexane-type hydrocarbons are allowed.

DesolventizingToasting

The extracted, drained marc contains approximately 25 percent hexane solvent holdup for soybean collets and 33 percent for flakes, which is vaporized in a DT. Some DT have cooling sections, but separate dryer-coolers (DCs) often are used in large installations. Steam is sparged into the marc as the heat source for volatilizing the solvent. The condensate must be removed subsequently by drying. The moist "toasting" operation destroys enzymes and anti-growth factors such as trypsin inhibitors and hemagglutinins in soybeans,51 and reduces meal protein solubility and digestibility by microorganisms, thus improving *rumen by-pass* or *rumen escape* in feeding cattle and sheep. Figure 8.12 shows a cutaway drawing of a Crown/Schumacher design desolventizertoasterdryercooler (DTDC) line.

Fig. 8.12 Schumacher type desolventizertoasterdryercooler (DTDC). (*Courtesy of Crown Iron Works Company, Minneapolis, MN*)



Miscella Refining

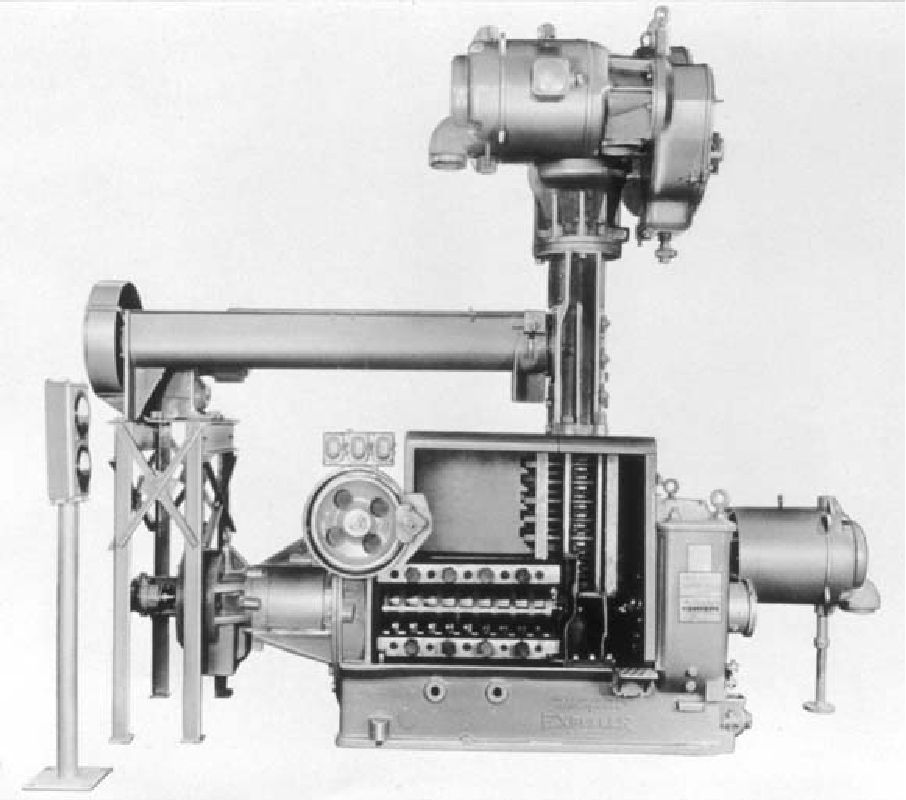
In processing most oilseeds, hexane is stripped from the miscella by distillation to produce a crude oil that subsequently is alkali or physically refined. However, gossypol and other pigments become extremely difficult to "bleach" if left in warm cottonseed oil for more than a few days. It is normal practice for cottonseed oil mills either to send their crude oil immediately to an alkali refinery or to have a miscella refinery on site, where phosphatides, free fatty acids, and color pigments are removed by alkali treatment of the oil while it is still mixed with the extraction solvent. Cooling the crude oil as produced, until refining, slows "fixing" of color.

During the process, miscella leaving the extractor contains 3035 percent oil and is raised to approximately 65 percent oil by evaporation. The FFA in the concentrate then are reacted with alkali (sodium hydroxide solution) to produce soaps that are removed with other water-soluble compounds by centrifugation. Next, the solvent is removed from the refined miscella by evaporation, and the soapstock is spread on the meal in the DT to recover its solvent. Hexane vapors from the miscella and the DT are condensed, and the solvent is recycled to the extractor for reuse. The noncondensable gases are passed through a mineral oil stripper to recover the last traces of hexane.

Screw Pressing

The main principle in screw pressing is to preheat (cook and dry) seed tissue to the point where the cell walls are brittle and rupture readily upon pressing. The seed generally is dehulled, tempered, flaked, and cooked to low moisture (24%) levels before hard pressing (Fig. 8.4) The high temperatures of hard screw pressing partially destroy the essential amino acid lysine and reduce the nutritional quality and economic value of protein meals. Pressing typically is done in a two-stage continuous screw press or expeller, as shown in Fig. 8.13. The oil usually goes to a settling tank. A layer of foam may be skimmed off. The midfraction (oil) is filtered and goes to refining. The settled solids (*foots*) are returned to the stock going into the screwpress.

Fig. 8.13 Andersen Expeller™ Press, model 55 in. TMSD. Note elevated conditioner carrying product from left to right across top of press; vertical screw press section (right side); second (higher pressure) screw press section carrying product from right to left. Other brands of hard presses have both stages on one shaft. (*Courtesy of Anderson International Corp., Cleveland OH*)



Decanters and Centrifuges

Two major types of centrifugal separators are used in some oils/fats extractions: horizontal decanters to separate liquids and solids, and vertical stacked conical disk centrifuges to separate immiscible liquids (oil and water) which also are able to remove very fine solids as sludge. A drawing of a decanter is shown in Fig. 8.14. They are built with different internal designs depending on the solidliquid ratio to be separated, and operate at •25003000× gravity *g*. When properly operated, a "dry" solids phase and a liquid phase of two immiscible liquidsaqueous and oilare separated. A cut-away of a three-phase stacked-disk centrifuge is shown in Fig. 8.15. These are vertical separators, operating at •6000*g*, which employ a spinning bowl and set of conical stacked disks. They are used primarily in three-phase mode to separate the liquid portion into two nonmiscible phases and occasional fine solids (*sludge*). The machines are not meant to handle significant amounts of coarse solids. The material to be separated enters at the bottom through a hollow spindle in the spinning platebowl assembly, and is discharged under the lower plates to be thrown against the side of the bowl by centrifugal action. The continual arrival of fresh liquid forces the earlier liquid into the plates, where the heavier aqueous phase stays at the outside of the bowl, and the oil phase moves into the center. Within the past decade, both major manufacturers of centrifuges in the Western world have introduced systems for adjusting the ratios of aqueous to oil phases exiting their machines while running. The solids, collected at the outer extremities of the bowl while the machines are running, are allowed to escape as heavy slurries by momentarily lowering the bottom of the bowl on a timed cycle.

Fig. 8.14 Schematic drawing of Westfalia Model CA 450 Continuous Clarifier Decanter. (*Courtesy of Westfalia Separator Company, Northvale, NJ*)

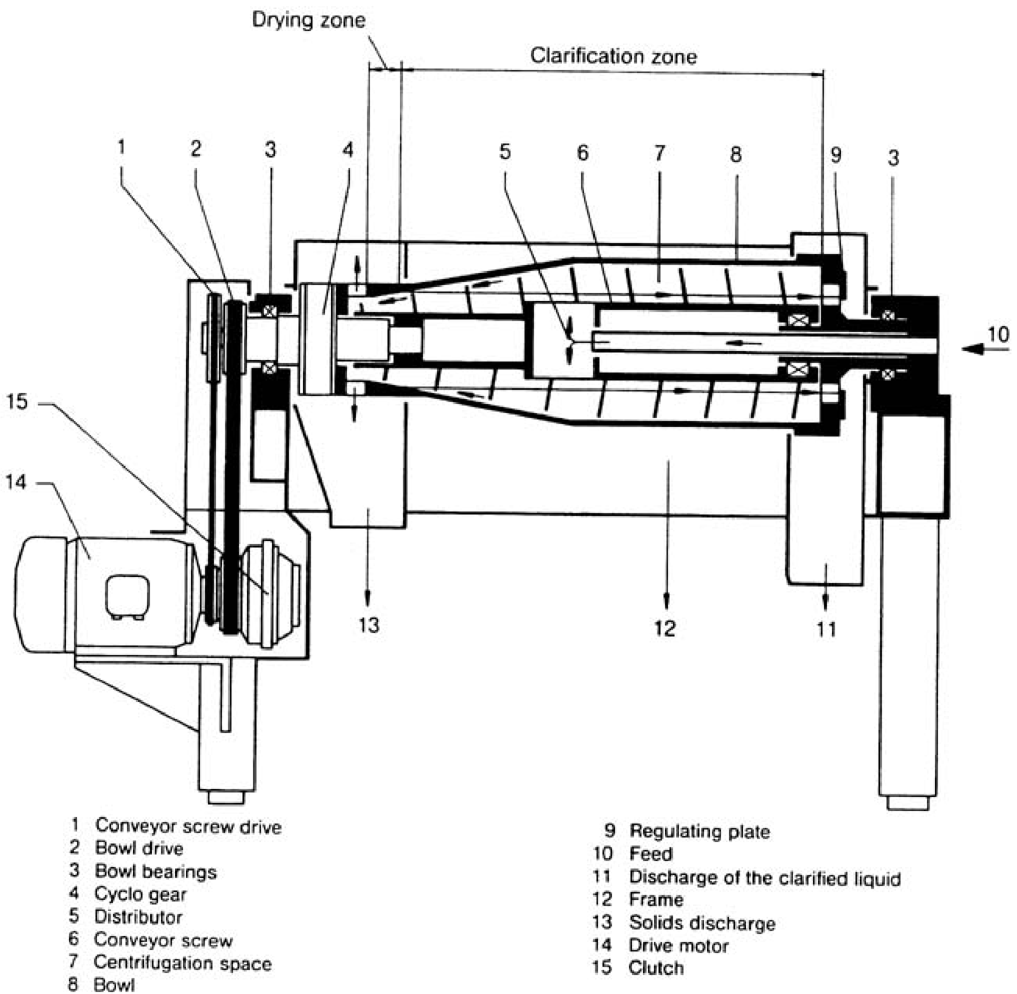
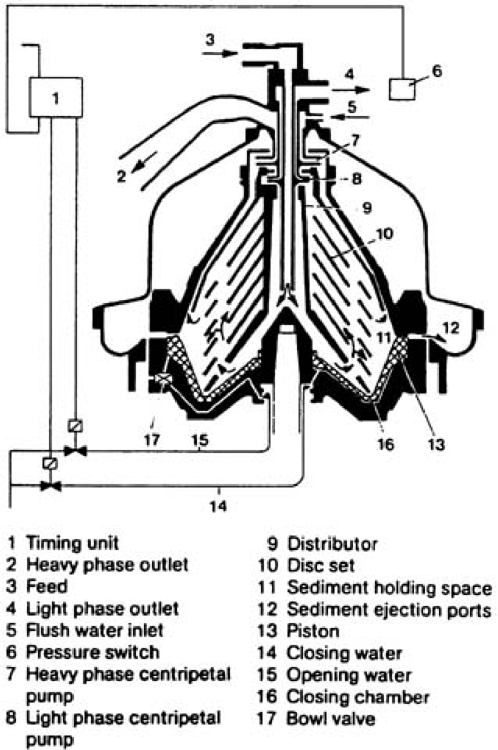


Fig. 8.15 Schematic drawing of Westfalia Model RSA 150 Refining Separator with sel-cleaning bowl. (*By permission of Westfalia Separator Company, Northvale, NJ.*)



Olive Oil

Oil is extracted from the fruit of the olive tree, *Olea europea*, by crushing cleaned, ripe olives, including the seed, using metal or stone edge rollers, mixing the paste while heating to coalesce the oil droplets, and pressing in hydraulic box-type presses. A second, lower quality oil may be obtained by returning the *pomace* (solids) for additional crushing, followed by repressing. The *musts* (liquids escaping the press) are centrifuged to separate the oil and water. Modern olive oil extraction plants crush the olives and seed using hammer mills, mix the paste while heating, and separate the must from the pomace using *decanters* (horizontal centrifuges). Virgin oils are separated by pressing, and usually are useable after filtration. Pomaces are dried and their remaining oil extracted by solvent. Many operations use the extracted, desolventized pomaces to fuel steam boilers.5254

Coconut Oil

Coconuts, *Cocos nucifera*, are hand-cut from the coconut palm with knives attached to long poles and dehusked before cracking the nut to drain away the coconut water. Then, the flesh is separated from the adhering shell and dried, raising the oil content from approximately 30 percent in the meat to 5065 percent in the resulting copra. Considerable amounts of copra are prepared by hand and sun-dried to supplement family income in tropical countries. Centralized coconut mechanical husking and dehulling facilities, which dry the coconut meat using heat generated from burning the husks and offer advantages of improved moisture control and reduced mold spoilage and aflatoxin problems, are increasing in number. In the oil extraction process, copra is: (1) cleaned by shaker screens to remove trash: (2) size-reduced by hammer or attrition mills, or fluted roller mills; (3) optionally flaked; (4) additionally dried if necessary; and (5) hard pressed or prepress-solvent-extracted.55

Palm Oil

Two distinctly different types of oils are produced from the fruit of the Southeast Asia and African oil palm, *Elaeis guineensis*, and its hybrids with the South and Central American palm, *E. oleifera*. Palm oil is obtained from the red-orange fleshy part of the fruit, which resembles an oversized olive about the size of a small chicken egg. Palm kernel oil is derived from the kernel within the nut. Well over 98 percent of the fatty acids in palm oil belong to the C:16 and C:18 group, whereas approximately 64 percent of the fatty acids in palm kernel oil consist of the C:12 and C:14 lauric group.

A palm tree produces 1015 fresh fruit bunches throughout the year, weighing 523 kg (1050 lb) each. The bunches are cut from the tree with knives attached to long poles and are transported to the oil mill, often by a small-gauge railroad. Although palm fruits somewhat resemble olives, they have a very strong lipase enzyme which is deactivated first by steaming the bunches at about 40 psig for 5075 min, and also loosens the fruits on the stalk. The fruits next are freed from the stalk in thresher drums, and passed through a digester to convert the fleshy pulp to mash. Then, the mash is pressed by twin-screw expellers or hydraulically to yield red crude oil. The nuts are dried, shells cracked, and the kernels separated and bagged or bulk-stored for sale and/or solvent extraction in a fashion similar to the processing of row crop oilseeds.56

Essential Oils

Essential oils are extracted in several ways: with fat; by distillation; by batch solvent extraction; and by critical CO2. The oldest method, used today only for extremely valuable essential oils, is *enf euragedef eurage*. This consists of layering blossoms of the material to be extracted between fat-coated glass plates and allowing the essential oil to perfuse into the fat. The blossoms are renewed daily. At the end of the season the fat is scraped from the plates, melted, poured into containers, and sold as *pomade*, or is batch-extracted with cold ethanol and sold as *extraits*. Another technique consists of macerating the material and extracting with hot fat. The most common extraction process for essential oils is steam distillation of a mash of the leaves or seeds to be extracted. Selective solvent batch extraction also is used.

8.6 Animal Fats and Fish OilsProduction and Utilization

Extraction of Lards and Tallows

A variety of methods, old and new, are used for extracting animal fats.57 The method used depends on the required properties of the nonfat (edible or inedible) product produced and the age of the facility since construction or the last modernization. In recent years, continuous processes generally have replaced batch processes for edible and inedible rendering to take advantage of heat recapture systems, economies and reduction of worker safety problems through automation, and the ability to contain vapors and minimize the discharge of offensive odors. Modern edible and inedible rendering plants located near urban areas often are under negative pressure, and the processing equipment under slight vacuum. Steam vapors from processing are condensed, and air taken into the building is water-scrubbed before release to the atmosphere. Edible and inedible animal fats and fish oils may be subjected to refinery processes similar in principle to those of vegetable oils. Industrial fats often are split into fatty acids for further processing by the oleochemicals industry.58

Edible products for human consumption are processed in facilities separate from inedible rendering plants. In the United States, construction of the processing area and designs of equipment selected are previewed and approved before startup by the Food Safety Inspection Service (FSIS) of the U.S. Department of Agriculture. Only raw materials passed by FSIS inspectors may be used. They may be chilled or frozen and shipped in refrigerated trucks or railroad cars under FSIS seal to the edible rendering facility. The rendering operation is monitored by on-duty FSIS inspectors. Edible rendering plants are limited to one species in the United States, and mixing fat of different species is not permitted.59

Edible beef tallow is produced from the fat of cattle (*Bos Taurus*). A counterpart, produced from sheep (*Ovis aries*) is known as *mutton tallow*. Lard is rendered from the fatty tissue of pigs (*Sus scrofa*). Edible chicken fat is rendered from skins and trimmings in some countries.60 Essentially, three methods are used for the separation of fat from animal tissues. In older *batch wet rendering* processes, an autoclave is filled with precut raw material, closed, and steam is injected to raise the temperature to about 140°C/280°F. After heating for 23 hr, the pressure is slowly reduced to atmospheric pressure to avoid emulsification. After a setting period, the free fat is decanted from the autoclave, the cracklings are pressed and sent to drying. Indirect heating is used in *batch dry rendering*. A reactor, equipped with a rotating agitator, and jacket at 67 bar (88103 psi) liberates the fat in 1.52 hr. On discharge, the fat is drained and the cracklings are pressed. Where batch-type cage presses are still used, the cake is formed in the shape of large wheels, sometimes called *greaves*. Much of the edible rendering industry has adopted *continuous wet rendering*, in which the minced raw material is heated first to 60°C/140°F, and then to 90°C/195°F, in an air-tight melting section within minutes. The heated material then is separated into solid and liquid fractions by a mechanical horizontal decanter (Fig. 8.14) or a dewatering screwpress, and the oil separated from the aqueous portion (stick water) and fines by a three-phase disk-type centrifuge (Fig. 8.15). The pure fat is cooled to approximately 10°C above melting point by passing through a tube-in-shell or plate heat exchanger before storage and shipping.61 Lower heating temperatures are used when also producing partially defatted beef fatty tissue (PDBFT), or similar edible products of other species, used in making processed meats, including frankfurters. Fatty tissues and trimmings are ground and heated to 43°C/110°F in a mixing tank to melt the fat but still retain the heatsetting properties of the tissues for later use. The slurry is passed through a disintegrator to rupture the fat cells, and then through a horizontal decanter for separation into solid and liquid phases. The solids are cooled for packaging and used in processed meats. The liquid is heated to 93°C/200°F to coagulate proteins, and passed through a desludging disk-type centrifuge to "polish" the oil by separating the stick water and removing any solids.62 In recent years, minimizing heat damage to fats and solids has been emphasized in edible and inedible rendering. Higher heat treatments using open cooking kettles produce darker and more strongly flavored fats and cracklings, also useable as edible food ingredients which are preferred in specific applications. But, for the most part, bland edible animal fats are preferred and may later be neutralized (alkali refined), bleached, hydrogenated, deodorized, rearranged (interesterified), and fractionated in ways similar to vegetable oils before incorporation into margarines/spreads and shortenings, or used for frying.63

Inedible animal products include viscera and nonedible parts, carcass parts condemned by inspectors, fatty trimmings from butcher shops, outdated fresh and processed meats, and dead animals that have been skinned. In modern rendering, the flesh and bones are ground or chipped, heated to denature the protein and release the aqueous and fat fractions; next, the slurry is passed through a decanter or dewatering press to separate the solids from the liquid which is further heated to reduce viscosity in separating fat, stickwater, and sludge. The meat/bone fraction is dried, reaching high temperatures specified by the federal government to destroy pathogenic bacteria. The solids then are pulverized by hammer mills and sieved with the separated large pieces recycled to the grinders. Inedible renders sometimes are called *the original recyclers*, an increasingly significant function as the world wrestles with problems of biodegradable waste disposal.

Restaurant Grease

Used frying oils typically are processed in a separate system. Restaurants and large commercial fryers dump the spent oils into on-site bins, equipped with steam heating coils provided by the grease processor. Pilferage of grease has been a long time problem in the industry, and covers on the bins are kept locked. The collector may attach a steam line to heating tubes in the bin to melt the fat if required; then, it is pumped to a tank on the collection truck. An exchange of fats occurs between the oil and the product during frying. For example, the used grease may contain substantial quantities of chicken fat. At the processing plant, the liquid grease is heated to 100°C/212°F and flashed into an evaporator flash chamber operating at 8285°C/180185°F and 21 in. mercury vacuum to remove the moisture. Next, the dried grease is passed through a decanter to remove any solids and partially cooled before storage.64

Animal Fat Specifications, Production, and Utilization

Specifications, or typical analyses, of edible, industrial and feed animal fats are shown in Table 8.8.59,65 Much emphasis is placed on titer, the solidification temperature of fatty acids in a saponified sample of the fat or oil (AOCS Method5 Cc 12-59). Feed ingredients in the United States may not be as aesthetically attractive as food ingredients, but are required to pass the same toxicology standards.

TABLE 8.8 Specifications, or Typical Analyses, of Edible and Industrial/Feeding Animal Fatsa, b

Grades	Titre Min. (°C)	FFA Max. (%)	FACc Color Max.	Moist. Max. (%)	Insoluble Impurities (%)	Unsaponifiables Max. (%)	Iodine Value	Total Fatty Acids (%)
Edibled								
TallowUSDA certified	41.0	0.75	3	0.20	0.05			
LardUSDA certified	38.0	0.50	e	0.20	0.05			
InedibleIndustriald								
Top white tallow	41.0	2	5	1 MIUf				
All-beef packer tallow	42.0	2	None	1 MIU				
Extra fancy tallow	41.0	3	5	1 MIU				
Fancy tallow	40.5	4	7	1 MIU				
Bleachable fancy tallow	40.5	4	None	1 MIU				
Prime tallow	40.5	6	13-11B	1 MIU				
Special tallow	40.0	10	21	1 MIU				
No. 2 tallow	40.0	35	None	2 MIU				
"A" tallow	39.0	15	39	2 MIU				
Choice white grease	36.0	4	13-11B	1 MIU				
Yellow grease	g	15	39	2 MIU				
Typical analysesFeed grade fats								
FGF (for all feeds)	3438	15	37	2 MIU			55	
FGF (for milk replacers)	3841	5	9	1 MIU			45	
All-beef tallow	3842	5	7	1 MIU			40	
All-pork fat	3238	15	37	2 MIU			58	
All-poultry fat	2835	15	19	2 MIU			65	
NRAh suggested standards for feeds								
Tallow		46	19	0.5	0.5	0.5	4858	90
Choice white grease		4	11A	0.5	0.5	0.5	5868	90
Yellow grease		15	3739	1.0	0.5	1.0	5879	90

Hydrolyzed animal/vegetable (A/V) blend	4050	45	1.5	0.5	2.5			85	90

[a] NRA, *Pocket Information Manual: Buyer's Guide to Rendered Products*, National Renderers Association, Inc., Alexandria, VA, 1993.

[b] Lusas, E. W., and Riaz, M. N., "Fats in Feedstuffs and Pet Foods," in *Bailey's Industrial Oil & Fat Products*, 5th ed., Vol. 3, *Edible Oil & Fat Products: Products and Application Technology*, Y. H. Hui (Ed.), pp. 255329, John Wiley & Sons, Inc., New York, 1996.

[c] FAC = Fat Analysis Committee.

[d] American Fats and Oils Association grading standards for tallow and grease.

[e] Lard color maximum = 1.5 red, Lovibond color, 5.25-in. cell.

[f] MIU = moisture, insoluble impurities, and unsaponifiables combined.

[g] When required, titer to be negotiated between buyer and seller on contract-by-contract basis.

[h] NRA = National Renderer's Association.

An estimated 4.18 million metric tons of animal fats were produced by rendering in the United States in 2000.[66] Of this amount, approximately 18 percent and 6 percent was edible tallow and lard, respectively, and 41 percent and 35 percent was inedible tallow and grease. Approximately 15 percent and 34 percent of the edible tallow and lard, respectively, and 37 percent of the inedible tallow and grease was exported. Of the inedible tallow and grease used in the United States, an estimated 75 percent was used as animal feed, 16 percent was converted to fatty acids by the oleochemicals industry, 4 percent was used in soaps, and 3 percent in lubricants. Nonedible animal fats are the lowest cost domestic fat sources. During 2000, their market price per pound was less than fuel oil, and some rendering plants chose to burn them as fuels. In 2001, animal fats were included with vegetable oils for federally supported trials of biodiesel.

Fish Oils

Raw materials for producing fish oils include: (1) *pelagic-type* (surface feeding) fish pursued for reduction to meat and oil; (2) waste products produced at facilities that process edible fish; and occasionally (3) *by-catch* species also netted with the primary catch. The type of processing used depends on geographic location, species of fish normally taken in the area, and whether at an on-shore plant or a factory ship. As in continuous wet rendering, whole fish or trimmings are ground, cooked, pressed, or decanted to yield solids and liquid, and the liquid fraction is centrifuged into stickwater and oil.[67,68] A significant fish protein industry, using trawlers for netting and mother ships for processing, has developed in Alaska. Local fish species have low oil content, and most of the oil produced is used on board to power steam boilers and engines.[69] Dried animal and fish meals do not contain natural antioxidants, and it is common practice to preserve their fat by using synthetic antioxidants. This is especially important in fishmeals, where rapid polymerization can generate sufficient heat to cause spontaneous combustion.

Fish oil may be alkali-refined, bleached, hydrogenated, deodorized, and used in making margarine/spreads, other food products, and nutraceuticals. Because it contains fatty acids with three or more double bonds, it readily polymerizes in the presence of air and is a major *drying oil* used for coatings. This property is further enhanced by *kettle bodying* (heating with mixing) and *blowing* (heating and mixing while bubbling in air). The fatty acid compositions of different fish oils, processing, and uses are presented by Bimbo,[70] and Bimbo and Crowther.[67] Nonfood uses include animal feeds, attractants and lures, automotive gaskets, caulking compounds, ceramic deflocculants, core oils, fatty acids, fatty chemicals, fermentation substrates, fire retardants, fuel oil, illuminating oil, insecticidal preparations, leather tanning, lubricants and greases, mold-release agents, mushroom culture, oil-field chemicals, oiled fabrics, ore flotation, plasticizers, polyurethane lures, pressed wood fiber boards, printing inks, protective coatings, refractory compounds, rubber compounds, rust proofing, soaps, specialty chemicals, and tin-plating oils.

Feeding Animal and Marine Fats

Animal and marine fats provide approximately 2.5 times more calories per unit dry weight than carbohydrates or proteins, and generally are economically advantageous. Digestive tract capacities limit the growth of broilers and turkeys, and the productivity of laying hens, and these animals respond well to *high-energy diets*, which are synonymous with fat-containing feeds. This also is true, to a lesser degree, for pigs and fish. However, high intakes of oil can disrupt normal function of rumens, and various by-pass techniques are used in feeding cattle. The most common has been to hydrogenate fatty acids to melting points above rumen temperatures. Because of concerns about bovine spongiform encephalitis (BSE or "mad cow disease"), the feeding of mammalian meat meals to ruminant animals has been outlawed in the United States and much of Europe, but these restrictions do not apply to feeding of tallows and greases.

Special needs must be addressed in pet foods. The lower melting point animal fats (choice white greases) provide more odor appeal than the tallows. However, the lower melting fats can wick through and disfigure paper packaging. Solutions have included using the higher titer fats, including nonpermeable plastic in multi-wall bags, laminating fat barriers onto paper, and filling into plastic bags. Some foods for working, guard duty, and sled dogs, and some fish feeds, require 3040 percent fat content. However highfat formulas do not extrude easily even on modern machinery. Solutions have included using twin-screw extruders, which convey better than single-screw machines, the inclusion of full-fat soybean meal where the oil still is bound within its natural matrix, and enrobing the products with fat after extrusion and drying.

In addition to increasing caloric density, feed palatability and appearance, feed efficiency, and reducing feed costs, the feeding of animal, marine and vegetable fats can:

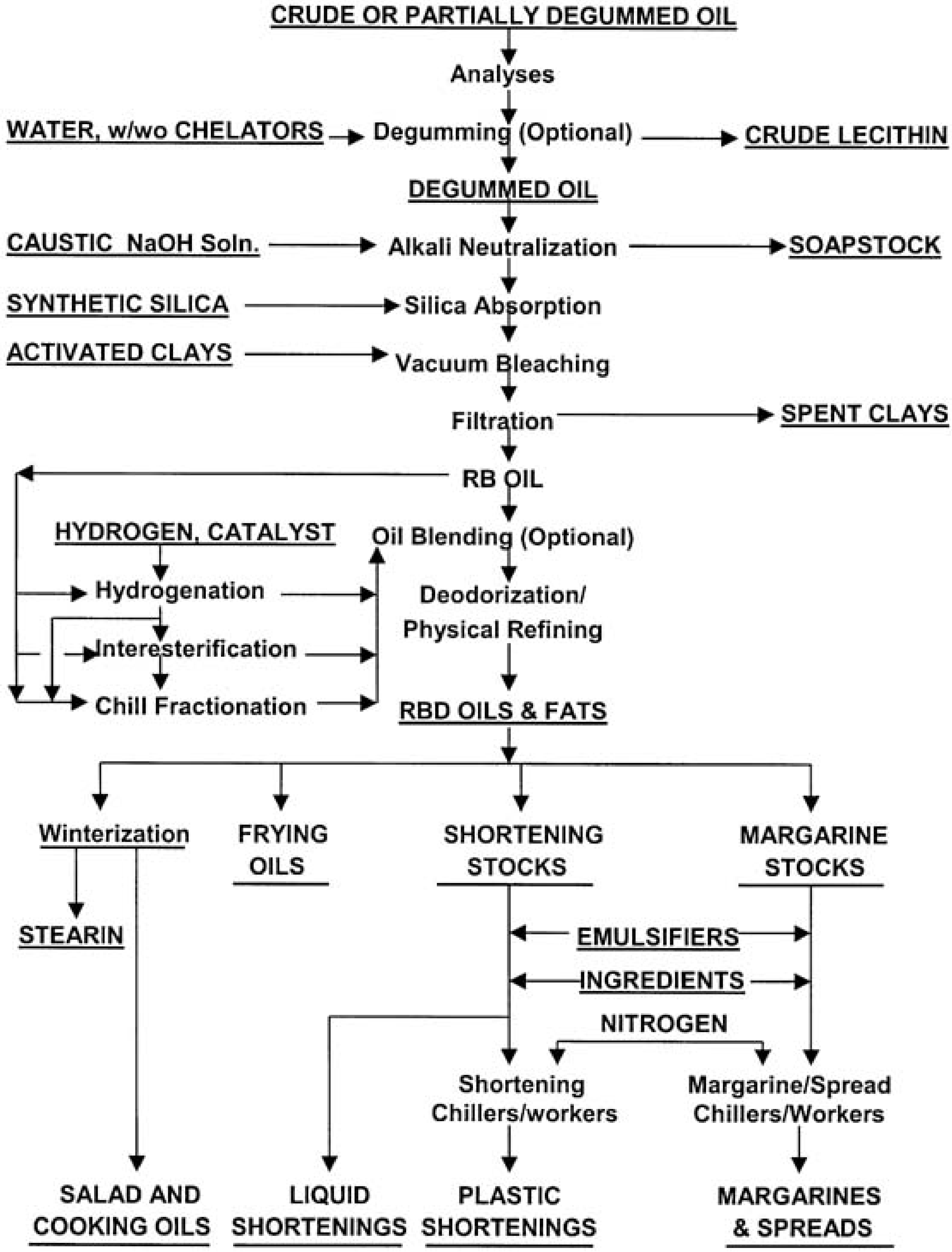
1. Provide needed molecular structures through dietary essential fatty acids (EFAs) and phospholipids.
2. Increase blood glycogen levels and endurance in working animals like horses and sled dogs.
3. Lower the heat of reaction during digestion and metabolism, thus increasing tolerance of heat by large animals in summer time or tropical areas.
4. Prevent dermatitis and improve the appearance of skin and hairan important effect of polyunsaturated fatty acids.
5. Carry fat-soluble vitamins and natural color compounds, for example, yellow colors to improve the color of egg yolks; red ("pink") colors for feeding salmon.
6. Prevent segregation of mixed feeds.
7. Lubricate feed processing machinery.
8. Bind heat-sensitive flavorings, vitamins, medications, "instant gravy mixes" to pet foods and feeds after extrusion and drying.
9. Reduce dustiness of feeds and improving animal health.[65]

8.7 Refining Fats and Oils

Technically, "refining" means alkali neutralization of free fatty acids in the oil. But over time, all post-extraction processing of oil ingredients has become know as *refining*, and the facility in which it is conducted a *refinery*. Converting the resulting ingredients into margarines and spreads, and bottling oils often are done at a different location or firm.

The objectives in refining and processing edible fats and oils include: (1) removal of free fatty acids, phospholipids (gums), color and off flavor/odor compounds, and toxic substances to produce light-colored, bland products with long shelf lives; (2) obtain a mixture of TAG with the desired solids content profiles over the temperature range of product use; and (3) preparation and storage of semi-solid products with desired textures. A flow sheet for refining and processing fats and oils is shown in Fig. 8.16. Refining procedures reliably purify oils extracted from cottonseed, peanuts, and corn germ that have been contaminated with (water-soluble) aflatoxins, although the resulting meals may have to be used as fertilizer or destroyed.

Fig. 8.16 Composite flow sheet of oils and fats refining and processing.



Oil Receiving and Handling

Maximizing yields of saleable oil requires even more detailed analyses and attention to lot-to-lot differences during refining than in preparing oilseeds for extraction. Using soybean oil as an example, the first priority on receiving a shipment, or the output of an adjoining extraction plant, is to characterize the overall quality of the oil and determine what needs to be done to prepare it for the market. The responsibil-ities of the refinery may include preparing freshly extracted oil ready for sale as Crude Degummed Soybean Oil, Once Refined Soybean Oil, or Fully Refined Soybean Oil for export under National Oil Processors Association (NOPA) Trading Rules,71 or other agreements made with the buyer.

To prevent the hydration and precipitation of phosphatides during storage and shipping, the phosphorous content of crude soybean oil must be reduced to less than 0.02 percent (200 ppm) before entering the trade. This usually is done by water degumming.

One of the first tasks, if the refinery purchases soybean oils for processing, is to check the composition of the received oil against the contract (usually NOPA Trading Rules), since this determines the final price paid:

1. Flash point (indicator of residual extraction solvent)AOCS Method5 Cc 9c-95(97) (>250°F)
2. Unsaponifiable matterAOCS Method Ca 6a-40(97) (<1.5%)
3. Free fatty acids, as oleicAOCS Method Ca 5a-40(97) (<0.75%)
4. Moisture and volatile matterAOCS Method Ca 2d-25(97) and insoluble impuritiesAOCS Method Ca 3a-46(97) (<0.3%)
5. PhosphorousAOCS Method Ca 12-55(97) (<0.02%)

Additionally, the analytical laboratory will estimate how much saleable oil can be produced from the lot usually by Neutral Oil and Loss analysis (AOCS Method Ca 9f-57). In this procedure, a solvent-diluted sample of the oil is poured over a column packed with activated alumina (aluminum oxide). After evaporation of the solvent, the weight percentage of the oil that passed through the column is considered "neutral oil," and the weight of oil retained is the "loss." The analyst may run a bleach test (AOCS Method Cc 83-63) or a refining test (AOCS Method Ca 9d-52) if concerns exist about reducing the color of the oil to the acceptable range.

Degumming, Phosphatides Processing, and Lecithin Uses

Phosphatide contents of common vegetable oils are shown in Table 8.9.72 Soy, corn, and canola phosphatides are separated in North America; some of the former

Soviet Union countries separate sunflower phosphatides. For soybean oil, the relationship between phosphatide and phosphorous content is

$$[\text{phosphatide }(\%) \times 10^4]/3.17$$
$$= \text{phosphourous (ppm)}$$

Phosphatides foul the bleaching earth, poison hydrogenation catalysts, and impart dark color to heated oils during deodorization and later in deep fat frying operations and are essentially removed in modern oil processing.

TABLE 8.9 Phosphatide Contents of Common Vegetable Oila

Type of Oil	Phosphatide Content (%)	As Phosphorousb (ppm)
Soybean	1.03.0	311940
Corn	0.70.9	220280
Safflower	0.40.6	130290
Sunflower	0.50.9	160290
Peanut	0.30.4	95190
Canola (super degummed)	0.16	50
Canola (crude)	1.03.0	311940

aFrom: Farr, W. E., "Refining of Fats and Oils," in *Introduction to Fats and Oils Technology*, 2nd ed., R. D. O'Brien, W. E. Farr, and P. J. Wan. (Eds.), pp. 136157, AOCS Press, Champaign, IL, 2000, with permission.

bPhosphorous calculated as: [phosphatide (%) × 104]/31.7 = phosphorous (ppm).

Often, phosphatides are the first component removed from crude oil. A reduction of phosphorous to less than 200 ppm (0.02%) in crude oil is required by NOPA Trading rules to prevent hydration and settling of gums during oil storage and shipping. If lecithin is saved at the refinery, the hydratable phosphatides are separated by simple *water degumming* (hydration with deionized water, followed by centrifugation). However, some of the phosphatides will have been converted to nonhydratable forms (NHP) by enzyme action. In this case, a chelating agent (usually phosphoric acid because of lower cost) will be added to the alkali in the FFA neutralization step to return the NHP to hydratable form. Palm oil has very low phosphatide levels, and is physically refined after acid degumming and bleaching. The addition of chelating acids to crude row crop oils before alkali refining will remove hydrated phosphatides with the centrifuged soapstock, but viscosity is high and appreciable losses of neutral oil can occur by occlusion. Thus, many soybean oil refineries, without markets for crude lecithin, still run preliminary degumming operations and spread the gums with the later removed soapstock over the desolventized marc for drying in the dryer cooler (DC) and sale as part of the soybean meal. Reportedly, a few refinery operators have developed successful physical refining methods for soybean oil, only to find no economic advantage over the alkali refining route.72

Commercial lecithin is produced by *water degumming* (precipitation from oil with ion-exchange-treated water), separation by stacked disk centrifuge, and vacuum drying to less than 1 percent moisture content. Crude lecithins contain 7072 percent acetone insolubles (AI) and are standardized to 6264 percent and an acid value of 30 by the addition of oil and fatty acids before trading. Crude lecithins may be treated with acetone to obtain free-flowing powders with 9598 percent AI. Lecithin can be additionally purified, bleached, fractionated, hydrogenated, hydroxylated, acetylated, sulfonated, and halogenated.73 One domestic company makes 13 lines of lecithin for food uses alone.

Food applications of lecithins include: emulsification, wetting and dispersing agents, modification of baking properties, pan release agents, viscosity reduction of melted chocolate, anti-spattering agents in margarine, antioxidant effects, and nutritional supplementation. Feed uses include wetting and dispersing agent in calf starters, and nutritional supplementation; choline is a recognized vitamin, and inositol has been found essential in feeding some species of fish. Industrial uses include: emulsification and dispersion of active agents in pesticides, dispersing agents and stabilizers in paints, dispersing agents and stabilizers for magnetic tapes, softening agents and penetrants for leather, and softening and lubrication of textiles. Cosmetic uses include: foam stabilizers and emollients in hair care, and emulsification, emollient, refatting and wetting agents in skin care. Pharmaceutical applications include: emulsifiers in parental nutrition, softening agents and carriers in suppositories, and emulsifier and penetration improvement in cremes and lotions.74

The process for removing phosphatides is known as *degumming*. There always will be some natural (NHP) in crude oil, but the development of additional NHP during seed extraction can be minimized, as explained earlier. In preparation for degumming, a tank large enough to supply the refinery with a uniform oil feed stock for a suitable period of time is filled and mixed. Samples are taken for free fatty acids and calcium and magnesium contents. Earlier practices of adding an amount of water equal to the weight of the phosphatides have been defined more precisely recently, with

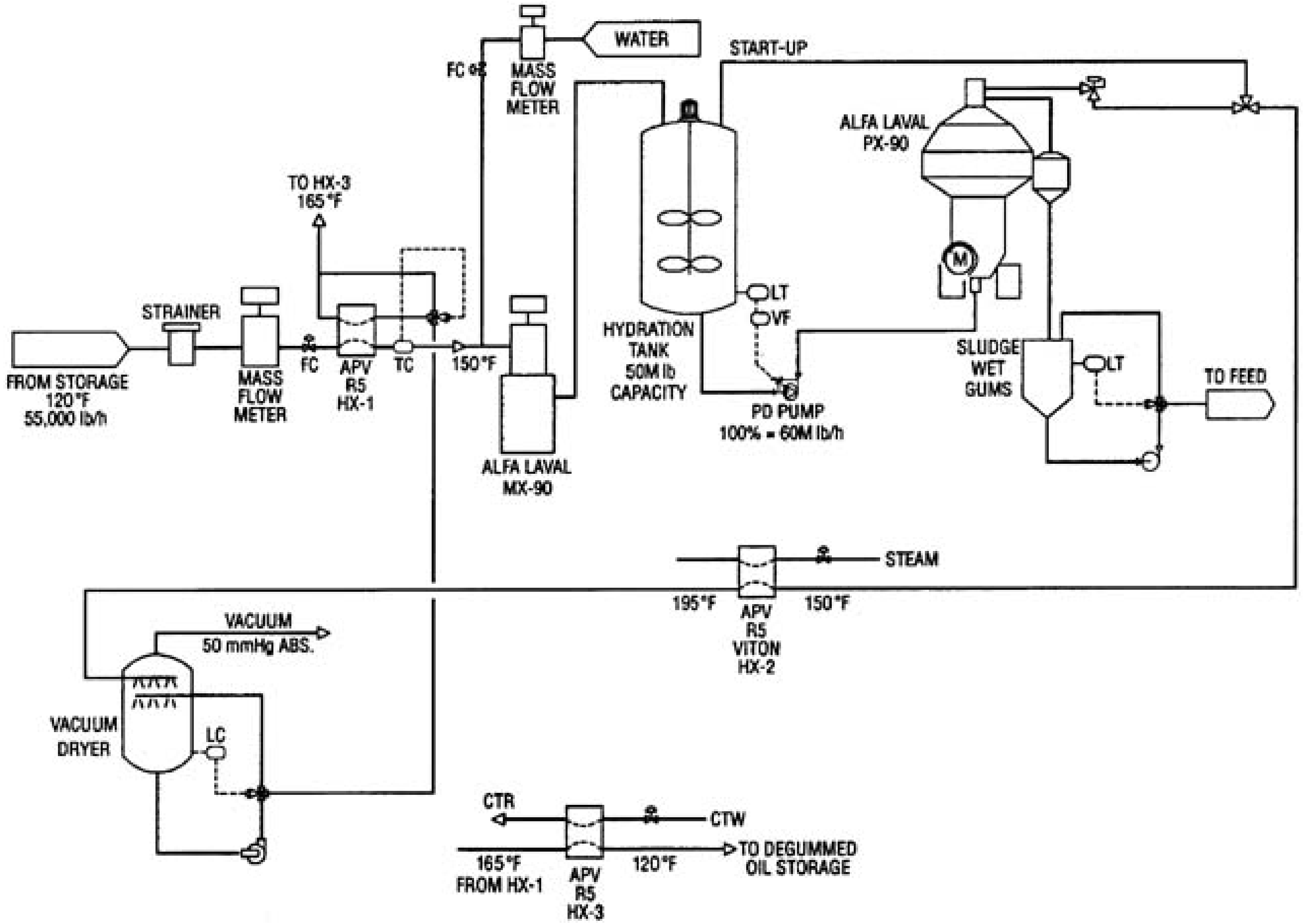
$$\text{added water} \quad = \quad (\text{ppm P} \times 3.17 \times 10^{-4})$$
$$\times\, 0.7 \text{ now recommended}$$

The amount of phosphoric acid used is

$$H_3PO_4 = [(Ca + Mg)/2] \times 10$$

with all components expressed in ppm. The phosphoric acid and water are added to the warmed (65°C/150°F) crude oil stream, pass through a high-shear mixer, and are pumped to the hydration tank. Details of the process, using an Alfa Laval (Lund, Sweden) PX-90 centrifuge rated at 33,000 kg/hr for degumming, are shown in Fig. 8.17. After degumming, the crude oil is vacuum dried to <0.3 percent moisture and volatiles content, and cooled to 50°C/120°F for storage or shipment, but this step can be omitted if the oil is refined next locally.72

Fig. 8.17 Flow sheet of a modern water or acid degumming line. (*From Farr, W. E., "Refining of Fats and Oils", in* Introduction to Fats and Oils Technology, *2nd ed., R. D. Brien, W. E. Farr, and P. J. Wan (Eds.), pp. 136157, AOCS Press, Champaign, IL., 2000. With permission.*)



It is noted that the objective in acid degumming is to chelate the calcium and magnesium ions and render the nonhydratable phosphatide forms hydratable. Phosphoric acid is most commonly used and is the least expensive, but citric and malic acids are effective, as well as ethylenediaminetetraacetic acid (EDTA). Acid-treated phosphatides are not used for the production of commercial lecithins. Extensive reviews on oil degumming have been prepared.75,76 Lurgi, a German equipment manufacturer, has developed an EnzyMax™ process that cleaves the nonhydratable phosphatides with a phospholipase-B at the triglyceride's second carbon to produce a lysophosphatide that is insoluble in oil and is removed by centrifuging.77

In earlier times, the nonhydratable phosphatides content was determined by analyzing a water degummed sample for phosphorous, but the procedure took too long for practical use. In reality, phosphorous is only a tag and loss of hydration results from the presence of divalent cations, primarily calcium and magnesium. Refineries now use induction coupled plasma (ICP) spectrographs for analyzing divalent cations rapidly in aspirated crude oil, and adjust the amount of phosphoric acid used for each "day tank lot" of analyzed oil. ICP units cost in the range of $70,00090,000, but, reportedly, pay for themselves through increased oil yields in as little as 6 months in mid-size (500 tons/day) refineries.

The crude oil from which gums were taken for lecithin production still contains nonhydratable phosphatides, but can be treated with a chelating agent before alkali neutralization and will be removed with the soapstock by centrifugation. Provision must be made for the added acid in calculating the amount of neutralizing alkali added.

Alkali Neutralization

The elimination of washing neutralized oil water is the major breakthrough of the decade in oils refining, with *Modified Caustic Refining* or *Silica Refining* processes still being optimized. Phosphatides are removed by degumming, as already described, and FFA in the crude oil are still neutralized with sodium hydroxide (caustic) solution. But, the refinery then has the choice of water-washing the oil after removal of soapstock, or adsorbing the residual soaps on silica hydrogel before bleaching, thus eliminating problems of disposing the wash water. Process demonstration by the W. R. Grace Company started in 1986, with increasing growth in the commercial use of Trisyl™ in the 1990s. Several suppliers now exist.72

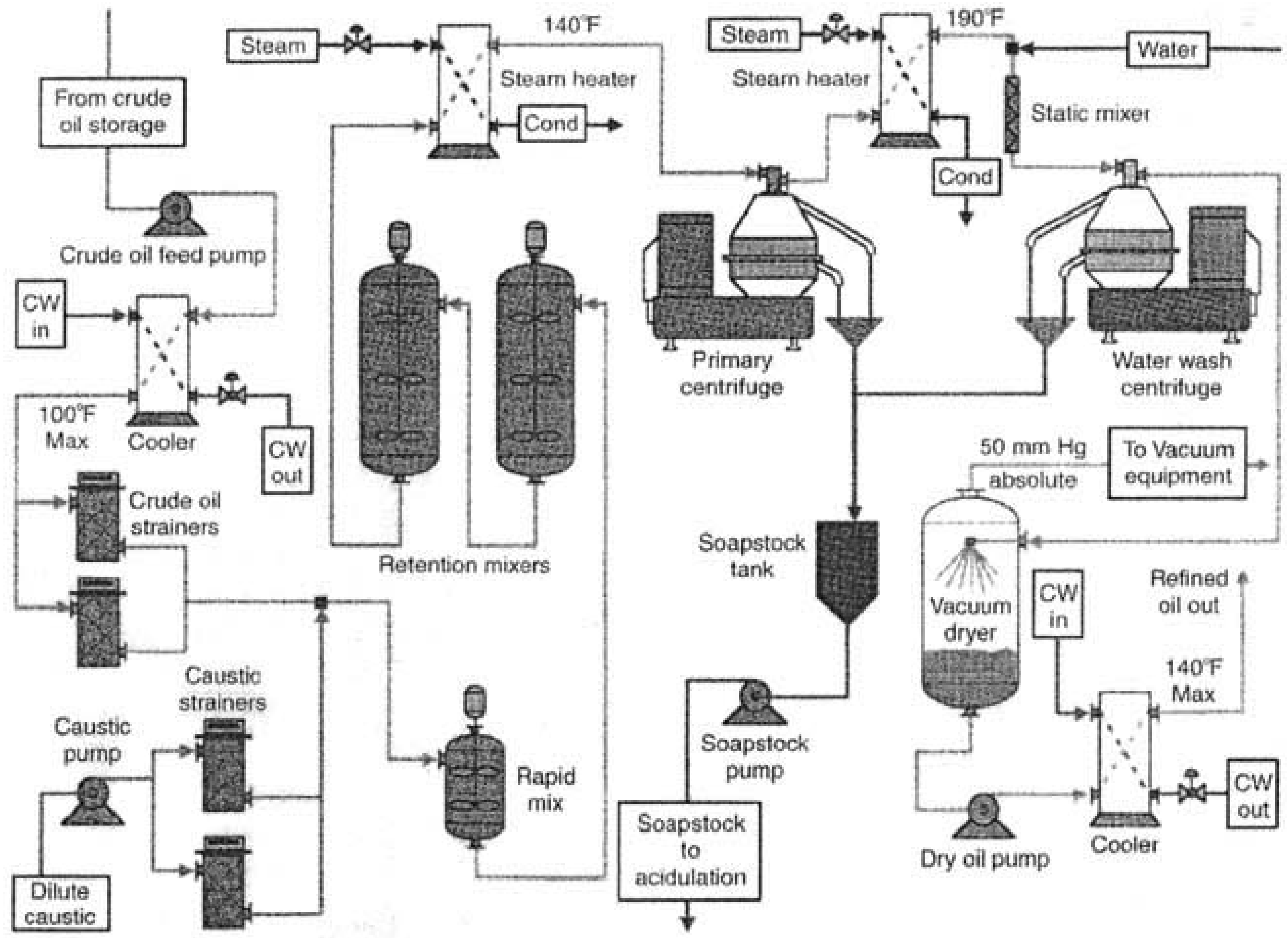
Two major processes have been used for alkali neutralization of FFA in row crop oilseed oilsthe "long mix" and the "short mix." The "short mix" process evolved in Europe, runs at a higher temperature, and reportedly is effective with a number of oils. The "long mix" process developed in the United States, and has been championed for refining soybean oil. It respects the principle that chemical reactions occur more rapidly, and are harder to control, at higher temperatures (van't Hoff rule that the speed of reaction doubles with each 10°C rise in temperature). Traditionally, the long-mix process starts with crude soybean oil at ambient temperature, uses a low-concentration caustic solution, and has a mixer retention time of 15 min after which the oilcaustic mixture is heated to 70°C/160°F to reduce its viscosity before centrifuging. In the short-mix process, crude oil is heated to 90°C/194°F, mixed with high-concentration caustic for 1 min and centrifuged.78 The throughput per hour of both systems is the same since the volume for the holding time is built into the long mix line. A flow chart of a long-mix neutralization process is shown in Fig. 8.18. The recommended retention time for soybean oil in this system is 6 min and had been shortened by a major improvement in causticoil mixers released by the Alfa Laval Company, Tumba, Sweden, in the mid-1990s.72

The amount of caustic to be added for neutralization (*treat*) is calculated as

$$\text{percent treat} = \Big[(\text{factor} \times \text{percent FFA}) \\ + (\text{percent excess})/\text{percent NaOH}\Big] \times 100$$

where factor = 0.142, NaOH is determined from the 20° Be of the caustic solution, and percent excess is selected from the following ranges based on experience: degummed soybean oil, 0.010.05; nondegummed soybean oil, 0.150.25. Continuing with Fig. 8.18, soybean oil from storage is adjusted to 38°C/100°F, passed through strainers, mixed with the treat in the rapid mixer, held in the retention mixers for 6 min, heated to 60°C/140°F, and passed through the primary (first) centrifuge.72 The soapstock is returned to the extraction plant with surplus or acid-degummed phosphatides to be spread over the meal before the dryer-cooler. The refinery then can water-wash and vacuum dry the oil by traditional methods, or treat it by silica gel adsorption. If a water wash is used, the oil is heated to 88°C/190°F with 1215 percent soft water, held in an agitated tank for •0.5 hr, and centrifuged to produce an oil with <20 ppm soap. The soap can be reduced to "0" by the addition of a small amount of phosphoric acid in the water-wash retention tank. The oil is next sprayed into a drying tank at 50 mm Hg absolute.72

Fig. 8.18 Flow sheet of continuous refining of soybean oil (long mix process) with water and wash option. (*From Farr, W.E., "Refining of Fats and Oils", in* Introduction to Fats and Oils Technology, *2nd ed., R. D. O'Brien, W. E. Farr, and P. J. Wan (Eds.), pp. 136157, AOCS Press, Champaign, IL, 2000. With permission.*



Sodium silicate neutralization also has shown promise as a potential commercial method. The resulting soaps form a granular agglomerate which can be removed by filtration to produce oil containing less than 100 ppm soaps. Thus, the costs of purchasing and maintaining primary and water-wash centrifuges, and wash water disposal are avoided. The filtered oil can be treated with neutral or activated clays to remove color, peroxides, residual phosphatides, and soaps to produce oils with free fatty acids contents of less than 0.05 percent and "0" PV (peroxide value).79 The method has been patented and is being evaluated on a small scale.

Various researchers have reported on attempts to remove fatty acids from oils by ultrafiltration membranes. In the absence of membranes that can withstand extraction solvents, success generally has been limited. However, progress has been made on degumming.80

Silica Gel Adsorption

Silica hydrogels are very effective in removing phosphatides, residual soaps, and metal ions (all poisonous to hydrogenation catalysts), thus reducing the amount of bleaching clay required (by 5075%) and leaving its function primarily to remove chlorophyll and secondary oxidation products. By eliminating soapy wash water, water discharged from refineries is reduced by •50 percent in volume and has much lower biological oxidative demand. Additionally, the cost and expenses of a second centrifuge are avoided.72

The method of using silica hydrogel has changed during perfection of the process. Variations in methods exist, but silica hydrogel can accept oil from the centrifuge at 0.20.4 percent water content. The current recommended process consists of blending the silica hydrogel with oil directly from the soapstock removal centrifuge, with minimum, if any, drying of the oil. Silica hydrogel then is removed by filtration before mixing the oil with the bleaching earth.72

Bleaching

Bleaching originally was meant as a step for reducing color in oils, but has become a misnomer of this industry. Although a limited amount of bleaching occurs during the process, the major reduction of red and yellow colors occurs during the high heat of the deodorization process. The current practical function of bleaching is to remove chlorophyll and oxidation-degraded compounds, and prepare the oil for hydrogenation by scavenging the remaining soaps, phosphatides, and minerals that would poison the catalyst. The theory and practical aspects of bleaching have been reviewed by several authors.8183

Bleaching earths are made from naturally occurring minerals, including palygorskitealso known as attapulgite, sepiolite, bentonite, montmorilloniteand other minerals belonging to the aluminum silicate family. They may be used as such, but typically are activated by treatment with hydrochloric or sulfuric acids which: (1) increase absorption by increasing surface area several fold; (2) provide acid centers that have catalytic properties; and (3) impart ion-exchange properties to the clay. These properties are important in adsorbing various undesirable impurities in the oil, and in rendering complex organic structures adsorbable. The ability of bleaching earths to adsorb phosphatides and soaps has been mentioned already; the same property also is important in removing pesticides, polycyclic aromatic hydrocarbons, *trans* and conjugated fatty acids, dimers, and polymers. Activated earths also provide a catalytic surface for breaking down peroxides. Decomposition is an exothermic reaction, with the heat possibly assisting the sometimes reported "press effect" on color reduction of carotenoids in the filter press. The cation exchange property of the activated earth is credited with removing magnesium from the center of the chlorophyll complex, rendering the pheophytin adsorbable to the clay. Cation exchange also is used for the removal of heavy trace metal pro-oxidants such as iron and copper, and for removing trace nickel in post-bleaching of hydrogenated oil.83 Although peroxides content is reduced, *p*-anisidine value (AV) (AOCS Method Cs 1890) increases. The AV is believed to estimate aldehydes (2-alkenals and 2,4-dienals) in animal and vegetable oils with the potential for later breakdown.

Bleaching clay load (typically 0.12.0%) and operating temperatures depend on the type and quality of oil processed.83 Modern bleaching processes are conducted under vacuum (50 mm Hg) to minimize later oil oxidation, and subsequent nitrogen blanketing is recommended.

Close coupling of the refining and bleaching operations is highly recommended, especially when using the Modified Caustic Refining or Silica Refining processes.72 Because of the high level of unsaturated oils and peroxides in spent bleaching earth, it is high susceptible to spontaneous combustion unless quenched with water. Disposal in land fills is becoming increasingly difficult. Spreading spent bleaching clay on soybean meal for animal feeding has been suggested, but caution should be taken in implementation since, by absorbing pesticides and mycotoxins, bleaching is one of the two safety valves in processing oils. The other is potential collection in condensed deodorizer distillate.

Oils and Fats Modification

If processing proceeded next to deodorization, the resulting oil would be called refined, bleached, deodorized (RBD). Modification processes often are conducted before deodorization/physical refining and include: increasing the stability of polyunsaturated fatty acids by hydrogenationespecially by saturating the third double bond of linolenic acid; removing TGA with higher melting points by *thermal fractionation* (chilling and filtering) to obtain clear table oils; and obtaining fat solids by thermal fractionation, hydrogenation, or interesterification of TAG, for compounding fats with specific softening and melting temperature profiles. Several modification and purification procedures may be included. As examples: (1) bleached oils often are also post-bleached after hydrogenation; (2) free fatty acids interfere with interesterification; thus, deodorized oils are used in the processes and the resulting oil bleached and again deodorized to remove free fatty acids that may have been created.

Various techniques are used to craft fats with temperature-solids profiles like the shortenings shown in Fig. 8.19,84 and margarines and spreads described later, from liquid oils. Two systems are used for characterizing solids content changes with temperature in fats. The Solid Fat Index (SFI) (AOCS Method Cd 10-57), using dilatometers, was developed in the United States. It is the older method, and is effective for fats containing up to 50 percent solids at 10°C/50°F. The Solid Fat Content (SFC) (AOCS Method Cd 16-81), using pulsed nuclear magnetic resonance (NMR) techniques, was developed more recently to accommodate palm oil and its products, and is used in the majority of other countries. SFC is effective for solids contents up to 95 percent; however, additional care in sample tempering may be required at the upper solids limits for SFI and SFC. In both methods, a sample of the fat is completely melted to destroy its crystal memory, then chilled to 0°C to set the crystals. Readings are then taken at selected temperatures (10°C/50°F, 21.1°C/70°F, 26.7°C/80°F,33.3°C/92°F,37.8°C/100°F, and 40°C/104°F) to develop SFI or SFC curves. In some industries, five-point or three-point readings are taken, or readings also made at higher temperatures. SFI and SFC give similar, but not identical results. A recent comparison of the two methods concluded that SFC reads higher at lower temperatures (10°C and 21.1°C), but similar to SFI at higher temperatures.85 The following equations for converting SFC to SFI, with correlation coefficients, R2, of 0.980.99 were offered:

commercial spreads:

$$\mathrm{SFI} = 1.98 + (0.72 \times \mathrm{SFC}) - (0.035 \times \mathrm{temp.})$$

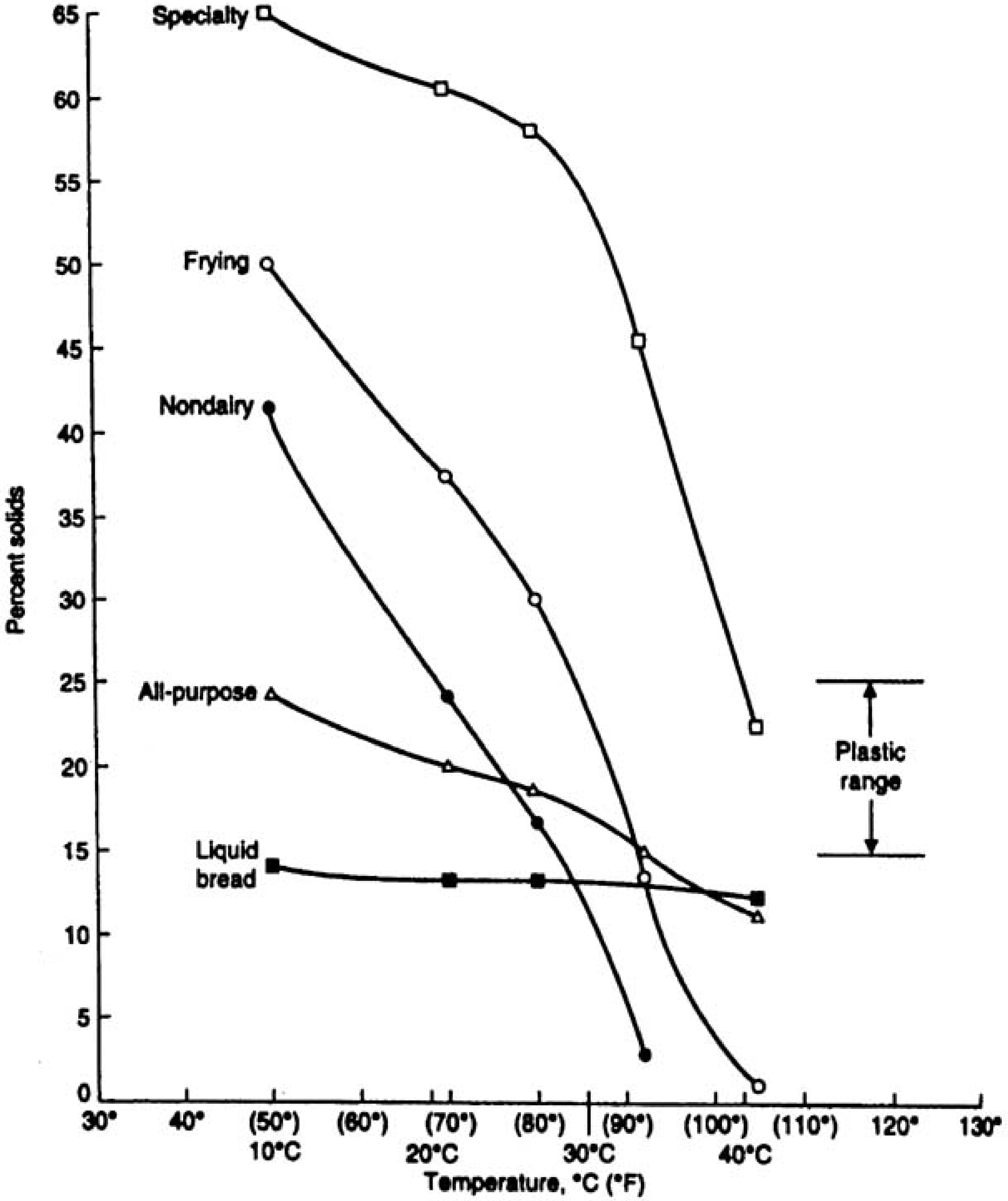
base stocks:

$$\mathrm{SFI} = -40.94 + (1.22 \times \mathrm{SFC}) + (1.03 \times \mathrm{temp.})$$

blends (base stocks/liquid oil):

$$\mathrm{SFI} = -0.94 + (0.82 \times \mathrm{SFC}) + (0.02 \times \mathrm{temp.}) + (0.02 \times \mathrm{percent})$$

Fig. 8.19 Solidstemperature curves for various types of shortenings. (*From O'Brien, R. D., "Shortenings", in* Practical Handbook of Soybean Processing and Utilization, *D. R. Erickson (Ed.), pp. 363379, AOCS Press, Champaign, IL, 1995. With permission.*)



The melting properties of fats in the mouth are important. Generally, consumers can detect a greasy characteristic if more than 3 percent solids remain at mouth temperature in spreads, or if more than 56 percent fat solids remain in baked, fried, and snack foods. Thus, doughnuts and bakery products are best eaten while warm to avoid greasy mouthfeel. The fried snack foods industry must choose between two alternatives: (1) Snacks fried in oil may: impart a cloudy appearance to see-through panels in the package; feel oily when picked up by the fingers; and stain clothing if dropped. This is expected in long time favorites like corn chips, but may not be acceptable in newly introduced snacks. (2) Processors who want a "dry feeling" on pickup of the snack may select a fat that is solid at room temperature, but melts rapidly in the mouth to avoid the greasy sensation. Some snacks leave a greasy lining in the mouth if they are eaten at the same time the consumer drinks a cold soda popthis has been a problem with olestra. A fat, such as the one marked "Nondairy" in Fig. 8.19, could be used in coffee whiteners.

Another important factor is "plastic range." Generally, 1525 percent solids at ambient temperature is considered acceptable for working a fat without making the product (dough/cake) too oily to handle. Doughs containing the "all-purpose shortening" shown in Fig. 8.19 can be machined over a wide range of temperatures in a warm bakery. Stick margarines often are compounded to be softer and easier to spread than butter when taken from the refrigerator, remain firmer at kitchen temperatures, and melt completely without greasy mouthfeel when eaten. Soft (tub) margarine is ready to spread as taken from the refrigerator.

Thermal Fractionation

The simplest type of thermal fractionation is *winterization* for cosmetic reasons to obtain salad and cooking oils that remain clear when stored in the refrigerator. The oil is chilled in tanks with mixing to crystallize the higher melting point waxes, or TAG which are natural or produced by light hydrogenation to delay oxidation of the oil. A filter aid is added to assist filtration. After separation of the liquid (*olein*), the filtering apparatus is heated to melt and recover the fat solids (*stearin*) which can be used in compounding shortenings and other products.86 The resulting oil is known as *RBWD* (refined, bleached, winterized, deodorized), and is expected to pass the "Cold Test" (AOCS Method CC 11-53) by resisting clouding for 5.5 hr at 0°C. Shallow, unstirred tanks in cool rooms were used for crystallization in earlier days. The rooms became warmer as crystallization progressed, demonstrating it is an exothermic reaction and oils will readily assume the more compact crystal forms provided they are able to shed the energy as heat.

Thermal fractionation technology is most developed in the palm oil industry, where most oils are fractionated before sales. Solids profiles of stearins that have been fractionated from crude palm oil by chilling to different temperatures is shown in Fig. 8.20.87 Thermal fractionation, and double fractionation can be useful tools in obtaining fat fractions with the specific desired characteristics (Fig. 8.21).88

Fig. 8.20 Examples of stearins separated from crude palm oil by chill fractionation. (From Tan, B. K., and Flingoh, C. H. OH. "PORIM Survey 1979/80: Oleins and Stearins from Malaysian Palm Oil Chemical and Physical Characteristics," PORIM Technology, No. 4, Palm Oil Research Institute of Malaysia, Kuala Lumpur, Malaysia, 1981. With Permission.)

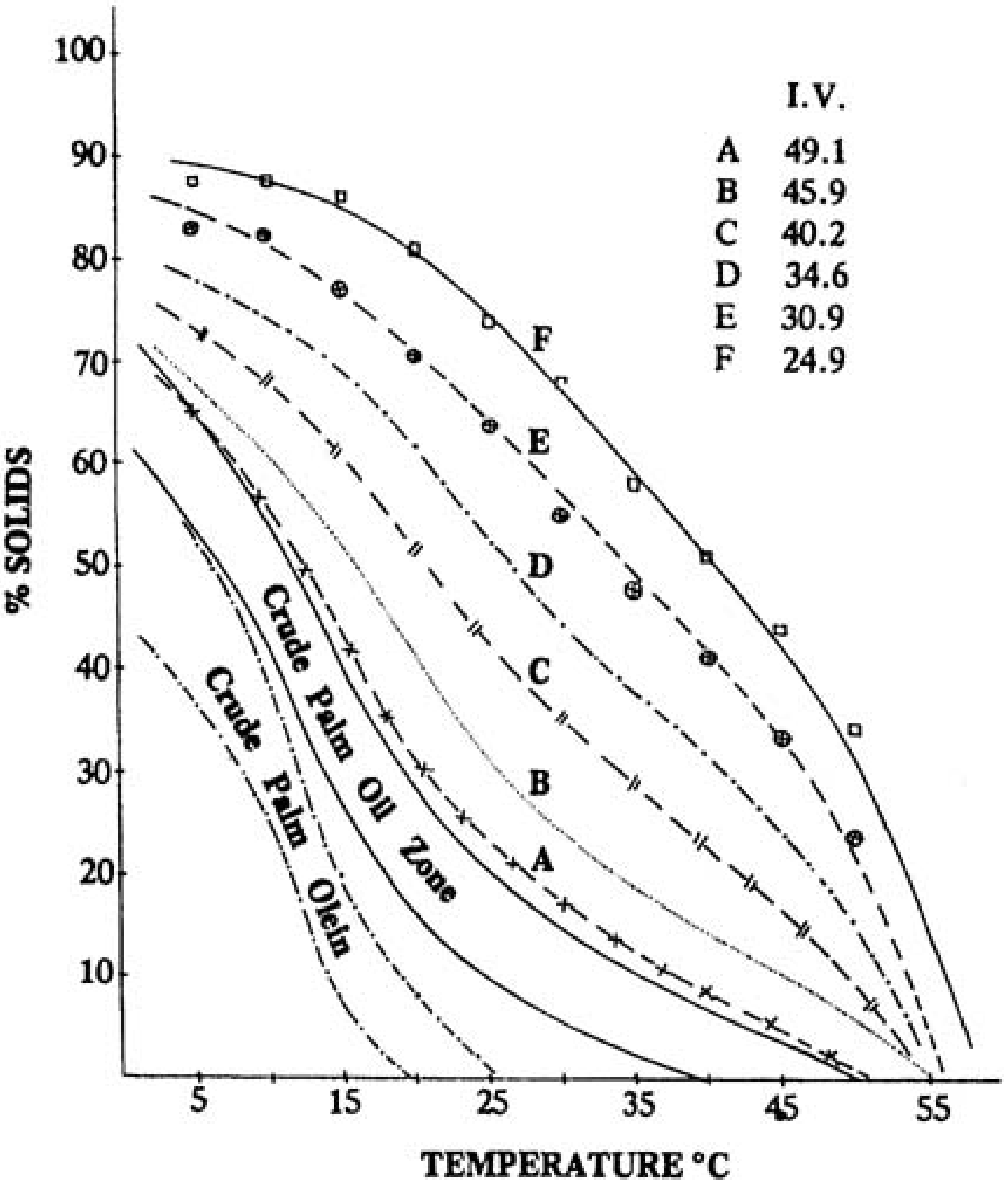
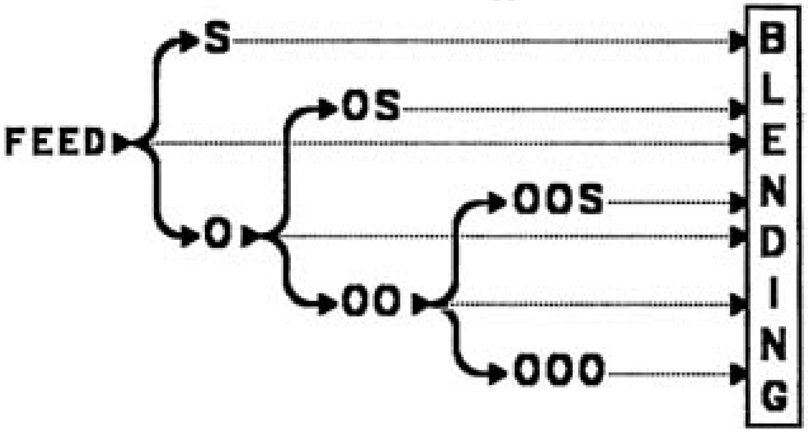


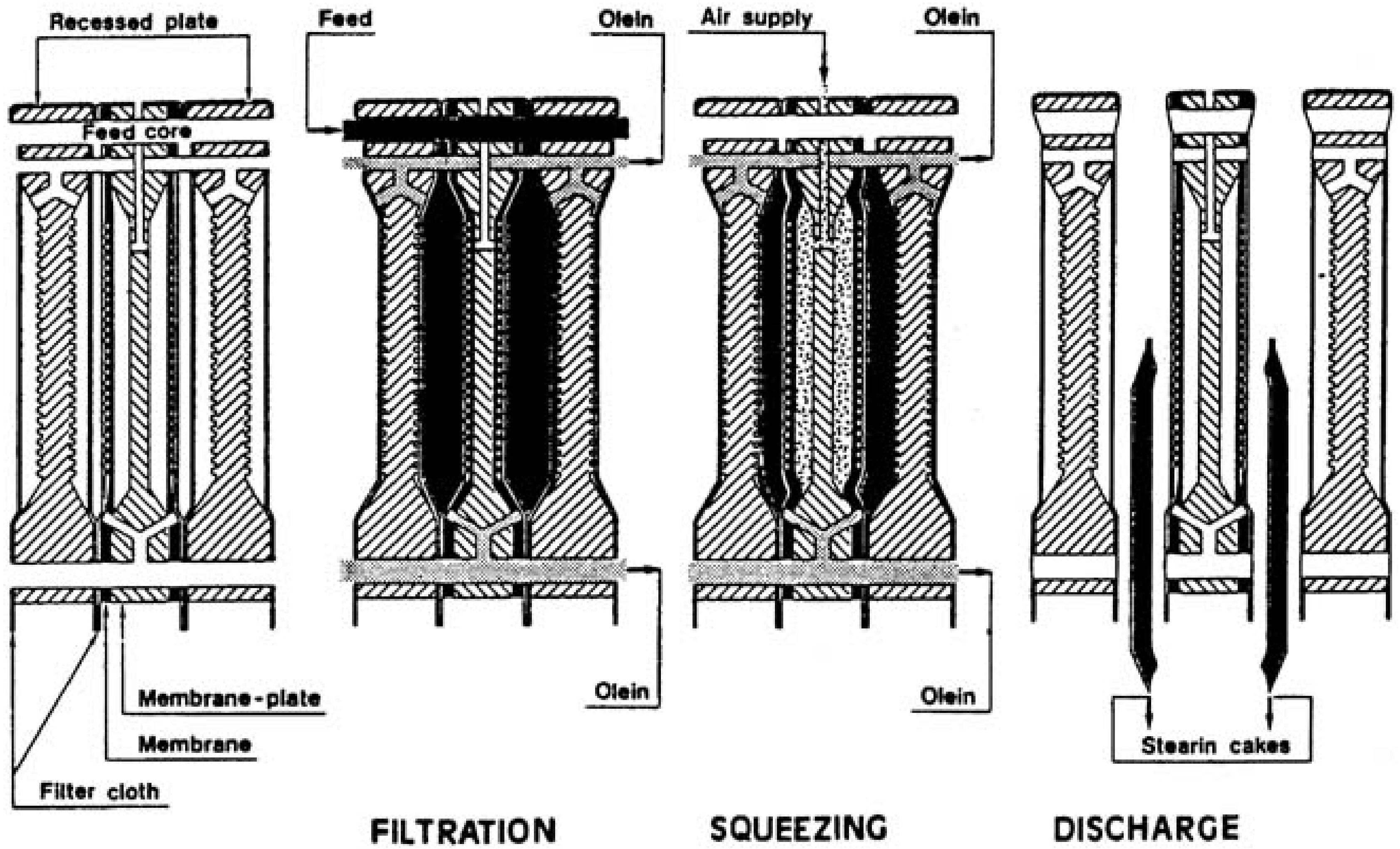
Fig. 8.21 Fat fractions from cascade fractionation. (*From Tirtiaux, A., in* World Conference Proceedings, Edible Fats and Oils Processing: Basic Principles and Modern Practices, *D. E. Erickson (Ed.), pp. 136141, AOCS, Champaign, IL, 1990. With permission*.)



Three major techniques are used in fractionation. In *dry fractionation*, oil is chilled with slow stirring and often seeded with crystals from an earlier batch. The resulting mixture can be quite viscous and present difficulties in draining oil from the crystals during filtration. In *solvent ("wet") fractionation*, the chilling and crystallizing of oil occurs in a solvent, often acetone. The viscosity of the liquid phase is greatly reduced, resulting in easier separation and improved purity of crystals. Solvent fraction must be conducted in explosion-proof facilities. In *detergent fractionation*, a surfactant is added to the chill-crystallized oilfat mixture, and greatly improves drainage.89,90 Significant progress in dry fractionation technology has been made in recent years, greatly simplifying processing and reducing problems of handling solvents and by-product streams.91 A variation of the detergent fractionation principle has been used in dewaxing sunflower oil. Advantage is taken of the soap content (10002000 ppm) and 34 percent water remaining in unwashed neutralized oil. After alkali neutralization, sunflower oil is pumped from the primary centrifuge directly to a heat exchanger for crystallization. The oil is held at 57°C/4145°F for about 4 hr for crystal growth, then heated by heat exchanger to 1215°C/5459°F, centrifuged to remove the wax crystals along with some water and soaps, washed with hot water, and centrifuged.92

Until the mid-1980s, continuous-belt vacuum filters mainly were used for separating oleins from stearins. These are expensive and complex machines, requiring controlled temperatures in various sections and reconditioning of the belt for the subsequent filtration cycle. Simpler membrane filters have gained in popularity. As shown in Fig. 8.22, the active members resemble a plate and frame filter. However, each cavity is lined on both sides with "membrane" filter stock, between which the crystaloil mixture is pumped. Initially, free oil escapes through the filter stock, until the cavity is filled with crystals. Addition of the oilcrystals slurry is then stopped, and compressed air applied between the membrane and the frame to squeeze the oil from the enclosed crystal mass. When completed, the frames open, allowing the crystal cake to fall into a take-away screw. The frames then reclose, and the cycle is repeated. Stearins produced by membrane filters contain significantly less free oil than vacuum belt filters.88

Fig. 8.22 Principle of membrane-type filtration of olein and stearin. (From *Tirtiaux, A., in* World Conference Proceedings, Edible Fats and Oils Processing: Basic Principles and Modern Practices, *D. R. Erickson (Ed.), pp. 136141, AOCS, Champaign, IL, 1990. With permission*.)



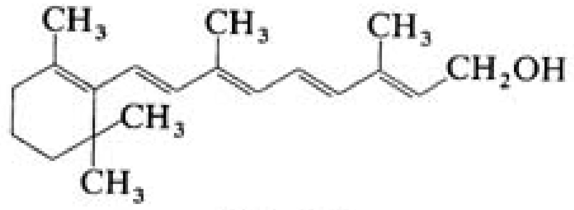
Oil is an effective solvent for higher temperature melting TAG, and results of thermal fractionation of common fats can be surprising. An example is milk fat (butterfat), which has been studied intensively.93 Fractions melting as high at 54°C/129°F have been recovered. Many tropical countries allow up to 5 percent high melting hydrogenated fats to raise the melting point of chocolate and prevent blooming (loss of sheen, paling of color, and development of coarse crystals) resulting from repeated melting and resolidification of bars while in nonair-conditioned stores. This also can be a problem during the summer months in temperate countries. However, these latter countries often prohibit inclusion of fats other than cocoa butter in chocolate. The problem has been reduced in milk chocolate by adding high melting fractionated milkfat. Milkfat also has been used in shortbread (brittle) cookies, laminated pastries, and other bakery products, where it produces a shortening effect, but never melts to appear greasy.

Hydrogenation

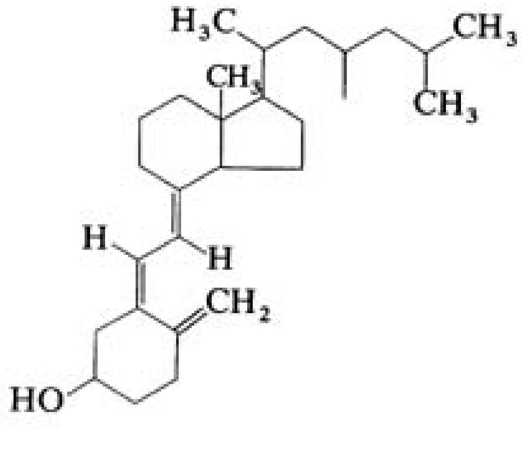
Hydrogenation is the catalytically assisted addition of hydrogen to carboncarbon double bonds. Its main uses are to: (1) increase fryer life of oils and shelf life of table oils and bakery products; and (2) produce solids for making shortenings, margarines and spreads, and various confectionery and specialty products. The process has been reviewed by many.94101

Food technologists would prefer to minimize the C18:3 content of oils to improve fryer and bottled oil shelf lives by reducing oxidation tendencies (Table 8.4) This has been partially accomplished by applying selective breeding and biotechnology to oilseeds, but the GMO products still are in the introductory phase. Since linoleic (C18:2) and C18:3 fatty acids are dietarily essential in small amounts, it is not desirable to eliminate them completely. For many applications, soy and canola oils are partially stabilized against oxidation by "brush" (light) hydrogenation.

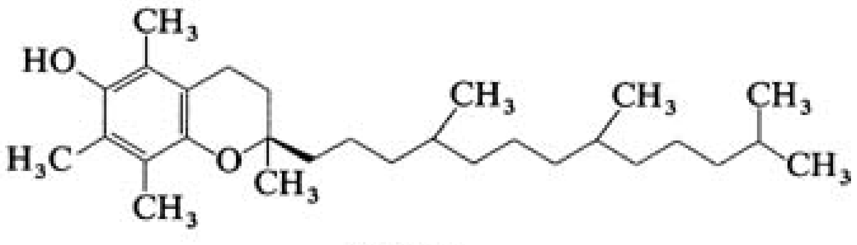
Selectivity, the tendency of the catalyst or process to remove one type of bond in favor of another, has several meanings in hydrogenation catalysts. *Preferential or saturate selectivity* indicates a focus on saturating a specific bond, for example the C18:3 bond of linolenic acid to form C18:2 linoleic acid. Preferential selectivity is estimated from the following kinetic relationships:



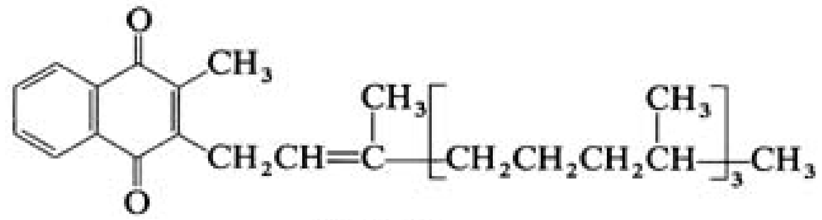
Vitamin A



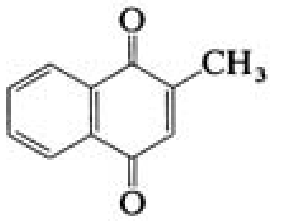
Vitamin D_3

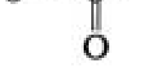


Vitamin E



Vitamin K_1





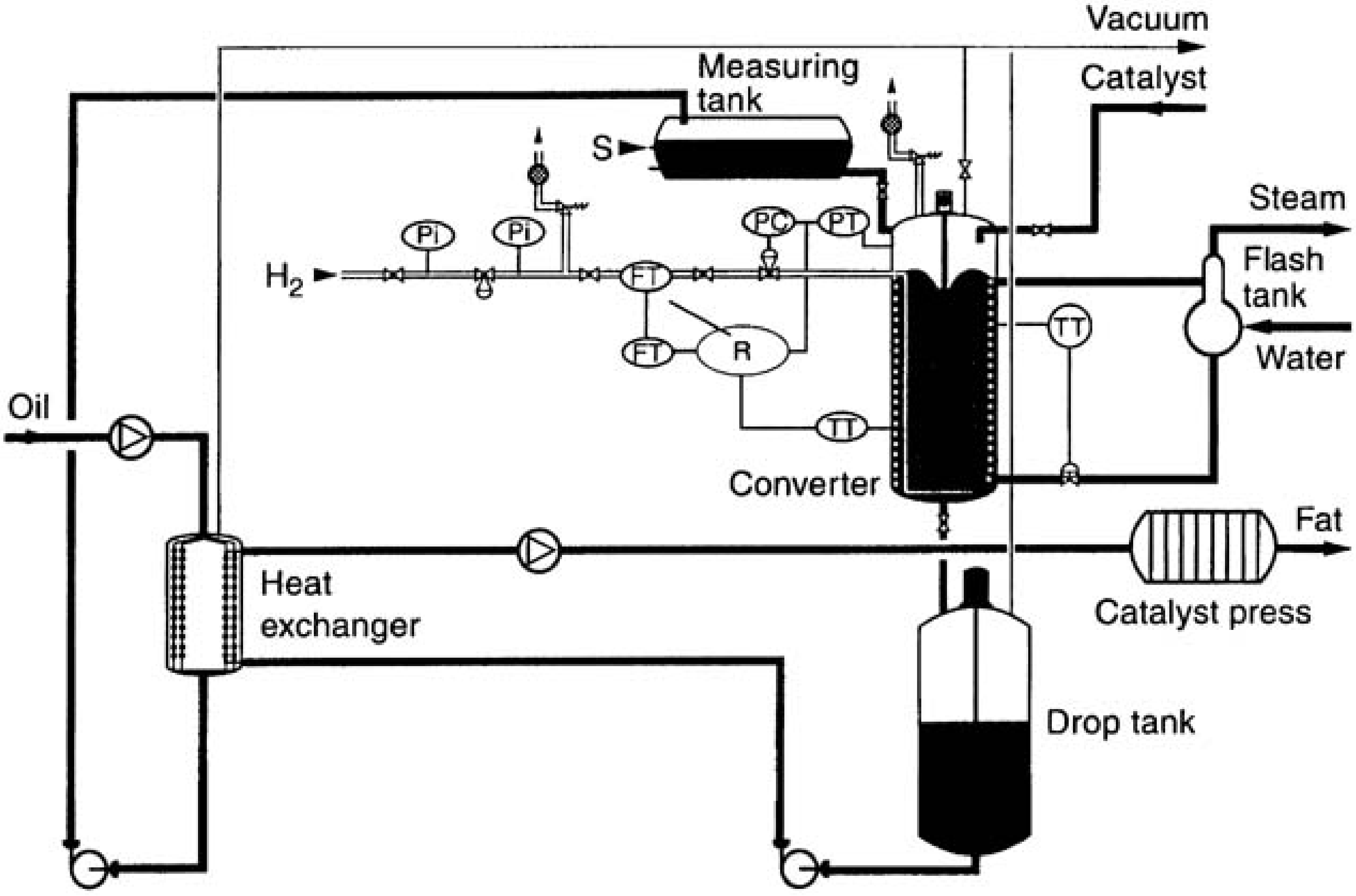
Vitamin K_3

with linolenic selectivity ratio (SR) defined as SR 1 = $k1/k2$ and linoleic selectivity as SR 2 = $k2/k3$. Linolenic acid selectivity for nickel catalysts varies from 2 to 3, and is as high a 6 for copper-based catalysts. Better control of linoleic selectivity, with SR 1 varying from 34 to 60, is achievable by the choice of catalysts, catalyst poisons, and reaction conditions. This gives the operator greater flexibility in the melting behavior and stability of the hydrogenated product.95

"*Trans*-isomer selectivity" is less directed and favors the formation of *trans* bonds. Catalysts vary in selectivity, with performance greatly affected by catalyst dosage, effectiveness of mixing, temperature, and hydrogen pressure in the reactor. Bonds become conjugated during hydrogenation, but not all atoms fall back neatly into their former positions after the process. The new types of TAGs produced increase the diversity, which is desirable in controlling later crystallization processes. The production of *trans* (partially saturated) bonds is intentional, and can be partially controlled.99

Hydrogenation is conducted in hardening plants (Fig. 8.23). The "converter" is a pressurized reaction tank, equipped with a high-speed mixer and assisting baffles, means for adding and removing the oil, a gas distributor, a means to add catalyst, and heating and cooling coils. The hardening plant additionally has means for premeasuring and heating the oil, a drop tank, heat exchangers, and a catalyst filter. The purity of the oil and hydrogen affect the life of the catalyst, typically a thin film of nickel on an inert carrier.100

Fig. 8.23 Hydrogenation line. (*From Hastert, R. C., "Hydrogenation," in* Introduction to Fats and Oils Technology, *2nd ed., R. D. O'Brien, W. E. Farr, and P. J. Wan (Eds.), AOCS Press, Champaign, IL, 2000. With permission.*)



In *brush hydrogenation*, a selective catalyst is used; the IV is reduced by 1525 to •115 units for soybean oil; about 15 percent *trans* isomers are formed; the C18:3 content is reduced to 3 percent maximum; and C18:0 content increases by •1 percent. In the preparation of *margarine bases*, selective catalysts are used; the IV is reduced to •70; and about 50 percent *trans* isomers are formed. In *shortening bases*, selective or nonselective catalysts are used; the IV is reduced to •75; and about 35 percent *trans* isomers are formed. In producing *stearin flakes*, a high activity catalyst is used and the IV is reduced to •510 with no *trans* isomers remaining. The flakes are scraped off the surface of a chilled roll, or beadlets can be produced using a shot tower with chilled air. Maintaining high levels of polyunsaturates is desirable in producing *coating fats*; the IV is reduced to •70; about 65 percent *trans* isomers are formed with the production of saturates (C18:0) minimized to a 24 percent increase; a sulfided nickel catalyst is used. Some plants run a second post-bleach on the oil after hydrogenation. Factors affecting the hydrogenation reaction include: reactor design, purity of feedstock and hydrogen, operating conditions, and choice and efficiency of the catalyst. Operating conditions include hydrogen pressure, reaction temperature, catalyst dosage, and agitation. Since *trans* formation is indicative of incomplete saturation, conditions that favor keeping the catalyst covered with hydrogen favor saturate selectivity. Progress in the hydrogenation process typically is determined on line by the refractive index, calibrated to the IV for each process and product.101

The preparation of temperature-profiled fat products, such as margarines/spreads and shortenings, often includes preparation of four to six base stocks hydrogenated to different IVs (Fig. 8.24), which are blended with oil and hardstocks to obtain the desired temperaturethe fat solids profile.102 An example of broadening the *plastic range* (1525 percent fat solids) of two base stocks by the addition of hardstocks is shown in Fig. 8.25. By itself, the 80 IV base stock remains in the 1525 percent fat solids (machinable) range only between the temperatures of 10°C and 20°C; but, by the addition of 8 percent hardstock (<5 IV), the range can be extended to •30°C. The addition of 12 percent hardstock to 85 IV base stock broadens the working range from •1012°C to 38°C, which is more typical of bakeries.94

Fig. 8.24 Solid Fat Index (SFI) profiles for six hydrogenated soybean oil (H-SBO) base stocks. *(From O'Brien, R. D., "Shortening Technology," in* Introduction to Fats and Oils Technology, 2nd edn., *R. D. O'Brien, W. E. Farr, and P. J. Wan (Eds.), pp. 421451, AOCS Press, Champaign, IL, 2000. With permission.*)

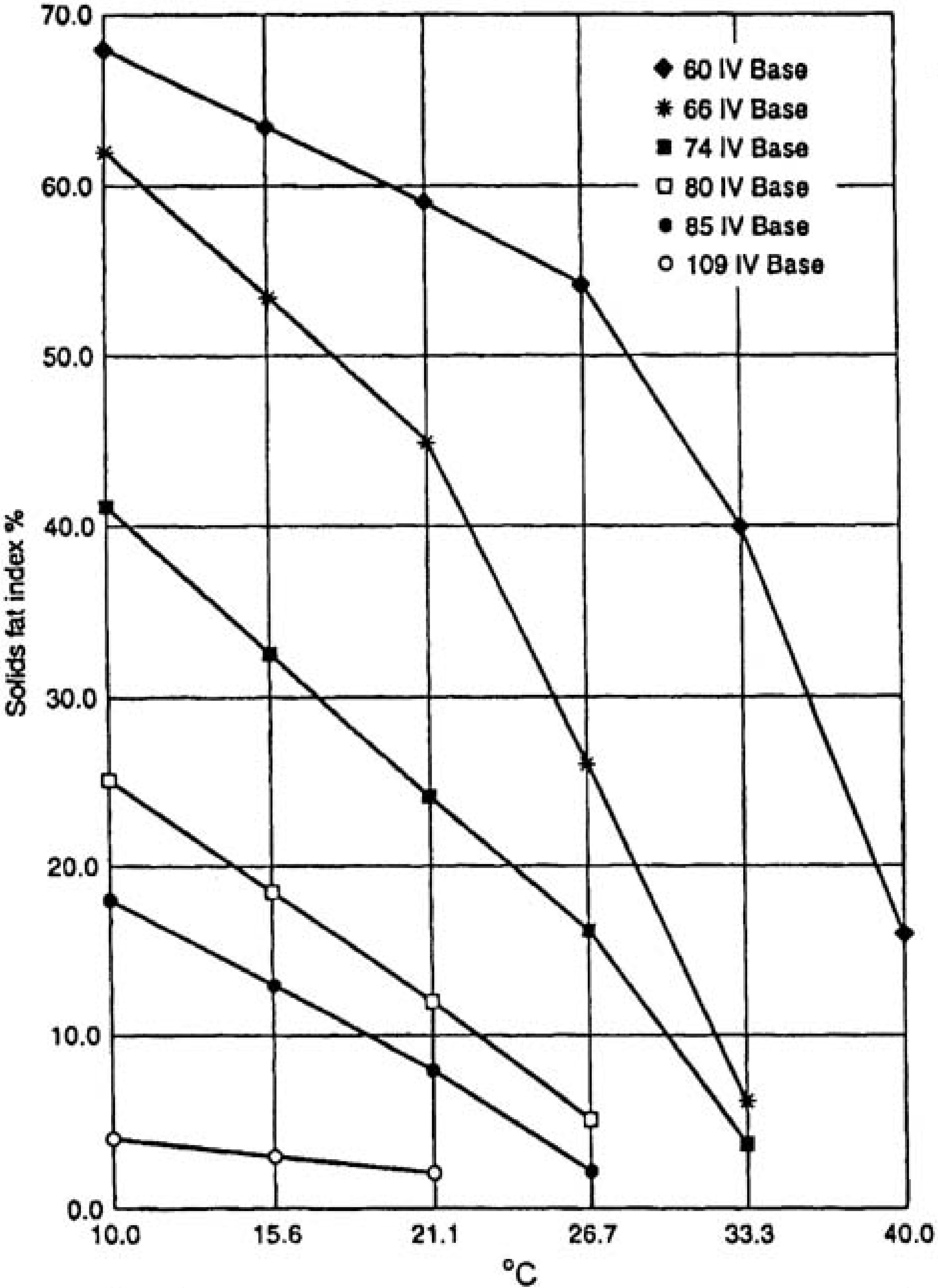
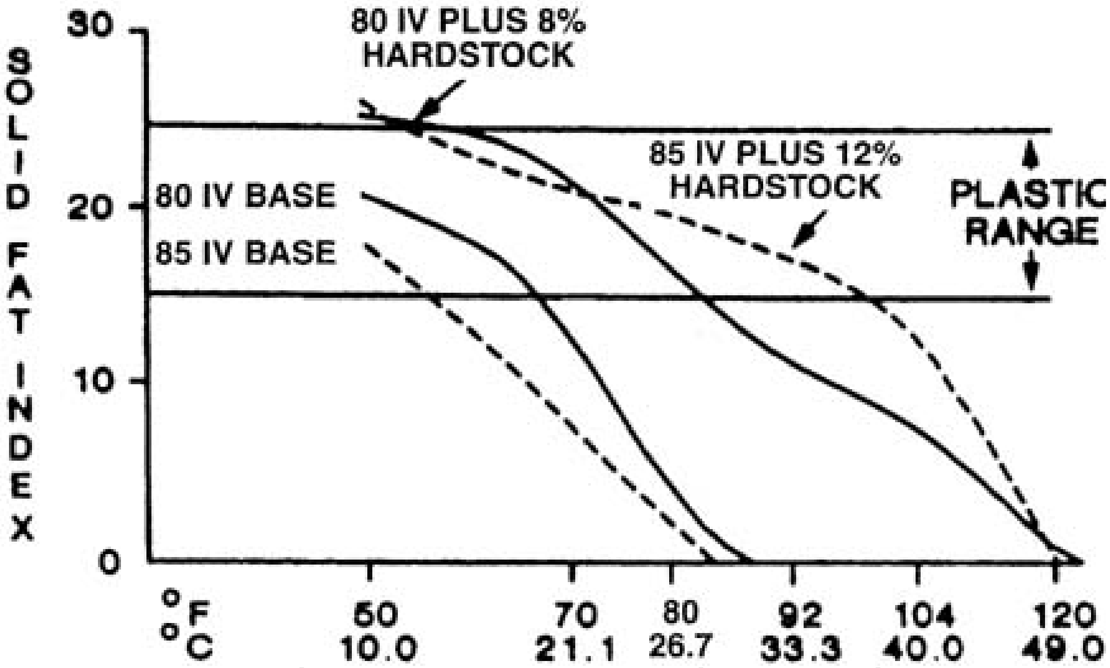


Fig. 8.25 Effects of adding hardstock (<5 IV) to broaden plastic range working temperatures of two base stocks. (*From Hastert, R. C., "Cost/Quality/Health: Three Pillars of Hydrogenation," in* World Conference Proceedings: Edible Fats and Oils Processing: Basic Principles and Modern Practices, *D. R. Erickson (Ed.), pp. 142152, American Oil Chemists' Society, Champaign, IL, 1990. With permission.*)



In response to a petition by a consumer group, on November 17, 1999, the U.S. FDA proposed to amend food nutrient labeling regulations to require that *trans* fatty acids present in foods and dietary supplements be included in the amount and percent Daily Value (% DV) declared for saturated fatty acids and, when present, the grams of *trans* fatty acids present in a serving of the product also be stated.103 Concerns about hydrogenated fats arose in the late 1950s, at which time they, tallow, and tropical oils were claimed to be atherosclerotic. After major research, tallow and tropical oils (especially high palmitic acid oils) were shown to be neutral in this regard. Nor was evidence of this effect found in hydrogenated oils. The current petition is based on concerns that labeled *trans* fatty acids raise blood cholesterol levels and may contribute to coronary heart disease (CHD).

A credible literature review is included in the FDA announcement, including historical studies of relationships between saturated dietary fats and blood cholesterol levels. Generally, it appears that the body handles higher melting point fats by increasing LDL cholesterol levels, with the partially hydrogenated vegetable oils not raising LDL levels as much as saturated fats when either replaces *cis* (lower melting temperature) fats. Consistent indications of *trans* fats being less healthy at their categorical melting points have not been reported; and numerous research reports of "no difference" exist.104 The comment period on the proposed ruling has been extended, and it does not appear that a final rule will be issued for several years. However, the U.S. oils/fats and related products industries are reviewing processes other than hydrogenation for sources of fat solids.

Interesterification

Interesterification (INES) is the exchange of acyl radicals between an ester and an acid (*acidolysis*), an ester and an alcohol (*alcoholysis*), or an ester and an ester (*transesterification*), and can be random, directed, or enzymatic. The process has been called *intraesterification* if an exchange of positions occurs within the same molecule, and *randomization* if exchange occurs between molecules.21,25 The principles can be used to position fatty acids on molecules with hydroxyl sites to produce monoglycerides (emulsifiers), fatty acid methyl esters (FAMEs) for analytical purposes, liquid fuels such as methyl soyate (biodiesel), specialty fats such as medium chain triglycerides (MCTs), and sugar-ester noncaloric fat substitutes such as Olean™. Research publications exist from at least as early as 1852, and U.S. patents from 1939.105

The basic *acidolysis* reaction is

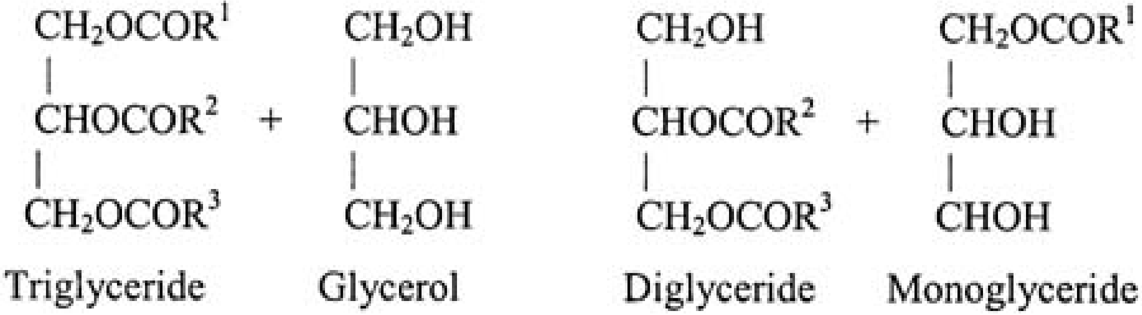
$$R^1COOR^2 + R^3COOH \quad R^3COOR^2 + R^1COOH$$

The *alcoholysis* reaction is

$$RCOOR^1 + R^2OH \quad RCOOR^2 + R^1OH$$

with methyl alcohol used for preparing FAMEs when determining the fatty acids components of TAGs, and much of the biodiesel made. Ethanol and other alcohols, up to four carbons in length (butanols), also have been used.21

Glycerolysis has been used to prepare mono- and diglyceride emulsifiers by reacting TAGs with an excess of glycerol.21



Standard "mono- and diglyceride" products contain 4050 percent monoglycerides, whose content can be raised to 53 percent without distillation. Distilled monoglycerides contain a minimum of 90 percent α-monoglycerides.21

In the *sucrosolysis* preparation of Olean™, the generic name for Proctor & Gamble Company's olestra, six to eight fatty acids, selected for functionality purposes, are positioned on sucrose. Although many noncaloric fats are limited or nondigestible, their producers are customers for the traditional vegetable oils. Uses of these and similar products have been summarized by Akoh.106

Except for enzyme-directed processes to place certain fatty acids in specific positions on TAGs, such as the production of coating fats, cocoa butter substitutes, or reduced-calorie fats,107 the majority of esterification processes are randomizations.

As mentioned earlier, nature arranges the fatty acids in plant oils to achieve the lowest melting point. The abbreviations S and U, for saturated and unsaturated fatty acids, respectively, also are used in indicating their prevalence on TAGs; thus: S3, S2U, SU2, and U3. Distributions of these combinations in nature and in randomized products are shown in Table 8.10.25,108 The literature indicates that, by withdrawing the unsaturated fatty acids from the *sn*-2 position, stability against oxidation often is reduced.21 Randomization always raises the melting point, more so for vegetable oils than for animal fats with an increase of 8.5°C for soybean oil. S3 TAGs hardly exist in natural vegetable oils, with only low concentrations occurring in animal fats. However, some are formed during randomization. *Directed randomization* is a specific technique that solidifies and precipitates S3 and S2U TAGs as formed, depending on the operating temperatures chosen. Removal of saturated fatty acids from cottonseed oil has been demonstrated recently.109 Directed random-ization can be used to modify either the olein or stearin fractions.

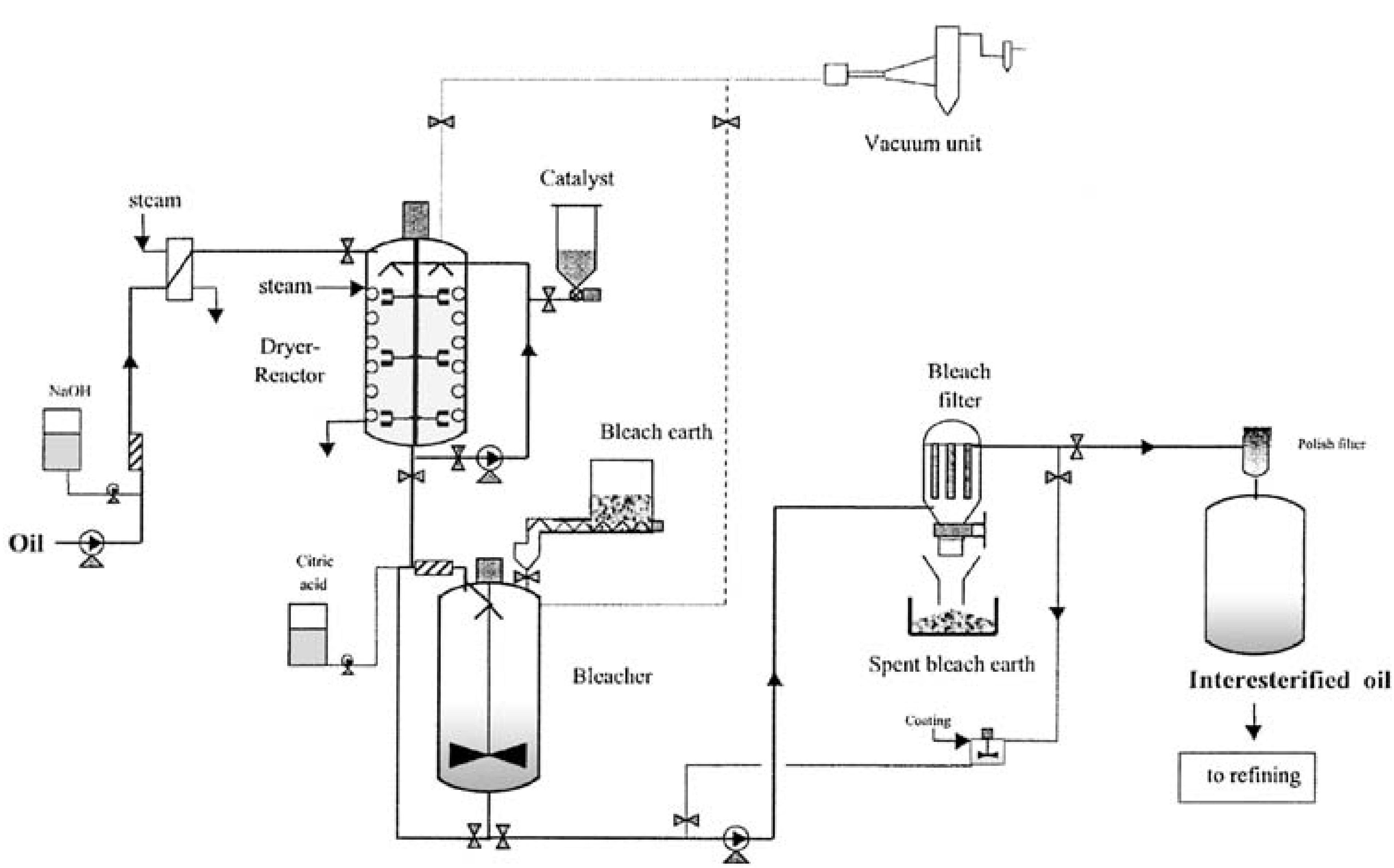
TABLE 8.10 Triglyceride Classes of Native and Interesterified Oils and Fats and of Noninteresterified and Interesterified Blends (1:1) of Fully Hydrogenated Soybean Oil with Vegetable Oilsa

		Single Oils					1:1 Blends with Hardened Soybean Oil			
		S3 (%)	S2U (%)	SU2 (%)	U3 (%)	M.P. (°C)b	S3 (%)	S2U (%)	SU2 (%)	U3 (%)
Palm oil	nc	6	50	38	6	39.8	57	13	20	10
	rc	13	39	37	11	47.0	41	43	14	2
Soybean oil	n	0	6	38	56	-7.0	50	2	17	31
	r	1	8	36	55	5.5	13	47	32	9
Cottonseed oil	n	<1	18	51	30	10.5	51	9	24	16
	r	3	18	44	35	34.0	25	34	31	10
Sunflower oil	n	0	1	24	75		51	0.3	11	38
	r	<0.2	4	27	69		11	47	34	8
Peanut oil	n	0	11	40	49		53	3	15	29
	r	1	10	38	51		16	47	31	6
Rapeseed oil	n	0	1	16	83		51	1	9	39
	r	0	1	17	82		10	44	37	9
Coconut oil	n	81	12	7	0	26.0	57	13	20	10
	r	74	24	2	<0.1	28.2	41	43	14	2
Palm kernel oil	n	76	15	9	0					
	r	53	37	9	0.7					
Cocoa butter	n	2	85	12	1	34.4				
	r	24	43	27	5	52.5				
Lard	n	8	30	50	12	43.0				
	r	10	32	40	18	43.0				
Beef tallow	n	22	60	18	0	46.2				
	r	13	38	17	12	44.6				

aFrom: Bookish, M., *Fats and Oils Handbook*, AOCS Press, Champaign, IL, 1998, with permission.
bFrom: Sonntag, N. V., "Fat Splitting, Esterification, and Interesterification," in *Bailey's Industrial Oil and Fat Products*, 4th ed., Vol. 2., D. Swern, Ed., pp. 97173, John Wiley & Sons, New York, 1982.
cn = native; r = randomized.

A flow sheet for a semi-continuous interesterification line with post-bleaching ability is shown in Fig. 8.26. The oil must be degummed, well-refined (<0.05% FFA), and free from peroxides and moisture that causes the production of soaps. Palm oil stearin and completely saturated (<5 IV) C18 hydrogenated fats generally are *trans* free and may be used as part of the feed. Reactions are run under vacuum, and oil is best stored under nitrogen in between processes.21

Fig. 8.26 Semi-continuous interesterification line with post-bleaching. (*Courtesy of Extraction De Smet, N. V./S.A., Edegem, Belgium*)



A variety of catalysts are available, including alkali metals, alkoxides, alkali hydroxides, sodium hydroxide and glycerol mixtures, metal soaps, and metal hydrides.21 Alkaline hydroxides (KOH or NaOH) in combination with glycerol were used in earlier days because of their low cost.110 Currently, sodium methylate and sodium ethylate are popular because of their efficiency. Rozendall has proposed that the actual interesterification catalyst is a sodium derivative of a diacylglycerol, activated during the process.111

The oil is loaded into the reactor, shown with both an agitator and a pumped circulation-spray loop, and heated under vacuum (110130°C) to reduce water and peroxide contents. Next, the oil is cooled to 7090°C and the catalyst is added as dry powder at 0.050.15 percent or suspended in dry oil. Randomization requires about 30 min, with an additional 1530 min allowed for completion. After the reaction is complete, the batch is transferred to a post-bleacher where the process is arrested by inactivating the catalyst by the addition of water or a (phosphoric or citric) acid solution. Bleaching earth is added to absorb the inactivated catalyst and soaps removed by filtration, and the oil sent to blending or deodorization. Losses from the formation of FFA and FAME are •10 times the catalyst weight, with an additional 0.51.0 percent lost in the formation of mono- and diglycerides, for a total of 1.52.0 percent. Continuous processes also are available. Interesterification progress can be monitored by on-line ultraviolet (UV) spectrophotometry, and completion of the process by melting point determination and other measurements.21

The main objective in interesterification is to produce solids free from *trans* fatty acids for later use in compounded fat products. Unlike hydrogenation, in which a series of base stocks are made, preparation of one interesterified batch for blending with oil and possibly hardstock is reported. Randomization can improve the functionality of a fat, as shown for lard in Fig. 8.27. Natural lard is within the 1525 percent solids plastic range between the temperatures of 9°C and 24°C; by randomizing, the temperature range is moved to 017°C, both conditions greatly restrict its use. But, by directed interesterification, the machinable range is moved to the acceptable ranges of 032°C. In contrast, natural cocoa butter (Fig. 8.28) is almost brittle at 26°C, melts rapidly by taking energy from the tongue, causes a pleasant cooling sensation in the mouth, and is completely liquid with good palate "cleanup" at 35°C. Randomizing destroys these properties, changing the fat into a softer mass which does not completely melt until 56°C.25

Fig. 8.27 Solids content of natural and random and directed interesterified lard. (*From Bockish, M.,* Fats and Oils Handbook, *AOCS Press, Champaign, IL, 1998. With permission.*)

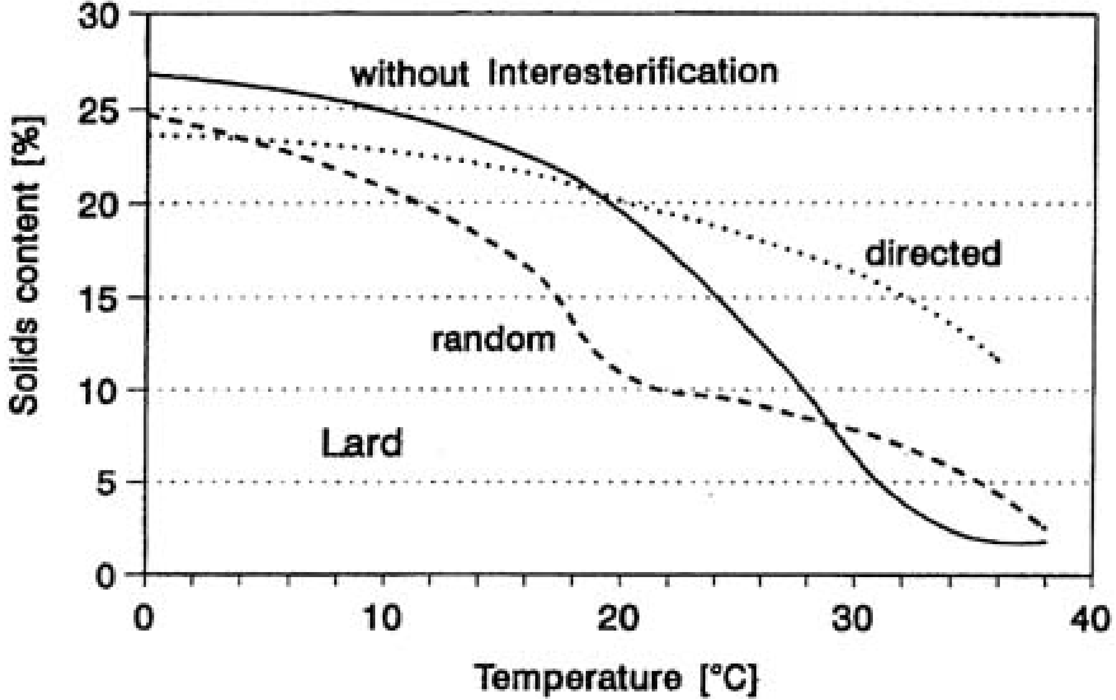
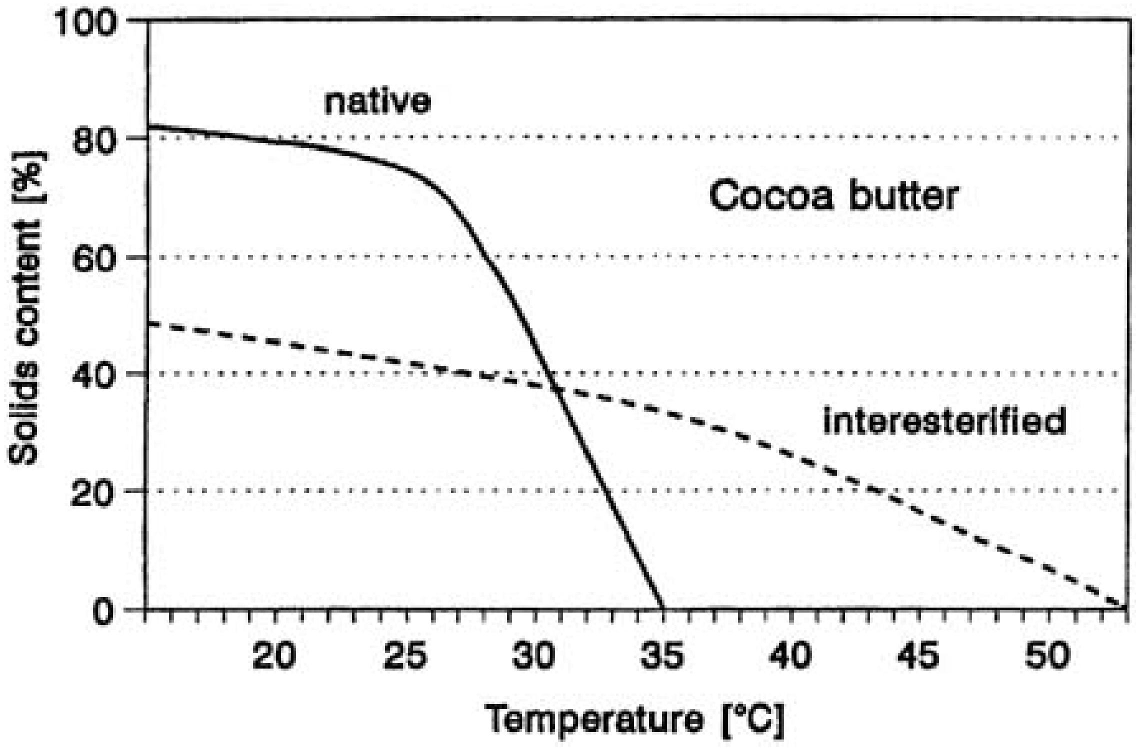


Fig. 8.28 Solids content of natural and randomized cocoa butter. (From *Bockish, M.,* Fats and Oils Handbook, *AOCS Press, Champaign, IL, 1998. With permission.*)



Interesterification is being pursued on the basis that *trans* production during hydrogenation increases the melting points of TAG and fatty acids, increases blood cholesterol levels, and increases atherosclerosis (plaque deposits in arteries), and CHD incidence. However, considerable documentation also shows that the presence of saturated fatty acids (primarily palmitic, and to a lesser degree myristic and stearic) at the TAG *sn*-2 position strongly favors atherogenesis. Pork lard contains 70 percent of its palmitic fatty acid in the *sn*-2 position, whereas only 17 percent of the palmitic acid in beef tallow is in this position. Lard is considerably more atherogenic to laboratory animals than tallow, with this property reduced by randomization; however, atherogenicity of tallow is increased by randomization. Levels of saturated fatty acids at the *sn*-2 position are very low in natural vegetable oils, but are increased by randomization, with accompanying atherogenesis in feeding trials.112,113 Continued documentation may show that health concerns about hypercholesterolemic and atherogenic effects of *trans* fatty acids are not left behind by switching to interesterification as a means of generating fat solids.

Deodorization/Physical Refining

Deodorization is the final step in the production of RBD oils and modified fats. The process is called *deodorization* if most of the FFA are removed by alkali neutralization as with row crop ("soft") oils, and *physical refining* if they are left to be mainly removed by steam distillation in the final step, as with palm and coconut oils.

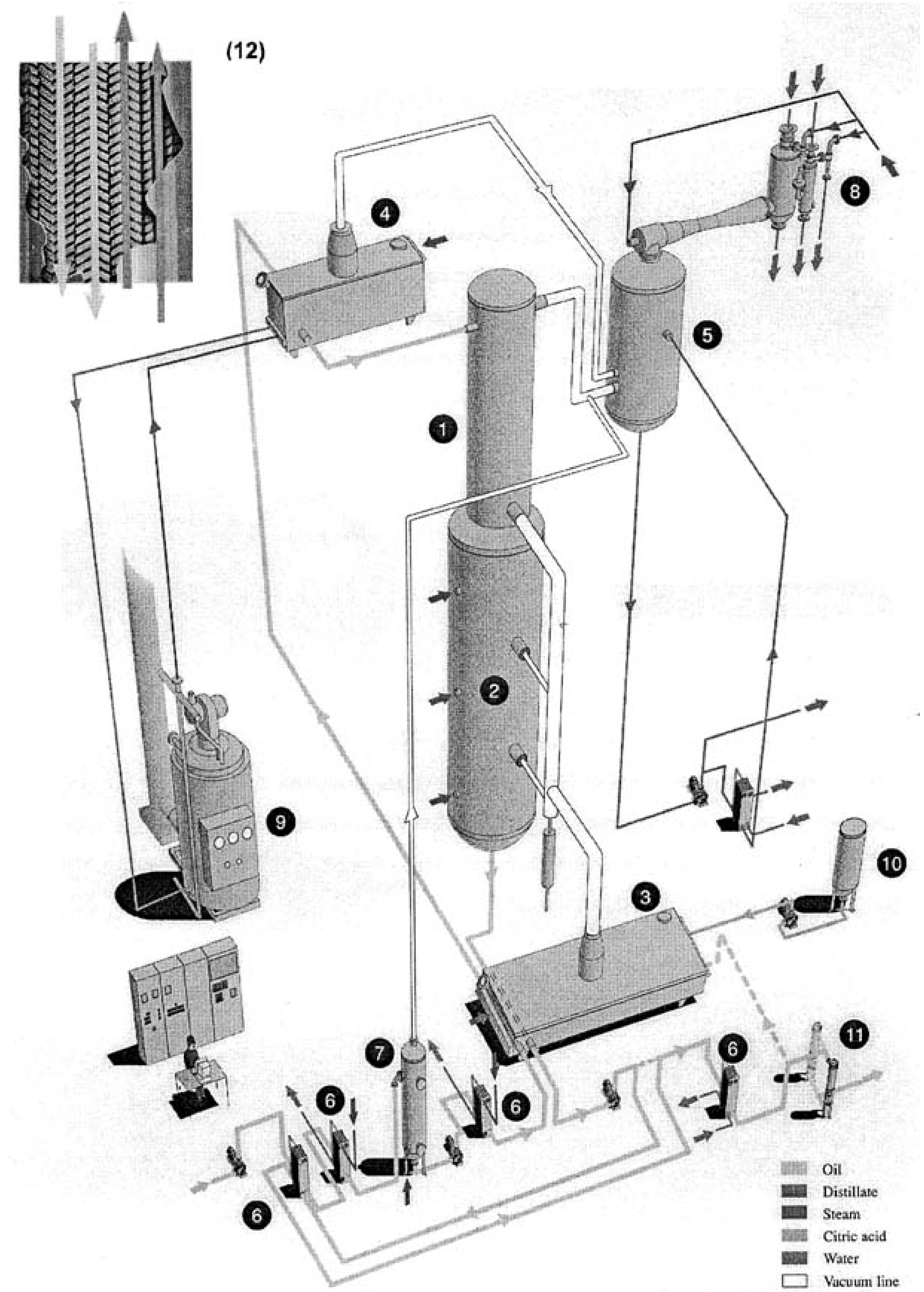
The volatile peroxides, other oxidation decomposition products, and odiferous compounds form reduced-boiling point azeotropes with water in the steam, at high temperatures, 250260°C/482500°F, and very low absolute pressures (•3 mbar) created by a series of steam ejectors. This is above the smoke point of soybean oil, but below the flash point, and oxygen must be excluded. Considerable heat bleaching of yellow-red carotenoids also occurs at this temperature. Typically, the deodorization process requires 2040 min after come-up time, uses 0.52.0 percent sparge steamthe higher level if tocopherols are removed, operates at between 2 and 4 mbar, and produces a product with about 0.030.5 percent FFA.114

Historically, the standard deodorizer held 60,000 lb of oil (one railroad tank car). Except for refineries making only a few kinds of oil, the building of continuous deodorizers slowed with the advent of Just-in-Time Delivery, supplier self-certification, and customers buying on the basis of their projected production schedules. This has led to the development of improved batch-continuous systems, which are designed to handle many batches of different oil blends per day, with minimum cross contamination and delays for process come-up.

For many years, deodorizers (operating at above 270°C/520°F) were heated by several types of mineral oil-like thermal fluids which, in turn, could by heated by direct-fire furnaces at 315°C/600°F in plumbing and coils at 3.2 bar (46 psig). During the 1970s, user health problems in Europe that were ascribed to the leakage of thermal fluids into oil during deodorization, although some details are still in question. As a result, the European market required that the local and imported oils industry shift to using high-pressure steam generators, operating at 80 bar (1150 psig) to provide a temperature of 295°C/560°F. Soon, other import countries also started adopting European Standards. This essentially meant that heating coils and jackets of deodorizers, owned by suppliers wanting to sell oil in Europe, had to be rebuilt or new deodorizers and high-pressure steam generators purchased.115

Deodorizers are built in many vertical and horizontal designs.114,115 A Soft Column™ deodorizer, introduced by the Alfa Laval Company in 1996, is shown in Fig. 8.29. It differs from traditional deodorizers in having a "packed column" that uses a structured packing over which deaerated and heated oil flows countercurrent to the stripping steam. It is claimed to have a highly efficient oil-to-oil heat recovery system of 75 percent, uses less deodorizing steam, and operates at lower temperatures.116 Figure 8.29 will serve for a description of what occurs in a deodorizer. Well-prepared RB oil is passed through a heater (6), then through a deaerator (7) to eliminate dissolved air, through the economizer (3) for additional heating, and to the final heater (4). It then passes through the stripping column (1), to the holding section with sparger pipes (2), and the economizer (3) for reclaiming some of the heat. After partial cooling, citric acid (2050 ppm) in solution is added (10) as a sequesterant for iron or copper that may be picked up later by the oil. The solution is added while the oil is hot and under vacuum to flash off the moisture. Next, the oil is cooled and passes through a polishing filter on the way to temporary storage and shipment under nitrogen. The liquid used to "scrub" the vapors (5) is previously condensed deodorizer distillate that is chilled, recycled, and drawn off as necessary to maintain a constant level in the scrubber. The condensate may entrap vapors of pesticides if they get this far in the refinery. Deodorizer distillate is rich in tocopherols, some of which have vitamin E activity. With the increasing consumption of vitamin E supplements in the United States, and strong markets for natural antioxidants, deodorizer distillate is very much in demand and the domestic supply from refineries has long been committed by contracts with vitamin E producers. As much as 60 percent of the tocopherols in soybean oil could be extracted at the deodorizer, but in doing so they no longer would be available to stabilize the shelf life of the oil or provide its full natural nutritional benefits.

Fig. 8.29 Soft Column™ deodorizer. Legends: (1) packed column; (2) oil holding section; (3) vacuum heat exchanger ("economizer"); (4) vacuum-sparged final heater; (5) condensate scrubber; (6) plate final heat exchanger; (7) deaerator before high-temperature oil heating; (8) steam ejection vacuum section; (9) high-pressure type steam boiler; (10) citric acid adding system; (11) polishing filters; (12) cutaway of steam stripping section of column; preformed plates provide large surface "packing" area; (13) steam rises on the right, oil (from (4)) descends on left. (*Courtesy of Alfa Laval Oil & Protein Technology A/S Warminster, PA*)



At this point in processing, the peroxide value in the oil should barely be detectable (<1ppm) but will soon start increasing again. Thus, margarine and other profiled-temperature fats are blended before shipment. A "rule of thumb" in the deodorization department is: "Don't deodorize oil until it is sold and ready to be shipped." After cooling to appropriate temperatures, processors may add additional oil-soluble ingredients, antioxidants, vitamins, colors, etc. that customers might have difficulty dispersing in the oilfat at their processing sites. Unless the fat needs to be texturized into a soft-plastic form by the addition of nitrogen and chilling to a shortening for specific handling requirements, it is shipped and kept melted at about 10°C above melting point. Nitrogen purging of the oil during pumping, and storage under a nitrogen blanket, are common if held for more than several days before use by the processor. Contact with copper- and iron-bearing materials, and contamination with water should be avoided. Typically, 13 percent *trans* fats are produced during deodorization, with lesser quantities formed at lower temperatures. Ultra-high vacuum steam distillation, operating at microbars rather than millilbars, has successfully deodorized oils and fats in laboratory and pilot plant scale, and is used in the production of higher-priced products. However, fats and oils scientists report that oil flavor is not as acceptable as traditional deodorization/physical refining. Several steam distillation methods for the removal of cholesterol from animal fats were developed and patented after the fast foods industry switched from tallow to vegetable oils for frying French fries in the 1980s. The ability of tallow to regain this former market seems doubtful.

Notes on Selected Oil Species

Soybeans (*Glycine max* L.) were domesticated in north China, probably during 17001100 BC. and the Shang Dynasty or earlier.117 Samples found their way to Europe during the 1700s, where they were grown as a curiosity in botanical gardens.118 Documentation exists for soybeans being brought, grown, and made into soy sauce near Savannah, Georgia, in 1765, by Samuel Bowen, an English seaman who had sailed to Canton, China, earlier as an employee of the East India Company. Benjamin Franklin sent seed from London to a botanist in Philadelphia in 1770.119 A sample, from a salvaged Japanese junk, was brought from San Francisco to Alton, Illinois, in 1851, where it was grown and also distributed to botanists in other states.120 Additional samples were brought back by Commodore Matthew Perry's expedition to Japan in 1854, sent by missionaries in the far East, and distributed in the United States before establishment of the Department of Agriculture in 1864.120 However, the plant mainly was grown for feeding cattle and as green manure to enrich soils.

The Treaty of Portsmouth (New Hampshire) to settle the 190405 RussoJapanese War gave Japan control of the Liaotung Peninsula in Manchuria. This area was leased previously from China by Russia during construction of the Trans-Siberian (Moscow to Vladivostok) Railroad for establishing all-weather naval and trading ports in Port Arthur (now Lüshun) and Dairen (Lütita, Talien, Dalian).121 The withdrawal of Japanese occupation troops from Manchuria resulted in a local surplus of the soybean crop, which had been expanded to feed the soldiers. Manchurian soybeans were shipped to England in 1907,122 and to Germany, other Northern

European countries, and the U.S. Pacific Coast, by 1910.123 Dairen, Manchuria, became the international soybean processing and export center for the early part of the 1900s. An English hydraulic press mill had been sold to China in 1868,124 but Manchurian and Japanese soybean extraction continued mainly by crushing seed with stone edge runners and pressing with wedge presses, which produced press cake that would not stand shipment to Europe.123,125,126

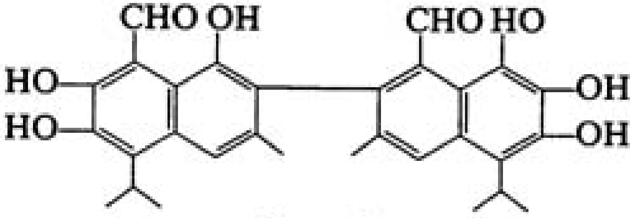
Europe wanted the soybeans for oil and the meal for animal feed, but was dissatisfied with leaving 56 percent oil in the cake, characteristic of box presses and expellers (screw presses). England and Germany initiated work on batch solvent extraction processes, but progress was interrupted by World War I. The Bollman basket countercurrent extractor was patented in Germany in 1919, followed by the Hildebrandt U-tube extractor in 1934.123

The first U.S. crushing of soybeans was in Seattle, Washington, in 1910, using seed imported from Manchuria.123 Domestically produced soybeans were crushed at cottonseed oil mills in the South and linseed (flaxseed) oil mills in the Midwest during the 19101920 era.123,127 The first commercially successful domestic soybean solvent continuous extraction plants were installed by Archer Daniels Midland (ADM) and the Glidden Company in Chicago in 1934, both using 100 tpd Hildebrandt U-tube extractors and commercial hexane as solvent. They also were equipped to produce phosphatides.124

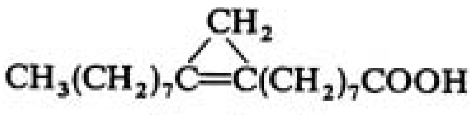
U.S. production of soybeans, for domestic processing and exporting to Europe, continued to grow until accelerated by World War II, which devastated the German oilseed crushing industry.128 This left the United States in the position of the largest soybean producer and processor. Soybean utilization increased during the reconstruction of Europe and Japan, by uses in famine abatement programs throughout the world, to meet growing world population food needs, and as feed to support world growth of the broiler industry which started about 1960. Eventually, land-rich countries in South America also became major soybean producers.

Cotton (*Gossypium arboreum* and *G. herbaceum*) was grown in the Indus River Valley (modern Pakistan) as early as 3000 BC., "New World" cottons were grown in Peru (*G. hirsutum*) in 2500 BC., and Sea Island cottons (*G. barbadense*) in the Caribbean Islands. Most of the world's cotton today is *G. hirsutum*, with fuzzy seed that requires saw delintering before dehulling and subsequent processing. Pima and Egyptian cottons (*G. barbadense*-type) have "naked" seed and can be processed directly. Cottonseed is among the first examples of a reclaimed by-product in our industrial world. Invention of the cotton gin by Eli Whitney in 1793 led to increased domestic production of cotton and also to seed disposal problems. The state of Mississippi passed the first anti-pollution law in the nation in 1857, prohibiting throwing cottonseed into rivers and requiring its orderly disposal from gins located less than one-half mile from a town.

Cottonseed contains the polyphenolic yellow-red pigment gossypol, 1,1′,6,6′,7,7′-hexa hydroxy-5,5′-diisopropyl-3,3′-dimethyl (2,2′-binapthalene)-8,8′-dicarboxaldehyde, in discrete bodies called *glands*. Processing challenges include removal of gossypol, or its 15 or more derivatives, from the oil to reduce color, and deactivating its toxic effects (through binding to the protein with moist heat) to enable feeding the meal to monogastric animals. Cottonseed oil also contains the cyclopropenoic acids, C18:CE malvalic acid (8,9methylene-8-heptadecenoic acid) and C19:CE sterculic acid (9,10-methylene-9-octadecenoic acid), which form the pink color in the Halphen test used to detect its adulteration of higher-priced olive oil.129 Earlier concerns about original adulteration of olive oil with cottonseed oil appeared again when fractionated palm oil was introduced to Europe.

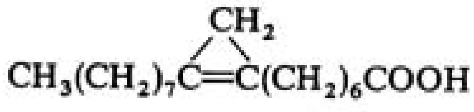


Gossypol



$$\mathrm{CH_3(CH_2)_7\overset{CH_2}{C{=}C}(CH_2)_7COOH}$$

Sterculic acid



$$\mathrm{CH_3(CH_2)_7\overset{CH_2}{C{=}C}(CH_2)_6COOH}$$

Malvalic acid

Cottonseed oil has long sold at a slight premium over soybean oil because of greater stability to oxidation, and desirable flavor in fried snack foods like potato chips. However, gossypol content, and lower protein quality put the meal at a price disadvantage. Feeding cottonseed to dairy cattle, whose rumen micro-organisms can detoxify limited amounts of gossypol, now utilizes about 60 percent of the crop and may portend eventual abandonment of extracting oil.

Rapeseed/Canola belong to the turnip rutabaga, cabbage, Brussels sprouts, and mustard family of crops that can be grown at low temperatures and moderate humidity. Three species have been grown as oilseeds: *Brassica napus*, known in Europe as rape, oil rape, Swede rape, and Argentine rape; *B. campestris*, known as rapeseed, oil turnip, turnip rape, and Polish rape; and *B. juncea*, known as leaf mustard, brown mustard, Oriental mustard, and Indian mustard. *B. campestris* was grown in India as early as 20001500 BC.

This crop is the classic example of an effective crop improvement program mobilized by one country in response to the impending loss of international markets. When it was reported in the early 1960s that the erucic acid (12-docosenoic, C22:1*n*9) in rapeseed oil can cause heart damage and other diseases, the Canadian government established a crash development program that led to the release of the first low-erucic-acid rapeseed (LEAR) in 1969. Continuing efforts to develop varieties with even lower erucic acid led to the later release of *double zero* (low-erucic, low-glucosinolate) varieties, which were named *canola* in 1980. Because of reduced glucosinolate levels, considerably more of the meal could be used in animal feed than earlier. Canola oil, which contains less than 2 percent erucic acid compared with 2040 percent in earlier rapeseeds, was granted GRAS (generally recognized as safe) status by the U.S. FDA in 1985.130

However, the introduction of canola left unmet needs for erucic acid in industrial markets. High erucic-acid rapeseed then was imported from Northern Europe for extraction, followed by efforts to develop industrial rapeseed with increased erucic acid content as well as the development of *crambe (Crambe abyssinica)* specifically for its erucic acid content. At the current state of development, equipment corrosion and poisoning of hydrogenation catalysts by sulfur in canola oil still are problems. Also, oil from the new canola varieties favors formation of β-crystals, in contrast to β″ crystal formation in the earlier rapeseed varieties.

Sunflower (Helianthus annus var. *Marcocarpus)* is a New World crop, known to have been grown in ArizonaNew Mexico in 3000 BC. and in the MississippiMissouri Basin at least since 900 BC. The crop was taken to Europe by early explorers and developed as a source of edible oil in Russia. The gray and white striped seed is known as *confectionery-type*. Its kernel is loose within the achene, and the seed usually is shelled and eaten. No major antinutritional factors are known to exist in the raw seed. *Oil-type* varieties predominantly are black-hulled, with seeds about one-third the weight of confectionery-types. The seed adheres tightly to the hull, and provides a challenge in dehulling. It has been repeatedly shown that, although the setting of seed by sunflower plants is not as sensitive to day length as that by soybeans, the oleic acid content is inversely related to the temperature of seed maturation. Thus, users desiring polyunsaturated oil with high linoleic acid content specified northern-grown sunflower oil, whereas those desiring extended oil stability, as in the production of fried foods, have preferred oil from seeds matured during the summer in the southern states. Sunflower varieties with high-oleic acid content (8592%) have been developed with the intention of using the oil as a feedstock in chemicals manufacturing processes.92 The industry currently is introducing "mid-oleic acid" oil, containing •65 percent oleic acid. Also, high-oleic confectionery-type sunflower seed has been developed, enabling extension of the shelf life of roasted seed.

Peanuts, *Ground Nuts (Arachis hypogae)* also are a New World crop and are known to have been grown in the Upper Plata Basin of Bolivia in 2000 BC. They were taken to Europe by early explorers, and were returned to the southeastern United States from Africa by slave traders. Broad cultivation did not occur domestically until the early 1920s, when the Southeastern United States was looking for a crop substitute to cotton, which was severely ravaged by the boll weevil.131 Most domestically grown peanuts are consumed as food, with over one-half of the crop production used in making peanut butter. The peanut crop is very susceptible to *Aspergillus flavus* invasion in the soil, which produces carcinogenic aflatoxins. World production of peanuts purely for oil has slowed because of limitations on feed uses of the meal. Direct food uses of peanuts in developing countries has increased.

8.8 Edible Uses of Oils and Fats

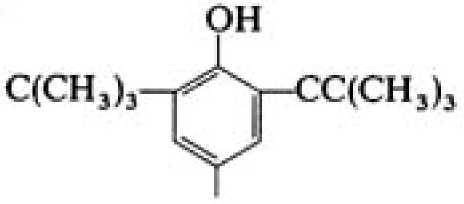
Antioxidants

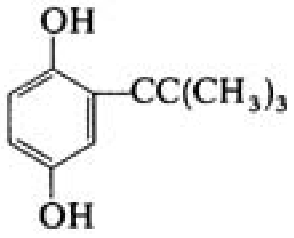
Fats and fat-containing products may be stabilized against oxidation by the addition of antioxidants as adjuvants. These compounds are believed to act as hydrogen donors or as free radical acceptors that intercept and hold quantums of energy that otherwise might induce oxidation. The major food grade synthetic antioxidants used include butylated hydroxyanisol (BHA), butylated hydroxytoluene (BHT), tertiary butylhydroquinine (TBHQ), propyl gallate (PG), and 2,4,5-trihydroxybutyrophenonone (THBP). Generally, they are allowed in food products at 0.01 percent of total fat weight each, with a total of 0.02 percent if a combination is used. Synergists used in combination with antioxidants include citric acid, isopropyl acid, phosphoric acid, ascorbic acid, ascorbyl palmitate, iso-ascorbic acid, tartaric acid, and lecithin, most of which act as chelators of pro-oxidative metals.

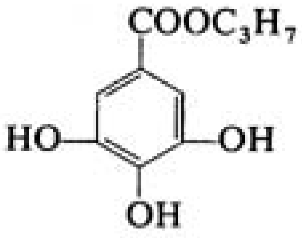
Natural antioxidants include gum guaiac, tocopherols (including vitamin E), and oil of rosemary (containing rosmaridiphenol). Because of cost, their use usually is reserved for premium-priced foods. Many types of tocopherols and associated compounds exist, and understanding of the mode of action is continually improving. The forms that give the most antioxidant protection do not always show the greatest vitamin E activity.132

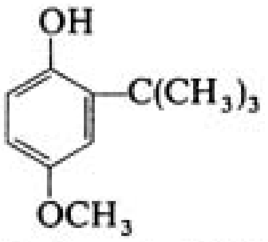
Assemblers of grocery store or food service convenience foods have the same general limitations. However, they also have access to natural cereal ingredients and herbs, herb extracts such as oils of rosemary, sage, and other materials in formulation that possess antioxidant properties but would not be counted in the maximum amounts of antioxidants permitted in the food. Types, mechanisms, and health effects of natural antioxidants are described in the references.133,134

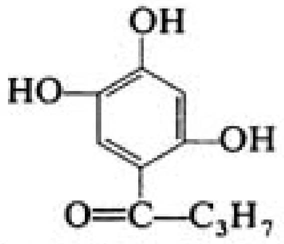
The crosslinking that occurs in drying oils also is an oxidation reaction. Polymerization of oils can occur rapidly in products such as fish meal that contain significant amounts of polyunsaturated triacylglycerolseven to the point of building up sufficient heat for a pile to spontaneously burst into flame. Thus, feed grade antioxidants, such as ethoxyquin, often are added before storage. Ethoxyquin also has been used for sparing the natural tocopherols for vitamin E activity in nutrition.

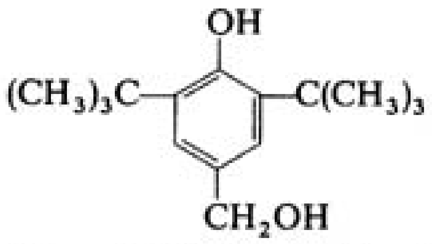

Butylated hydroxytoluene (BHT)


Tertiary butylhydroquinone (RBHQ)


Propyl gallate (PG)


Butylated hydroxyanisol (BHA)


Trihydroxybutyrophenone (THBP)


Hydroxymethylditertiarybutylphenol

Synthetic and natural antioxidants are known to be heat- and steam-distillable, and preferably should be added to oil products after the maximum heat encountered in processing. It further is known that many natural phenolic compounds in cereal-oilseed products, as well as reaction products of Maillard (nonenzymatic amino acid-reducing sugar) browning and natural wood smoke, have antioxidant effects. Additional techniques for reducing oxidative activity include maintaining nitrogen blankets on oils stored in tanks, and distributing oils in opaque containers or brown bottlesalthough this is not popular with many marketing departments who want to display clear, light-colored oils.

Nonionic Surfactants and Emulsifiers

Nonionic surfactants and emulsifiers, whose molecules have both aqueous (polar) and alkane (nonpolar) compatible sectors, also are common adjuvants. Their molecules have regions that are sufficiently similar to become part of either the water or oils phases, and other regions sufficiently dissimilar to repel that phase. For example, when added to a crystallizing fat, some nonionic surfactants may orient themselves to become part of the crystal, thus preventing further replication and limiting crystal size. Likewise, some will react with gelatinized starch in aqueous and bread systems and prevent its recrystallization (retrogradation). Furthermore, surfactants may orient around discrete droplets to stabilize water in oil (W/O) or oil in water (O/W) emulsions. When this occurs, the compatible end becomes associated with the discrete droplet, leaving the other end turned outward to associate with the compatible continuous phase.135

The major groups of commercial emulsion stabilizers include: (1) glycerol esters, favoring W/O emulsions; (2) esters of monoacylglycerols with hydroxycarboxylic acids (including lactic, succinic, malic, and tartaric); (3) sodium stearoyl-2-lactylate (SSL), favoring O/W emulsions; (4) fatty acid monoesters of ethylene glycol; (5) sorbitan fatty acid esters, known as SPANS favoring W/O, and TWEENS favoring O/W emulsions; (6) phospholipids; (7) water-soluble gums, including gum arabic, tragacanth, xanthin, agar, pectin carrageenan, and methyl- and carboxymethylcellulose; and (8) proteins.135,136

Glycerol has three exposed hydroxyl groups, resulting in a compound completely miscible in water at all concentrations. Fatty acids have both hydrophilic (water-attracting) and lipophilic (oil-attracting) ends. As they are esterified to glycerol, the molecular structure that originally was primarily watersoluble becomes increasingly oil-soluble. Finally, when all three positions are esterified with fatty acids, the molecule is nonpolar and soluble only in organic solvents.

The hydrophilicity of nonionic surfactants can be characterized numerically as their hydrophilelipophile balance (HLB). An HLB value of 36 indicates that the compound is a likely W/O emulsifier; 79, a wetting agent; 813, an O/W emulsifier; 1315, a detergent; and 1518, a solubilizer (of oil or other nonpolar compounds) into water. The HLB values of some common compounds are presented in Table 8.11.135 An HLB value of 8.0 is shown in Table 8.11 for lecithin, but manufacturers are able to supply lecithins with values of 212.

TABLE 8.11 HLB (HydrophileLipophile Balance) Numbers for Some Surfactants

Name	CAS Number	HLB Value
Oleic acid	112-80-1	1.0
Acetylated monoglycerides (film formers)		1.5
Sorbitan trioleate (SPAN 85a)	26266-58-6	1.8
Glycerol dioleate	25637-84-7	1.8
Sorbitan tristearate (SPAN 65a)	26658-19-5	2.1
Glycerol monooleate	25496-72-4	3.4
Glycerol monostearate	31566-31-1	3.8
Sorbitan monooleate (SPAN 80a)	1333-68-2	4.3
Sorbitan monostearate (SPAN 60a)	1338-41-6	4.7
Soy lecithin	8020-84-6	8.0
Sodium stearoyllactylate (anionic type)	18200-72-1	8.3

POE sorbitan monooleate (TWEEN 81a)	9005-65-6	10.0
POE sorbitan monostearate (TWEEN 60a)	9005-67-8	14.9
POE sorbitan monooleate (TWEEN 80a)	9005-65-6	15.0
POE stearic acid (monoester) (TWEEN 20a)	9004-99-3	16.9

aBrand names by Atlas.

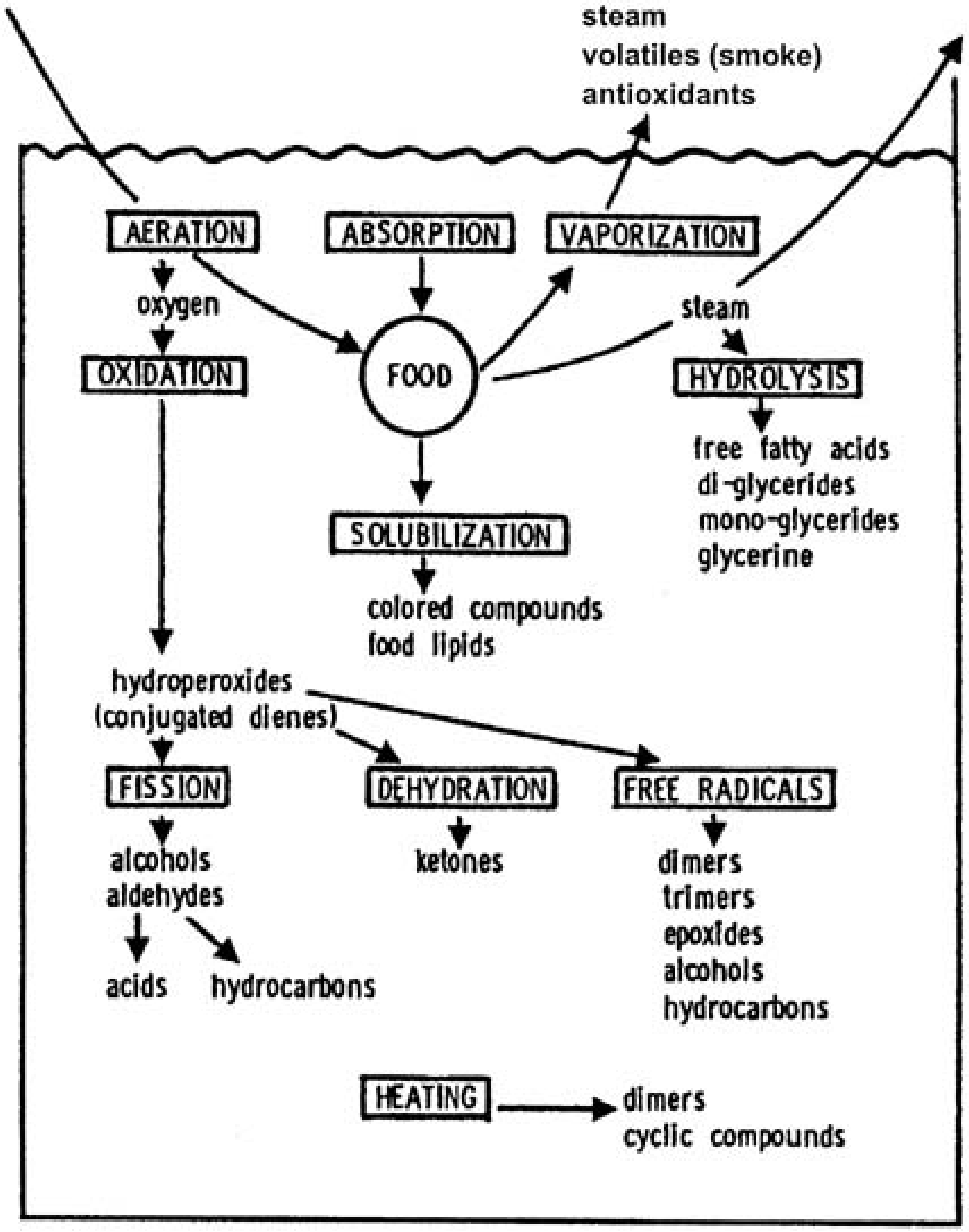
TABLE Oils

The processing of RBWD table oils has been mentioned throughout the text. Depending on positioning in the marketplace, these products may also serve as light duty frying oils and may be brush hydrogenated and contain added natural or synthetic antioxidants and methyl silicone at 0.53.0 ppm to slow breakdown, initiation of foaming, and smoking during frying.137 The U.S. and North European markets consider light product color and bland flavor as an indication of quality. Olive oil aficionados often prefer the green color and its stronger taste as an indication of "virgin" or minimum processing. All oils revert to stronger flavors in time. Slight flavor reversion generally is acceptable in cottonseed oil, but not in soybean oil. Many years of research were required to reduce the problem in soybean oil, and the mechanism still may not be fully understood. Converting the industry to only stainless steel oil contact surfaces (specifically avoiding iron and copper), reducing residual phosphatides content to essentially "zero," early inactivation of phospholipases and associated enzymes when preparing seed for extraction, nitrogen blanketing and reduced storage temperatures of bulk oil, and the use of antioxidants are believed to have helped.

Frying Oils

Oil acts as a heat transfer medium during the frying process which rapidly heats, cooks, and sometimes puffs, dehydrates, or forms a crust around the product. The mechanics of deep fat frying are shown in Fig. 8.30.138 The entering material carries oxygen that temporarily aerates the hot oil, and water that forms steam, with both rapidly swept out because of low solubility of gases in the hot liquid. Besides enzymes, which are likely to be inactivated rapidly, raw materials also carry oxidation catalysts like iron in hemoglobin of fresh meats, chlorophyll, and color pigments. Alkali may accompany the material, especially in chemically leavened doughnut batter and alkali-treated corn snacks. In short, almost every concern mentioned thus far in seed preparation, extraction, and oil processing occurs during frying. The high temperatures of frying, 177204°C/350400°F, are deleterious to oil quality. Reportedly, *trans* fats have been produced, and some of the degradation-prone anisidine value products remaining after refinery deodorization deteriorate. Oils start breaking down into simpler compounds, and also form cyclic compounds and polymers, a process that continues while fryers are left hot during inactive afternoons between preparation of lunch and dinner meals. However, limited self-cleansing occurs in the system by steam distillation, as noted by the reduction of carbonyl compounds in oils after inactive fryers are put into service again. An excellent review on frying has been edited by Perkins and Erickson.139

Fig. 8.30 Changes occurring during deep fat frying. (*From Fritsch, C. W., "Measurements of Frying Fat Deterioration,"* J. Am. Oil Chem. Soc., 58, *272274 (1981). With permission*.)



Unlike in some countries, few U.S. households own dedicated frying pots containing oil in readiness for meal preparation. Most deep frying is done by commercial operations like snack foods processors, producers of convenience fried foods like Chinese egg rolls and frozen meals, and fast food vendors who prepare French fries, fried chicken, and other products. These operations can be assigned to two groups: those who sell all the purchased oil with the product, and those who must recondition and occasionally dispose of oil. Requirements for the selection and use of frying oils in these applications differ.140

Generally, industrial snack food fryers are designed to use all the oil, and with clean-out systems and in-line filters. In theory, they should be able to operate indefinitely, only adding make-up oil as needed. These operations are of two kinds: "clean label," where company policy dictates that no, or minimum, additives can be used, and those allowing additives. "Clean label" operations carefully inspect and monitor the refinery operations of their self-certified suppliers. Antioxidants or silicone defoamers are not permitted in these oils. Oils are received in bulk, typically at less than 0.05 percent FFA, <1.0 PV, <4 ppm phosphorous, and <0.75 ppm chlorophyll, and are kept in stainless steel tanks under nitrogen.140 Stability against oxidation of soybean oil has consistently increased in trials as linolenic acid content decreased, whether by plant breeding, mixing with other oils, or by hydrogenation.141143 Less than 2 percent linolenic acid content has been a long-term industry goal, with the way by which linolenic acid is reduced (hydrogenation or breeding) appearing to have little effect on fried product stability.144

Oils with PVs higher than 2.5 are not used in well-disciplined frying operations. The freshness of the fried snack foods is dependent on the use of moisture and oxygen impermeable packaging, including laminates of windowless aluminized films that block 99+ percent of the light to prevent photo-induced oxidation. Additionally, the pouches are nitrogen flushed before sealing, thus creating pillow packs that protect the product against crushing. The snacks are handled by company delivery personnel who place them on store shelves or racks. Products may have shelf lives of 68 weeks, but inventories are carefully managed to ensure rapid turnover and fresh products.140

Normally, the warm surface oil serves as the binder ("tacking agent") for adsorbing salt and dry flavorings to snack foods. The concept of reducing or entirely eliminating oil from snack foods appeared during the early 1990s. Rather than deep fat frying, snacks were dried at high temperatures in fluidized bed continuous dryers. A far smaller amount of oil was then sprayed onto the dried product for flavor. Where a "fat free" snack food was desired, solutions of edible gums or specialty starches were sprayed onto the snack to serve as tacking agents, for water-soluble flavors; the product then required an additional drying step (to less than 1.52.0 moisture content to ensure crispiness).140

Industrial fryers, who cannot turn the oil over completely, have lengthened its life by 310 times by using polydimethylsiloxane (methyl silicone), which is not allowed in some countries. Levels as low as 0.20.3 ppm have been found effective, with commercial usage of 0.55.0 ppm reported. Users are advised to minimize levels of usage to 1.03.0 ppm. Dispersion of polydimethylsiloxane in oil is difficult. The compound operates by suppressing foaming and polymerization, and increasing smoke points of oils by up to 13.9°C/25°F. Antioxidants steam distill out of the oil during normal frying. Their initial inclusion essentially protects the oil only until the time of use, but some operators insist on periodically adding them to the fryer. For greater effectiveness, antioxidants are best sprayed onto the product in fresh oil after frying, or included in dry seasonings.140 Additional steps to prolong the use of frying oil include: in-line filters, periodic cleaning of fryers to remove settled charred product, neutralization of fatty acids, and refreshing the oil by passing through adsorbent earth filters continuously or at the end of the day. Numerous kits and advisory services are available.

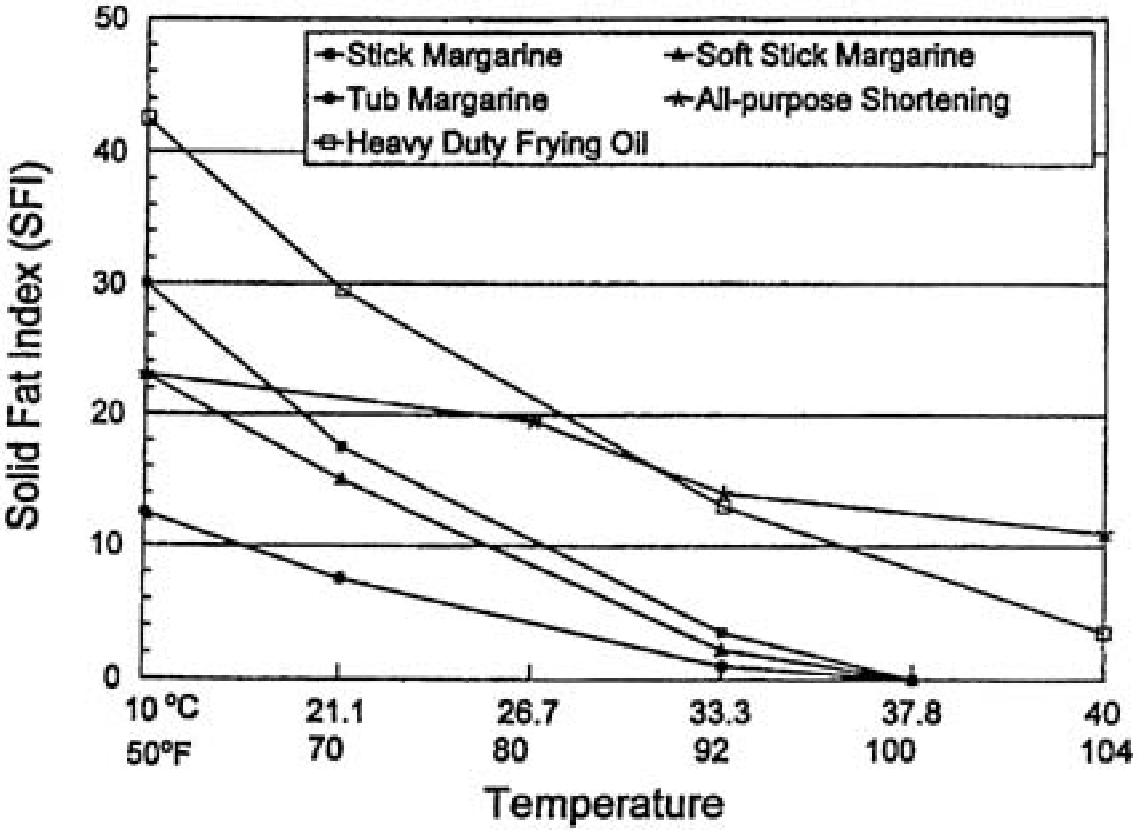
Large commercial frying operations, and fast food franchises that prepare French fries and chicken, have provisions and trained personnel to care for frying oils. The greatest food safety concerns are about small restaurants that do occasional frying during mid-day and evening meals. Several countries have imposed standards on the quality of frying oil in use. For example, Germany requires that the smoke point be not lower than 170°C/338°F, and total polar compounds not exceed 24 percent.145 Products fried in such oils usually would be objectionable in taste to most Americans long before they reach the unusable specifications. Yet, with exceptions of reduced digestibility and depletion of vitamin E (which can be supplemented), laboratory animals directly fed thermally degraded fats have not done as poorly as anticipated from the history of the oil and known presence of mutagens. Obviously, gaps exist in our toxicology and nutrition knowledge.

8.9 Margarines and Spreads

Margarine is one of the major temperature-profiled fat products. It was invented in France in 1869 intentionally as a butter substitute, and was first produced in the United States in 1873. Originally, it was made from animal fat; coconut oil became the lead fat in margarine in 1917, partially hydrogenated cottonseed oil in 1934, and partially hydrogenated soybean oil in 1956. Various legal principles have been tested through margarine. Among the first was protectionism for butter and the dairy industry, with the requirement that a sign "Oleomargarine Sold Here" be posted on the door of every store offering it. (In retrospect, modern marketers might see this as the best free advertising possible.) In response to dairy industry claims of product inferiority, a law was passed in 1923 requiring that margarine be fortified to the same level of vitamin A as butter; vitamin D was added later. The principle that an intended direct substitute be at least as nutritious as the replacement, has become part of the FDA's expectations for new products. Sales of colored margarine were approved on a state-by-state basis in the 1950s and 1960s, and U.S. per capita consumption of margarine surpassed butter in 1956. The major enabling technical breakthroughs were: (1) the invention of the internal scraped surface heat exchanger (SSHE, "Votator") in 1937; and (2) the lifting of emulsifier restrictions in 1992, which permitted the development of a wide range of reduced-fat spreads. Margarine typically has had the same fat content as butter in most nations, and is 80 percent in the United States. Spreads with as little as 20 percent fat are sold nowthe minimum limitation being technical rather than legal.

SFI profiles of several margarine types, an all-purpose shortening, a frying oil, and two shortenings are shown in Fig. 8.31.98 The SFI profile of butterfat varies with the season and feed, and typically is in the 4050 percent solids range 10°C/50°F. The soft stick margarine is softer than butter when taken from the refrigerator and, with tub margarine, is spreadable.

Fig. 8.31 Solid Fat Index (SFI) profiles for hard stick, soft stick and tub margarines, and for all-purpose shortening and heavy duty frying oil. (*Data plotted from Erickson, D. R., and M. D. Erickson, "Hydrogenation and Base Stock Formulation," in* Practical Handbook of Soybean Processing and Utilization, *D. R. Erickson (Ed.), pp. 218238, AOCS Press, Champaign, IL, 1985*.)



The basic sequence for making margarines/spreads includes:

1. Formulation of an oil mixture from hydrogenated base stock or interesterified fats, hard stock, and oil which has the desired SFI or SFC profile.
2. Ensuring that sufficient diversity occurs in the species and fractions to provide a variety of TAG that will form small β′ crystals.
3. Compounding the margarine oil blend at the refinery; final deodorization; shipping the melted blend to the margarine plant; and storage under a nitrogen blanket.
4. Preparing an oil-soluble additives mixture containing mono- or diglycerides and other emulsifiers, flavorings, oil-soluble Vitamins, and yellow color at the margarine/spread plant.
5. Preparing a water-soluble additives mixture containing water, salt, antimicrobial compounds, viscosity thickeners, and water-soluble flavorings. Formula amounts of thickeners and water-soluble flavorings increase as the fat content of the spread is reduced. The water should be microbiologically potable and deionized to remove calcium, magnesium, iron, and copper.
6. Blending the heated oil-soluble additives mixture into the warm oil in a mixing tank. Slowly adding the warmed water-soluble additives mixture into the blend to produce a water-in-oil emulsion.
7. Pasteurization; partial cooling.
8. Chilling and working the emulsion in a series of (typically ammonia-chilled) internal scrape-surface heat exchangers ("A" units) and "picker bar" workers ("B" units), with resting tubes interspersed to ensure adequate time for the establishment of the desired crystal structures.
9. Filling.
10. Holding under controlled temperature conditions for the establishment of the desired crystals.
11. Shipping.

Much of the physical chemistry for making temperature-profiled fats has already been reviewed, including Figs 8.1 and 8.2. Additional references are available.30,146 We cannot change the laws of thermodynamics, but can slow or hasten arrival at equilibrium. The margarine literature often shows a table similar to Table 8.12, but does not explain the practical implications. Introducing a diversity of TAG makes it more difficult for the same species to align and crystallize. It is easier to make whole cottonseed oil margarines/spreads than whole soybean oil counterparts since the former contains about 2.5 times more palmitic acid than soybean oil. Thus, when soybean oils are hydrogenated they tend to increase the preponderance of C18 fatty acids. For years, it was common practice to include about 10 percent cottonseed hardstock in soybean margarines. A reduction in TAG diversity increased the problems of making margarine when erucic acid was eliminated from canola (rapeseed) and other species, but increasing diversity, as by the rearrangement of lard and increased presence of *trans* fats, has assisted in making smooth-textured margarines. A similar lack of diversity is encountered in making palm oil margarines; thus, the addition of C18 hardstocks and longer pin working and resting times have been employed. Crystals grow by adding identical compounds side by side. Emulsifiers, such as mono- or diglycerides, can align next to the crystal and be accepted as part of the lattice. But, not being identical, they foul the surface and stop crystal growth

TABLE 8.12 Crystal Forming Tendencies of Hydrogenated Oils, Collected and Updated from Various Sources

Beta Prime (β′) Type	Beta (β) Type
Cottonseed	Soybean
RapessedHEARa	CanolaLEARa
Lardmodifiedb	Lardnonmodifiedb
Palm	Palm kernel
Rice bran	Sunflower
Herring	Olive
Menhaden	Corn
Milk fat (butter fat)	Peanut
Tallow	Safflower

Cocoa butter
Coconut
Sesame

[a] By elimination of erucic acid from traditional high erucic acid rapeseed, the C22 fatty acid family also was eliminated. The replacement canola (low-erucic-acid rapeseed) had one less variety of fatty acids, and became a β crystallizer.

[b] A reverse situation occurred in lard, which has few predominant TAG types. Rearrangement created additional types which led to β′ crystallization.

Margarine, like butter, is a water-in-oil emulsion; that is, the water, and all ingredients soluble in the water are encapsulated in a continuous phase of oil. This has many practical advantages, since bacteria can grow in their individual droplets only to the point that local nutrition permits, but cannot cross over to a second nearby food supply. If the product is salted at 1.5 percent, the water droplets will have a salt concentration of 7.5 percent, which is inhibitory to many bacteria. If the emulsion has been properly made and pasteurized, many droplets will be small and not contain bacteria cells. Lecithin is included in margarines/spreads at 0.10.5 percent as an antispattering agent used for frying; it is introduced with the oil-soluble additives initially added to the base oil. The margarine/spread production flow sheet in Fig. 8.32 shows how these principles are applied. The figure represents a multipurpose plant, in which various products can be made. The legend in the upper right corner shows the effects of the operations employed. The emulsions are prepared in tanks (A), although the provisions for pasteurization and cooling before processing are not shown. Consumer margarines must meet FDA Standards of Identity which specify the permissible additives. Industrial buyers may arrange to have other FDA-approved ingredients added as a means to simplify their dispersion in manufactured products. The emulsion is pumped to an internal scrape-surface chiller, and exits at 1019°C/5065°F as a soft glass (stiffening when held in the hand). Since this is not the type of final crystal desired, the chilled emulsion is then sent through a series of pin workers (tubes through which the margarine is pumped while a shaft with finger-like pins mixes it), chillers, and resting stops to craft the desired crystal. The final temperatures for table margarines and margarinebutter blends are -18°C to -20°C (014°F), and -5°C to -1°C (1223°F) for low-calorie spreads.147

Fig. 8.32 Flexible Perfector Plant™ for making stick, tub and pastry margarines, and shortening. (*Courtesy of Gerstenberg & Agger A. S. Brondby, Denmark*)

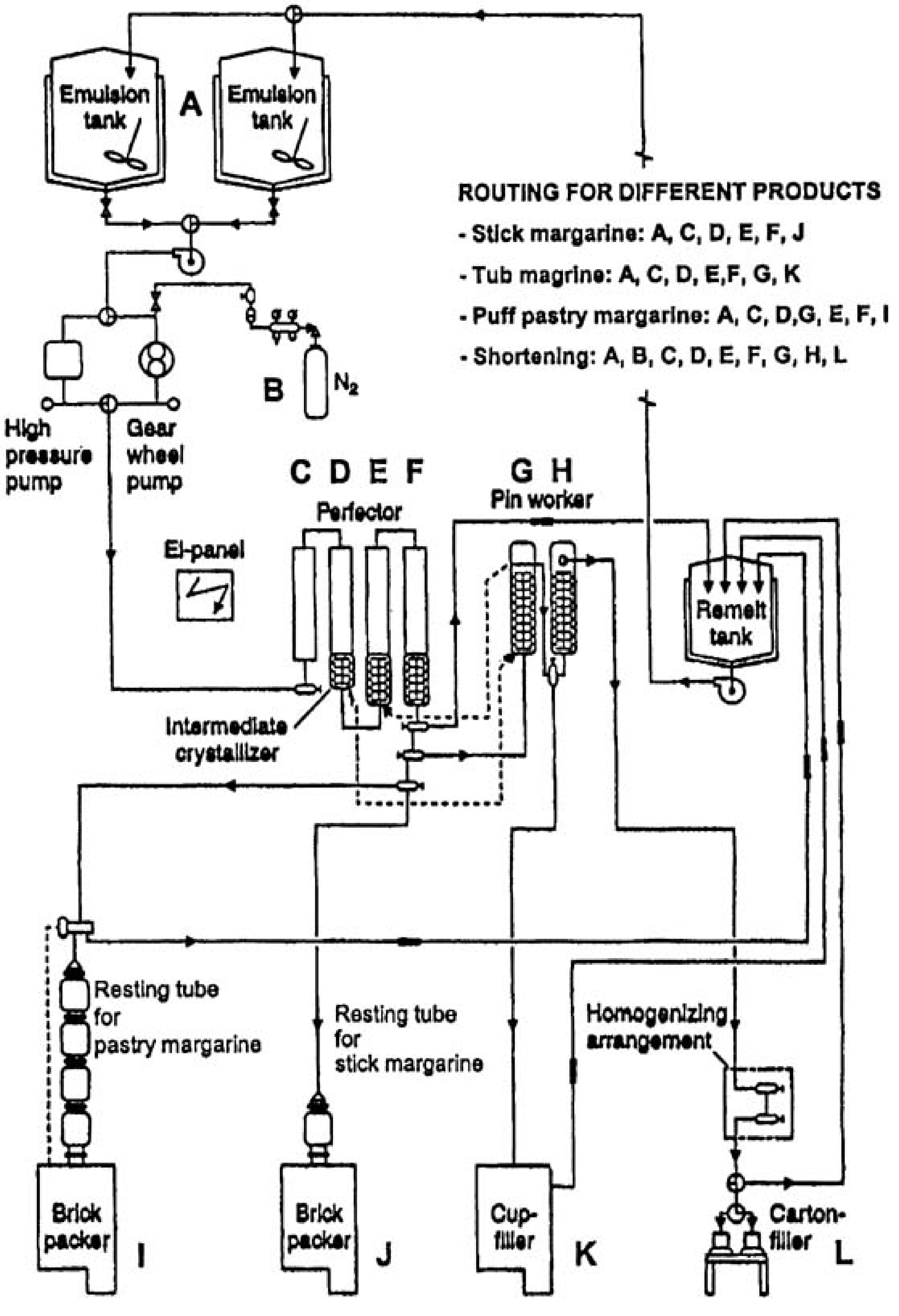


Figure 8.32 shows that tub margarine is pin-worked considerably more to keep it more liquid until filling. The solids content profile also is different from the stick margarine, and a smooth-textured product with minimum stiffening is desired.

Some bakery products ("puff pastries", phyllo) consist of flaky layers, made by rolling the dough thinly, covering it with shortening, folding, rerolling, and repeating the process many times. On baking, they puff up. The gluten layers must stay intact, and the shortening also must be flexible but not give off free oil during the machining. Puff pastries, such as croissants, whose fat melts and cleans up in the mouth, can be made from butter, but must be prepared at low temperatures (10°C/50°F) in refrigerated rooms using chilled equipment. A very flat SFI curve (30.033.5 at 10°C; 28.030.0 at 21.1°C; 24.526.5 at 33.3°C and 19°C minimum at 40.0°C), is recommended for tallowvegetable oil puff pastry margarines. They melt slightly above body temperature, but greasiness is noticed by some people. Historically, puff pastry margarine was made using chilling rolls, but processes exist now for using scrape-surface heat exchangers, "B" units, and large resting tubes.147

< previous page

next page >

8.10 Shortenings

Typically, the function of fats in baking is to "shorten" or control the development of wheat flour gluten and avoid or limit toughness in the product. In doing so, starch becomes the predominating matrix, and tenderizing complexes can be established between the fat and carbohydrates. A fat with a flatter solids temperature profile than the "all-purpose shortening" in Fig. 8.32 can accomplish its "shortening effect" during machining of the dough and baking even if it does not melt substantially during eating.

Numerous emulsifiers are available for dough (continuous gluten) and cake (sugar, corn sweetener, starch, fat emulsion) systems. Emulsifiers act as conditioners in dough systems by: improving tolerances to variations in flour and other ingredients; increasing resistance to mixing and mechanical abuse; providing increased gas retention, shorter proof times, and increased product volume in yeast-leavened systems; improving uniformity in cell size, finer grain, stronger cell walls; improving slicing; and by delaying staling and starch retrogradation to lengthen product freshness.136

Cakes essentially are emulsified slurries before baking. Considerable use is made of sugar (with corn sweeteners increasingly used), starch in low-protein-content flours, and fat. Emulsifiers have three functions in cake systems: (1) to improve air incorporation; (2) to disperse shortening into smaller particles to maximize the number of air cells; and (3) to improve moisture retention. Complexes occur between fats, emulsifiers, and starches that result in smooth, tender, moist cakes. Emulsifiers include broad classes of lecithin and lecithin derivatives, mono- and di-fatty acid glycerol esters, hydroxycarbocylic acid and fatty acid esters, lactylated fatty acid esters, polyglycerol fatty acid esters, ethylene or propylene glycerol fatty acid esters, ethoxylated derivatives of monoglycerides, and Sorbitan™ fatty acid esters.136 Many oil processors sell company brands of shortenings containing emulsifiers, and large bakers can arrange to have emulsifiers added to their melted fat mixtures before shipping.

As shown in Fig. 8.32, nitrogen is injected into the first scrape-surface chiller with the shortening oil mixture to give an opaque white appearance, and added plasticity. Some bakers prefer to use semi-plastic shortenings in applications where oil may shorten the dough too rapidly, and for "creaming" of sugar and shortening as in making cake frostings. Some small restaurants prefer semi-solid "shortenings" and have fryers able to accept standard 18.2 kg (14 lb) cubes. Care should be taken to avoid development of large air spaces when melting cubes over the (electric) heating rods to avoid burning the shortening.

Other Edible Applications

Cocoa butter (CB) has a challenging chemistry and has attracted many efforts to develop cheaper, acceptable alternatives. The following definitions provide a quick introduction to this field: (1) *cocoa butter equivalents* (CBEs) are compounded mostly from tropical oils other than palm. Because their melting and crystallization properties closely resemble CB, they are compatible as diluents at all levels of substitution; (2) *cocoa butter replacers* (CBRs) are made from nonlauric oils (typically soybean, cottonseed, or palm) partially hydrogenated for maximum *trans*-C:18 isomer formation to acquire a steep melting profile. They are best used for enrobing bakery products, but their melting profiles can be improved by chill fractionation; and (3) *cocoa butter substitutes* (CBSs) are made primarily from lauric-type fats (C:8C:12 of palm kernel and coconut origin), are hydrogenated under conditions that favor *trans* formation, but may contain a limited amount of hydrogenated C18 triglycerides. CBS limitations include: (1) they must be used alone because of incompatibility with cocoa butter; (2) they require the use of cocoa for "chocolate" flavor; and (3) all ingredients with active lipase systems must be avoided to prevent the development of a soapy flavor.148

The reader is referred to the volume by O'Brien137 for other food applications of fats, including: icings for sweet goods; spray-dried nondairy creamers; coffee whiteners; aerated whiteners with encapsulated air to produce a cappuccino effect when added to coffee; dried powders designed to be reconstituted, pasteurized, homogenized, and packaged for restaurant use as creamers; vending machine dry creamers; whipped topping shortenings, also used for making bakery cream pie fillings and cake toppings, aerosol toppings, powdered toppings, and frozen ready-to-use toppings; cheese analog shortenings, frozen dessert (mellorine) shortenings, sour cream analog, and dip base fats; and sweetened milk and sweetened condensed milk analog fats.

8.11 Industrial Uses of Oils

Timeline

Industrial uses typically means nonfood, nonfeed applications, although feed uses have not been excluded consistently. No one knows when man first used oils or fats for lighting, medical, and cosmetic applications, lubricants, or combined them with wood ashes to make soaps. The following sequence has been published for soybean oil:4

1. <AD 980China: soy oil likely used for illumination in lubricating fluids and coatings.
2. AD 980China: documented use of soy oil boat caulking materials.
3. 1908Europe: soy oils used in soaps; glycerine sought for making explosives for Panama Canal project and printer's inks; oil additionally used in rubber substitutes and linoleum flooring.
4. 1910United States: soy oil classified as a drying oil; used as cheap replacement for linseed oil in paints.
5. 19141918United States: largest soy oil industrial market is soaps, followed by paint, varnish, enamel, linoleum, oilcloth, asphalt, and other waterproofing materials.
6. 1919Blowing warm air through heated oil found to increase viscosity by initiating oxidation and polymerization; blown soy oil improves the properties of printing inks.
7. 1926Soy oil used for plasticizing and increasing the elongation of rubber.
8. 1930sUnited States: Kienle and Hovey of General Electric developed soybean oil alkyd resins used in paints to improve drying, adherence, endurance, and color; Ford Motor Company used soybean oil and its derivatives in enamel paints for automobiles; Du Pont's "four-hour enamel," based on soybean oil, considered the most important factor in furthering soy oil usage in paint.
1. Quality of soy oil first reported suitable for food use and hydrogenation; replacements of cottonseed oil begin.
2. Strong chemurgic movement established; credited with nearly 200 industrial uses during its 10-year existence. Ford Motor Company uses significant amounts of soybean oil for enamel paint, glycerine for shock absorbers, in 1937 automobiles.
9. 1945Chemurgic movement dies after World War II.
10. 1950sD. Swern, USDA scientist, develops epoxy plasticizers from oils or monohydric fatty esters for use in plastics.
11. Late 1970sConverting oil triglycerides to methyl- or ethyl-fatty esters by transesterification ("alcoholysis") reduces injector fouling, carbon deposits, and degradation of lubricating oils in diesel engines, compared with direct use of vegetable oils or diesel fuelvegetable oil mixtures.
12. Early 1980sAmerican Newspaper Publishers Association develops "first-generation soybean oil-based inks" to replace uncertain mineral oil supplies.
13. Mid 1980sDegummed, alkali-refined cottonseed and soybean oils used in pesticide aerial spays.
14. 1987USDA Federal Grain Inspection Service allows the use of soybean and other edible oil sprays to reduce the risk of grain dust explosions in elevators. Use of 12 percent soybean oil in livestock feeds reduces dust in pig-rearing facilities, improves animal health; gives 510 percent increase in weight gains.

Chemurgy Revisited

Chemurgy is the use of replenishable farm crops as feed stocks for industrial processes. The concept became very popular with onset of the Great Depression in the early 1930s because it promised new uses for agricultural crops and the development of a self-sustaining national economy. Famous advocates of the era included Henry Ford, industrialist, and George Washington Carver, crop scientist. The movement was heavily subsidized by the federal government, slowed as World War II approached, and ended in the prosperity of the post-war era.

Currently, the United States again has agricultural surpluses, especially soybean oil. Naturally, the following question arises: "Are business prospects for chemurgic products different now than in 1937, especially with the United States committed to an open global trade policy?" New factors include: (1) Agricultural products offer the promise of biodegradability, which is becoming increasingly popular with an environment-concerned public. (2) The global politics of petroleum carry many hidden costs, including threats of continued supply, occasional price fixing, and potential involvement in wars. (3) The petroleum companies are driven by economies of scale, and have limited flexibility in responding to small, although reliable, markets. (4) Biotechnology offers the promise of tailored crops, able to produce high levels of specific chemical feed stocks, previously unavailable. But, (5) the portion of the public engaged in agriculture and agribusiness is the smallest ever, and federal subsidies may not be easily obtainable in the future. Nevertheless, current research in chemurgic applications is the highest ever, and many products have been launched by small entrepreneurs to regain markets previously lost by vegetable oils to petroleum feed stocks (Table 8.13).

TABLE 8.13 Examples of Soybean Oil and Lecithin Industrial Usesa

Soybean Oil Technical Uses	Oleochemicals from Oil	Soybean Lecithin
Anti-corrosion agents	Methyl esters Soy diesel fuel Solvents	*Wetting agents*
Anti-static agents		Dry powders
Candles		Cosmetics
Caulking compounds		Plant pigments
Composite materials	Fatty acids	
Concrete release agents	Fatty alcohols	*Nutritional*
Core oils	Glycerinindustrial and explosive uses	Medical
Crayons		Vitaminsanimal feeds
Dust control agents		
Electrical insulation		*Anti-foaming agents*
Epoxy resins		Alcohol
Fungicides		Yeast
Hydraulic fluids		
Printing inks		*Dispersing agents*
Linoleum backing		Ink
Lubricants		Pesticides
Metal casting/working oils		Magnetic tape
Oiled fabrics		Paint
Paints		Paper
Pesticide carriers		Synthetic rubber
Plasticizers		
Protective coatings		*Other*
Putty		Viscosity modifications, concrete, drilling muds
Soaps/shampoos/detergents		
Solvents		Softening and curing leather
Vinyl plastics		
Wallboard		
Waterproof cement		

aCourtesy of American Soybean Association, St. Louis, MO.

Industrial Soybean Oil Consumption Statistics

Estimates for recent industrial soy oil consumption are shown in Table 8.14. The most rapid current growth is in sales of methyl esters and construction of biodiesel plants. The category "fatty acids, soap, and feed" is one of convenience. U.S. Bureau of Census figures on soaps are not published, but research activity is underway to improve methods for producing sulfonated fatty acids for use in detergents. The chemistry of fatty acids processing and uses has been summarized by Johnson and Fritz.149

TABLE 8.14 Domestic Uses of Soybean Oil, 2000a

	Production		
Type of Use	Amount	Units	Percent of Category
U.S. soybean productionb	75.39	mmtc	
U.S. soybean crush	43.28	mmtc	
Oil produced	8.1	mmtc	
Edible soybean oil consumedd	6.24 t	mmtc	100
Salad or cooking oil	2.95	mmtc	47
Baking and frying fats	2.24	mmtc	36

Margarines and spreads	0.73	mmtc	12
Other edible products	0.06	mmtc	<1
Inedible products	0.27	mmtc	4
Industrial soybean oil consumptione	288	tmtf	100
Fatty acids, soap, and feed	164	tmtf	58
Resins and plastics	52	tmtf	18
Inks	49	tmtf	17
Paints and varnishes	18	tmtf	6
Biodiesel	4	tmtf	1
Solvents	0.5	tmtf	<1

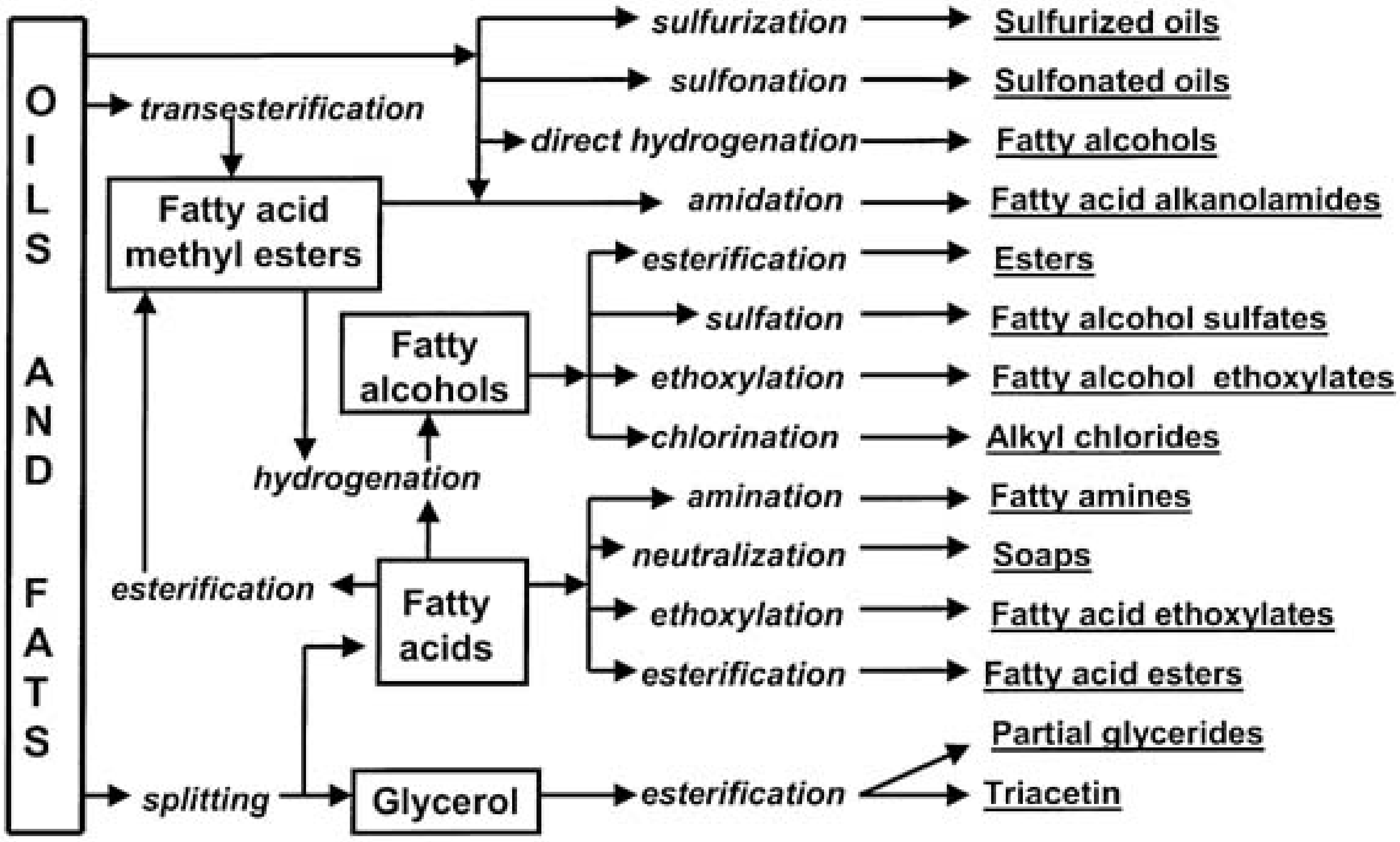
aFrom Soy Stats 2002, United Soybean Board, St. Louis, MO;
bUSDA statistics;
cmmt = million metric tons;
dU.S. Census Bureau statistics;
eUnited Soybean Board/Promar 1999/2000;
ftmt = thousand metric tons.

Crude soybean oil has limited uses beyond sprays to keep down dust. Even there, the spray nozzles are in danger of clogging by phospholipids, which also leave sticky surfaces if an area or material is sprayed too often. Generally, once-refined (degummed, alkali-neutralized, water-washed, and dried) oil is the minimum quality used for dust control and aerial spraying of pesticides. Anti-corrosion and anti-polymerization agents are added in lubricants and also in hydraulic fluids. Methyl and ethyl esters of fatty acids do not have the glycerol component, and are an even more "pure" form in that they do not leave carbon residues when used as fuels.

Fatty Acid Methyl Esters

Pathways for converting oils and fats into various oleochemicals are shown in Fig. 8.33.4,150 Recent technical summaries include inks, paints and coatings, biodiesel fuels and additives, lubricants, and associated processing.151,152 Fatty acid methyl esters (FAME) are the gateway to many products. Glycerol, produced as a by-product of alcoholysis interesterification, is in great demand. Current uses include: cleaning graffiti, stains, sticky deposits; light uses as lubricants; degreasing baths; inclusion in penetrating oils; asphalt and concrete mold release agents; and adjuvants in various applications. Considerable investments in expanding manufacturing facilities have been made in the past year in expectation of rapid growth of this industry.

Fig. 8.33 Oleochemical derivatization pathways. (*Modified from Zoebelein, H., "Renewable Resources for the Chemical Industry,"* INFORM, 3*, 721725 (1992), and Johnson, L. A. and Meyers, D. J., "Industrial Uses for Soybeans," in* Practical Handbook of Soybean Processing and Utilization*, D. R. Erickson (Ed.), pp. 380427, AOCS Press, Champaign, IL, 1995*.)



Vegetable oil fuels have been prepared by various methods, including microemulsification, transesterification, and pyrolysis. The fuel properties of some vegetable oils and soybean esters are presented in Table 8.15.153 The heat content of various vegetable oils is nearly 90 percent that of vegetable oils. Problems encountered in handling vegetable-oil-based fuels include: higher viscosity, and higher cloud and pour points, which may require supplemental heating of fuel tanks in cooler weather. Also, raw material costs are high. Methyl esters currently are made from spent frying oils and inedible animal fats and attract much publicity. In fact, supplies of such materials are limited compared with the amount of TAG that would be required if significant quantities of methyl esters were used in fuels. Government support of biodiesel use has consisted of: (1) individual states mandating that specific amounts of agriculturalorigin methyl esters be added to diesel fuel sold locally; and (2) reducing or eliminating taxes collected on such fuels. In turn, this limits the amounts of methyl esters that can be blended with diesel fuels sold competitively with nonblended fuels.

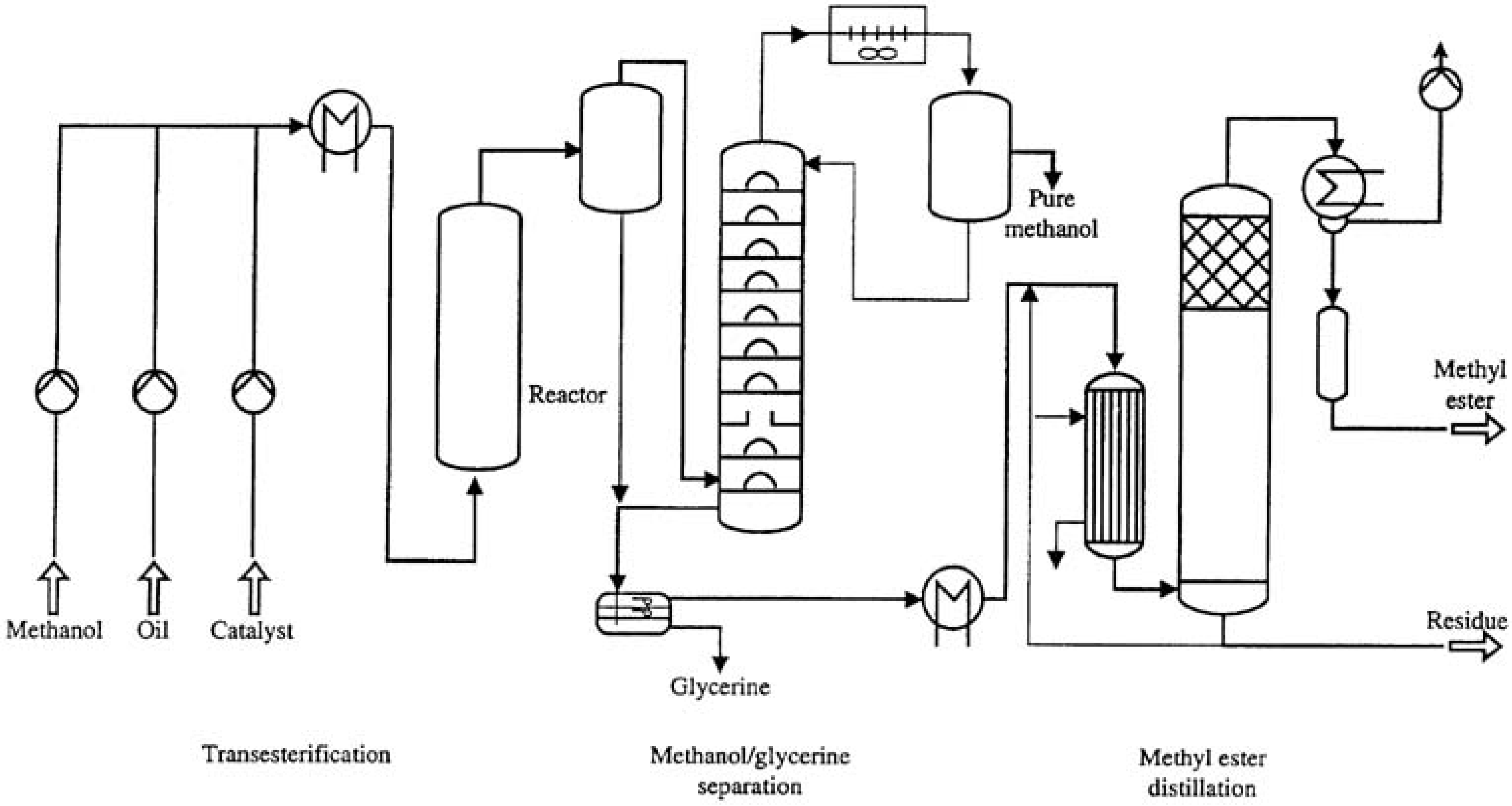
TABLE 8.15 Fuel Properties of Some Vegetable Oils and Soybean Estersa

Oil or Ester	Viscosity (mm2/S)	Cetane No.	Gross Heat of Combustion (kJ/kg)	Cloud Point (°C)	Pour Point (°C)
Oilsb					
Castor	297.0		37,274	None	-31.7
Corn	34.9	37.6	39,500	-1.1	-40.0
Cottonseed	33.5	41.8	39,468	1.7	-15.0
Crambe	53.6	44.6	40,482	10.0	-12.2
Linseed	27.2	34.6	39,307	1.7	-15.0
Peanut	39.6	41.8	39,782	12.8	-6.7
Rapeseed	37.0	37.6	39,709	-3.9	-31.7
Safflower	31.3	41.3	39,519	18.3	-6.7
High oleic safflower	41.2	49.1	39,516	-12.2	-20.6
Sesame	35.5	40.2	39,349	-3.9	-9.4
Soybean	32.6	37.9	39,623	-3.9	-12.2
Sunflower	33.9	37.1	39,575	7.2	-15.0
Soybean estersc					
Methyl soyate	4.1	46.2	39,800	2	-1
Ethyl soyate	4.4	48.2	40,000	1	-4
Butyl soyate	5.2	51.7	40,700	-3	-7
No. 2 diesel fuel	2.7	47.0	45,343	-15.0	-33

aFrom: Foglia, T. A., Nelson, L. A., Marmer, W. N., Knothe, G. H., Dunn, R. O., and Bagby, M. O., "Improving the Properties of Vegetable Oils and Fats for Use as Biodiesel," in *Emerging Technologies, Current Practices, Quality Control, Technology Transfer, and Environmental Issues*, S. S. Koseoglu, K. C. Rhee, and R. F. Wilson (Eds.), pp. 121125, AOCS Press, Champaign, IL, 1998, with permission.
bViscosity determined at 38°C.
cViscosity determined at 40°C.

Biodiesel is also called "methyl soyate" if made from soybean oil. Tests during the mid-1970s, early 1980s, showed that diesel engines can initially run on vegetable oils or animal fats, on their mixtures with diesel fuel; but, despite various additives to the fuel and engine oil, problems eventually were encountered with fuel injection valve clogging, cylinder head carbon deposits, and engine oil fouling by fuel blow-by. An example of processes that were developed for making methyl or ethyl fatty acid esters is shown in Fig. 8.34.154 Various techniques were tried to reduce the cold temperature viscosity problems of methyl soyate. Winterization lowered the crystallization temperature by 7.1°C.155 Increasing the molecular diversity by the addition of isopropyl and 2-butyl (branched alcohol) esters lowered the crystallization temperature of soybean methyl esters by 711°C and 1214°C, respectively.156 More recently, it has been reported that this molecular diversity effect is significantly diluted in 20:80 mixtures of soyate and No. 2 diesel.153

Fig. 8.34 Flow diagram for manufacturing methyl esters by interesterification. (From *Krawczyk, T., "Biodiesel,"* INFORM, 7, *800801, 804808, 810824 (1996). With permission*.)



Other Industrial Uses of Soybean and Other Oils

Other current areas of soy oil industrial applications research include plastics, coatings, lubricants, and hydraulic fluids. Potential applications are only limited by imagination, economics, and the business skills of the respective entrepreneur. The United Soybean Board maintains a website (www.unitedsoybean.org) of current soybean oil-based industrial products manufacturers, listed under the categories of: adjuvants, alternative fuels and fuel additives, building and construction, cleaners, concrete, dust suppressants, engine oils, hydraulic fluids, ingredients, metal working fluids, printing, and miscellaneous. The categories list suppliers and as many as several hundred products each. In many cases, suppliers list their web sites for interested persons to learn more about uses and specifications of products.

Oilseed crops, currently studied extensively as potential sources of specialty fatty acids, include: *Crambe abyssinica* for erucic acid; *Limnanthes alba* for very long-chain fatty acids; *Dimorphotheca pluvialis* for dimorphecolic acid; *Lesquerella fendleri* for lesquerolic acid; *Calendula officinalis* for calendic acid; and *Euphorbia lagascae* and various *Vernonia* species for vernolic acid.152

8.12 Analytical Methods

Selected analytical methods, adopted by the American Oil Chemists' Society (AOCS)5 for characterizing the composition, structure, physical properties, and stability of fats and oils, are summarized below. Prescribed equipment must be used and conditions followed. Some of these techniques are limited to specific oil species, but adaptations are available for other species. Procedures for sample drawing and preparation also are specified.

Active Oxygen Method for Fat Stability (AOM) (Cd 12-57)
determines the time (in hours) for a sample of fat or oil to attain a predetermined peroxide value (PV) under the conditions of the test. The method is used to estimate the comparative oxidative stability of fats and oils. The method has been placed in surplus, in favor of Cd 12b-92 (Oil Stability Index), but retains official status and is still used in domestic industry.

p-Anisidine Value (AV) (Cd 18-90)
determines the amount of aldehydes (principally 2-alkenals and 2,4-dienals) in animal and vegetable fats and oils. These are degradation products of peroxides, which are not removed by bleaching. Some fats and oils chemists propose increased use of this method in purchase specifications.

Bleaching Test for Soybean Oil (Cc 8e-s63)
determines the color of a sample of soybean oil after treatment with a specified bleaching earth. Specific methods exist for other oil species.

Boemer Number, Foreign Fats in Lard (Cb 5-40)
estimates the presence of tallows and similar fats, based on differences in melting points of foreign glycerides and fatty acids as compared with pure pork fat.

Cloud Point Test (Cc 6-25)
determines the temperature at which a cloud first forms in a sample of melted fat in the first stage of crystallization.

Cold Test (Cc 11-53)
measures the relative resistance of a sample to crystallize at an established temperature in terms of time. The test often is used as an index of stearin removal in the winterization of salad oils.

Color Measurement by LovibondWesson (Cc 13b-45)
determines the color of clear oil samples by comparison with glasses of known color characteristics.

Fatty Acid Composition by Gas Chromatography (Ce 1-62)
quantitatively determines saturated and unsaturated fatty acids with 824 carbon atoms in animal fats, vegetable oils, marine oils, and fatty acids after conversion to their methyl ester forms.

Fatty Acid Composition by GLC (Ce Ic-89)
measures the fatty acid composition and levels of *trans* unsaturation and *cis, cis* methylene-interrupted unsaturation of vegetable oils using capillary gas liquid chromatography.

Flash PointPensky-Martens Closed Cup for Fats and Oils (Cc 9b-55)
determines the temperature at which an oil sample will flash when a test flame is applied. This technique sometimes is used to estimate levels of residual hexane and to ensure the safety of workers handling the oil. Some refineries use gas chromatography methods instead.

Free Fatty Acids in Crude and Refined Fats and Oils (FFA) (Ca 5a-40)
determines FFAs, as oleic acid, by ethanolic sodium hydroxide titration.

Halphen Test for Detecting Cottonseed Oil (Cb 1-25)
estimates the presence of cottonseed oil in vegetable or animal fats or oils as the result of a pink color formed between the reagent and cyclopropenoic fatty acids (sterculic and malvalic) normally present in cottonseed oil.

Hexane Residues in Fats and Oils (Ca 3b-87)
determines, by gas chromatography, the "free" volatile hydrocarbons remaining in fats and oils after extraction with hydrocarbon solvents. The results are expressed in terms of hexane.

Insoluble Impurities in Fats and Oils (Ca 3a-46)
determines dirt, meal, and foreign substances that are insoluble in kerosene and petroleum ether.

Iodine Value of Fats and OilsCyclohexane Method (IV) (Cd Ib-87)
measures the unsaturation of fats and oils in terms of centigrams of iodine absorbed per gram of sample. The method is applicable to all normal fats that do not contain conjugated double bonds. It often is used to estimate the degree of hydrogenation of oils.

Melting PointCapillary Tube Method (Cc 1-25)
determines the temperature at which a sample of fat in a closed capillary becomes completely clear and liquid; broadly applicable; popular for tropical fats.

Melting PointMettler Dropping Point (Cc 18-80)
determines the temperature at which a sample becomes sufficiently fluid to flow in a specified apparatus.

Melting Point, Open Tube-Softening Point (Cc 3-25)
determines the temperature at which a solidified fat, in an open capillary tube, softens sufficiently to slip and rise to the top of the heating bath. This method is applicable to fats such as coconut oil, stearin, hydrogenated fats, and hard tallows. The results sometimes are reported as the "melting slip point," but the method is different from the AOCS Slip Point (Cc 4-25).

Melting PointWiley Method (Cc 2-38)
determines the temperature at which a sample disc of solidified fat assumes a spherical shape while suspended in a heating bath with an alcoholwater density gradient. A seldom used method, primarily replaced by the Mettler Dropping Point.

Moisture-Distillation Method (Ca 2a-45)
determines only moisture in triacylglycerols and emulsions by distillation with an immiscible solvent (toluene).

Moisture and Volatile MatterAir Oven Method (M&V) (Ca 2c-25)
determines the moisture and volatile matter by heating in a hot air oven. This method is applicable to animal and vegetable fats, but not to drying oils, coconut group fats, or oils with added monoacylglycerols.

Oil (Aa 4-38)
determines oil content in a dried sample of oil-bearing material by extraction with petroleum ether. This method is specific for cottonseed, which first must be fumed with hydrochloric acid to prevent oil adsorption to the fiber. Additional methods exist for other oilseeds.

Oxygen Stability Index (OSI) (Cd 12b-92)
measures the oxidation induction period of fat sample (essentially the time for a sample to exhaust is antioxidation properties) under conditions of the test.

Peroxide Value, Fats and Oils (PV) (Cd 8-53)
determines all substances, in terms of milliequivalents of peroxide per 1000 g of sample, that oxidize potassium iodide (KI). These substances generally are assumed to be peroxides or products of fat oxidation.

Phosphorus in Oils (Ca 13-55)
estimates the phospholipid content of crude, degummed, and refined vegetable oils in terms of phosphorus. Refineries often use induction coupled plasma (ICP) spectrographs to analyze divalent cations rapidly in aspirated crude oil. The calcium and magnesium measured are mainly responsible for nonhydratable phosphatides and are determined directly. An AOCS method for analysis by ICP does not exist yet.

Refining Loss, Vegetable Oils, Crude (Ca 9a-52)
determines the loss of free fatty acids and impurities when crude oils are refined under specified procedures.

Residual Lint (Aa 7-55)
determines the lint content of cottonseed by fuming (digesting) with hydrochloric acid.

Saponification Value of Fats and Oils (Cd 3-25)
determines the number of milligrams of (alcoholic) potassium hydroxide necessary to saponify a 1-g sample of a fat or an oil.

Schaal Test (Schaal Oven Method)
an accelerated test for determining the oxidative stability of a fat or a fat-containing food product. Results are reported as the time elapsed until a rancid odor is detected. Not an AOCS method; see American Association of Cereal Chemists' Method Manual. Modifications reported using OSI apparatus.

Smoke, Flash, and Fire PointCleveland Open Cup Method (Cc 9a-48)
determines the temperatures at which fats and oils smoke, flash, or burn. Smoke point determinations sometimes are used to follow degradation of frying oils with use.

Solid Fat Content of Fats and Oils by NMR (SFC) (Cd 16-81)
estimates the percent of solids in a semisolid fat on the basis of the pulsed nuclear magnetic resonance (NMR) signal of hydrogen in the liquid fraction. The method is used in the palm oil industry and widely throughout the world.

Solid Fat IndexDilatometric Method (SFI) (Cd 10-57)

estimates the percent of solids in a semisolid fat on the basis of changes in volume with temperature. This method utilizes glass dilatometers, and is the primary method in the United States.

Totox Value

an estimate of the degree of oxidation of a fat or oil, calculated as

$$\mathrm{totox} = 2 \times (\mathrm{PV}) + \mathrm{AV}.$$

Triglycerides by GLC (Ce 5b-86)

quantitatively determines triglycerides (triacylglycerols) in liquid vegetable oils in terms of molecular weight and degree of unsaturation as a function of their equivalent carbon number using high-pressure liquid chromatography.

Unsaponifiable Matter in Fats and Oils, Including Marine Oils (Ca bb-5,3)

determines substances dissolved in fats and oils that cannot be saponified (turned into sodium salts) by the usual caustic treatment, including higher aliphatic alcohols, sterols, pigments, and hydrocarbons. This method is not suitable for marine oils or feed grade fats.

8.13 References

Ege, S. N., Organic Chemistry, D.C. Heath and Company, Lexington, MA, 1984.

National Research Council, Recommended Dietary Allowances, 10th ed., National Academy Press, Washington, DC, 1989.

Foreign Agricultural Service. Oilseeds: Markets and Trade. FOP 08-01, 2001, U.S. Department of Agriculture. Washington, DC, August.

Johnson, L. A., and Meyers, D. J., "Industrial Uses for Soybeans," in Practical Handbook of Soybean Processing and Utilization, D. R. Erickson (Ed.), pp. 380427, AOCS Press, Champaign, Illinois, 1995.

AOCS Methods, Official Methods and Recommended Practices of the AOCS. AOCS Press, Champaign, Illinois, 2000.

National Research Council, Nutrient Requirements of Fish, National Academy Press, Washington, DC, 1993.

Robertson, J. A., Morrison, W. H., III, and Wilson R. L., Effects of Planting Location and Temperature on the Oil Content and Fatty Acid Composition of Sunflower Seeds, USDA Agric. Results ARR-S-3, U.S. Dept. Agriculture, Washington, DC, 1979.

Robertson, J. A., "Use of Sunflower Seed in Food Products," Crit. Rev. Food Sci. Nutr., 6, 201240 (1975).

Pryde, E. H., Fatty Acids, AOCS Press, Champaign, IL, 1979.

Pryde, E. H., Princen, L. H., and Mukherjee, K. D., New Sources of Fats and Oils, AOCS Press, Champaign, IL, 1981.

Cahn, R. S., Ingold, C. K., and Prelog, V., "The Specification of Asymmetric Configuration in Organic Chemistry," Experientia, 12, 8194 (1956).

Hirschman, H., "The Nature of Substrate Asymmetry in Stereoselective Reactions," J. Biol. Chem., 235, 27622767 (1960).

Nawar, W. W., "Lipids," in Food Chemistry, 3rd ed., O. R. Fennema (Ed.), pp. 225317, Marcel Dekker, Inc., New York, 1996.

Beckman, H. J., "Hydrogenation Practice." J. Am. Oil Chem. Soc., 60, 234A242A (1983).

Sevanian, A., Lipid Peroxidation in Biological Systems, AOCS Press, Champaign, IL, 1988.

Frankel, E. N., Lipid Oxidation, The Oily Press, Dundee, Scotland, United Kingdom, 1988.

Min, D. B. and Smouse, T. H. (Eds.), Flavor Chemistry of Lipid Foods, AOCS Press, Champaign, IL, 1989.

Nawar, W. W., "Lipids," in Food Chemistry, 3rd ed., O. R. Fennema (Ed.), pp. 225319, Marcel Dekker, Inc., New York, 1996.

Min, D. B., Lee, S. H., and Lee, E. C., "Singlet Oxygen Oxidation of Vegetable Oils," in Flavor Chemistry of Lipid Foods, D. B. Min and T. H. Smouse (Eds.), pp. 5797, AOCS Press, Champaign, IL, 1989.

Yurawecz, M. P., Mossoba, M. M., Kramer, J. K.G., Pariza, M. W., and Nelson, G. J. (Eds.), Advances in Conjugated Linoleic Acid Research, Vol. 1, AOCS Press, Champaign, Illinois. 1999.

Lampert, D., "Processes and Products of Interesterification," in Introduction to Fats and Oils Technology, 2nd ed., R. D. O'Brien, W. E. Farr, and P. J. Wan (Eds.), pp. 208234, AOCS Press, Champaign, Illinois, 2000.

Hagemann, J. W., Tallent, W. H., Barve. J. S., Ismail, L. A., and Gunstone, F. D., "Polymorphism in Single Acid Triglycerides of Positional and Geometric Isomers of Octadecanoic Acid," J. Am. Oil Chem. Soc., 52, 204207 (1975).

Rossell, J. B., "Phase Diagrams of Triglyceride Systems," in Advances in Lipid Research, Vol. 5, R. Paoletti and D. Kritchevsky (Eds.), pp. 353408, Academic Press, New York, 1967.

Small, D. M., Handbook of Lipid Research. The Physical Chemistry of Lipids: 4, Plenum Press, New York, 1986.

Bockish, M. Fats and Oils Handbook., AOCS Press, Champaign, IL, 1998.

Hagemann, J. W., "Thermal Behavior and Polymorphism of Acylglycerides," in Crystallization and Polymorphism of Fats and Fatty Acids, N. Garti and K. Sato (Eds.), pp. 995, Marcel Dekker, Inc., New York, 1988.

Hernqvst, L., "Chocolate Temper," in Industrial Chocolate Manufacture and Use, S. T. Beckett (Ed.), pp. 153171, AviVan Nostrand Reinhold, New York, 1988.

Hernqvst, L., "Crystal Structures of Fats and Fatty Adds," in Crystallization and Polymorphism of Fats and Fatty Acids, N. Garti and K. Sato (Eds.), pp. 97137, Marcel Dekker, Inc., New York, 1988.

Shipley, C. G., "X-Ray Crystallographic Studies of Aliphatic Lipids," in Handbook of Lipid Research. 4: The Physical Chemistry of Lipids, D. M. Small (Ed.), pp. 97147, Plenum Press, New York, 1986.

Sato, K., "Crystallization of Fats and Fatty Acids." in Crystallization and Polymorphism of Fats and Fatty Acids, N. Garti and K. Sato (Eds.), pp. 228263, Marcel Dekker, Inc., New York, 1988.

Widlak, N., Hartel, R., and Narine, S. (Eds.), Crystallization and Solidification Properties of Lipids, AOCS Press, Champaign, IL, 2001.

Poot, C., Dijkshoorn, W., Haigton, A. J., and Verburg, C. C., "Laboratory Separation of Crystals from Plastic Fats Using Detergent Solution," J. Am. Oil Chem. Soc., 52, 6972 (1975).

Rivarola, G., Anon, M. C., and Calvelo, A., "Crystallization of Waxes during Sunflowerseed Oil Refining," J. Am. Oil Chem. Soc., 62, 15081513 (1986).

Rivarola, G., Anon, M. C., and Calvelo, A., "Crystallization of Waxes in Sunflowerseed Oil," J. Am. Oil Chem. Soc., 65, 17711773 (1988).

Nichols, P. D., Nichols, D. S., and Bakes, M. J., "Marine Oil Products in Australia," INFORM, 5, 254261 (1994).

Cherry, J. P., and Kramer, W. H., "Plant Sources of Lecithin," in Lecithins: Sources, Manufacture & Uses, B. F. Szuhaj (Ed.), pp. 1631, AOCS Press, Champaign, IL, 1989.

Foreign Agricultural Service, "Oilseeds: World Markets and Trade," FAS Circular Series FOP 02-02. U.S. Department of Agriculture, Washington, DC, February, 2002.

Rudbeck, J., "Market Report 2000," Render, 30(2), 1015 (2001).

Technical CommitteeISEO, Food Fats and Oils, Institute of Shortening and Edible Oils, Inc., Washington, DC, 1988.

Nutrition Insight 5, Is Total Fat Consumption Really Decreasing?, USDA Center for Nutrition Policy and Promotion, U.S. Department of Agriculture, Washington, DC, April 1998.

Unidentified, "State of the Industry Report," Snack Food & Wholesale Bakery, 89(6), SI-3SI-74 (2000).

A Brief History of Olestra, Center for Science in the Public Interest (www.cspinet.org/olestra/history.html). Downloaded, March 28, 2002.

List, G. R., and Freidrich, J. P., "Oxidative Stability of Seed Oils Extracted with Supercritical Carbon Dioxide," J. Am. Oils Chem. Soc., 66, 98101 (1989).

Moore, H. N., "Oilseed Handling and Preparation Prior to Solvent Extraction," J. Am. Oil Chem. Soc., 60, 141A144A (1983).

Johnson, L. A., and Lusas, E. W., "Comparison of Alternative Solvents for Oils Extraction," J. Am. Oil Chem. Soc., 60, 181A194A (1983).

Galvin, J. B., "Toxicity Data for Commercial Hexane and Hexane Isomers," in Technology and Solvents for Extracting Oilseeds and Nonpetroleum Oils, P. J. Wan and P. J. Wakelyn (Eds.), pp. 7585, AOCS Press, Champaign, IL, 1997.

Wakelyn, P. J., "Regulatory Considerations for Extraction Solvents for Oilseeds and Other Nonpetroleum Oils," in Technology and Solvents for Extracting Oilseeds and Nonpetroleum Oils, P. J. Wan and P. J. Wakelyn (Eds.), pp. 4874, AOCS Press, Champaign, IL, 1997.

Sullivan, D. A., Campbell, B. D., Conway, M. F., and Grimsby, F. N., "Isopropanol Extraction of Oilseeds," Oil Mill Gazetteer, 87(10), 2428 (1982).

Baker, E. C., and Sullivan, D. A. "Development of a Pilot Plant Process for the Extraction of Soy Flakes with Aqueous Isopropyl Alcohol," J. Am. Oil Chem. Soc., 60, 12711277 (1983).

Lusas, E. W., and Hernandez, E., "Isopropyl Alcohol," in Technology and Solvents for Extracting Oilseeds and Nonpetroleum Oils, P. J. Wan and P. J. Wakelyn (Eds.), pp. 199266, AOCS Press, Champaign, IL, 1997.

Liener, I. E., "Implications of Antinutritional Components in Soybean Foods," Crit. Rev. Food Sci. Nutr., 34(1), 3167 (1994).

Martinez Suarez, J. M., and Mendoza, J. A., "Olive Oil Processing and Related Aspects: Extraction and Refining: General Aspect," in Proceedings World Conference on Emerging Technologies in the Fats and Oils Industry, A. R. Baldwin (Ed.), pp. 299305, AOCS Press, Champaign, IL, 1968.

Boskou, D., Olive Oil Chemistry and Technology, AOCS Press, Champaign, IL, 1996.

Firestone, D., Fedeli, E., and Emmons, E. W., "Olive Oil," in Bailey's Industrial Oil & Fat Products, 5th ed., Vol. 2, Edible Oil & Fat Products: Oils and Oilseeds, Y. H. Hui (Ed.), pp. 241270, John Wiley & Sons, Inc., New York, 1996.

Canapi, E. C., Agustin, Y. T.V., Moro, E. A., Pedrosa, E., Jr., and Bendaño, M. L. J., "Coconut Oil," in Bailey's Industrial Oil & Fat Products, 5th ed., Vol. 2, Edible Oil & Fat Products: Oils and Oilseeds, Y. H. Hui (Ed.) pp. 225254, John Wiley & Sons, Inc., New York, 1996.

Baisron, Y., "Palm Oil," in Bailey's Industrial Oil & Fat Products, 5th ed., Vol. 5, Industrial and Consumer Nonedible Products from Oils and Fats, Y. H. Hui (Ed.), pp. 271376, John Wiley & Sons, Inc., New York, 1996.

Tuft, L. S., "Rendering," in Bailey's Industrial Oil & Fat Products, 5th ed., Vol. 5, Industrial and Consumer Nonedible Products from Oils and Fats, Y. H. Hui (Ed.), pp. 120, John Wiley & Sons, Inc., New York, 1996.

Wellons, F. H., "Industrial Uses of Rendered Fats and Oils," in Bailey's Industrial Oil & Fat Products, 5th ed., Vol. 5, Industrial and Consumer Nonedible Products from Oils and Fats, Y. H. Hui (Ed.), pp. 2132, John Wiley & Sons, Inc., New York, 1996.

NRA, Pocket Information Manual: Buyer's Guide to Rendered Products, National Renderers Association, Inc., Alexandria, VA, 1993.

Piette, G., Hundt, M., Jacques, L., and Lapoint, M., "Influence of Extraction Temperature on Amounts and Quality of Rendered Chicken Fat Recovered from Ground or Homogenized Skin," Poultry Sci., 80, 496500 (2001).

S⊘bstad, G. E., "Animal Fats: The Effect of Process Technology on Fat Quality," in World Conference Proceedings: Edible Fats and Oils Processing: Basic Principles and Modern Practices, D. R. Erickson (Ed.), pp. 3136, AOCS Press, Champaign, IL, 1990.
Prokoph, W. H., "Rendering Systems for Processing Animal By-Product Materials," J. Am. Oil Chem. Soc., 62, 805811 (1985).
Love, J. A., "Animal Fats," in Bailey's Industrial Oil & Fat Products, 5th ed., Vol. 1, Edible Oil & Fat Products: General Applications, Y. H. Hui, (Ed.), pp. 118, John Wiley & Sons, Inc., New York, 1996.
Schottelcotte, W., "State of the Art Grease Processing," Render, 24(3), 10, 11, 21 (1996).
Lusas, E. W., and Riaz, M. N., "Fats in Feedstuffs and Pet Foods," in Bailey's Industrial Oil & Fat Products, 5th ed., Vol. 3, Edible Oil & Fat Products: Products and Application Technology, Y. H. Hui, (Ed.), pp. 255329, John Wiley & Sons, Inc., New York, 1996.
Rudbeck, J., "Market Report 2000," Render, 30(2), 1012, 1315 (2001).
Bimbo, A. P., and Crowther, J. B., "Marine Oils: Fishing for Industrial Uses," INFORM, 3, 988996, 9981001 (1992).
Pigott, G. M., "Marine Oils," in Bailey's Industrial Oil & Fat Products, 5th ed., Vol. 3, Edible Oil & Fat Products: Products and Application Technology, Y. H. Hui (Ed.), pp. 225254, John Wiley & Sons, Inc., New York, 1996.
Burnham, F., "Alaska's Billion Pounds of Protein," Render, 17(5), 812 (1989).
Bimbo, A. P., "The Emerging Marine Oil Industry," J. Am. Oil Chem. Soc., 64, 706, 708715 (1987).
NOPA, Yearbook & Trading Rules 20002001, National Oilseed Processors Association, Washington, DC, 2000.
Farr, W. E., "Refining of Fats and Oils," in Introduction to Fats and Oils Technology, 2nd ed., R. D. O'Brien, W. E. Farr, and P. J. Wan (Eds.), pp. 136157, AOCS Press, Champaign, Illinois, 2000.
List, G. R., "Commercial Manufacture of Lecithin," in Lecithins: Sources, Manufacture and Uses, B. F. Szuhaj (Ed.), pp. 145161, AOCS Press, Champaign, IL, 1988.
Erickson, D. R., "Degumming and Lecithin Processing and Utilization," in Practical Handbook of Soybean Processing and Utilization, D. R. Erickson (Ed.), pp. 174183, AOCS Press, Champaign, IL, 1995.
Segers, J. C., and van de Sande, R. L. K. M., "DegummingTheory and Practice," in World Conference Proceedings, Edible Fats and Oils Processing: Basic Principles and Modern Practices, D. R. Erickson (Ed.), pp. 8893, AOCS Press, Champaign, IL, 1990.
Dijkstra, A. J., "Degumming, Refining, Washing and Drying Fats and Oil," in Proceedings of the World Conference on Oilseed Technology and Utilization, T. H. Applewhite (Ed.), pp. 138151, AOCS Press, Champaign, IL, 1993.
Dahlke, K., and Eichelsbacher, M., "EnzyMax™ and Alcon™Lurgi's Route to Physical Refining," in Emerging Technologies, Current Practices, Quality Control, Technology Transfer, and Environmental Issues, Vol. 1, S. S. Koseoglu, K. C. Rhee, and R. F. Wilson (Eds.), pp. 5359, AOCS Press, Champaign, IL, 1998.
Erickson, D. R., "Neutralization," in Practical Handbook of Soybean Processing and Utilization, D. R. Erickson (Ed.), pp. 184202, AOCS Press, Champaign, IL, 1995.
Hernandez, E., "Latest Developments in Refining and Filtration of Specialty Oils with Sodium Silicates," Texas A&M University, College Station, TX. Paper presented at American Oil Chemists' Society Annual Meeting, Minneapolis, Minnesota, May 1317, 2001.
Lin, L., Rhee, K. C., and Koseoglu, S. S., "Recent Progress in Membrane Degumming of Crude Vegetable Oils on a Pilot-Plant Scale," in Advances in Oils and Fats, Antioxidants, and Oilseed By-Products, S. S. Koseoglu, K. C. Rhee, and R. F. Wilson (Eds.), pp. 7682, AOCS Press, Champaign, IL, 1998.
Patterson, H. W.B., Bleaching and Purifying Fats and Oils, AOCS Press, Champaign, IL, 1992.
Taylor, D. R., "Adsorptive Purification," in Proceedings of the World Conference on Oilseed Technology and Utilization, T. H. Applewhite (Ed.), pp. 152165, AOCS Press, Champaign, IL, 1993.
Zschau, W., "Bleaching," in Introduction to Fats and Oils Technology, 2nd ed., R. D. O'Brien, W. E. Farr, and P. J. Wan (Eds.), pp. 158178, AOCS Press, Champaign, IL, 2000.
O'Brien, R. D., "Shortenings," in Practical Handbook of Soybean Processing and Utilization, D. R. Erickson (Ed.), pp. 363379, AOCS Press, Champaign, IL, 1995.
List, G. R., Steidley, K. R., Palmquist, D., and Adlof, R. O., "Solid Fat Index vs. Solid Fat Content: A Comparison of Dilatometry and Pulsed Nuclear Magnetic Resonance for Solids in Hydrogenated Soybean Oil," in Crystallization and Solidification Properties of Lipids, N. Widlak, R. Hartel, and S. Narine (Eds.), pp. 146152, AOCS Press, Champaign, IL, 2001.
O'Brien, R. D., "Liquid Oils Technology," in Introduction to Fats and Oils Technology, 2nd ed., R. D. O'Brien, W. E. Farr, and P. J. Wan (Eds.), pp. 463485, AOCS Press, Champaign, IL, 2000.
Tan, B. K., and Fling Oh, C. H., PORIM Techno. Palm Oil Res. Inst. Malaysia No. 4, 1981.
Tirtaux, A., "Dry Fractionation: A Technique and an Art," in World Conference Proceedings: Edible Fats and Oils Processing: Basic Principles and Modern Practices, D. R. Erickson (Ed.), pp. 136141, AOCS Press, Champaign, IL, 1990.
Basiron, Y., "Palm Oil," in Bailey's Industrial Oil & Fat Products, 5th ed., Vol. 2, Edible Oil & Fat Products: Oils and Oilseeds, Y. H. Hui (Ed.), pp. 271377, John Wiley & Sons, Inc., New York, 1996.
Kellens, M., and Hendrix, M., "Fractionation," in Introduction to Fats and Oils Technology, 2nd ed., R. D. O'Brien, W. E. Farr, and P. J. Wan (Eds.), pp. 194209, AOCS Press, Champaign, IL, 2000.
Tirtiaux, A., and Gibon, V., "Dry Fractionation: The Boost Goes On," in Emerging Technologies, Current Practices, Quality Control, Technology Transfer, and Environmental Issues, S. S. Koseoglu, K. C. Rhee, and R. F. Wilson (Eds.), pp. 9298, AOCS Press, Champaign, IL, 1998.
Davidson, H. F., Campbell, E. J., Bell, R. J., and Pritchard, R., "Sunflower Oil," in Bailey's Industrial Oil & Fat Products, 5th ed., Vol. 2, Edible Oil & Fat Products: Oils and Oilseeds, Y. H. Hui (Ed.), pp. 603689, John Wiley & Sons, Inc., New York, 1996.
Kaylegian, K. E., and Lindsay, R. C., Handbook of Milkfat Fractionation Technology and Applications, AOCS Press, Champaign, IL, 1995.
Hastert, R. C., "Cost/Quality/Health: The Three Pillars of Hydrogenation," in World Conference Proceedings: Edible Fats and Oils Processing: Basic Principle and Modern Practices, D. R. Erickson (Ed.), pp. 142152, AOCS Press, Champaign, IL, 1990.
Ariaansz, R. F., "Hydrogen Theory," in Proceedings of the World Conference on Oilseed Technology and Utilization, T. H. Applewhite (Ed.), pp. 166172, AOCS Press, Champaign, IL, 1993.
Kokken, M. J., "Hydrogenation in Practice," in Proceedings of the World Conference on Oilseed Technology and Utilization, T. H. Applewhite (Ed.), pp. 173179, AOCS Press, Champaign, IL, 1993.
Patterson, H. W.B., Hydrogenation of Fats and Oils, AOCS Press, Champaign, IL, 1994.
Erickson, D. R., and Erickson, M. D., "Hydrogenation and Base Stock Formulation," in Practical Handbook of Soybean Processing and Utilization, D. R. Erickson (Ed.), pp. 218238, AOCS Press, Champaign, IL, 1995.
Ariaansz, R. F., and Okonek, D. V., "Isomer Control During Edible Oil Processing," in Emerging Technologies, Current Practices, Quality Control, Technology Transfer, and Environmental Issues, S. S. Koseoglu, K. C. Rhee, and R. F. Wilson (Eds.), pp. 7791, AOCS Press, Champaign, IL, 1998.
Hastert, R. C., "Hydrogenation," in Introduction to Fats and Oils Technology, 2nd ed., R. D. O'Brien, W. E. Farr, and P. J. Wan (Eds.), pp. 179193, AOCS Press, Champaign, IL, 2000.
Okonek, D., "Hydrogenated Oil End Products," in Vegetable Oils Processing Practical Short Course Manual, E. Hernandez, S. J. Rathbone, and K. C. Rhee (Eds.), Texas A&M University, College Station, TX, 2001.
O'Brien, R. D., "Shortening Technology," in Fats and Oils Technology, 2nd ed., R. D. O'Brien, W. E. Farr, and P. J. Wan (Eds.), pp. 179193, AOCS Press, Champaign, IL, 2000.
Federal Register, "Food labeling: *trans* Fatty Acids in Nutrition Labeling, Nutrient Content Claims, and Health Claims; Proposed Rule," Federal Register 64 (221), 6274562825, November 17, 1999.
Kritchevsky, D., "*Trans* Unsaturated Fat in Health in Nutrition and Health," in World Conference Proceedings, Edible Fats and Oils Processing: Basic Principles and Modern Practices, D. R. Erickson (Ed.), pp. 158165, AOCS Press, Champaign, IL, 1990.
Going, L. H., "Interesterification Products and Processes," J. Am. Oil Chem. Soc., 44, 414A, 416A, 418A, 420A, 422A, 454A456A (1969).
Akoh, C. C., "Reduced-calorie Fats and Fat Replacement," in Emerging Technologies, Current Practices, Quality Control, Technology Transfer, and Environmental Issues, S. S. Koseoglu, K. C. Rhee, and R. F. Wilson (Eds.), pp. 223234, AOCS Press, Champaign, IL, 1998.
Fomuso, L. B., and Akoh, C. C., "Enzymatic Modification of Triolein: Incorporation, of Caproic and Butyric Acids to Produce Reduced-Calorie Structured Lipids," J. Am. Oil Chem. Soc., 3, 269272 (1997).
Sonntag, N. V., "Fat Splitting, Esterification, and Interesterification," in Bailey's Industrial Oil and Fat Products, 4th ed., Vol. 2, D. Swern (Ed.), pp. 97173, John Wiley & Sons, Inc., New York, 1982.
Hernandez, E., and Lusas, E. W., "Trends in Transesterification of Cottonseed Oil," Food Technol., 51(5), 7276 (1977).
Sreenivasan, B., "Interesterification of Fats," J. Am Oil Chem. Soc., 55, 798805 (1978).
Rozendaal, A., "Interesterification of Oils and Fats," in World Conference Proceedings: Edible Fats and Oils Processing: Basic Principles and Modern Practices, D. R. Erickson (Ed.), pp. 152157, AOCS Press, Champaign, IL, 1990.

Kubow, S., "Triglyceride Structure and Lipoprotein Metabolism," in Structural Modified Food Fats: Synthesis, Biochemistry, and Use, A. B. Christophe (Ed.), pp. 149159, AOCS Press, Champaign, IL, 1998.

Kritchevsky, D., "Triglyceride Structure and Atherosclerosis," in Structural Modified Food Fats: Synthesis, Biochemistry, and Use, A. B. Christophe (Ed.), pp. 183188, AOCS Press, Champaign, IL, 1998.

Kellens, M., and De Greyt, W., "Deodorization," in Introduction to Fats and Oils Technology, 2nd ed., R. D. O'Brien, W. E. Farr, and P. J. Wan (Eds.), pp. 235268, AOCS Press, Champaign, IL, 2000.

Zehnder, C. T., "Deodorization," in Practical Handbook of Soybean Processing and Utilization, D. R. Erickson (Ed.), pp. 239257, AOCS Press, Champaign, IL, 1995.

Stenberg, O., and Sjöberg, P. "A Novel Deodorization Plant Design for Seed Oils," in Emerging Technologies, Current Practices, Quality Control, Technology Transfer, and Environmental Issues, S. S. Koseoglu, K. C. Rhee, and R. F. Wilson (Eds.), pp. 103108, AOCS Press, Champaign, IL, 1998.

Hymowitz, T., "On the Domestication of the Soybean," Econ. Bot., 24, 408421 (1970).

Hymowitz, T., and Newell, C. A., "Taxonomy of the Genus Glycine, Domestication and Uses of Soybeans," Econ. Bot., 35, 272288 (1981).

Hymowitz, T., and Harlan, J. R., "Introduction of Soybean to North America by Samuel Bowen in 1765," Econ. Bot., 37, 371379 (1983).

Hymowitz, T., "Introduction of the Soybean to Illinois," Econ. Bot., 41, 2832 (1987).

Britannica CD, Encyclopedia Britannica CD 99 Multimedia Edition, Encyclopedia Britannica, Chicago, IL, 1999.

Bowdidge, E., The Soya Bean: Its History, Cultivation (in England) and Uses, Oxford University Press, London, 1935.

Markley, K. S., and Goss, W. H., Soybean Chemistry and Technology, Chemical Publishing Company, Inc., Brooklyn, NY, 1944.

Staff, "Continuous Solvent Extraction: The Early Beginnings of a Giant Industry," J. Am. Oil Chem. Soc., 54, 202A207A (1977).

Govier, G. E., "European Methods of Crushing Soya Beans," Oil and Fat Industries, 5, 251254 (1928).

Burtis, E. L., "World Soybean Production and Trade," in Soybeans and Soybean Products, Vol. 1, K. L. Markley (Ed.), pp. 61117, Interscience Publishers, Inc., New York, 1950.

Markley, K. S. (Ed.), Soybeans and Soybean Products, Vols. 1 and 2, Interscience Publishers, Inc., New York, 1951.

Goss, W. H., The German Oilseed Industry, Hobart Publishing Company, Washington, DC, 1947.

Jones, L. A., and King, C. C., "Cottonseed Oil," in Bailey's Industrial Oil & Fat Products, 5th ed., Vol. 2, Edible Oil & Fat Products: Oils and Oilseeds, Y. H. Hui, (Ed.), pp. 159240, John Wiley & Sons, Inc., New York, 1996.

Michael Eskin, N. A., McDonald, B. E., Przybylski, R., Malcolmson, L. J., Scarth, L. J., Mag, T., Ward, K., and Adolph, D., "Canola Oil," in Bailey's Industrial Oil & Fat Products, 5th ed., Vol. 2, Edible Oil & Fat Products: Oils and Oilseeds, Y. H. Hui (Ed.), pp. 197, John Wiley & Sons, Inc., New York, 1996.

Young, C. T., "Peanut Oil," in Bailey's Industrial Oil & Fat Products, 5th ed., Vol. 2, Edible Oil & Fat Products: Oils and Oilseeds, Y. H. Hui (Ed.), pp. 377392, John Wiley & Sons, Inc., New York, 1996.

Gregory, III, J. F., "Vitamins," in Food Chemistry, 3rd ed., O. R. Fennema (Ed.), pp. 531617, Marcel Dekker, Inc., New York, 1996.

Packer, L., Traber, M. G., and Xin, W. (Eds.), Proceedings of the International Symposium on Natural Antioxidants: Molecular Mechanisms and Health Effects, AOCS Press, Champaign, IL, 1996.

Shahidi, F. (Ed.), Natural Antioxidants: Chemistry, Health Effects, and Applications, AOCS Press, Champaign, IL, 1997.

Becher, P., and Schick, M. J., "Microemulsions," in Nonionic Surfactants: Physical Chemistry, M. J. Schick (Ed.), pp. 435491, Marcel Dekker, Inc., New York, 1987.

Artz, W. E., "Emulsifiers," in Food Additives, A. L. Branen, P. M. Davidson, and S. Salminen, (Eds.), pp. 347417, Marcel Dekker, Inc., New York, 1990.

O'Brien, R. D., Fats and Oils: Formulating and Processing for Applications, Technomic PressCRC Publishing Co., Boca Raton, FL, 1998.

Fritsch, C. W., "Measurements of Frying Fat Deterioration," J. Am. Oil Chem. Soc., 58, 272274 (1981).

Perkins, E. G., and Erickson, M. D., Deep Frying: Chemistry, Nutrition, and Practical Applications, AOCS Press, Champaign, IL, 1996.

Banks, D. E., and Lusas, E. W., "Oils and Industrial Frying," in Snack Foods Processing, E. W. Lusas and L. W. Rooney (Eds.), pp. 137104 Technomic PressCRC Publishing Co., Boca Raton, FL, 2001.

Evans, C. D., Warner, K., List, G. R., and Cowan, J. C., "Room Evaluation of Oils and Cooking Fats," J. Am. Oil Chem. Soc., 49, 578582 (1972).

Moulton, K. J., Beal, R. E., Warner, K., and Bounty, K. B., "Flavor Evaluation of Copper-nickel Hydrogenated Soybean Oil and Blends with Unhydrogenated Oil," J. Am. Oil Chem. Soc., 52, 469472 (1975).

Mounts, T. L., Warner, K., and List, G. R., "Performance Evaluation of Hexane-extracted Oils from Genetically-modified Soybeans," J. Am. Oil Chem. Soc., 71, 16161 (1994).

Tompkins, C., and Perkins, E. G., "Frying Performance of Low-linolenic Soybean Oils," J. Am. Oil Chem. Soc., 72, 223229 (2000).

Firestone, D., "Regulation of Frying Fat and Oil," in Deep Frying: Chemistry, Nutrition, and Practical Applications, E. G. Perkins and M. D. Erickson (Eds.), pp. 323334, AOCS Press, Champaign, IL, 1996.

Garti, N., and Sato, K., Eds., Crystallization and Polymorphism of Fats and Fatty Acids, Marcel Dekker, Inc., New York, 1988.

Bumbalough, J., "Margarine Types and Preparation Technology," in Fats and Oils Technology, 2nd ed., R. D. O'Brien, W. E. Farr, and P. J. Wan (Eds.), pp. 452462, AOCS Press, Champaign, IL, 2000.

Wainwright, B., "Specialty Fats and Oils," in Introduction to Fats and Oils Technology, 2nd ed., R. D. O'Brien, W. E. Farr, and P. J. Wan (Eds.), pp. 486514, AOCS Press, Champaign, IL, 2000.

Johnson, R. W., and Fritz, E. (Eds.), Fatty Acids in Industry, Marcel Dekker, Inc., New York, 1988.

Zoebelein, H., "Renewable Resources for the Chemical Industry," INFORM, 3, 721725 (1992).

Johnsen, P. B., "Emerging Industrial Uses for Soybeans," in Emerging Technologies, Current Practices, Quality Control, Technology Transfer, and Environmental Issues, S. S. Koseoglu, K. C. Rhee, and R. F. Wilson (Eds.), pp. 109112, AOCS Press, Champaign, IL, 1998.

Derksen, J. T.P., and Cuperus, F. P., "Processing of Oilseed Constituents for Nonfood Applications," in Emerging Technologies, Current Practices, Quality Control Technology Transfer, and Environmental Issues, S. S. Koseoglu, K. C. Rhee, and R. F. Wilson, Eds., pp. 113120, AOCS Press, Champaign, IL, 1998.

Foglia, T. A., Nelson, L. A., Marmer, W. N., Knothe, G. H., Dunn, R. O., and Bagby, M. O., "Improving the Properties of Vegetable Oils and Fats for Use as Biodiesel," in Emerging Technologies, Current Practices, Quality Control, Technology Transfer, and Environmental Issues, S. S. Koseoglu, K. C. Rhee, and R. F. Wilson, Eds., pp. 121125, AOCS Press, Champaign, IL, 1998.

Krawczyk, T., "Biodiesel," INFORM, 7, 800801, 804808 810824 (1996).

Lee, I., Johnson, L. A., and Hammond, E. G., "Reducing the Crystallization Temperature of Biodiesel by Winterizing Methyl Soyate," J. Am. Oil Chem. Soc., 73, 631636 (1996).

Lee, I., Johnson, L. A., and Hammond, E. G., "Use of Branched-chain Esters to Reduce the Crystallization Temperature of Biodiesel," J. Am. Oil Chem. Soc., 72, 11551160 (1995).

< previous page

next page >

9.1 Introduction

Sugar and starch are among the most abundant plant products available, and huge industries exist worldwide to extract and process them from agricultural sources. The world production of sugar (sucrose from cane and beet) in 1999/2000 was 136 million metric tons, raw value, with 27.4 percent of that being beet sugar and 72.6 percent being cane sugar.1 The proportion of beet sugar to cane sugar has fallen steadily since about 1971, when it constituted 42.8 percent of total sugar production. The total production of sugar has also risen dramatically since then, when it was 71.7 million tons in 1971/72.1

The primary use of sugar is as a sweetener in the manufacture of food and in household use. When used for such purposes, most of it is highly refined or purified, but considerable quantities are consumed in some areas of the world as a crude product, as in India where as much as 29 percent production is consumed as uncentrifuged sugar (gur, also known as jaggery, and khandsari) or as cane juice.2 Sugar is used to a limited extent in the production of other chemicals, such as sucrose esters, and in the form of by-product molasses as a substrate for fermentation processes and to produce alcohol, both for consumption and power, usually mixed with gasoline.

Starch is widely used in the textile and paper industries, as well as for food. A major use of starch is its hydrolytic conversion to glucose and enzymatic conversion to fructose for use as a sweetener in the food industry. Technological developments and governmental price-support programs for the domestic (United States) cane and beet sugar industries begun in the mid-1970s greatly magnified the importance of starch as a source of nutritive sweeteners.3 Starch-derived sweeteners compete directly and successfully with sugar in the United States, and increasingly in other countries.

The sugar industry has been characterized by steady, small incremental improvements in technology and production over the years, with subsequent improvements across the process, from field practices to new products.4

9.2 Sugar

History

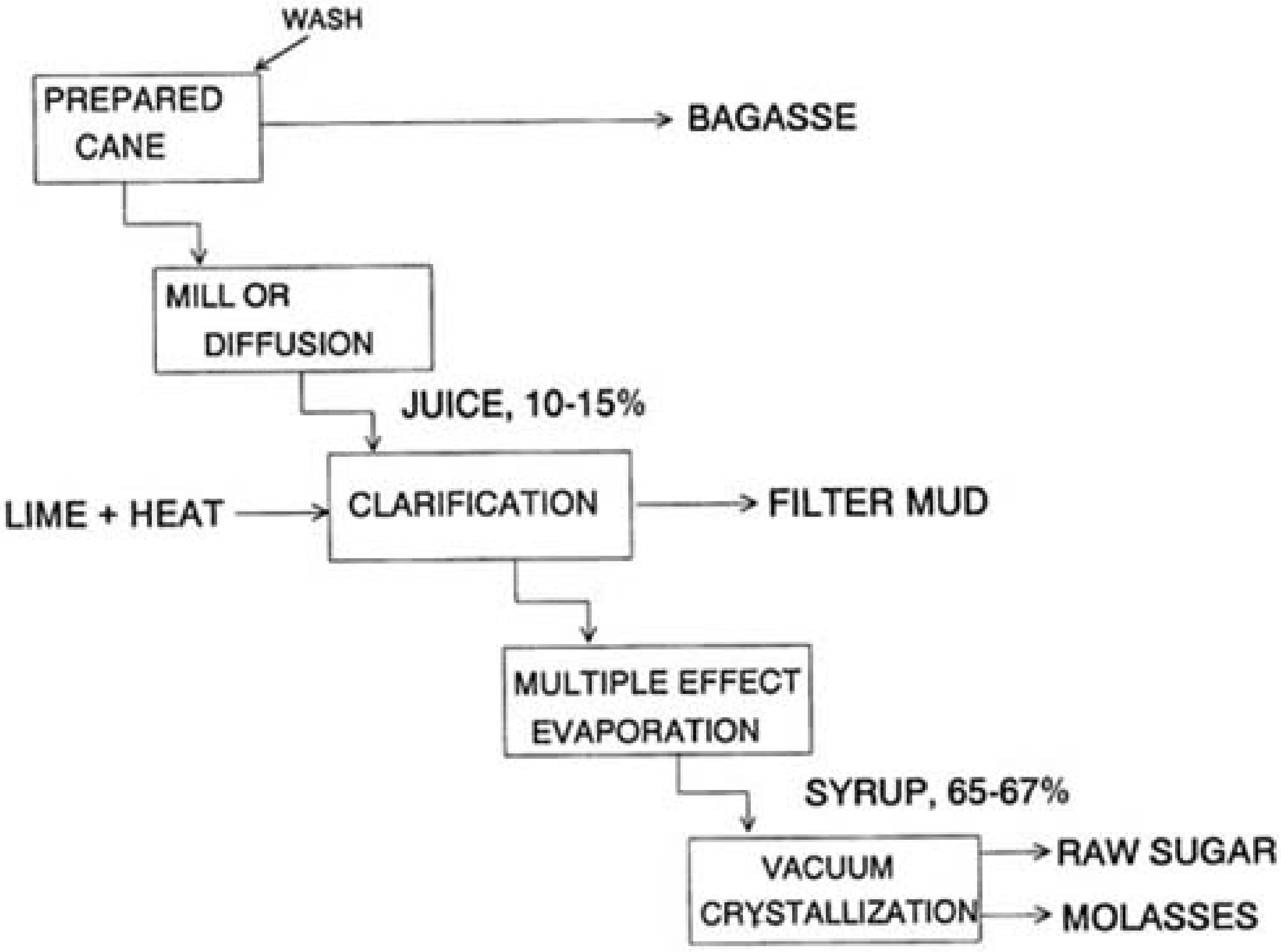
The ancestry of the sugar cane and its use as a food have been traced to the island of New Guinea. (An interesting legend related by J. A. C. Hugill[5] associates sugar cane with the origin of the human race.) Around 8000 B.C., the plant started on its migration from New Guinea to the many areas of southeastern Asia, Indonesia, the Philippines, Malay, Indochina, and eastern India, with humans probably acting as its main dispersal agent.[6] It was in Bengalese India that sugarcane first was cultivated as a field crop and the juice manufactured into various solid forms. A general knowledge of sugar was prevalent throughout India by 400 B.C. By the A.D. tenth century, sugar cultivation and manufacturing had become important industries in Persia and Egypt. The early Islamic movement spread the knowledge of the sugar industry throughout the Mediterranean area. On the second voyage of Columbus to America, in 1493, sugarcane was introduced in Santo Domingo. It spread rapidly through the West Indies and Central America. Cortez brought cane to Mexico, and Pizarro introduced it in Peru. By 1600, the sugar industry was the largest industry in tropical America.[7]

The modern sugar industry dates from the end of the eighteenth century, when steam replaced animal energy and made possible the development of larger and more efficient production units. The vacuum pan appeared in 1813, bag filters in 1824, multiple-effect evaporators in 1846, filter presses in 1850, centrifugals in 1867, dryers in 1878, and packaging machines in 1891.[8] Cultivation of the sugar beet plant and the manufacture of sugar from the beet developed in the industrial nations of Europe during the eighteenth and nineteenth centuries. In 1747, the German chemist, Marggraf, established that sugar from beets was the same as sugar from cane. His pupil, Achard, in 1799, demonstrated that sugar can be commercially prepared from beets. During the Napoleonic wars, a short-lived beet sugar industry was established in France.[9] Today, beet is the major source of sugar in Europe.

9.3 Raw Sugar Production

Traditionally, raw sugar is produced at the cane mill and refined sugar is produced at a refinery, usually far removed from the source of the raw sugar. Figure 9.1 is a generalized flow diagram for the manufacture of raw cane sugar. The several steps are described in the sections that follow.

Fig. 9.1 Generalized flow diagram of steps in raw cane sugar manufacture. (*Adapted from Clarke*.38)



Agriculture

The sugarcane is a large perennial tropical grass belonging to the genus Saccharum. Three basic species, *S. officinarum, S. robustum,* and *S. spontaneum,* make up the sugarcane complex, with a large number of varieties. Sugarcane is propagated commercially by cuttings, each cutting consisting of portions of the cane plant having two or more buds, or "eyes." The buds sprout into shoots from which several other shoots arise below the soil level to form a clump of stalks known as a "stool." From 12 to 20 months are required for the crop to mature from new plantings and about 12 months for ratoon crops (i.e. cane stalks arising from stools that have been previously harvested). Most of the field operations have been mechanized, but planting is still done by hand in some cane-producing areas. Fields are replanted after two to five or more ratoons (cuttings made from the original planting.). As cane grows, its foliage is largely limited to the upper one-third of the plant (the lower leaves die off from shading as the stalk grows taller). The cane stalk is round and jointed, and grows to more than 10 ft when mature (Fig. 9.2). It is covered with a hard rind that varies in color from light brown, green, yellowish green, purple, or striped, depending on the variety. There is usually a fine film of waxy material on the surface of the rind, heavier in some varieties than others. The stalk is divided into a series of internodes, from 3 to 10 in. in length, separated by nodes. The inter-nodes contain the pithy material from which most of the sucrose-containing cane juice is obtained. The nodes, somewhat woody in nature, contain one bud or "eye" on alternating sides, which sprout when planted.

Fig. 9.2 Sugarcane in the field.

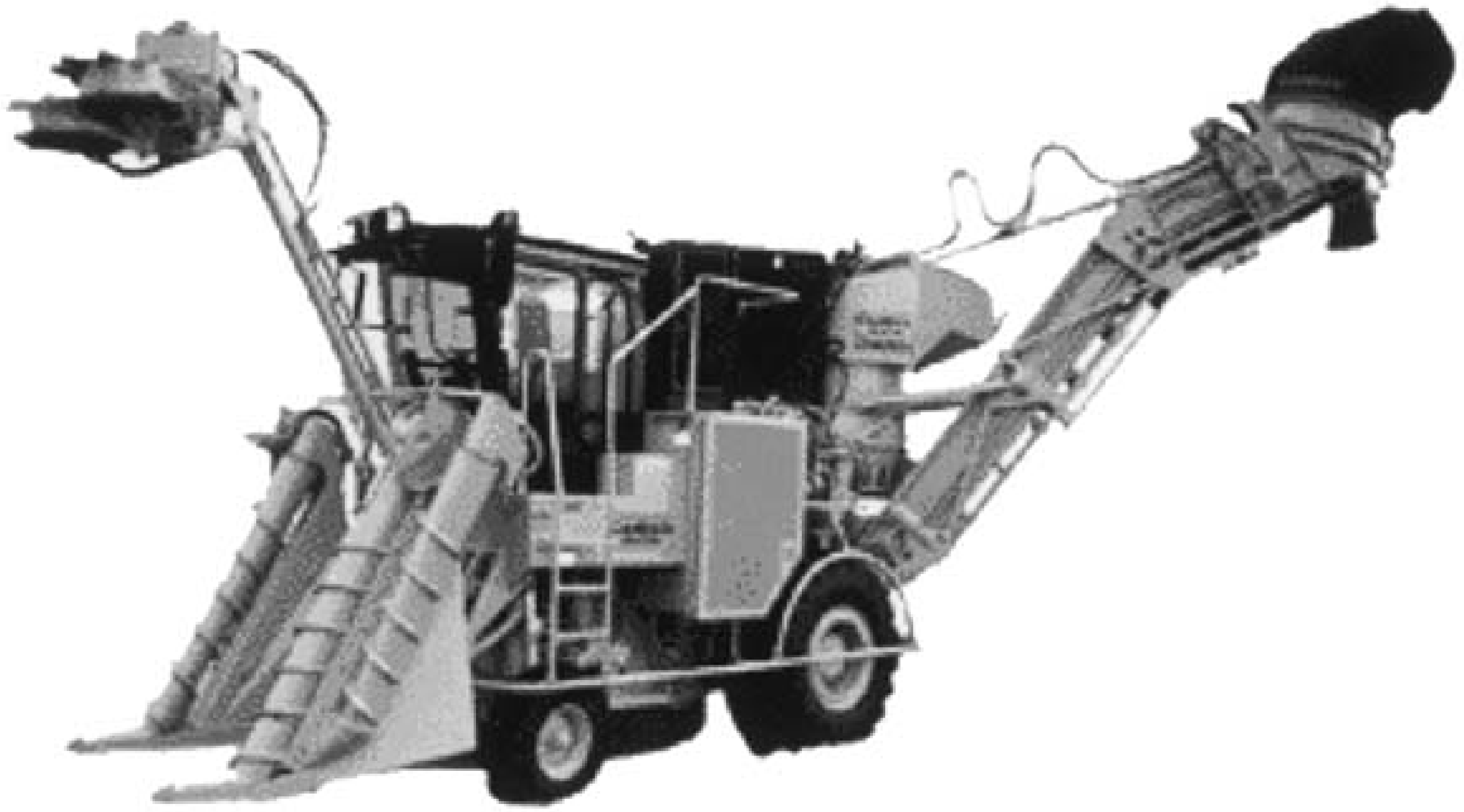
Breeding programs throughout the world develop cane varieties that are suited to their particular climatic conditions, which give the highest yield, balanced with disease resistance and other desirable agronomic traits. In Louisiana, for example, cane varieties are bred for cold tolerance, among other traits, because of the short growing season and the possibility of freezing weather during the harvest. Varieties retain vigor for 1015 years and then need to be replaced by new varieties.

There is research ongoing in genetic manipulation of sugarcane for herbicide and disease resistance, but no commercial varieties exist to date.

Harvesting

Sugarcane is cut by hand with machete-type knives in a few producing areas. In hand-cutting, the tops and leaves are removed, and the cane is cut at ground level. In more industrialized areas, cane is harvested by machines. Two types of machine harvesting are in widespread use: whole-stalk harvesters, also known as soldier harvesters, and chopper harvesters, also known as combine harvesters, in which the cane is cut into smaller pieces, known as billets, about 812 in. in length. Combine harvesters exist as single-row (Fig. 9.3) and two-row models (Fig. 9.4). The whole stalk cane may be transported immediately to the mill where it is dumped in large cane yards for future processing, or left in the field for future pick-up. Billeted cane should always be processed as soon as possible because of the faster deterioration rate of the juice quality caused by the shorter pieces. In many areas, cane is burned before harvesting to help remove leaves and trash. The practice of burning has become less environmentally and socially acceptable in recent years, and, in spite of many efforts by the industry to control adverse effects, the practice is in decline. The practice of harvesting unburned cane is called green cane harvesting, and it introduces a host of challenges to the processor because of the heavier load of impurities, such as color-producing molecules and polysaccharides, that are expressed into the juice.10,11 Cane may be transported to the mill by railcars, trucks, trailers and carts.

Fig. 9.3 Cameco single-row harvester.



A new development in harvest control for better quality of cane (and more so, beet) has been the use of global positioning systems (GPSs) based on satellite navigation facilities.12

Fig. 9.4 Cameco two-row harvester.

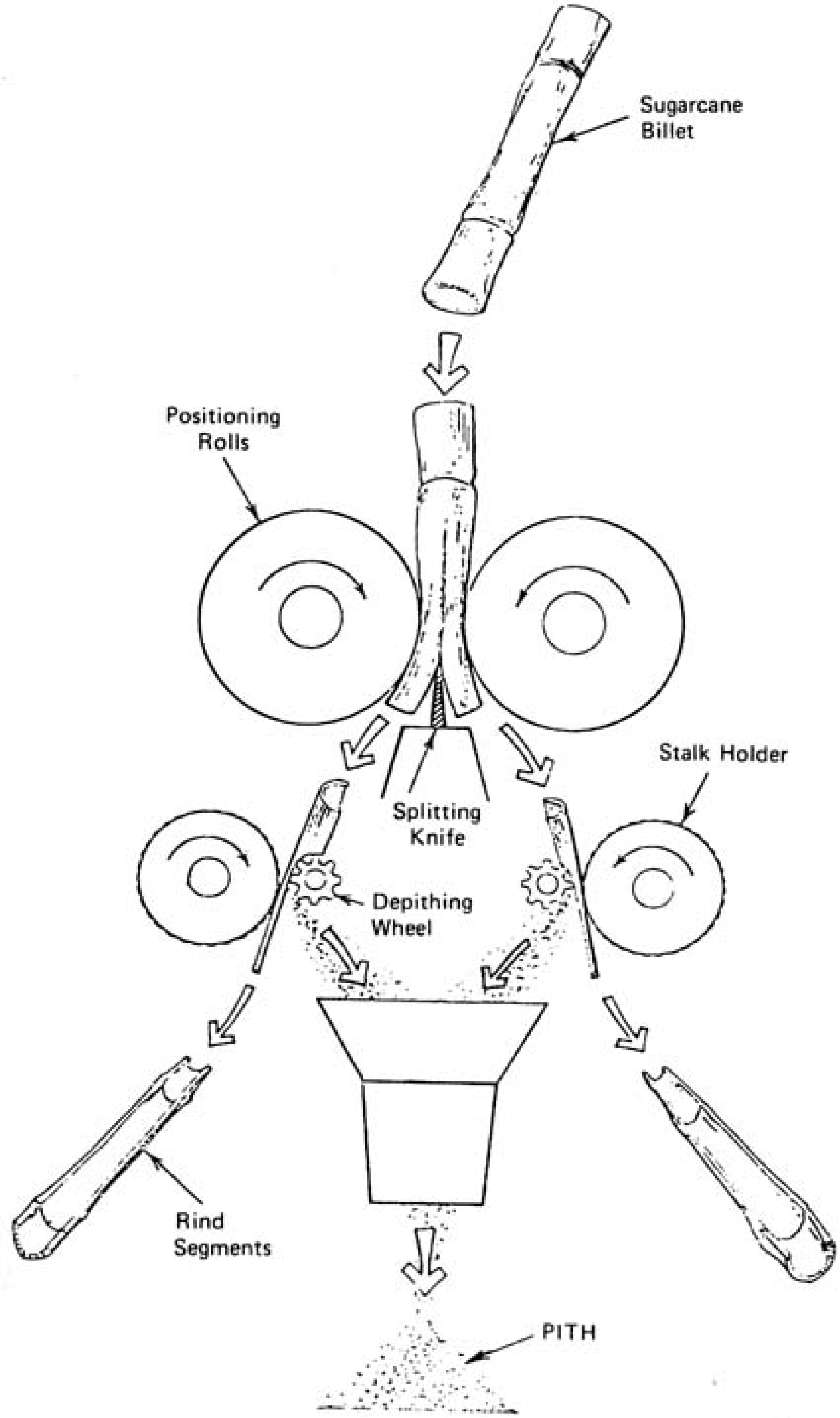
Preparation of Cane for Milling

Because of the presence of adhering soil, rocks, and field trash, mechanically harvested cane usually must be washed before milling. In some carefully controlled situations, when the weather has been dry, washing can be eliminated. Washing systems range from a simple spraying with warm water on the carrier or table to an elaborate system consisting of conveyors with water jets, stripping rolls, and baths for the removal of stones. In most mills, a biocide is added to prevent bacterial growth in the cane. Mill sanitation is important to control bacterial activity, which can result in direct sucrose loss as bacteria consume the sugar, and in the production of undesirable by-products, such as dextran, a polysaccharide that has deleterious effects in the production of sugar. With the advent of billeted cane, more sucrose can be lost in the wash water due to the multitude of cut ends, resulting in some reduction in yield.

The cane is prepared for grinding/milling by knifing or shredding or a combination of the two to chop the cane into smaller pieces. It is important for the mill to pay attention to good cane preparation, since this is one of the most important parameters affecting extraction of the juice.13 The Unigator, developed in the late 1970s, combines the action of the chopper and the shredder into a single unit.14 The advantages gained are a greater increase in juice extraction by the mills, an increase in the grinding rate, and, in some cases, a decrease in power consumption by the milling operation. The preparation index (PI) is a measure of how well the cane has been prepared for juice extraction, and represents the percentage of cane plant cells that are ruptured by the cane preparation equipment. It is measured as brix (percent of sucrose) in the ruptured cells as a proportion of the total brix in the cane.

A novel approach to the preparation of sugar cane for the extraction of sugar was demonstrated in a machine called a Tilby separator. Cane stalks, cut into 812 in. segments, are fed into the separator, split in half, depithed by a depithing wheel, resulting in a clean, sucrose-containing pith and rejected rind segments (Fig. 9.5). This approach has not been adopted on any large scale, but is used in some tropical countries to produce edible cane juice and also has some uses for research purposes.

Fig. 9.5 Tilby separator process. (*Courtesy Ander-Cane, Inc., Naples, FL*)



Juice Extraction

The shredded cane passes through a series of three horizontal rollers (mills) arranged in a triangular pattern, with the toproll rotating counterclockwise and the bottom two rollers clockwise. A series of three-roller mills, numbering 37, is called a tandem. The pressure on the top roll is regulated by hydraulic rams and averages about 300 tons. Below each mill is a juice pan into which expressed juice flows. The crusher and first mill extract 6070 percent of the cane juice, and the remaining mills take out 2225 percent. The fibrous cane residue remaining after juice is extracted is called bagasse. When the fiber content of the bagasse reaches about 50 percent, extraction by conventional milling approaches zero. Because the juice remaining with the fiber contains the same proportion of sucrose as the original cane juice, the yield of sugar would be substantially lowered if extraction were terminated at this point. Consequently, a process called "compound imbibition" is used to reduce the sucrose in the bagasse by repeated dilution and milling. In a five-mill tandem, water is added to the fourth mill, and the expressed juice from that mill is brought back to the second mill. The expressed juice from the third mill is recirculated to the first mill and the fifth mill to the third mill. In this way, the juice in the bagasse is always diluted ahead of each mill. The amount of imbibition applied at each mill is approximately equal to the amount of water applied to the fourth mill or to the penultimate mill in a differently numbered tandem. The expressed juice is screened through perforated metal screens with 1-mm diameter openings. Additional screening may be provided by stationary or vibrating metal cloths.

Diffusion

An alternative way to obtain juice from the cane is diffusion. In diffusion, the cane is prepared in such a way as to disrupt the cells for maximum expression of juice. Diffusers have capacities ranging from 100 to 300 tons of cane per hour and may be used in conjunction with part of a milling tandem, the crusher and the first mill. Preparation of the cane for diffusion is essential for good extraction. In the diffuser, the crushed cane is countercurrently washed with imbibition water at temperatures ranging from 50 to 75°C. These high temperatures help to prevent microbial growth in the cane juice. The last stage of a diffuser, the bagasse exit, receives water that gradually increases in sugar content as it proceeds to the final stage. Most of the sugar is extracted in the first four or five stages by simple displacement of sucrose from the ruptured cane cells. In later stages, diffusion appears to take place in unruptured cane cells. One or two mills are used to express water from the bagasse after discharge from the diffuser. The crusher and one mill extract 70 percent of the juice from the cane, and the diffuser removes 30 percent of the expected yield. Sucrose extraction for a milldiffusion system averages 9798 percent, compared with the 9091 percent with a straight milling system.15

Cane diffusers have been installed in cane mills throughout the world and are manufactured by a number of companies. Most large cane mill installations have diffusers, the major exceptions being in the United States, where only one mill has a diffuser.

Biocides

Microbial inhibitors, or biocides, mentioned briefly above, are very important for control in both beet and cane processing, as the juice contains sucrose and other nutrients in the right concentration for promoting microbial growth. The biocides in most common use around the world are dithiocarbamates, glutaraldehyde, ammonium bisulfite, formalin, and hydrogen peroxide. Combinations of these may act synergistically.16 Hop acids are the latest in the arsenal to fight microbial infection.17

Dextran and Other Polysaccharides

The major microbiological losses in sugar processing are caused by the common soil bacterium *Leuconostoc mesenerioides*, which uses sucrose as a food source, and produces a long-chain, glucose-containing polysaccharide, known as dextran, as a waste product. Dextran, in high enough concentration, causes gummy solutions and difficulties in processing, along with sucrose loss. Additionally, it interferes with the polarimetric analysis of sucrose because, having a positive rotation, it is interpreted as sucrose, resulting in falsely high estimates of the sugar content in the cane or beet juice. Various methods exist for dextran measurement, all with some limitations. The most recent, and promising, though expensive, method, is the use of a monoclonal antibody nephelometric method.18

Juice Purification (Clarification)

The cane juice obtained from milling or diffusion is an acidic (pH around 5.55.7), turbid, greenish-brown liquid containing 1218 percent by weight sucrose (depending on the maturity of the cane), in addition to soluble and insoluble impurities, such as soil, protein, waxes, gums, starch, fine bagasse (bagacillo), soluble salts, and pigments. The first stage of purifying the juice, designed to remove the maximum amount of impurities, is called clarification or defection and employs lime and heat. Lime serves the dual purpose of raising the pH and stabilizing the juice against hydrolysis of the sucrose in the acid conditions and helping to form a precipitate to remove impurities. The heat plays a dual role also in helping to disinfect the juice of harmful bacteria and improving the coagulation and precipitation of the impurities.

In simple lime defecation, milk of lime is added to the cold juice in amounts to bring the pH of the juice to around 6.77.5. The limed juice is pumped through heaters in which the juice is heated to 90115°C. Many modifications to this process exist, especially with respect to the heating of the juice. Fractional liming with double heating involves liming the cold juice to pH 6.4, heating, liming to pH 7.6, and heating again before sending the juice to the clarifiers. Polyacrylamide flocculents are added to aid in the coagulation of the precipitate.

The combination of lime and heat forms a flocculent precipitate with various components in the juice, consisting mostly of insoluble lime salts, coagulated protein, and entrapped colloidal and suspended matter. The precipitate is removed by sedimentation or settling in continuous closed-tray clarifiers. The juice leaving the clarifier is a clear brown liquid. The flocculent precipitates that settle on the clarifier trays are called "muds." They contain about 5 percent solid matter. Entrained sugar is recovered from the mud by means of rotary vacuum filters equipped with a perforated metallic screen cloth. The turbid filtrate is returned to the clarification system, and the press cake or filter mud is usually sent to the fields as fertilizer. Good clarification depends upon the formation of a stable flocculent precipitate that settles rapidly. Juices low in phosphates (less than 300 ppm P2O5 on juice solids) and other minerals, such as magnesium salts, tend to be difficult to clarify and phosphate may need to be added. Very fine silty soil carried in from the field may also interfere with clarification. The main change in clarifiers over the last 20 years has been the development of the trayless SRI-type clarifier, which can handle much higher throughputs than the older-type clarifiers.19

Sulfitation

In many, if not most, parts of the world, with the exception of the U.S. cane sugar industry, sulfur dioxide is an important chemical used in processing. The process of adding sulfur dioxide is known as sulfitation (alternate spelling, sulphitation). Sulfur dioxide gas is produced by burning elemental sulfur. The gas is added to the cane juice during clarification, in the presence of lime, and many of the complex, little understood, reactions that produce color in cane juice are inhibited by the presence of SO2 gas. For example, enzymes responsible for the enzymatic browning reaction are inhibited, and reducing sugars (glucose and fructose) are blocked by sulfur addition at the aldehyde group and cannot react with amino acids to form browning polymers in the Maillard reaction. There are a number of variations for the addition of SO2 to juice, based on temperature, order of addition, and pH levels. The advantages of using SO2 are much lower color in the raw sugar and the possibility for making a white sugar directly (see the section "Direct Consumption Sugars" below). Disadvantages include increased scaling problems in the evaporators, high sulfite content in the final product, and secondary color formation (color return) on storage. Recently, the use of liquid SO2 has been shown to have many advantages over the gas, especially in a significant reduction in gas consumption (up to 69%) and a better quality of product.20

Evaporation

The clarified juice (about 85% water) is pumped to evaporators where it is concentrated to a clear heavy syrup containing about 65 percent solids. Evaporation is carried out in multiple-effect evaporators in order to achieve maximum steam economy. Each effect is arranged in series and operated so that each succeeding one operates under a higher vacuum (lower pressure). This arrangement allows the juice to be drawn-from one vessel to the next and permits it to boil at a low temperature. The concentrated juice (final evaporator syrup) is removed from the last effect by a pump. Triple, quadruple-, and quintuple-effect evaporators are used, with the quadruple-effect type plus a pre-evaporator, being a common configuration. In a four-effect (quadruple) evaporator, 1 lb of steam evaporates 4 lb of water. Several types of evaporators exist. The older, more traditional evaporators are long tube climbing film, or Kestner evaporator, and the Roberts evaporator. The Kestner evaporator consists of numerous long vertical tubes inside a cylindrical shell. The juice to be concentrated is fed to the bottom of the tubes and heated, causing the juice to boil and rise through the vessel. At the top, concentrated juice and vapor are separated in separators. The Kestner evaporator has a short residence time and a high heat transfer coefficient, but is more subject to scale deposits. The Roberts evaporator has short tubes, a wider diameter, and does not have a juice/vapor separator.

Newer types of evaporators are the falling film tubular evaporator and the plate evaporator, either rising film or falling film configuration. Advances and innovations in evaporation are driven by the need for energy efficiency and to improve heat transfer coefficients. The goal in evaporation is to move the liquid through as quickly as possible at the highest temperature without causing degradation of the sucrose. The newest of the evaporators is the falling film plate evaporator (e.g. the BalckeDür), the first one being installed in 1992. Recently, technology has become available that allows the retrofitting of existing Robert evaporators with falling film plate packs and some peripheral equipment, allowing the total steam demand of a cane sugar factory to be decreased by almost 30 percent on cane.21

Evaporator scaling

A severe problem in cane sugar processing is the deposition of scale on the evaporator tubes or plates. Scale deposits consist of calcium and silicate salts, enmeshed in a matrix of organic material22,23 and can accumulate very rapidly, causing a severe drop-off in heat transfer and efficiency of evaporation. Cane mills are forced to clean their evaporators on the order of every 1015 days, using mixtures of acid and alkali and mechanical cleaning. Various anti-scalants are available, with some, though not a great deal, of efficacy. Evaporator scaling is a problem in cane sugar refining and beet sugar production as well, but not to the extent that it is in raw sugar production.

Crystallization

The final evaporator syrup is pumped to a vacuum pan, in which it is evaporated to supersaturation to cause sugar crystallization. The vacuum pan is a single-effect evaporator designed to handle viscous materials. It is a vertical cylinder with its bottom designed to allow easy removal of the crystallized mass. The heating elements used in vacuum pans either are short, large-diameter vertical tubes (calandria pans) or coil around the inner surface of the truncated cone of the pan (coil pans). A typical vacuum pan with a "catchall" or entrainment separator for separating syrup from vapors measures approximately 25 ft in height and 15 ft in diameter. The working capacity is about 1000 ft3 of massecuite (mixture of crystals and syrup or mother liquor). The shape of the pan and the positioning of the heating elements within the pan are important design factors in maintaining good circulation of the massecuite. For example, floating calandrias (calandria not attached to the shell of the pan), horizontal pans, and pans having mechanical circulators are used. The process of crystallization is called boiling or pan boiling in the industry.

The move toward continuous processing has resulted in the development of continuous vacuum pans, and low grade continuous pan boiling is well established,24 and raw sugar boiling with continuous pans is increasing as well. The benefits of continuous pan boiling include control simplicity, improved product quality, low product losses, steam savings, steadier demand, and energy savings.25 Most continuous pans are a variation of the original 1931 Werkspoor patent, of a horizontal cylindrical pan divided by 818 stationary partitions or compartments, each heated by steam. Continuous vacuum pans are used in both the cane and beet sugar industries.

It should be noted that batch pans are still required for the production of high-quality refined sugar. One version of the continuous vacuum pan is a horizontal cylinder with compartments in its lower part. The pan is provided with an additional evaporator called the concentrator, where the density of the syrup is raised to 7880 degrees Brix.*

*The Brix scale is a density scale for sugar (sucrose) solutions. The degrees Brix are numerically equal to the percentage of sucrose in solution (wt/wt). The term Brix solids refers to the solids in solution as determined by a refractometer.

Seed is added in the first compartment, and the resulting massecuite moves progressively through the compartments of the pan. Additional syrup is added to each compartment to control the fluidity of the massecuite. Approximately 20 percent of the syrup is introduced at the concentrator, and 80 percent of the syrup is fed to the various compartments through special feed headers.

The crystallization of sucrose in vacuum pans is called "sugar boiling" and each boiling is termed a "strike." Because a single crystallization does not recover all the sucrose from the syrup, mother liquor from a strike is recycled for recovery of additional sugar. There may be as few as two and as many as four strikes to recover the maximum amount of sugar. A three-boiling system is the most popular. The strikes are designated by letters. Thus, the first strike is the A boiling, and the products of the A massecuite after centrifugation, are A sugar and A molasses. The boiling systems are based on "purity," the amount of sucrose remaining in solution, and usually are expressed as the ratio of the polarization value to the total solids as measured by a Brix refractometer. In a three-boiling system, the purity of the A massecuite is set between 80 and 85 by blending syrup with A molasses; the purity of the B massecuite is set between 70 and 75 by blending with syrup and A molasses; the purity of the C massecuite is set between 55 and 60 by blending with syrup and B molasses. Other variations are possible.

The A and B massecuites, after being discharged from a vacuum pan, are sent to centrifugals for separation. The A and B sugars are combined to become commercial raw sugar, the principal product of a cane sugar factory. The C massecuite is a low-purity, highly viscous material that is not immediately sent to the centrifugals because of the large amount of recoverable sucrose remaining in solution. Instead, it is placed in crystallizers, U-shaped horizontal containers equipped with coils attached to a hollow rotating shaft through which water circulates. The massecuite remains in the crystallizer from 1 to 4 days to allow additional crystallization to take place. Centrifugation of the C massecuite yields a final molasses (blackstrap molasses) and a C sugar that is used for seeding the A and B boilings.

Centrifugation

Massecuite from the vacuum pans is sent to centrifugal machines in which the crystals are separated from the mother liquor. A centrifugal consists of a cylindrical perforated basket lined with a screen of perforated sheet metal. The basket, enclosed in a metal casing, is mounted on a vertical shaft that rotates the basket, imparting centrifugal force to the massecuite. In a batch centrifugal, hot massecuite is fed into the basket through a short chute from the holding vessel. As the basket rotates, the massecuite forms a vertical layer on the screen lining. When the machine reaches operating speed (10001800 rpm), the syrup flows through the perforated lining and basket and is removed through an outlet at the bottom of the casing. The sugar on the lining is washed with a spray of water to decrease the amount of molasses adhering to the crystals. The basket continues to rotate until the sugar is fairly dry, at which time the machine is switched off and brakes applied. The sugar is discharged by an automatic plough.

In continuous centrifugals, the machines do not stop, but continue in motion while receiving fresh supplies of massecuites. The rate of feed must be carefully regulated to obtain the optimum separation of molasses from crystals, while at the same time not abrading or breaking too many crystals.

Raw Sugar

The final product of the mill is raw sugar, about 97.598.5 purity, a bulk commodity shipped around the world to be refined by importing countries. All raw sugar produced in the United States is refined within the United States, and none is shipped offshore.

Control

Factory control is based on temperature, pH, and polarization. Cane is tested by various means as it comes to the mill to determine its characteristics and to determine the amount to pay the farmer. Control of microbial infection, discussed above, is important. The starch content of juice from immature cane can be significant (e.g. in Louisiana and South Africa) and amylase enzyme may be added continuously to the clarified juice to reduce it. Starch will cause severe viscosity problems if not removed. If there is a problem with dextran, dextranase enzyme can be added, but it is expensive and not as effective as desired.

Molasses Exhaustion

Molasses contains a significant amount of sucrose and its maximum exhaustibility (removal of sucrose) has always been a priority for the sugar industry. Well-exhausted cane molasses contains about 3032 percent sucrose. Many factors contribute to poorly exhausted molasses, besides inefficiency of process, and these include the impurities present in the cane juice. Many of these impurities, such as polysaccharides, ash components, and organic acids are melassigenic (i.e. they increase the solubility of sucrose), making sucrose more difficult to crystallize, and therefore going preferentially into the molasses.26

Bagasse

Bagasse is the fibrous portion of the cane stalk that remains after extraction of the cane juice. When discharged from the milling train, bagasse contains about 50 percent by weight of water. The average fuel value of ash-free dry bagasse is 8300 BTU/lb (gross calorific value = 19,400 kJ/kg dry ash free). It is a valuable by-product because it is used to fire the boilers in cane mills, making cane factories essentially energy-sufficient. Sometimes excess energy is produced which can be sold to the power companies. A sugar mill that processes one million tons of sugarcane per year produces in bagasse fuel the equivalent of about 420,000 barrels of imported crude oil, more than enough to operate the sugar mill and cogenerate electricity.27 Excess bagasse is used for paperboard and wallboard and as a chemical feedstock for conversion to furfural.28 It is made into a number of paper products in China.29

Direct Consumption Sugar

Many cane sugar mills in tropical countries produce white sugar by using combinations of sulfitation, syrup clarification, or carbonation. These edible sugars are known as mill white sugar or plantation white sugar. The sugars usually are slightly off-white, when compared with refined sugar, and have a higher ash and moisture content. Because sulfur is used in the production, the finished product may also have sulfite in it.

Syrup Clarification

In some factories a process known as syrup clarification is used to produce a higher-quality raw sugar and better yield. Evaporator syrup is treated with phosphoric acid, lime, and a polymer flocculent; insoluble impurities are floated to the surface of the syrup with aeration and removed by skimming. The clarified syrup goes to the vacuum pan for crystallization. Syrup clarification reduces the color and turbidity of the sugar produced, and can also be used in conjunction with sulfitation to produce an off-white, edible sugar.30

Talodura

The Talodura process, used in syrup clarification, consists of using a combination of sulfite clarification followed by subsequent clarification of the evaporated liquor with lime and phosphoric acid to which a flocculant is added. It has made improvements in the color of the mill white sugar, boiling house recovery, and chemical costs per ton of sugar.31

Sulfitation

The raw cane juice is heated to 75°C, clarified with lime and sulfur dioxide, and in some cases filtered and then evaporated. The syrup may be treated with sulfur dioxide again, after which it goes through a three- or four-boiling system. The A and B sugars are mixed with high-purity syrup, centrifuged, dried, and screened for size and distributed as white sugar.

Blanco Directo

Blanco Directo is a trademark of Tate & Lyle Process Technology. The process and similar processes (as described above) are used at sugarcane factories, in conjunction with raw sugar production, to produce a high-grade white sugar suitable for use in food and beverages. This product is economically competitive with imported refined sugars. Although the term "Blanco Directo" is the trade name for one particular process, it appears to be used somewhat generically for similar processes.

Carbonation

In this process the mixed juices from the mills are heated, clarified with lime, and evaporated to about 35 degrees Brix. The syrup is re-limed and treated with carbon dioxide, filtered, re-carbonated, reheated and re-filtered. After carbonation, the syrup is given a double sulfur treatement and filtered. The resulting syrup is subjected to a three- or four-boiling system with the A and B sugars used as the refined product.

Edible Products from the Mill

Raw sugar is not considered an edible food in the United States, but it is in many other countries. There has been a growing trend to produce food-grade sugar at the cane mill (see discussion about plantation white sugar above). The growing societal interest in "unrefined" or "natural" products has led to a huge growth industry for products from sugar mills.32 These include such products as demerara, Barbados, muscovado, and turbinado sugar, which are various grades of brown/raw sugar. Turbinado sugar is raw sugar, which has been washed (centrifuged in the presence of steam) to remove some of the colored molasses coating on the surface of the crystal, resulting in a light golden sugar with large crystals and a mild cane taste.

Organic Sugar

The largest area of growth for edible mill products is organic sugar. This sector is growing at 2530 percent per year.33 There are many regulations and restrictions in place for a mill to be certified to produce organic sugar; these deal with land preparation and the use of herbicides, fertilizers, and chemicals. The yields from organic production are decreased 3060 percent from conventional sugar production, resulting in a product which is about four times more expensive than refined sugar.

New Technology

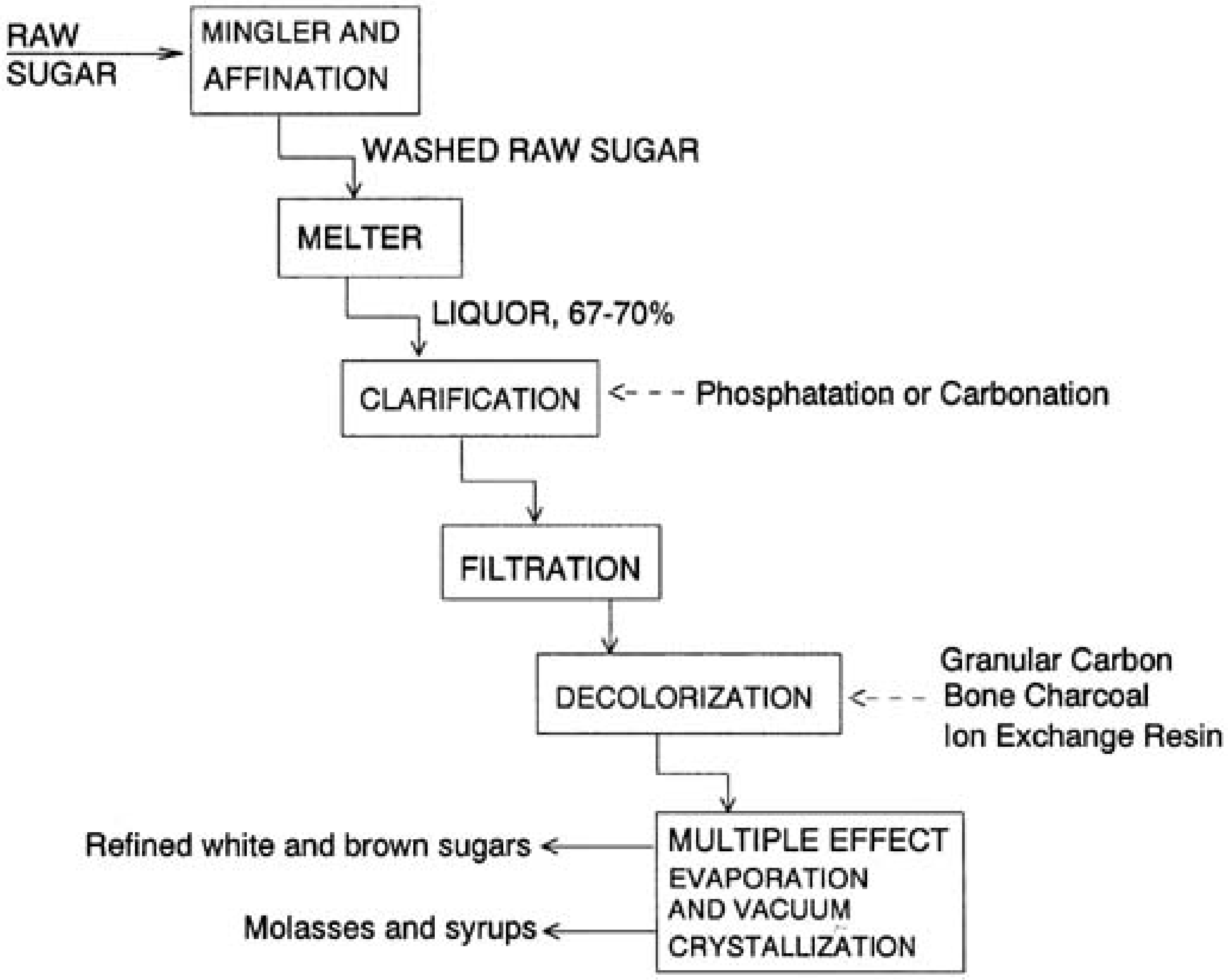
In recent years, a number of new technologies have been explored for the production of raw sugar, including ultra- and membrane filtration, ion-exchange, juice softening (the use of weakly cationic resin to remove calcium), and chromatographic methods, which are, in the words of one author, "beginning to blur the demarcation of steps like clarification."34 These techniques are not limited to raw sugar production but are also being explored in the beet industry; less so in cane refining. The goals in the use of these technologies are to reduce or eliminate the use of lime and/or sulfur, to produce a better quality sugar, or to produce refined sugar directly from cane juice at the mill.

Membrane filtration technology has not yet proven economically feasible in the industry, but continued research may eventually lead to viable commercial products.35 Carousel chromatographic systems, such as the continuous ion-exchange separation (ISEP) and continuous chromatographic separation (CSEP) systems offer some promise in the sugar industry. They have been used successfully in other industries, and may be tailored to obtain a number of different fractions or products.36,37

9.4 Cane Sugar Refining

The raw sugar produced at the factory is light brown in color and contains about 98 percent sucrose. As described above, in most instances this product is not considered an edible food, but rather is a bulk commodity. The process of purifying raw sugar to make a white sugar is called refining, and the sugar produced is called refined sugar. Refining traditionally involves several unit stages, as shown in Fig. 9.6. Because of the increased quality of the raw sugar produced in some areas, not all the stages are used, and various combinations of process are used at different refineries.

Fig. 9.6 General flow diagram of sugar refining. (*Adapted from Clarke*.38)



Affination and Melting

The first step in refining raw sugar is called affination (or washing), and consists of removing the film of molasses, which contains a large portion of the impurities in the sugar, from the surface of the raw sugar crystal. The raw sugar is mixed with a saturated sugar syrup (7275 degrees Brix) at about 50°C in a U-shaped trough called a mingler. Saturated syrup is used to avoid dissolving the raw sugar. The mingler has a rotating agitator to maintain maximum contact between the sugar and the syrup. The mixture is centrifuged to separate the crystals from the syrup, and the crystals are then washed with a small spray of hot water or steam. The "washed" or "affined" sugar is "melted" or dissolved in water to a density of 5560 degrees Brix. Potable water, steam condensate, and "sweet waters" are used for dissolution. The liquor from the melter is screened to remove insoluble matter, such as sand, stones, wood, cane fibers, and lint. Screening is done by vibrating screens, cyclonic separators, or centrifugal screens.

In recent years, some refiners have eliminated the affination stage by buying very high-purity, low-color raw sugarin effect, requiring the raw sugar mill to do the affination at the mill.

Purification

The melted liquor is then purified by any of several combinations that include clarification and decolorization. In clarification, the liquor is treated chemically to remove suspended matter, colloids, and other impurities that contribute to turbidity. Clarified liquor is optically clear, but is still highly colored. Decolorization procedures include treating the clarified liquor with various filtration media, such as bone char, granular activated carbon, or ion exchange resin, either alone or in some combination. It is necessary to clarify the liquor before the filtration medium because the particulate impurities will quickly "blind" the filtration material and render it useless.

Clarification

All clarification treatments include the addition of lime. Lime increases the pH, which stabilizes the sucrose against hydrolysis at high temperatures, and the calcium forms insoluble precipitates with many of the impurities in the raw liquor. Simple liming is rarely used in refining. The important clarification processes are *phosphatation* and *carbonatation*, which include combinations of lime with either phosphoric acid or carbon dioxide.

Phosphatation

In phosphatation, the screened liquor is heated to 6070°C and mixed with phosphoric acid (0.0050.025% P_2O_5 on solids). The mixture is immediately limed to pH 7.08.0, aerated with compressed air, and sent to a clarifier, a tank equipped with heating coils. The liquor enters the clarifier at one end and is heated to 88°C while flowing to the outlet at the opposite end. A flocculent precipitate of calcium phosphate forms that reacts with or entraps colloids, organic acids, some coloring matter, and suspended particles. Air flotation is used to separate the precipitate from the liquor, causing it to rise to the top of the clarifier as a blanket of scum, which is skimmed off by moving paddles. (This process is sometimes known as phosflotation.) About 2530 percent of the color is removed in this process.

Talofloc Process (Color Precipitation)

The addition of flocculants and surfactants has greatly improved the simple phosphatation process. The Talofloc process (developed by Tate & Lyle) consists of adding a cationic surfactant (Talofloc) to the raw melt liquor in concentrations ranging from 0.02 to 0.07 percent on solids. This is followed by the addition of lime and phosphoric acid and the incorporation of air. Just before the liquor enters the clarifying chamber, a polyacrylamide flocculant (Taloflote) is added that induces aggregation of the calcium phosphate precipitate, causing a faster and more complete separation of the impurities. As much as 60 percent of the color can be removed by this process. Other similar additives are commercially available.

Carbonatation

In carbonatation (sometimes also known as carbonation), the melter liquor, heated to about 6080°C, is limed to about pH 10 (0.40.8% CaO on solids), gassed with carbon dioxide, heated to about 85°C, and re-gassed until the pH drops to between 7.7 and 8.5. Because of the two-stage gassing, carbonatation requires two clarifiers, a primary and a secondary. Washed flue gases are the source of carbon dioxide. The carbonated liquor resides in tanks called saturators for about 3 hr to allow growth of the calcium carbonate. The calcium carbonate precipitate that forms entraps some colored matter, colloids, organic acids, and inorganic compounds. The precipitate, known as mud, is removed by filtration, such as Sweetland presses. Carbonatation alone achieves about 50 percent decolorization and, with the addition of polycationic additives, can achieve as much as 70 percent.39

Filtration

The liquor from phosphatation and carbonatation contains small amounts of finely dispersed participate matter that require filtration for removal. The filtration process is similar for both types of clarified liquor. A precoat of filter aid (diatomaceous earth or perlite) is first deposited on the filter surfaces of the press. Additional filter aid is added to phosphate-clarified liquors to improve press runs. Liquor at 7085°F is fed to the press at a pressure of about 60 psig until the flow rate drops below a predetermined level, ending the filtration cycle. The calcium carbonate particles in carbonated liquor act as a fairly good filter aid medium, as long as the particles are developed to optimum size during carbonatation. Large quantities of sugar polysaccharides (starch, dextran, and others) in the melt liquor will restrict the growth of calcium carbonate, requiring the use of diatomaceous earth. A filter cycle may last from 2 to 12 hr depending upon the quality of the feed liquor.

New Developments

At least one refinery has, in recent years, been able to omit all chemical clarification by using only very high-purity, low-color raw sugar.Clarification is accomplished by a two-stage coarse filtration, followed by a finer filter-aid filtration. Again, this development is possible only because of the high quality of the incoming raw sugar.

Scums and Muds

Both phosphatation and carbonatation produce scums or muds that have a moderate level of entrained sucrose, which must be recovered. Various schemes exist for removing the sugar, such as secondary clarification of phosphatation scums and sluicing of carbonation muds with dewatering on rotary vacuum filters.

Decolorization

Filtered clarified liquor is a clear, brown liquid having a solids content between 55 and 65 percent, pH of 6.77.2, and temperature between 65 and 85°C. The next stage of refining involves removing this color to produce a clear, nearly colorless solution from which white sugar can be crystallized. To decolorize it, the liquor is passed through one or a combination of decolorizing adsorbents, such as bone char, granular activated carbon (GAC), or ion-exchange resins. Bone char, consisting of the sintered long bones of cattle, is composed of calcium hydroxyapatite (8085%) in a carbon matrix. It removes colorants, colloidal material, and a considerable amount of the ash components. GAC by itself only removes color; it is usually treated with magnesium oxide in order to buffer the sugar liquor and may remove a small amount of ash. Ion exchange resins absorb color and change the composition of the ash.

Bone char and granular carbon generally are used in fixed beds or in cylindrical columns 2025 ft high and about 10 ft in diameter. Liquor flow through bone char is about 1500 gal/hr over 3060 hr. Liquor flow through GAC is 3000 gal/hr for 2030 days. After the decolorizing cycle is completed, the adsorbent is sweetened off (sugar removed) by displacing the liquor with water. After washing, the spent carbonaceous adsorbent (bone char or GAC) is transferred to regenerating equipment consisting of dryers, kilns, and coolers. Bone char is regenerated at 540°C in a controlled amount of air. GAC is revivified at 950°C in a limited oxygen and steam atmosphere. After regeneration, the adsorbent is returned to the system for a new decolorizing cycle.

Ion-exchange resins are used in columns 810 ft high and 610 ft in diameter, holding between 100 and 300 ft3 of resin. Flow rates are rapid (30004500 gal/hr), and the cycle is short (816 hr). Regeneration is accomplished in the column with a 1015 percent salt solution at 4560°C. The chloride form of a strong anionic resin decolorizes the liquor and the sodium form of a strong cation resin softens the liquor. The sugar in the sweet water from washing the various adsorbent columns is recovered by returning it to the melter.

Crystallization

Decolorized liquor is a pale yellow liquid (ICUMSA color 200300) with solids content of 5565 percent (see *Process Control and Instrumentation* for further information on ICUMSA). This liquor goes to multiple-effect evaporators to be concentrated to a solids content of about 72 degrees Brix, and then pumped to the "white pans" for crystallization. There are four stages in the crystallization process: (1) seeding or graining; (2) establishing the seed; (3) growth of crystals; and (4) concentration. Crystallization is governed by the degree of supersaturation, which is the state in which more solids are in a solution than can be dissolved by water at a given temperature. There are three phases of supersaturation:40

1. The metastable phaseexisting crystals grow in size, but no new ones form.
2. The intermediate phaseexisting crystals continue to grow and new ones form.
3. The labile phasecrystals form spontaneously without the presence of others.

Sugar-boiling for proper crystallization is done in the metastable and intermediate phase; below the metastable zone, crystals dissolve and inefficiencies result; above the intermediate phase, uncontrolled "false grain" or extra, small, and agglomerated crystals form, resulting in poor quality crystals. When the syrup is in the metastable phase, it is "seeded," that is, a predetermined amount of very small crystals is added, which act as nuclei for crystal growth. A sufficient quantity of evaporated liquor is drawn into the pan to cover the heating elements, and water is evaporated from the syrup until its supersaturation approaches 1.25. At this point the steam pressure is lowered, and seed crystals are added. The seed consists of finely pulverized sugar dispersed in isopropyl alcohol or sugar liquor. This method is called shock seeding because addition of the seed induces an immediate formation of crystal nuclei throughout the supersaturated syrup. The nuclei are grown to a predetermined size or grain. Once the grain is established, the crystals are grown to size by maintaining supersaturation between 1.25 and 1.40 through control of the steam pressure, vacuum, and the feed rate of the evaporated liquor. Adequate circulation during crystal growth is important.

When the volume of the massecuite reaches the maximum working capacity of the pan, the syrup feed is shut off, and evaporation is allowed to proceed until a thick massecuite is formed. When the massecuite concentration is considered "tight," the steam and the vacuum are shut off, and the massecuite is dropped into a holding tank equipped with agitators, where it is kept in motion until discharged into the centrifuges.

The boiling system in a refinery is straightforward. The first strike is boiled from evaporator liquor, the second strike is boiled from the runoff syrup of the first strike, continuing on to three or four strikes. These strikes are blended to form the final white sugar product. The runoff of the last strike, which is quite high in color and ash, can be used in a variety of ways: as syrup for affination, as syrup for remelt, or as syrup for producing brown sugar, also called "soft" sugar in the industry.

Centrifuging

Refined sugar crystals are recovered from the mother liquor by centrifuging the massecuite in equipment similar to that used for affining raw sugar. However, at this point, the crystals are washed with a greater quantity of hot water. The washed crystals are discharged into a holding bin supplying a dryer.

Drying and Conditioning

Sugar from the centrifugals contains 12 percent moisture and is too wet to be placed in storage or packaged. The wet sugar is fed to drying equipment called granulators, which usually are large rotating drums, 1535 ft long and 67 ft in diameter, inclined slightly for gravity discharge. Screw conveyers, scrolls, vibratory "grasshoppers," or linear belts are used to transport wet sugar from the centrifuge to the drier, where heated air is blown through the dryer, concurrent with the flow of the sugar. Sugar exiting the granulators has a moisture of about 0.03 percent.

Remelt Sugar

The sugar contained in affination syrup must be recovered for economy, which is accomplished by crystallizing the sugar in a vacuum pan. The area of the refinery where various "low-grade boilings" take place is called the remelt house. The resulting sugar is "raw" in composition and is returned to the refinery at the affination station, hence it is "remelted." The residual syrup from the remelt station is known as refiners' blackstrap, from which no further sugar can be recovered. Various grades or streams of remelt are produced, depending on the refinery.

Packaging and Storing Refined Granulated Sugar

Sugar leaves the dryer at around 5255°C and is allowed to cool to about 45°C. Sugar leaving the dryer still contains enough moisture to cause problems with caking on storage, and the proper bulk storage and conditioning of sugars in silos is a widely discussed topic.41,42 One of the most important considerations for refined sugar is its ability to store well, without caking or darkening in color.

Granulated sugar remains free flowing for a longer period of time if it is "conditioned," as conditioning further reduces the moisture level, which is necessary to prevent caking. Conditioning takes place in huge vertical silos, with up to 3 million pounds capacity. Conditioned dehumidified air is percolated into the sugar from the bottom of each silo. Residence times in silos vary, from about 24 to 72 hr. Proper control of air flow, humidity of the air, and temperature are essential for good conditioning. Sugar leaves the silos with a moisture content of 0.025 percent or less. Following conditioning, the sugar is screened to meet the particle size requirements of customers. Screened sugar is then sent for packaging or for transport in bulk.

Sugar that is not immediately used is placed in storage silos for periods which can be as long as 2 months. Sugar storage is another area where problems with caking and solidification can occur due to moisture migration, and newer systems are beginning to appear where sugar in storage is treated with a slow stream of conditioned air.

Granulated sugar is packaged in 1, 5, 10, 25, 50, and 100 lb bags, 2000 lb sacks, as well as cartons and heat seal packets. Other packaging configurations are possible.

Specialty Sugars

Refineries produce a number of specialty sugars and syrups designed to meet the needs of customers. These include sugars with various crystal sizes, powdered sugars, brown sugars, and liquid sugars.

Liquid Sugars

Liquid sugars (sugar dissolved in water) are economically important because food manufacturers use sugar in the form of a syrup and because of the ease and efficiency of handling a liquid product. Liquid sugars are produce by one of two methods: (1) dissolving refined sugar in water; or (2) further purification of in-process liquors by ion-exchange treatment to remove minerals, and further decolorization using bone char and pulverized or granular carbon. Liquid sugar can consist only of sucrose, as a mixture of sucrose and invert sugar, or as a blend with various glucose syrups from starch hydrolysis. Sucrose liquid sugars are usually distributed at a 67 percent sugar concentration, whereas invert sugar and mixtures of invert and sucrose are distributed at concentrations of 7277 percent sugar. Partially inverted sugar syrups are commercially preferable because of the higher solids content and greater microbiological stability.

Partially inverted sugar syrups are made by either of three hydrolytic methods: (1) acid hydrolysis with mineral acids (sulfuric or hydrochloric); (2) hydrolysis by cation ion-exchange resin; or (3) inversion by enzymes. Acid hydrolysis by mineral acids has the disadvantage that the resulting syrups have a high ash content because the solution is neutralized with sodium or potassium hydroxide. Enzymatic conversion has the disadvantage of being expensive and not very efficient at the high temperature and densities of the solutions being hydrolyzed. Cationic resin treatment offers the best alternative, resulting in an almost ash-free invert syrup. The most popular invert syrup is a 50 percent mixture of sucrose and invert.43

Liquid brown syrups are also available, known as refiners syrups, and can be adjusted in color, flavor, ash level, and sucrose content according to the specifications of the customer.

Brown or Soft Sugars

Brown sugars range in color from light to dark brown. They have very small crystals and a high moisture content, ranging from 0.5 to 3 percent. The combination of very small crystals and high moisture gives the sugars the "soft" appearance and texture. Brown/soft sugars also contain higher ash, invert sugar, and an array of compounds that contribute to the characteristic flavors. They are made in two ways. The traditional method is to boil them from a low-purity process liquor to obtain the desired color, flavor, and composition. The newer and more common method is to "paint" or coat a lesser quality (third or fourth strike) refined sugar with refiner's syrup or molasses to produce a product with similar appearance and characteristics. A typical brown sugar composition is 8590 percent sucrose, 25 percent invert, 24 percent moisture, and 13 percent ash.

Screened Sugars

These are sugars that have been separated on various sieves to give a narrow, defined range of particle size. Baker's special, known as caster sugar in England, has a fine grain size, about 75150 µm, and is desirable because it mixes, blends, and dissolves evenly. Fruit sugar is another screened sugar with a slightly larger grain.

Pulverized Sugars

A range of products are manufactured by milling granulated sugar to a desired size. Examples include powdered or confectioner's sugar, with particle sizes in the 4575 µm, and characterized by the proportion of the sugar passing through certain sieve sizes. In the United States, these sugars usually are mixed with 3 percent starch to prevent caking. Elsewhere, tricalcium phosphate may be added for the same purpose.

Fondant Sugar

Fondant sugar is a special form of pulverized sugar, which exists as a blend of sugar (in the 9095% range) and another ingredient, such as maltodextrin, invert, glucose, or starch. It may be dry and free-flowing or a paste. It is also called icing sugar or glazing sugar. Dry fondants usually are agglomerated forms.

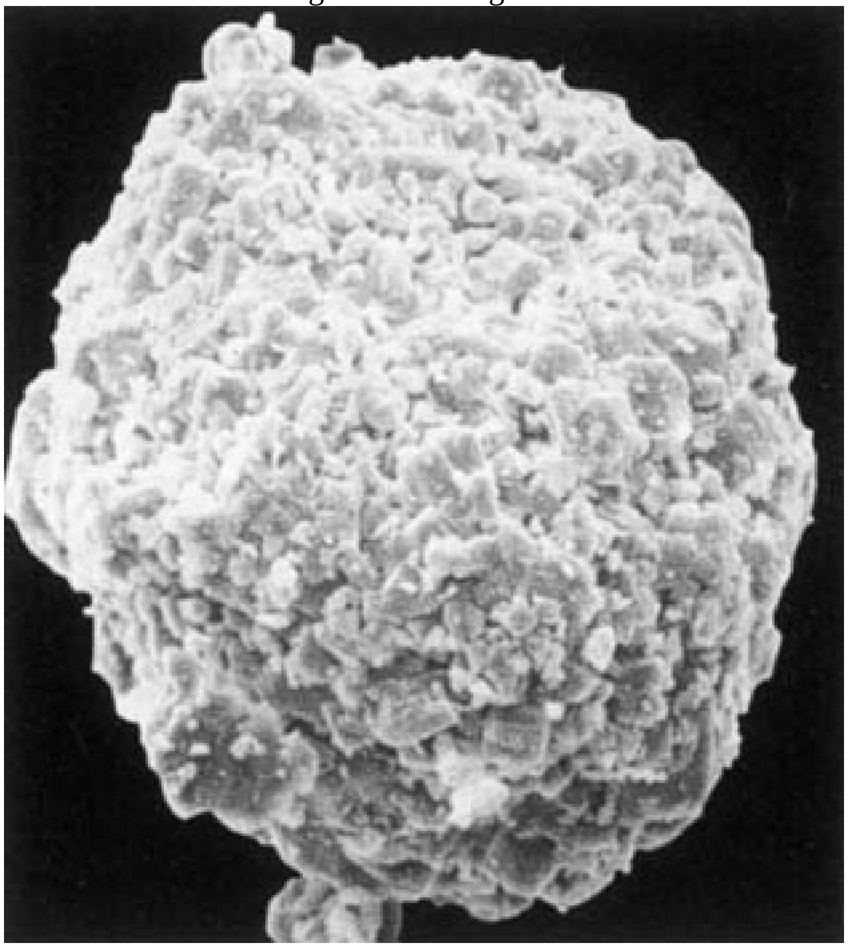
Agglomerated Sugars

The agglomeration of sugar crystals helps to prevent caking, so products, such as soft sugars that are subject to caking, can be converted into a free-flowing dry product. Several companies manufacture "dry fondant" by agglomerating mixtures of invert sugar and very fine (less than 40 µm) sucrose particles. Particles below 40 µm cannot be distinguished by the tongue, so the product is perceived as a smooth paste.

Compound Products

Compound products are made by manipulating the crystal size and shape and incorporating another ingredient in a process called co-crystallization. A very hot (120°C), supersaturated sucrose solution is allowed to cool with agitation in the presence of a second ingredient, resulting in a dry, free-flowing, agglomerated, porous granule, shown in Fig. 9.7.44 Typical products include granulated brown sugar, powdered brown sugar, molasses granules, honey granules, and some fondants. Flavors, such as apple or peanut butter, may be co-crystallized with sugar. The process is said to protect the flavors from loss of volatiles and from oxidation.

Fig. 9.7 Example of a co-crystallized sugar product. The ball represents many microsized sugar crystals agglomerated into a porous granule, which allows inclusion of a second ingredient throughout the structure.



9.5 Beet Sugar

Like cane sugar processing, the main stages in the extraction and production of sugar from sugar beets have remained basically unchanged for the last one hundred years, with cumulative improvements in sucrose yield and processing taking place over the years. The process is a simple one, with many variations: beets are harvested, stored until needed, washed, sliced, and the sugar-laden juice extracted by diffusion. The juice is purified using lime and carbon dioxide, sometimes also sulfur dioxide, filtered, concentrated, and crystallized. The goal is to produce a good quality white sugar in the highest possible yield. Europe and the United States are the major beet sugar producing areas of the world. Sugar beets are grown in 14 states, with Minnesota, North Dakota, Idaho, and Michigan the major producers.45

9.6 Agriculture

The sugar beet, *Beta vulgaris*, is a temperate zone root crop grown largely in the Northern hemisphere (Fig. 9.8). It is adapted to a wide range of climatic conditions, growing, for example, in the short, cold summers of Finland and the hot dry climates of the Imperial Valley of California and southern Spain. Sugar beet agriculture began in earnest in Europe around 1830 and in the United States around 1880. Better yielding beet varieties, mechanization of the harvest, and increasing efficiency in processing have helped the beet sugar industry grow to the extent that it competes in parts of the world with cane sugar.

Fig. 9.8 A sugar beet. (*Courtesy of American Sugarbeet Glowers' Association*)

The original sugar beet seed is of the multigerm type, which results in several plants germinating in one spot, necessitating laborious hand thinning. In 195052, two Russian scientists, V. F. and Helen Savitsky, working in the United States, developed two lines of monogerm seeds. These seeds develop only one plant per seed, which reduced or eliminated the need to thin the young beet plants. Today all sugar beet cultivation uses monogerm hybrids.

Harvesting and Beet Handling

After the plants have reached maturity in late fall, they are harvested by machines that remove the top growth of leaves, lift the roots from the ground, and deliver them to a holding bin or a truck.

Storage

Because the harvesting season is shorter than the processing season, beets are stored in piles or "clamps" at the factory or at outlying points near transportation. Proper storage of sugar beets is very important to prevent deterioration of the beets. Frequently, storage piles are ventilated to lower the temperature of the beets, thus reducing sugar loss due to respiration during storage. Sucrose losses inevitably occur if the temperature fluctuates widely and if there are freezethaw cycles.

In the Red River Valley area of Minnesota and North Dakota, storage sheds for sugar beets were constructed in the mid-1980s that measure 210 ft wide by 520 ft long and hold 50,000 tons of beets. The temperature in these storage facilities is controlled at not much above freezing to minimize degradation of the sucrose content.

Washing

Beets are transferred from storage to the factory in water flumes. These flumes lead directly into a rock-catcher, which allows rocks to settle out, and then on into a trash-catcher and a rotary washer. A large amount of soil usually is entrained with the beet roots, and processes for recovering and returning the soil to the field are in use.

Slicing

The washed beets are sliced into thin V-shaped cossettes by means of specially shaped knives set in frames mounted around the periphery of a rotating drum. Good removal of rocks and trash is essential for a reasonable life for the knives.

Extraction of the Juice

Diffusion

The process of extracting the juice from the beet is called diffusion. The cossettes are weighed and transferred continuously into a diffuser where water passes countercurrent to the movement of the cossettes and where, by diffusion, sugar and some of the nonsugars of the beet are extracted. For optimum extraction of juice, the sucrose-containing plant cells need to be fully disrupted to allow the escape of the juice. This is done mechanically (515% of cells are disrupted by the slicers) and with heat. A balance between temperature and the time of heating is necessary: it should be high enough to maximally extract the juice, but not so high as to cause extraction of undesirable nonsugars, such as pectins, or to cause color formation and sugar degradation. The temperature of the juices during diffusion is 6075°C. Anti-foam agents are used to control foam in continuous diffusers, and bactericides are added for microbiological control. Continuous diffusers have largely replaced batch diffusers. Continuous diffusers come in a variety of forms and shapes, but all employ the same principle, namely the movement of juice countercurrent to the movement of the cossettes.

The RT-type diffuser is a large revolving drum with an internal helix that separates the drum into moving compartments. As the drum revolves, the cossettes travel by the action of the moving helix from one end of the drum to the other end, while the juice moves in the opposite direction. The BMA diffuser is a cylindrical tower with a conveyor mechanism attached to a central rotating shaft. Guide plates on the shaft direct the cossettes upward, while the juice exits through screens at the bottom.

Beet Pulp

Beet pulp is the solid residual portion remaining after diffusion, and is a valuable by-product. Wet pulp, containing 8590 percent moisture, is dewatered in horizontal double-screw presses, reducing the moisture content to about 75 percent. Pulp is further dried to about 810 percent moisture in pulp dryers, direct fired horizontal rotating drums. Newer pulp-drying technology, utilizing a pressurized fluid bed, results in 40 percent more capacity and better energy efficiency.46 The main use of beet pulp is as a pelletized animal feed, either alone or mixed with molasses.

In the 1980s, several beet companies developed edible fiber from beet pulp to take advantage of interest in the health benefits of dietary fiber.47 The fiber has a high water-holding capacity and other interesting properties, but the market has remained small and specialized, with uses as a fiber additive in both human and pet food.

Juice Purification

The diffusion juice (also called raw juice) is a turbid liquid ranging in color from lavender to nearly black, containing 1518 percent sucrose and 13 percent nonsugars, including proteins, nitrogenous bases, amino acids, amides, inorganic material, and pectinaceous matter. It also contains 0.25 percent sodium and potassium. These impurities are removed by a series of purification processes using lime and carbon dioxide (carbonation), and sulfur dioxide (sulfitation). In recent years, there has been some use of ion exchange resins to produce high-quality refined granulated sugar.

A knowledge of the chemistry of the raw beet juice is important in controlling the phases of juice purification. The addition of increasing amounts of lime and carbon dioxide should precipitate pectinaceous materials and insoluble anions such as phosphate, sulfate, oxalate and others, as calcium salts, cause alkaline degradation of invert sugar into lactic and other acids, and flocculation of colorants. Lime is also essential for promoting good filtration.

Classic juice purification consists of preliming, main liming, first carbonation, first sludge separation, second liming, second carbonation, second sludge separation, sulfitation, and filtration. pH control is very important throughout the process and depends on the quality of the beet juice, which is determined in the lab.

Preliming is a partial operation in which a small amount of lime (0.120.3% CaO is added to the juice, increasing the juice pH from about 6.2 to 10.811.4. The lime reacts with nonsugar impurities in the juice to make insoluble precipitates and soluble products. Insoluble precipitates are formed quickly. The addition of more lime (main liming) to an alkalinity of 0.81.3 percent CaO on juice results in a second set of reactions that produce soluble products. During these reactions, the small amount of invert sugar present in the raw juice is destroyed and amides are saponified. These reactions help to stabilize the juice.

Carbonation

In this process, heated, limed juice is sent to a carbonator for gassing with carbon dioxide. The carbon dioxide reacts with the lime to form calcium carbonate, which helps to absorb the insoluble impurities, and some of the soluble impurities as well. The resulting mixture, containing insoluble lime salts, is pumped to subsiders (thickeners, clarifiers) to remove the insolubles by settling, and clarified juice is decanted. From the subsider, the partially clarified filtrate is recarbonated (second carbonation) to lower the pH further, and the residual lime is precipitated. The juice is then treated with sulfur dioxide gas, which helps to inhibit some color-forming reactions in the juice, removes any remaining traces of calcium, and provides the final pH adjustment, to a pH of around 88.5. This treatment is followed by press filtration. The sludge from the subsiders is filtered on rotary-drum filters, and the sugar-containing filtrate is returned to the first carbonation step. The carbonation process can be either continuous or batch. In some factories the clarified juice is also sent to softeners (cation resins) to remove any remaining calcium.

Beet processing uses a large amount of lime (22.5% on beet) and efforts are being made to reduce the usage by effective control of temperature, pH, and other parameters48 which can bring down usage to about 1.01.2 percent and inclusion of innovative processes, such as membrane filtration49 which can further reduce lime usage to around 0.8 percent on beets. The recovery of lime by reburning is also being studied. The clarified juice is called "thin juice" and contains about 15 percent solids.

Evaporation and Standard Liquor

The thin juice discharged from the filter presses after clarification is evaporated to "thick juice" or evaporator syrup (5065% solids) in multiple effect evaporators, containing five, six, or even seven effects. If decolorizing absorbents are used, they are usually added to the thick juice. Granular and powdered carbon have been used for this purpose. Thick juiced that has been treated with powdered carbon usually must be double filtered to remove all the carbon. Low raw sugars (sugars with purities below that of refined sugar) and white centrifugal wash water are added to the thick juice in the melters to make "standard liquor" from which white sugar is crystallized. Standard liquor is usually filtered with diatomaceous earth before going to the vacuum pan.

Crystallization, Centrifuging, and Drying

Crystallization practices in the U.S. beet sugar factories are similar to those in a cane sugar refinery and result in a white granulated sugar comparable in quality with refined cane sugar. However, some European factories may make "raw sugar" as a separate product. White sugar centrifuge stations as well as drying operations are comparable with those operations described earlier for cane sugar factories.

A major difference between cane and beet sugar refining is that beet refining is done in one stage, with refined sugar produced directly from the sugar beets, whereas cane sugar is produced first as a raw sugar and then refined. It is possible for beet sugar to be produced directly because the nature of the coloring material in beet sugar is fundamentally different from that of cane. A beet evaporator syrup with color over 3000 ICUMSA color units produces a white sugar with an ICUMSA color of 2540. However, for a cane white sugar of the same color to be produced, the color of the evaporator syrup must be only 200300 color units. The nature of these differences is not yet well understood.

9.7 Molasses Desugarization

Sugar Recovery from Beet Molasses

Molasses is the viscous, dark-colored material that remains when no more sugar can be economically extracted by crystallization. Beet molasses is generally high in sucrose, 60 percent dry basis, and represents considerable value. Sugar is now routinely recovered from beet molasses by various ion exclusion chromatography systems.

Historically, recovery of sucrose from molasses was accomplished by using the Steffan process, in which sucrose was separated by means of a lime salt known as calcium saccharate. Another process, known as the barium process, precipitated sucrose in the form of barium saccharate. Descriptions of these processes can be found in McGinnis.50 These processes have been replaced by ion exclusion. Molasses desugarization by ion exclusion utilizes a bed of ion-exchange resin operated in such a way as to simulate a countercurrent moving bed system. Its rationale and theory are discussed in detail in the literature.51,52 Up to 90 percent of the sucrose in the molasses can be recovered and other valuable products, such as betaine, can also be recovered.

A strong acid cationic resin, cross-linked with divinylbenzene to form a sturdy polymer network, is used in the monovalent salt form. It is placed in a column alternately fed with molasses and water. The height-to-diameter ratio of the column is important (not less than 4:1 and preferably around 20:1), and flow rates from 0.4 to 0.6 gallons per square foot per minute have been suggested. The clarity and temperature (7080°C) of the feed are also important. Sugar and nonsugar components of the molasses are separated chromatographically. Four fractions are obtained: the first fraction contains most of the inorganic substances, colorants, and organic nonsugars; the second fraction is similar to the first in composition of the feed but contains about 25 percent more water; the third fraction is a high-purity sucrose solution between 10 and 20 degrees Brix; and the fourth fraction is an extremely dilute solution of sucrose and invert sugar (25 degrees Brix). The sucrose fraction either goes to pan crystallization or is further purified with activated carbon and ion-exchange resins and sold as liquid sugar. The first or waste fraction (the raffinate) is concentrated to 70 degrees Brix and marketed as "fodder molasses." After a number of cycles the resin has to be regenerated with salt to remove the accumulated calcium and magnesium ions, which render the resin ineffective. In some cases, feed molasses is softened prior to ion-exclusion treatment.

Desugarization of Cane Molasses

Desugarization of cane molasses is more difficult and less economical for several reasons: it has a lower content of sucrose (about 35%) and a higher content of invert sugar (about 10%); it is more viscous because of higher concentrations of polysaccharides; it has a lot more suspended material; and it has a higher salt and calcium concentration which has to be "softened" before it can be put on a resin. Therefore, cane molasses must be clarified to remove suspended matter and gums, and prior softening is necessary to extend the life of the resin. The resulting product fraction from cane molasses contains a mixture of sucrose and reducing sugars.53

Sugar Recovery from SyrupThick Juice Storage

In 1962, a system for storing thick juice produced in excess of crystallizing capacity was developed in the United States. Storing thick juice serves two purposes: (1) it enables the "beet end" of the factory to operate at full capacity even though the "sugar end" cannot handle the output; and (2) it provides raw material for the sugar end to operate at full capacity when the beet end is in trouble. Syrup storage and sugar recovery is now common practice in the beet sugar industry.

9.8 Process Control and Instrumentation

The sugar industry has always been in the forefront of chemical control. The most important areas of control include measurement of sugar content (done by polarization), pH control (because sugar is so unstable at low or high pH, and more so when high temperatures prevail), color measurement, and control of vacuum pans (for good crystallization).

At the end of the nineteenth century, the International Commission for Uniform Methods of Sugar Analysis (ICUMSA) was formed. This organization, which continues today, is dedicated to all aspects of sugar analysis, validation of methods, and promulgating official methods that are scientifically valid. A good early history of ICUMSA, covering the first session in 1897 up to 1936, can be found in Bates.54

Raw sugar quality characteristics that are measured include polarization, color, grain size, moisture, invert, dextran, starch, and surface color (color that could be removed in affination). New techniques, such as gas chromatography, high-pressure liquid chromatography (HPLC), and near infrared (NIR) analysis are in use for determining various parameters, particularly sucrose, glucose, and fructose. NIR is thought to be particularly promising because it is nondestructive and can theoretically measure a wide range of parameters, such as moisture in bagasse and polarization of juice. These methods are coming to the forefront because the older method of determining sugar content, polarization, has several drawbacks. Originally, lead acetate was used to clarify solutions forpolarization measurement; with lead use restricted in most countries, aluminum and zinc salts have replaced it, but these are not as effective clarifiers. Polarization is also not a direct measurement of sucrose, so it can be inaccurate in low-purity materials, such as molasses.

Instrumentation is used extensively in sugar boiling to control the process of crystallization. Some of the principles used to provide control of vacuum pans for boiling a strike of sugar are listed below.55

1. Boiling point rise (BPR): thermometers are used to measure the temperature of the massecuite and its vapors (the difference between these two temperatures is the observed BPR).
2. Electrical conductivity of the massecuite, based on the principle that conductivity is inversely proportional to the viscosity of the solution, which in turn has a similar relationship to the water content and thus the degree of supersaturation.
3. Fluidity of the massecuite: An ammeter is used to measure the current used by the motor of a mechanical circulator; changes in current indicate changes in the fluidity of the massecuite.
4. Soluble solids measured by a refractometer.
5. Radiofrequency.
6. Microwave density probe.56

9.9 Cane and Beet Sugar Production

Cane sugar is produced in tropical and subtropical areas of the world, whereas beet sugar is produced in temperate areas. The United States is third among the world's sugar producers and is the only country with significant cane and beet sugar industries.57 Cane sugar production is approximately 72 percent of the total world sugar production and beet sugar is 28 percent. The major producers of cane and beet sugar are shown in Table 9.1.58

TABLE 9.1 Major Sugar-Producing Countries, 2000/2001 (in Units of One Million Metric Tons)58

Cane Sugar Production		Beet Sugar Production		Major World Sugar Producersa	
India	20.110	Germany	4.751	India	20.110
Brazil	17.036	France	4.685	Brazil	17.036
China	5.798	United States	4.128	United States	7.604
Mexico	5.226	Turkey	2.755	China	6.739
Thailand	5.220	Poland	2.185	Mexico	5.226
Australia	4.350	Russia	1.707	Thailand	5.220
Cuba	3.585	Ukraine	1.696	Australia	4.350
United States	3.476	Italy	1.687	Cuba	3.585

aIncludes sum of beet and cane production.

In the United States, beet sugar is produced in Michigan, Minnesota, North Dakota, Colorado, Montana, Nebraska, Wyoming, California, and Idaho. Cane sugar is produced in Florida, Louisiana, Texas, and Hawaii.59

9.10 Sugar Consumption and Usage

Per capita consumption of sugar varies widely from country to country and is often a measure of the affluence and level of industrial development of a country. Table 9.2 shows the per capita consumption of selected areas.60

TABLE 9.2 Per Capita Consumption of Sugar (Cane and Beet) in Selected Countries60

Country/Area	Per Capita Consumption (lb/year)
Australia	106
Brazil	114
Canada	91
China	15
Cuba	130
European Union	85
Former Soviet Union	73
India	34
Indonesia	33
Mexico	97
United States	72

Sugar usage patterns, particularly in developed countries, have not changed markedly over the years except in the United States, where high fructose corn syrup (HFCS) has replaced sugar in the manufacture of beverages (see Table 9.3). Nonfood uses of sucrose constitute a very small portion of total use.61

TABLE 9.3 U.S. Sugar Deliveries to Industrial and Nonindustrial Users, Selected Years (1000 Short Tons, Raw Value)

Use	1990	1993	1995	1997	2000
Bakery and cereal products	1,608	1,785	1,905	2,161	2,264
Confectionery products	1,279	1,292	1,372	1,350	1,328
Ice cream and dairy products	462	424	452	436	499
Beverages	228	158	169	158	168
Canned, bottled, and frozen foods	332	336	279	308	330
All other food uses	642	725	864	793	817
Nonfood use	102	85	64	66	85
Total industrial use	4,660	4,805	5,103	5,272	5,491
Hotels, restaurants and institutions	108	108	103	78	71
Wholesale grocers, jobbers, sugar dealers	2,130	2,075	2,173	2,283	2,241
Retail grocers, chain stores	1,077	1,235	1,236	1,281	1,242
All other deliveries	76	171	189	186	339
Total nonindustrial use	3,391	3,589	3,701	3,828	3,893
Total U.S.	8,051	8,394	8,804	9,100	9,383

9.11 Derivatives of SucroseSucrochemistry

Because of its high purity and wide availability, sucrose has been viewed as an attractive feedstock for organic chemical synthesis. Sucrochemistry has been pursued for many years, and many products have been produced.6264 However, the economics of production have not been good and the reactivity of sucrose is difficult to control. Among the derivatives of sucrose are ethers, fatty and other esters, acetals, and reduction products. Successful products from sucrose include surfactants, surface coatings, food and feed additives, polymers, textile chemicals, pharmaceuticals, and pesticides. A promising new product from sucrose is a series of epoxies and plastics with various functional properties, developed at the Southern Regional Research Center of the U.S. Department of Agriculture.65 Other sucrose-based products in successful commercial distribution include olestra, a liquid sucrose polyester fat substitute, sucrose acetate isobutyrate, and sucrose-based detergents.66 Sucralose, a sucrose derivative containing 3 chlorine molecules, was developed in 1976 by Tate & Lyle, and received approval for food use in the United States in 1998. It is 600 times sweeter than sucrose, is noncaloric, and stable to both high temperature and low pH.

9.12 Sweeteners Derived from Starch

Starch is hydrolyzed to produce a wide range of sweeteners, including syrups and crystalline products. The bulk of starch-derived sweeteners comes from corn (maize) starch. Other starchy products are also used to produce starch-derived sweeteners, and small specialty markets exist for syrups from rice, barley, wheat, tapioca, and cassava.

The United States is the largest producer of corn sweeteners and the largest market for high-fructose corn syrup. The major use of HFCS is in beverages and soft drinks, which account for 7580 percent of usage. Important milestones in the development and growth of HFCS are shown below.67

1. late 1950sdiscovery that glucose isomerase enzyme converts glucose into fructose
2. 1967first commercial U.S. production
3. 1972immobilization of glucose isomerase for continuous production of HFCS-42
4. 1977commercial production of HFCS-55
5. 1980HFCS approved as 50 percent of the sweetener in Coca Cola
6. 1983HFCS approved as 50 percent of the sweetener in Pepsi Cola
7. 1984100 percent level of HFCS approved in Coca Cola and Pepsi Cola

The 1984 approval of 100 percent HFCS as a sweetener in soft drinks resulted in phenomenal growth in the HFCS industry. In 1985, HFCS production was 5.3 million short tons (dry weight) and in 2000, it was 9.3 million short tons, with 9.6 million forecast for 2002.67,68 In 2001, consumption of sugar and HFCS were almost equal in the United States.

In Europe, where other starch sources, such as wheat and potato, are used, these are referred to as high fructose syrups (HFSs) or high fructose starch-based syrups (HFSSs).

The corn industry makes widespread use of enzymes for carbohydrate conversion. The advent of enzyme technology in the corn industry in the 1960s dramatically changed the starch industry and allowed the development of new products. Today, enzyme hydrolysis of starch has largely replaced acid hydrolysis, which is used as an adjunct in starch conversion. Enzymes used to make corn syrups and HFCS include the following:

1. alpha amylaseconverts starch into dextrin; used in the manufacture of glucose syrup
2. glucoamylaseconverts dextrin into glucose; specific for amylose
3. pullulanaseconverts the amylopectin fraction of starch into linear segments
4. beta amylaseReleases maltose units from the nonreducing end of a polysaccharide
5. glucose isomeraseconverts glucose into fructose; used to make HFCS

Starch Conversion

Starch conversion refers to the process of converting starch into other products. It involves gelatinization, liquefaction, and saccharification. Liquefaction refers to the acid- or enzyme-catalyzed conversion of starch into maltodextrin. Starch, usually from wet milling of corn, is pumped as a slurry to the conversion plant, where it undergoes one or more hydrolytic processes to yield mixtures of various carbohydrates in the form of syrups or as crystalline dextrose. The kind and amount of the various carbohydrates obtained depend upon the type of hydrolysis system used (acid, acidenzyme, or enzyme enzyme), the extent to which the hydrolytic reaction is allowed to proceed, and the type of enzyme(s) used. The fact that most starches consist of two different kinds of polymers (amylose and amylopectin) also has an effect on the nature of the products obtained.

The extent to which starch is converted into simpler carbohydrates is indicated by the dextrose equivalent (DE), a measure of the reducing sugar content calculated as dextrose and expressed as a percentage of the dry substances. Hydrolyzates having dextrose equivalents ranging from 5 to 100 are produced. Those having a low dextrose equivalent frequently are referred to as dextrins and are not sweet. They are produced by minimal acid hydrolysis or roasting. Starch-hydrolyzate syrups commonly are produced as "low," "intermediate," "high," or "very-high" conversion products, as more or less standard products, as shown in Table 9.4. High fructose corn syrups are classified according to the fructose content, as a percent of solids (i.e. 42%, 55%, 90%), rather than by DE. Table 9.5 shows the composition of various corn syrups. In addition to carbohydrates, the syrups contain some sodium chloride and traces of nitrogenous substances. Maltodextrin is a dried corn syrup having a DE below 20, while the dry dextrins with a DE above 20 are called corn syrup solids.

TABLE 9.4 Conversion Groups for Starch Hydrolyzates

Conversion	DE
Low (Type I)	2038
Intermediate (Type II)	3858
High (Type III)	5873
Very high (Type IV)	73 and above

TABLE 9.5 Carbohydrate Composition of Glucose Syrups (Saccharides as a Percent of Total Carbohydrates)69

DE	Glucose	Maltose	Tri-a	Tetra-a	Penta-a	Hexa-a	Hepta-a	Higher
15	3.7	4.4	4.4	4.5	4.3	3.3	3.0	72.4
35	13.4	11.3	10.0	9.1	7.8	6.5	5.5	36.4
45	21.0	14.9	12.2	10.1	8.4	6.5	5.6	21.3
55	30.8	18.1	13.2	9.5	7.2	5.1	4.2	11.9
65	42.5	20.9	12.7	7.5	5.1	3.6	2.2	5.5

aRefers to the number of glucose units in the oligosaccharide: tri = maltotriose; tetra = maltotetraose; penta = maltopentaose, etc.

Acid Hydrolysis of Starch

As mentioned above, this process now serves mainly as an adjunct to enzymatic conversion of starch and is rarely used alone. A starch slurry containing 3545 percent solids is acidified with hydrochloric acid to about pH 1.81.9. The suspension is pumped into an autoclave (converter) where live steam is gradually admitted to a pressure of 3045 psi. The conversion time largely determines the DE of the hydrolyzate; for example, 8 min may produce 42 DE syrup, and 10 min may produce 55 DE.70 Converted liquors are neutralized with sodium carbonate to a pH of 57, with coagulation of insoluble protein, fats, and colloidal matter. The scum is removed by centrifugation.

The dark-colored clarified liquor is pressure-filtered and concentrated to 60 percent solids in multiple-effect evaporators. The concentrated liquor is decolorized with granular carbon in columns 12 ft in diameter and 30 ft high in a countercurrent manner; that is, liquor flows upward in the columns, while a portion of the carbon is removed from the bottom periodically. It has been reported that carbon is used at a rate of 2.5 percent of dry solid processed and that 5 percent of carbon is lost during revivification.71 In some cases, granular carbon has been replaced by synthetic polymeric adsorbents to decolorize the syrup. "Low ash" syrups usually are deionized with ion-exchange resins. The processed liquor is evaporated to a final solids content of 7585 percent in a single-effect evaporator.

AcidEnzyme Hydrolysis

Starch is first liquefied and hydrolyzed to specific dextrose equivalents with hydrochloric acid. After evaporation to 60 percent solids, a saccharifying enzyme (fungal α-amylase) is added to continue hydrolysis to the desired level. By choosing two or more types of enzymes (such as α-amylase, β-amylase, glucoamylase, pullulanase) and adjusting the initial acid hydrolysis, syrups with different ratios of dextrose, maltose, and higher saccharides can be obtained.72

EnzymeEnzyme Hydrolysis

Enzymeenzyme conversion employs heat and an enzyme for starch liquefaction in place of acid. This is the most common form of corn processing today. Subsequent hydrolysis is by enzymes, as above. The choice of hydrolytic system depends upon economics and the kind of end product desired. Enzymes are usually inactivated by heating the syrup to 7580°C, with the exception of the heat-stable α-amylases mat have come on the market in the last 1015 years.

A starch slurry of 3040 percent dry matter is gelatinized and liquefied in a single dose jet-cooking process. Heat-stable α-amylase enzyme is metered into the starch slurry after pH adjustment to 6.06.5, and the slurry pumped through a jet cooker. Steam is injected to raise the temperature to 105°C and held for about 5 min, which is sufficient to gelatinize the starch (break up the starch granules). The partially liquefied starch is then cooled to 95100°C and the enzyme is reacted at this temperature for 12 hr, until the required DE is obtained.72 Following liquefaction, the starch undergoes saccharification (conversion to sugars) using glucoamylase and pullulanase enzyme, resulting in maltose syrups, glucose syrups, and mixed syrups, as described above. Two enzymes are required at this point because starch is made up of two types of glucose polymersamylose, which is a linear polymer, and amylopectin, a branched polymerand each requires a different enzyme to break it up. Glucoamylase hydrolyzes the linear chains of amylose, while pullulanase specifically attacks the amylopectin at the branching points. The proper combination and time of hydrolysis will result in the desired range of products, all the way to syrups with as much as 97 percent glucose content. The glucose syrup may also undergo isomerization with immobilized glucose isomerase to make fructose syrups.

Crystalline Dextrose

Highly purified, high-dextrose syrup is crystallized to produce crystalline dextrose. The high-purity liquor is pumped to insulated crystallizers fitted with slowly moving agitators for crystallization of dextrose monohydrate. A heavy seed base (about 25%) from a previous batch is mingled with the syrup and cooled to about 38°C. The seeded liquor is held at this temperature for several days until about 60 percent has crystallized. The mixture then is centrifuged to separate the crystals from the mother liquor. The wet sugar is dried in rotary dryers or recrystallized into anhydrous dextrose. The monohydrate also may be converted to anhydrous dextrose by drying in hot air. A second crop of crystals is taken from the mother liquor, and the runoff syrup from this step is final molasses or "hydrol."

Crystalline Fructose

The newest sweetener from the corn industry is crystalline fructose, which came on the market around 1987, and is now available in retail outlets as a sweetener as well as used in commercial products.

High Fructose Corn Syrup (HFCS)

HFCS is produced from very pure glucose syrup using glucose isomerase. Commercial glucose isomerase is produced from a variety of microorganisms: *Bacillus coagulans*, *Actinplanes missourensis*, and several Streptomyces species. Glucose isomerase is always used in immobilized form because it is an intracellular enzyme. Immobilization is generally done by cross-linking with glutaraldehyde after the cells have been disrupted by passing through a homogenizer. The isomerization reaction is carried out at 60°C at pH 78. Reaction time is limited to prevent degradation of the fructose. All isomerization reactions today are done in continuous mode, rather than batch mode. The operating lifetime of a column of immobilized enzyme can be as long as 200360 days, depending on the robustness of the enzyme used.72

The basic feature of the isomerization process is a series of reactors containing the immobilized enzyme fixed in a packed bed. The feed material is of the highest practical purity possible because enzymatic activity is closely related to the purity of the glucose syrup. Glucose syrup of about 94 DE is filtered, treated with activated carbon to remove residual color, and deionized with ion-exchange resin to lower the ash content, particularly the calcium ion, which interferes with the magnesium enzyme activator. The dissolved oxygen is reduced by flash evaporation, which also concentrates the feed stock to 4045 percent glucose and raises the temperature to 6065°C. At this point, prior to entering the reactors, magnesium ions are added as enzyme activators. The pH is adjusted to about 8. Isomerized liquor is removed from the process when the equivalent of 42 percent fructose on a dry basis is reached. After isomerization, the pH of the syrup is lowered to 45, and it is purified by ion-exchange resin and activated carbon. This results in a 42 percent HFCS. For use in soft drinks, it is converted to 55 percent HFCS by passing through large chromatographic columns of zeolites or the calcium salts of cation-exchange resins, which absorb and separate the fructose from the other components. This produces a stream of about 90 percent fructose, which is then blended with 42 percent fructose syrup to produced the desired 55 percent fructose syrup product.

HFCS technology and products have spread to Canada, Japan, and several western European nations.

9.13 Molasses

Molasses, defined as the residual mother liquor from which little or no additional sugar can be recovered economically, is a by-product of the cane, beet, and dextrose industries. Each type of manufacturer has designated this liquid by-product with its own name. The molasses from both raw cane sugar production and cane sugar refining is commonly called blackstrap molasses. In the cane mill, it is also called final molasses, and in the refinery it is sometimes called refinery molasses. Molasses from beets is simply called beet molasses, and that from starch hydrolysis is known in the United States as hydrol. The composition of molasses varies a lot depending on location, varieties harvested, and the efficiency of the operation. The amount of sucrose/sugar remaining in molasses is a measure of its "exhaustion." The amount of refinery molasses is usually small, since as much as possible is recycled and recovered and a lot of it goes into the manufacture of brown sugar. Some cane refineries have no molasses output at all. The general composition of each type of molasses is shown in Table 9.6.

TABLE 9.6 Composition of Molasses from Various Sources. Values are on a Dry Basis

Constituent	Cane Final Molasses	Refinery Blackstrap	Beet Molasses	Hydrol
Sucrose	3140	5060	6063	0
Reducing sugars	2030	510	0.51.5	55.0
Higher saccharidesa	1.54.0	0.51.0	1.02.0	<1
Ash	1014	510	1012	7.2
Nitrogen	0.4	0.10.3	2.0	0.07

aIncludes polysaccharides.

Blackstrap molasses from the mill contains significant quantities of sucrose and reducing sugars, some of it produced during the processing of the cane juice, caused by hydrolysis of sucrose, and some of it from the native cane juice. Refinery blackstrap has fewer impurities because the raw sugar starting materials has fewer impurities than the cane juice but more sucrose. Beet molasses contains primarily sucrose and little or no reducing sugar because of the lower content of reducing sugars in beet juice and because the highly alkaline processing conditions destroy most of the reducing sugars. Hydrol contains no sucrose. Higher saccharides result from incomplete hydrolysis of starch. Cane and beet molasses contain polysaccharides that come from the plant (starch and cell wall polysaccharides) as well as some oligosaccharides, that either occur naturally or are formed during processing.

The main use of molasses is for cattle feed along with some minor chemical production such as citric acid by fermentation. The sucrose in beet molasses is recovered by chromatographic processes described earlier. Cane molasses can also serve as a fermentation substrate for rum and "power" or fuel ethanol and for the production of baker's yeast.

Brazil has been in the forefront of producing fuel ethanol from cane sugar. Brazil unveiled the world's first alcohol-powered car after the 1970s oil crisis, using its abundant supplies of sugar cane. This innovative program was very successful for almost 20 years, but poor distribution networks and frequent strikes by sugar cane workers caused shortages of alcohol in 1989, sending prices soaring. By 1997, there was a question whether the alcohol program would survive, when the Brazilian state oil company, Petrobras, ended its monopoly over ethanol distribution. However, the Brazilian ethanol program has continued, and the Brazilian government set total alcohol content in gasoline at 22 percent as of May 31, 2001. Total alcohol production for 2000/2001 was anticipated to be 10.6 billion liters, but difficulties persist with the cost of ethanol production.73

9.14 Other Sweeteners

Several carbohydrate-based sweeteners exist on the market today. Although sucrose is the "gold standard" for sweeteners, other sweeteners gain commercial viability because they are less expensive (such as corn-based sweeteners), have fewer calories, or are better suited for diabetic diets. Unlike the synthetic sweeteners, most of the carbohydrate-based sweeteners are less sweet than sucrose, with the notable exception of sucralose, winch is 600 times sweeter than sucrose and is non-nutritive. Fructose and xylitol are reported to be sweeter than sucrose. Table 9.7 shows the relative sweetness of some of these compounds.74,75

TABLE 9.7 Relative Sweetness of Carbohydrate-based Sweeteners

Crystalline fructose	180
Fructose in solution (515%)	115125
High fructose corn syrup	100130
Invert syrup	105
Sucrose	100
Tagatose	92
Xylitol (<10% solution)	85
Xylitol (10% solution)	100
Xylitol (>10% in solution)	120
Glucose (2% solution)	50
Glucose (810% solution)	6070
Glucose (50% solution)	90100
Erythritol	70
Sorbitol (hydrogenated glucose)	70
Maltitol (hydrogenated maltose)	68
Maltose	50
Trehalose	4550
Mannitol	40
Lactitol (hydrogenated lactose)	3040
Lactose (milk sugar)	1530

An important group of carbohydrate-based sweeteners are the polyols, also known as sugar alcohols. These are not considered sugars for labeling purposes, so foods containing them as sweeteners can be called "sugar free." The polyols are usually 1530 percent sweeter than their corresponding sugar. Some of the polyols give a cooling sensation on the tongue. Several of the new carbohydrate sweeteners are described below.

Erythritol, 70 percent as sweet as sucrose, is a polyol (sugar alcohol), currently used as a bulk sweetener in reduced-calorie foods. It has only 0.2 cal/g (sucrose has 4 cal/g). It was given Generally Recognized as Safe (GRAS) status by the Food and Drug Administration (FDA) in 1997, which allows it to be used in foods in the United States.76 Erythritol is produced by fermentation with the fungus *Moniliella pollinis*.77

Tagatose, a monosacharide similar to fructose, received GRAS status from the FDA in April 2001. This sugar, 92 percent as sweet as sucrose, with 1.5 cal/g, is produced from galactose obtained from whey, using immobilized enzymes.78 It is claimed to have the same bulking properties as sucrose.

Trehalose, a disaccharide sweetener, 4550 percent as sweet as sucrose, was given GRAS status in 2000. It is naturally found in mushrooms, honey, lobster, shrimp, and foods made with yeast. It has been used in Japan for decades, and is commercially produced from starch by bacterial enzymes.79 Besides its mild sweetening power, it maintains cell structure during freezing and dehydration of foods. It is a nonreducing sugar, so it does not participate in the Maillard reaction (will not brown) and helps to protect the color of processed foods.80

Other naturally occurring sweeteners include palm sugars, maple sugar, honey, and agave syrups (composed of about 90% fructose).

9.15 Regulation and Trade in Sugar

Raw cane sugar, refined sugar, sugar syrups, specialty sugars, and sugar-containing products enter the United States under a variety of tariff-rate quotas (TRQs). Low within-quota tariff rates facilitate access to the U.S. market. The minimum sizes of the TRQs and the corresponding duties reflect obligations made by the United States. When Congress ratified the NAFTA and WTO (Uruguay Round) treaties, these TRQs became part of the U.S. Harmonized Tariff Schedule.

Most of the sugar produced in the world (about 70%) is used for local or domestic consumption. The remaining 30 percent is in excess of local demand and makes up the global trade in sugar. In 1999/2000, 40.6 million metric tons, raw value, of centrifugal sugar were exported globally, out of a total production of 136 million tons. In recent years, the trend has been toward the export of more white sugar instead of raw sugar. Most European countries are beet sugar producers, and many became self-sufficient in the 1970s, and have subsequently become exporters of refined sugar. Britain, France, and Finland continue to import raw cane sugar to be refined. The European Union is currently the world's largest exporter of refined white sugar, at 8.1 million tons in 1999/2000, followed by Brazil at 3.0 million tons. Brazil also exports 5.4 million tons of raw sugar, making it the largest net exporter of sugar in the world.1

Trade in sugar throughout much of the world is controlled by agreements and special trading arrangements. The United States regulated sugar production, distribution, and importation by means of the Sugar Act of 1948 as amended (U.S. Congress, Public Law 339), but the act was allowed to expire in 1974. From 1981 to the present, the United States has had a price support program, connected to farm legislation. In 1982, an import quota system went into effect to supplement the price support program. In 1988, Australia filed a complaint against the United States, charging that the import restriction on sugar violated the General Agreement on Tariffs and Trade (GATT) rules. When the United States was found to be in violation on some of these rules, it implemented a change from an absolute quota to a tariff rate quota. The United States also has a program called the General System of Preferences (GSP) that exempts countries (mostly in the Caribbean area) from any duty. The EU countries import raw sugar within a special agreement with ACP (Africa, Caribbean, and Pacific) countries.

Raw cane sugar prices are based on the price of sugar delivered to New York and is quoted on the New York Coffee, Sugar, and Cocoa Exchange as Sugar Number 14 Contract (domestic). There is no future market for U.S. refined sugar. The world refined sugar price is based on the Number 5 Contract on the London International Financial Futures and Options Exchange (LIFFE), which is based on the London daily spot market price for refined sugar in free-on-board ships in European ports.81

Copies of the Number 11 (world sugar) and Number 14 contracts (domestic) for raw sugar can be found in issues of *F.O. Lichts World Sugar and Sweetener Yearbook*.

The USDA's Sugar-Containing Products Re-Export Program is designed to put U.S. manufacturers of sugar-containing products on a level playing field in the world market. The Refined Sugar Re-Export Program is designed to facilitate the use of domestic refining capacity to export refined sugar into the world market. The program establishes a license against which a refiner can export domestically produced refined sugar and later import world raw sugar, import world raw sugar for refining and distribution into the domestic market, and later export refined sugar, or import raw sugar, refine it and export it into the world market. The program was implemented to mitigate the imposition of restrictive quotas, which reduced the quantity of raw sugar allowed to enter the U.S. domestic market.

9.16 Environmental Concerns

The sugar industry, in common with all other sectors of the food and chemical processing industries, is concerned with protecting the environment. This means that close attention must be paid to all gaseous and liquid effluents, with the aim of minimizing their production and ensuring that they meet all minimum standards before being discharged into the environment. The goal of the sugar industry is to be ecologically sustainable and economically viable and efficient. It also must take into account community concerns, such as the question of whether to continue the pre-harvest burning of sugarcane in various areas. Another factor of community concern is the odor produced by beet factories, and odor abatement programs are in place. It is certain that present standards will not be relaxed and that they can be expected to become even more restrictive as time goes on. It is incumbent on the industry that older plants and operations be reviewed continually to determine what further steps can be taken to control emissions and reduce their volume. It is assumed that any changes made to existing operations, as well as all new designs, would include careful evaluation of their implications for waste management requirements. The question of lead usage in polarization analysis was discussed earlier, and most sugar producers have taken steps to eliminate its usage by using alternatives (aluminum salts) or newer methods of analysis (HPLC and NIR).

Issues of Genetic Engineering

In the United States, herbicide-tolerant genetically engineered sugar beet was approved in 1998, and has been available to the market since 1999. However, farmers have been wary about planting it because soft drink, food, and other industrial users are concerned about growing consumer doubts. As of this writing, no genetically engineered commercial canes exist, but research is under way in various countries to develop disease-resistant varieties. It is not expected to appear commercially for about 10 years. Modified corn has been grown for many years, and several of the enzymes used are from modified organisms. Because these have been in place for some years, there is more acceptance.

9.17 References

F.O. Lichts World Sugar Statistics, F.O. Licht GmbH, 2001.
Sawhney, D. H., "The Indian Sugar Industry," Zuckerind., 127, 186194 (2002).
Tegge, G., "Development and Prospects of Sweeteners Obtained by Glucose Isomerization," in International Sweetener Report No. 1, H. Ahlfeld (Ed.), pp. 38, F.O. Licht, GmbH., Ratezeburg, Germany, Dec. 1978.
"Recent developments in the sugar industry," by the Editor, Internat. Sugar J., 102, 4360 (2000).
Hugill, J. A. C., "History of the Sugar Industry," in Sugar: Science and Technology, G. G. Birch and K. J. Parker (Eds.), pp. 1537, Applied Science Publishers Ltd., London, 1979.
Meade, G. P., The Sugar Molecule, Sugar Research Foundation, N.Y., 11(1), 1119 (195960).
Deere, N., The History of Sugar, Vol. 1, Chapman & Hall, London, 1949.
Hugot, E., and Jenkins, G. H., Handbook of Cane Sugar Engineering, American Elsevier, New York, 1960.
McGinnis, R. A., Beet Sugar Technology, Van Nostrand Reinhold, New York, 1951.
Eggleston, G., Legendre, B. J., and Richard, C., "Effect of Harvest Method and Storage Time on Sugarcane Deterioration," Proc. Conference on Sugar Processing Research, pp. 205221, 2002.
Godshall, M. A., Legendre, B. L., Richard, C., and Triche, R., "Effect of Harvest System on Cane Juice Quality," Proc. Conference on Sugar Processing Research, pp. 222236, 2002.
Kördel, P., "New Equipment and Processes in the Sugar Industry," in F.O. Licht's Guide to Equipment, Products and Services for the Sugar and Allied Industries, Directory of Suppliers, pp. E3E25, 1998/1999.
Rein, P., "Cane Preparation," Sugar Bulletin, 8, 911 (Mar. 2002).
Ducasse, V., "Heavy Duty Chopper Fiberiser," Proc. International Society of Sugar Cane Technologists, 16th Congress, pp. 15681601, 1978.
BMA Cane Diffusion, Sugar y Azucar, July 1970, p. 22.
Fowers, M., and Lloyd, L. E., "Biocide Synergy: Reducing Costs and Losses Through Physical and Chemical Interactions," Presented at the 30th General Meeting of the American Society of Sugar Beet Technologists, Orlando, USA, Feb. 1999.
Pollach, G., Hein, W., and Beddie, D., "Application of Hot Beta Acids and Rosin Acids in the Sugar Industry," Paper Presented at the Sugar Processing Research Conference, March 2002, New Orleans, Proceedings in press.
Rauh, J. S., Cuddihy, J. A., and Falgout, R., "Analyzing Dextran in the Sugar Industry: A Review of Dextran in the Factory and a New Analytical Technique," Proc. Sugar Industry Technologists, Vol. 58, pp. 1728, 1999.
Steindl, R. J., "The Clarification of Cane Juice," Internat. Sugar J., 101, 213215 (1999).
Moodley, M., Bekker, D. J., Pienaar, P. J., and Pillay, R., "Replacement of Sulphur Burner with Liquid Sulphur Dioxide System at the Malelane Refinery," Proc. South African Sugar Technologists Association, Vol. 71, pp. 142145, 1997.
Grant, G., Anders, M., Morgenroth, B., Niepoth, K., and Ehrenreich, "Application of the New Falling-film Plate Technology in the Cane Sugar Industry," Internat. Sugar J., 103, 7073 (2001).
Walthew, D. C., "Aspects of Evaporator Scale Formation and Control in the South African Sugar Industry," Proc. Conference on Sugar Processing Research, pp. 2243, 1996.
Godshall, M. A., and Whartell, L., "Composition of Scale in Louisiana Mills," Poster presented at the Conference on Sugar Processing Research, March 1013, 2002, New Orleans, Louisiana, USA.
Rein, P. N., "A Review of Experience with Continuous Pans in TongattHulett Sugar," Proc. South African Sugar Technologists Association, 1986.
Moor, B., "Continuous Pans: Part 1History, Theory and Experience Applied in a New Design," Internat. Sugar J., 103, 542546 (2001).
"Influence of Individual Nonsugars on Molasses Formation," in Sugar Technology, Beet and Cane Sugar Manufacture, G. Mantovani et al. (Eds.), Verlag Dr Albert Bartens, Berlin, pp. 685686, 1998.
Coerver, J. F., "Is Bioenergy Opportunity Knocking?," The Sugar Bull., 80(4), 1314 (Jan. 2002).
Manohar Rao, P. J., Industrial Utilization of Sugar Cane and Its Co-Products, ISPCK Publishers, Delhi, India, 1997.
Yang, W., Lan, C., and Lian, H., "High Quality Paper Products from Bagasse," Proc. Conf on Sugar Processing Research, pp. 274285, 2000.
Cane Sugar Handbook, 12th ed. (Section 5.77) "Raw Syrup Clarification Processes," pp. 156162, J. C. P. Chen and C. C. Chou (Eds.), John Wiley & Sons, Inc., 1993.
Bennett, M. C., Elvin, J. R., Heineman, H. W. B., Pottage, P. R., and Rundell, J. T., "Syrup Clarification for Improving Sugar Quality and Yield," XVI Congress of International Society of Sugar Cane Technologists, pp. 27972809, 1978.
Bosworth, M., "Billington'sthe Unrefined Sugar Specialists," Internat. Sugar J., 102, 654655 (2000).
Buzzanell, P. J., "Organic Sugar: Short Term Fad or Long Term Growth Opportunity?," Paper presented at International Sugar Organization, 9th International Seminar, "Hot Issues for Sugar," London, Nov. 2000. http://www.organic-research.co.uk/research/Papers/sugar.htm
Blackwell, J., "Recent Developments in Sugar Processing," Internat. Sugar J., 104, 2842 (2002).
Kochergin, V., "Practical Applications of Membrane Filtration in the Sugar Industry," Internat. Sugar J., 103, 529532 (2001).
Rossiter, G. J., "Chromatography in Sugar Processing," Symp. on Advanced Technology for Raw Sugar and Cane and Beet Refined Sugar Production, pp. 181193, 1999.
Rossiter, G., Jensen, C., and Fechter, W., "WSM ProcessIntegrating UF and Resin Processes into a Cane Mill Flowsheet," Paper presented at the Sugar Processing Research Conference, March 2002, New Orleans, Proceedings in press.
Clarke, M. E., "Cane Sugar Process and Byproduct Streams," Paper presented at Iowa Biotechnology Consortium, March 1991.
Cane Sugar Handbook, 12th ed. (Section 12.15) "Modification of Phosphatation and/or Carbonation," p. 458, J. C. P. Chen and C. C. Chou (Eds.), John Wiley & Sons, Inc., 1993.
Cane Sugar Handbook, 12th ed. (Section 7) "Pan Control and Use of Instruments," J. C. P. Chen and C. C. Chou (Eds.), pp. 236248, John Wiley & Sons, Inc., 1993.
Mikuz, O., and Budizeck, L., "Sugar Storage in Silos," Sugar Technol. Rev., 13, 53129 (1986).
Burke, R. T., "Conditioning of White Sugar in Silos," Proc. Sugar Industry Technologists, Vol. 55, pp. 181195, 1996.
van der Poel, P. W., Schiweck, H., and Schwartz, T. (Eds.), "Liquid Sugars: Manufacture and Properties," in Sugar Technology, Beet and Cane Sugar Manufacture, pp. 951956, Verlag Dr Albert Bartens KG, Berlin, 1998.
Awad, A., and Chen, A. C., "A New Generation of Sucrose Products Made by Cocrystallization," Food Technol., 47, 146148 (1993).
Sugar and Sweetener Situation and Outlook Report, SSS-232, USDA, Economic Research Service, Sept. 2001.
Jensen, A. S., "Steam Drying of Beet Pulp and Bagasse," Paper presented at Conference of Sugar Processing Research, New Orleans, LA, March 2002, Proceedings in press.
Tjebbes, J., "Utilization of Fiber and Other Nonsugar Products from Sugarbeet," in Chemistry and Processing of Sugarbeet and Sugarcane, pp. 139145, M. A. Clarke and M. A. Godshall (Eds.), Elsevier, Amsterdam, 1988.
Sargent, D., Philip, E. J., and Cubero, M. T. G., "Lime Reduction in Juice Purification," Proc. Conf. on Sugar Processing Research, pp. 2139, 1998.
Tebble, I., "Use of Membrane Filtration to Reduce Lime Usage in Purification," paper presented at the Conference on Sugar Processing Research, New Orleans, LA, Mar. 2002.
McGinnis, R. A., Beet-Sugar Technology, Beet Sugar Development Foundation, Colorado, 1982.
Kearney, M., "Simulated Moving-bed Technology Applied to Chromatographic Recovery of Sucrose from Sucrose Syrups," pp. 291304, Proc. Conference on Sugar Processing Research, 1990.
Paananen, H. A., "Trends in the Chromatographic Separation of Molasses," in Separation Processes in the Sugar Industry, M. A. Clarke (Ed.), Proceedings of a Workshop, April 1819, 1996, New Orleans, published by Sugar Processing Research Institute, Inc., New Orleans, Louisiana, USA, Oct. 1996.
Hongisto, H., and Hikkala, H., "Desugarization of Cane Molasses by the Finnsugar Chromatographic Separation Process," pp. 30313037, XVI Congress, International Society of Sugar Cane Technolgoists, Sao Paulo, Brazil, 1978.
Bates, F., "Polarimetry, Saccharimetry and the Sugars," pp. 767779, Circular C440 of the National Bureau of Standards, 1942.
Cane Sugar Handbook, 12th ed. (Section 7) "Pan Control and Use of Instruments," pp. 236248, J. C. P. Chen and C. C. Chou (Eds.), John Wiley & Sons, Inc., 1993.
Theisen, K.-H., Senkowski, K., and Diringer, T., "Concentration Measurement for Continuous Sugar Crystallization Processes," Paper presented at 31st American Society of Sugar Beet Technologists, Vancouver, 2001.
Anon., Sugar and Sweetener: Background, Economic Research Service, USDA, May 10, 2001; http://www.ers.usda.gov/briefing/sugar/background.htm
F.O. Lichts World Sugar Statistics, pp. 1519, F.O. Licht, GmbH., 2002.

Sugar and Sweetener, Situation and Outlook Yearbook, SSS-231, Economic Research Service, U.S. Department of Agriculture, May 2001.
Koo, W. W., "The U.S. Cane and Beet Sugar Industry Under Alternative Trade Liberalization Policy Options," Agricultural Economics Report No. 434, Jan. 2000.
Sugar and Sweetener Situation & Outlook Report, SSS-231, Economic Research Service, USDA, May 2001.
Hickson, J. L., "The Potential for Industrial Uses of Sucrose," in Sugar: Science and Technology, G.G. Birch and K.J. Parker (Eds.), pp. 151180, Applied Science Publishers, London, 1979.
Kollonitsch, V., Sucrose Chemicals, pp. 17105 and 173221, The International Sugar Research Foundation, Inc., 1970.
Lichtenthaler, F. W. (Ed.), Carbohydrates as Organic Raw Materials, VCH, Germany, 1990.
Sachinvala, N. D., et al., "Use of Sucrose-based Epoxy Formulations and Cellulosic Fibers in the Design, Preparation and Screening of New Composite Insulation Materials," Polymers for Advanced Technologies, 13, 6679 (2002).
Godshall, M. A., "Future Directions for the Sugar Industry," Internat. Sugar J., 103, 378384 (2001).
Buzzanell, P. J., "U.S. Corn Sweeteners: Recent Trends and Future Prospects," F.O. Lichts World Sugar and Sweeter Yearbook 1997/98, pp. D20D34.
Sugar and Sweetener Situation and Outlook Report, SSS-232, Economic Research Service, U.S. Department of Agriculture, Sept. 2001.
Corn Industries Research Foundation, Corn Syrups and Sugars, Washington, DC, 1965.
Kooreman, J. A., "Physical and Chemical Characteristics of Enzyme Converted Syrup," Manufacturing Confectioner, June, 3590 (1955).
MacAllister, R. V., Advances in Carbohydrate Chemistry and Biochemistry, Vol. 36, R.S. Tipson and D. Horton (Eds.), Academic Press, New York, 1979.
Olsen, H. S., "Enzymes in Starch Modification," in Enzymes in Food Technology, R.J. Whitehurst and B.A. Law (Eds.), pp. 200228, CRC Press, 2002.
Bolling, C., and Suarez, N. R., "The Brazilian Sugar Industry: Recent Developments," Sugar and Sweetener Situation and Outlook Report, SSS-232, pp. 1418, Economic Research Service, USDA, September 2001.
Godshall, M. A., "The Role of Carbohydrates in Flavor Development," Food Technol., 42, 7172, 74, 76, 78 (1988).
Godshall, M. A., "Use of Sucrose as a Sweetener in Foods," Cereal Foods World, 35, 384389 (1990).
Calorie Control Council. http://www.caloriecontrol.org/erythritol.html
Embuscado, M. E., "Erythritol, A Unique Low-calorie Alternative Sweetener," 2000 IFT Annual Meeting, Abstract No. l 72-2.
Roh, H. J., Kim, P., Park, Y. C., and Choi, J. H., "Bioconversion of D-galactose into D-tagatose by Expression of L-arabinose Isomerase," Biotechnol. Appl. Biochem., 31, 14 (2000).
Chaen, H., "Production of Trehalose by a Novel Enzymatic Method and Its Applications," J. Appl. Glycosci., 44, 115120 (1997).
Ennen, S., "Trehalose gets GRAS," Food Processing, August, 44 (2001).
Anon., Sugar and Sweetener: Background, Economic Research Service, USDA, May 10, 2001; http://www.ers.usda.gov/briefing/sugar/background.htm

10.1 Introduction

Phosphates, compounds of the element phosphorous, are produced from relatively abundant supplies of phosphate rock.

The major use of phosphate is to supply phosphorous, one of the three essential plant foods, nitrogen, phosphorus, and potassium. Phosphate rock extraction from its ore, and its subsequent conversion into fertilizer materials and industrial chemicals, is a relatively mature art. Single superphosphate, a mixture of monocalcium monohydrate and gypsum formed by the reaction of sulfuric acid with phosphate rock, has been used as a fertilizer since the mid-1800s.

Phosphoric acid, derived by the treatment of phosphate rock with sulfuric acid so as to produce gypsum in a separable form, was manufactured in many locations by batch and countercurrent decantation methods in the 1920s.

Phosphoric acid produced by the later process is called the "wet process acid" to distinguish it from "furnace phosphoric acid," which is produced by hydrating the phosphorous pentoxide formed by burning elemental phosphorous in air. The "wet" processes utilized for the bulk of today's phosphoric acid and fertilizer production trace their origins to the early 1930s.1

Higher-purity industrial and food-grade phosphates, until recently, were most often derived from furnace processes. New plants recover purified phosphoric acid suitable for food-grade uses from relatively impure wet process acid, using solvent extraction technology.

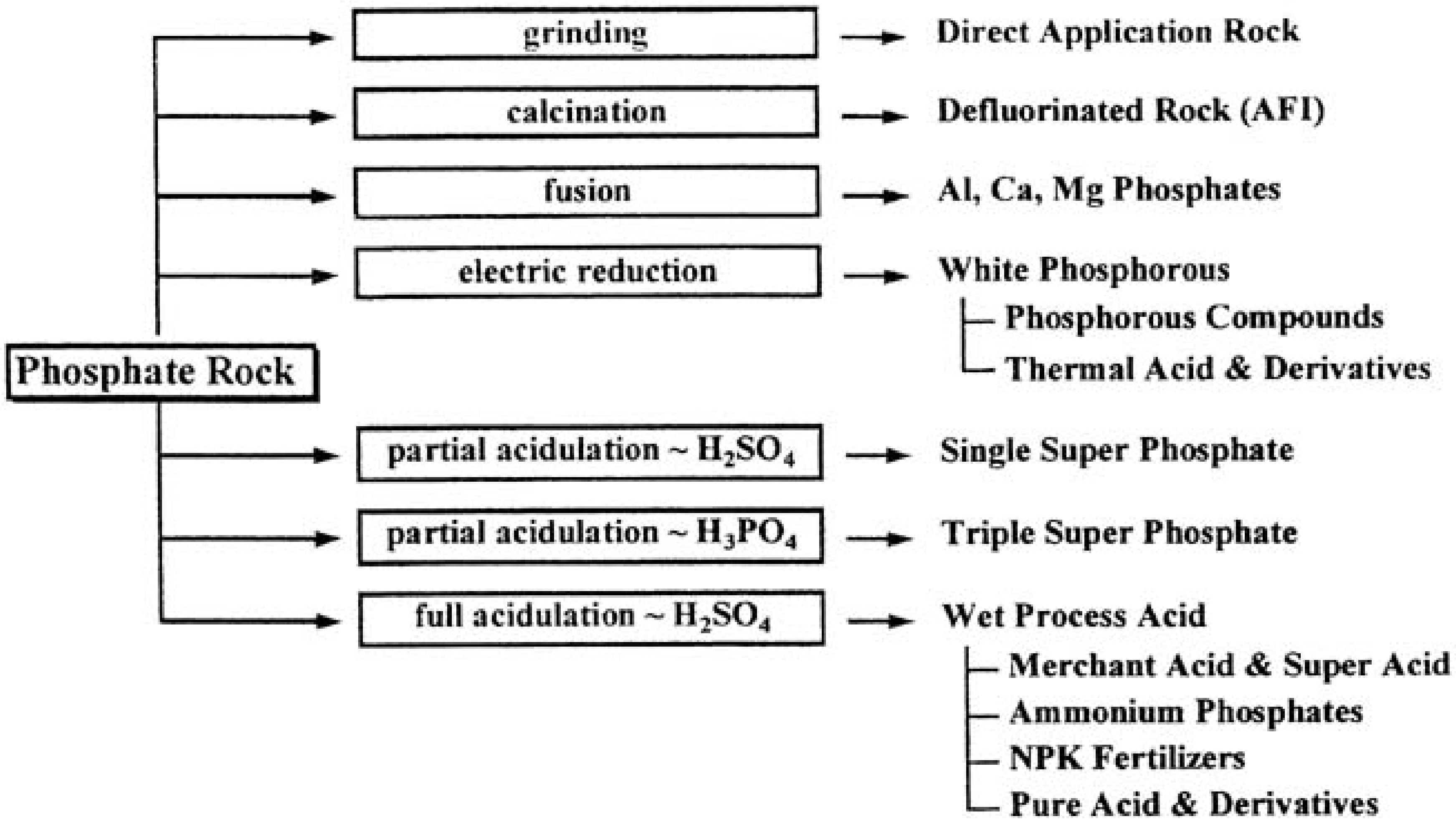
10.2 Phosphate Rock

Naturally occurring mineral products having sufficient phosphate content to be of commercial value are classified as phosphate rock. The grade or phosphate content of these products has been traditionally reported as percent $Ca_3(PO_4)_2$, which is referred to as bone phosphate of lime (BPL), tricalcium phosphate (TCP), or triphosphate of lime (TPL). Stoichiometric factors relating traditional rock analysis to other commonly used analytical terms are 100.00 percent BPL = 45.77 percent P_2O_5 = 19.96 percent P.

International trade of phosphate rock is based on dry metric tons, whereas U.S. domestic sales are in short tons, on a dry or as is moisture basis. Specifications for purchased rock may address grade, particle size, moisture content, and chemical impurities such as CaO, MgO, SiO_2, Al_2O_3, Fe_2O_3, F, Cl, Na_2O, and K_2O. The content of organic material and heavy metals is also of importance.

The treatments by which phosphate rock is commonly converted to fertilizers and chemicals are summarized in Fig. 10.1.

Fig. 10.1 Phosphate rock treatments and end products.



Minerals

The most common and widely distributed phosphate minerals are the apatite group, with the general formula $Ca_{10}(PO_4)_6(X)_2$. The apatite is designated as fluorapatite, hydroxyapatite, or chlorapatite, when X = F, OH, or Cl, respectively,. The most abundant sedimentary apatite is carbonate fluorapatite (francolite). Relative to pure fluorapatite, francolite is characterized by the substitution of Na and Mg for Ca and of carbonate and fluoride for phosphate. An empirical formula for francolite is given below and the chemical compositions of the end members of the fluorapatitefrancolite series, as quantified by McClellan,2 are given on Table 10.1.

Francolite: $Ca_{10-a-b}Na_aMg_bPO_4)_{6-x}(CO_3)_xF_{0.4x}F_2$

TABLE 10.1 Composition of the FluorapatiteFrancolite End Series

Constituent	Fluorapatite	Francolite
%CaO	55.60	55.40
%P_2O_5	42.20	34.00
%CO_2		6.30
%F	3.77	5.04
%Na_2O		1.40
%MgO		0.70

Source: McClellan.2

Van Kauwenberg3 has described the mineralogy and alteration of phosphate ores in Florida. Mineralogical composition varies by particle size. Francolite and then quartz are the most abundant minerals for plus 20-mesh particles (pebble), while the reverse is true for particles in the 20200-mesh fraction (flotation feed). In the minus 200-mesh size fraction (clay waste) quartz, francolite, wavellite, crandallite, goethite, dolomite, and a variety of clay minerals such as smectite, kaolinite, illite, and palygorskite occur.

Resources and Ores

Naturally occurring phosphates exist, or originated, as accessory minerals in igneous rocks. Prolonged weathering gradually converts the water-insoluble apatite into dissolved compounds that accumulate in the world's oceans. Sedimentary marine deposits (phosphorites) are formed when phosphorous compounds are precipitated by chemical or biological reactions. Bernardi4 describes secondary enrichment as an important aspect in the formation of sedimentary deposits.

McKelvey5 reported that the earth's crust contains an average of 0.27 percent P_2O_5, most of which occurs as apatite species. Sedimentary rocks, which predominate at the earth's surface, host the majority of commercial phosphate deposits discovered to date. Igneous rocks, which make up about 95 percent of the earth's crust, contain few phosphate deposits of commercial value. Guano deposits formed from the droppings of sea birds or bats are of minor importance, as are guano-related deposits. Northolt6,7 describes known phosphate deposits in the world, and estimates that identified phosphate resources in North America total more than 35 billion metric tons.

Resources are typically quantified as in situ tons of phosphatic material, without regard for economic criteria. However, it is preferable to quantify phosphate reserves as tons of phosphate rock recoverable according to specified economic, chemical, and regulatory criteria.8 The definition of phosphate reserves therefore requires an integrated program of geological exploration, laboratory testing, and classification using applicable criteria.

Significant commercial deposits of sedimentary phosphate ore occur in the United States, the Former Soviet Union, Morocco, China, Jordan, and Tunisia, and lesser deposits are mined in many other countries. Although phosphorite ores generally are classified as having siliceous or carbonate gangue minerals, soluble salts and organic material are also of concern. The phosphate content of the ores, depending on conditions of deposition and secondary enrichment, ranges from 10 percent to more than 70 percent BPL. The recovery of by-products from phosphorite ores is uncommon; however, uranium has been extracted commercially from phosphoric acid.

The types of igneous rock in which commercial deposits of apatite have been found are nephelineseyenite and carbonatites. The apatite deposits in Russia's Kola Peninsula are associated with nepheline seyenite. Carbonatite deposits that are mined for their phosphate content include Siilinjarvi in Finland, Jacupiranga and Araxa in Brazil, Phalaborwa in South Africa, and Kapuskasing in Canada. By-product recovery from igneous phosphate ores is common. Nepheline ($NaAlSiO_4$) is recovered from the Russian ore, calcite from the Finnish ore, copper concentrate and baddelyite (ZrO_2) from the South African ore, and barite from Brazilian ore. A minor percentage of the world's phosphate rock production is recovered from guano-related phosphate deposits.

Mining

Phosphate ores are extracted from deep deposits by underground mining methods and from shallow deposits by surface mining methods. Underground mining tends to be more costly and therefore less common for phosphate deposits than surface mining. Because 1 ton of phosphate rock has only about 10 percent of the commercial value of 1 oz of gold, low-cost mining is imperative.

Sedimentary phosphate deposits are exploited by underground mining in China, Mexico, Morocco, and Tunisia. Generally, the flat laying ore is most economically extracted by room-and-pillar mining or long wall mining. Ore from deep sections of the igneous phosphate deposits in Russia's Kola Peninsula is mined by a block caving technique.

Many shallow deposits have unconsolidated ore covered by unconsolidated overburden. Large electric walking draglines are ideally suited for such deposits, as evidenced by their use at large capacity phosphate mines in the southeastern United States, Morocco, Jordan, Senegal, and Togo. Other deposits have overburden and ore that may be partially or fully consolidated. For these deposits, ripping or drilling and blasting are required to fragment the overburden and ore to the extent that they can be excavated. Power shovels, backhoes, and wheeled loaders are also commonly used for excavation. Figure 10.2 shows a hydraulic shovel loading phosphate ore into a haul truck in Jordan.

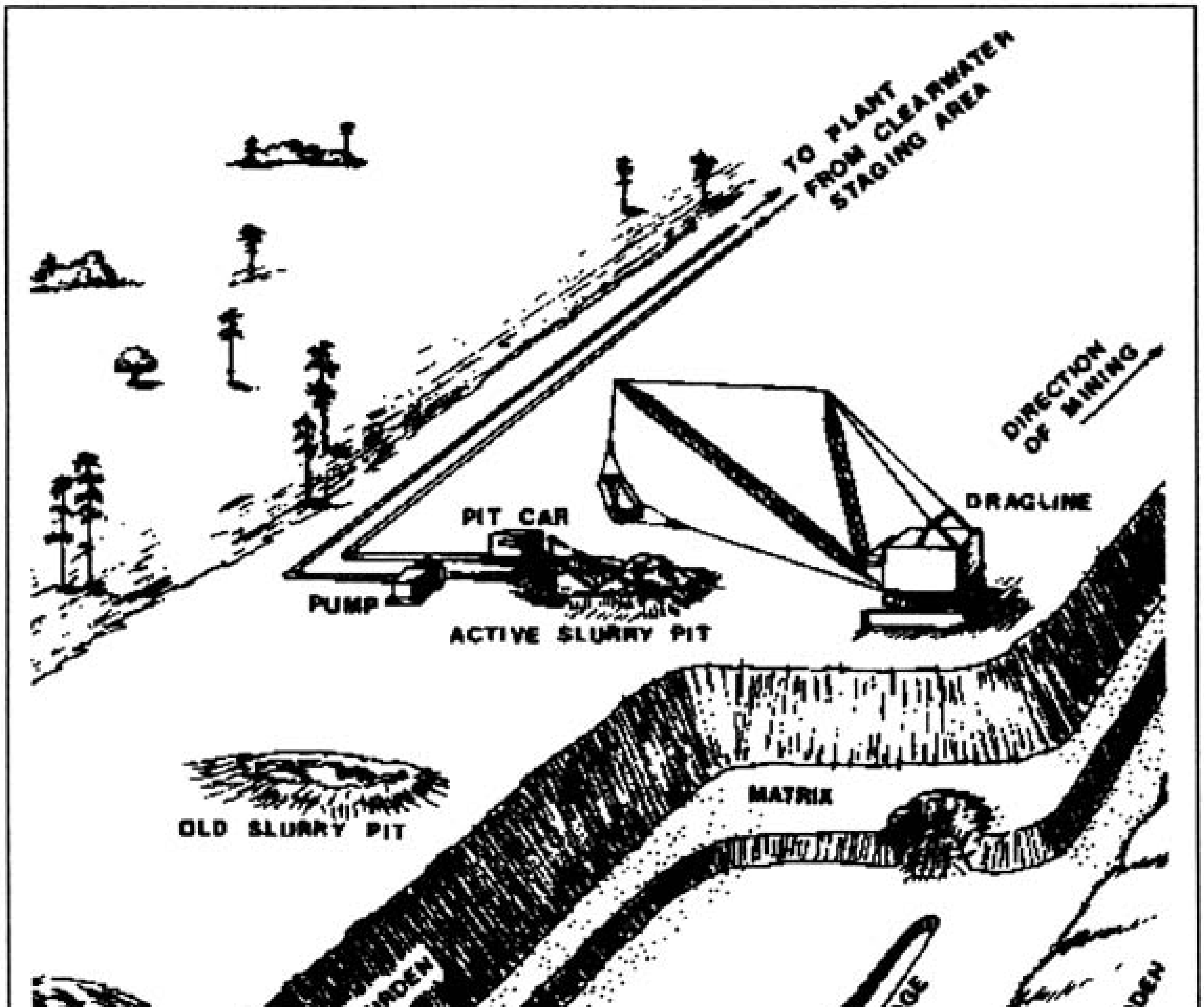
Fig. 10.2 Loading trucks with phosphate ore in Jordan.

Bucketwheel excavators are used for overburden removal at phosphate mines in eastern North Carolina, Senegal, and Togo. One mine in central Florida has used cutter head dredges for both overburden removal and ore excavation.

The method of transporting ore from the mine to the beneficiation plant depends on ore characteristics, mining methods, and local infrastructure. Railroad transport has been practiced in Russia and Iraq. Haul trucks and belt conveyors are commonly used in China, Jordan, Mexico, Morocco, Russia, Syria, Tunisia, and the western United States. Slurry pipelines of 1822 in. diameter, operating at less than 100 psig, are used exclusively in central Florida, north Florida, and eastern North Carolina. The pipelines may be extended up to 10 miles or more, by installing a series of centrifugal slurry pumps at 4000 ft intervals.

A typical Florida phosphate mining scheme, utilizing a dragline and slurry pipeline, is illustrated in Fig. 10.3. The dragline first exposes the phosphate ore (matrix) by stripping and casting the overburden into the adjacent mined area. The matrix is then dug by the dragline and placed in a slurry pit, where it is gunned with high-pressure water. Gunning the matrix, as shown in Fig. 10.4, transforms the unconsolidated ore into a slurry which is pumped to the beneficiation plant.

Fig. 10.3 Typical Florida phosphate mine.



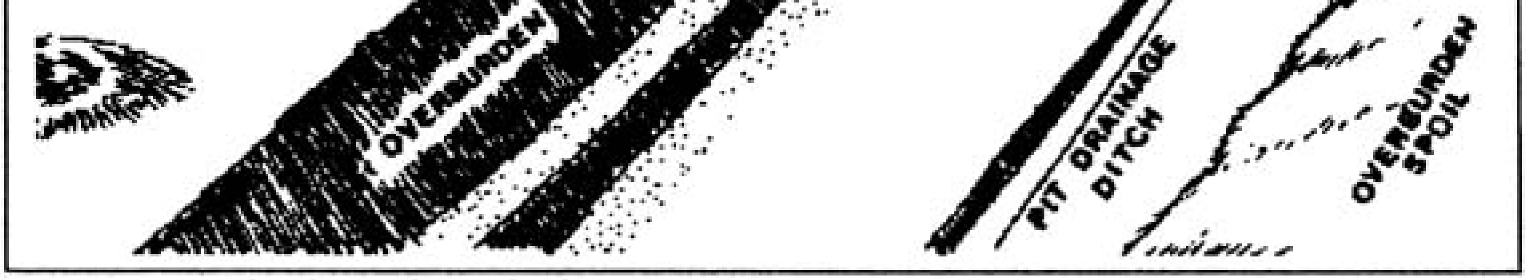


Fig. 10.4 Gunning phosphate rock in a Florida mine.

Beneficiation

Beneficiation, also known as mineral dressing or ore processing, may involve a variety of operations such as size reduction, size separation, mineral separation, dewatering, and thermal processing. Almost all phosphate ores require beneficiation to meet commercial specifications concerning particle size, moisture content, or chemical analyses.

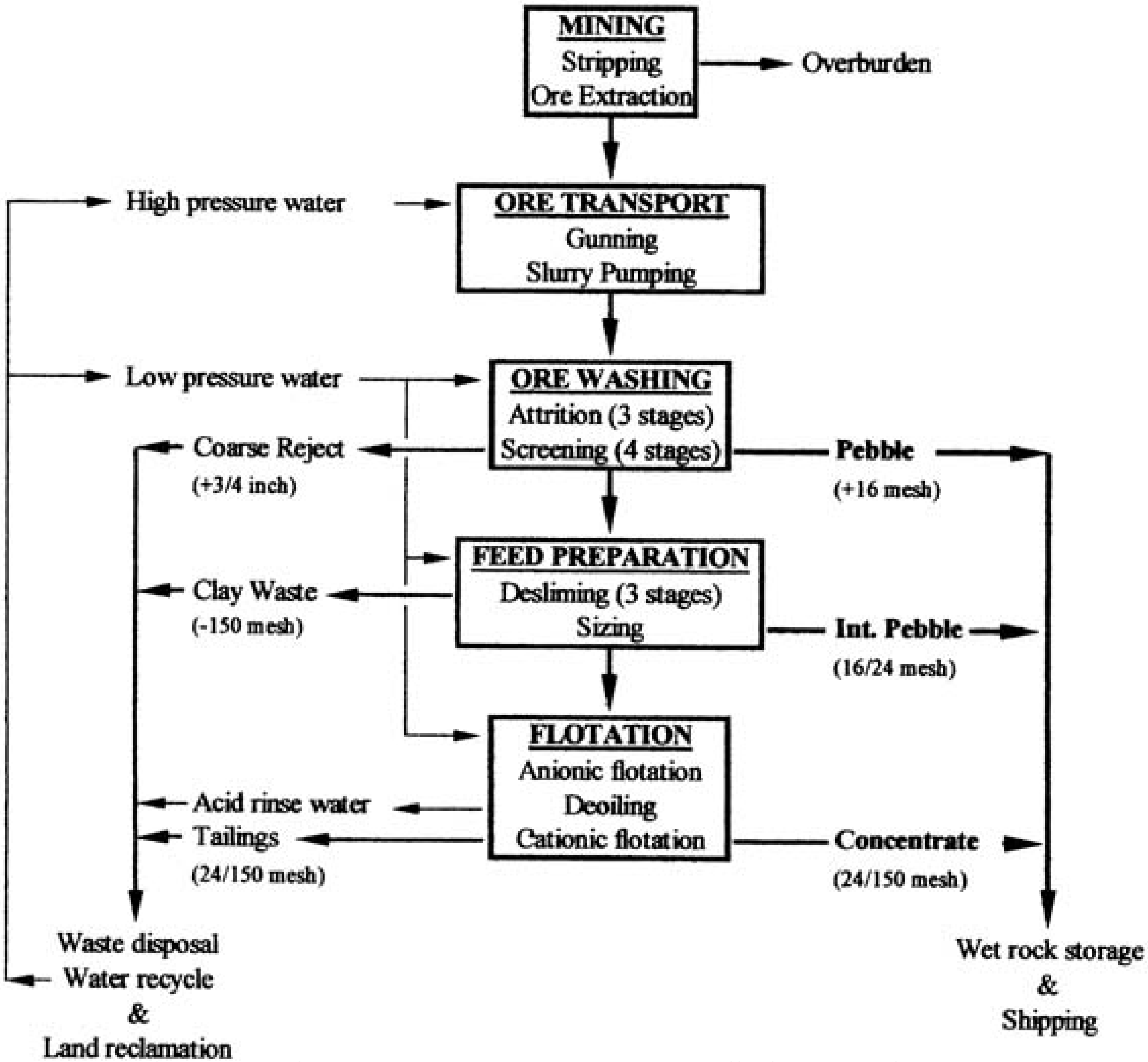
The usual first beneficiation operation is size reduction, which may be achieved by crushing, grinding, or disaggregating by scrubbing and washing. Particle size reduction liberates mineral species so that they can be separated. Size separation usually follows size reduction. When gangue minerals are more indurated than the phosphate, it is often practical to reject coarse waste material by wet or dry screening. Similarly, when gangue minerals are microcrystalline or softer than the phosphate, fine waste material may be rejected by wet or dry classification. Soluble chlorides, when present, must be removed from phosphate rock by washing with fresh water followed by dewatering.

Gangue minerals frequently have the same particle size as the phosphate mineral grains, and techniques such as heavy media separation, magnetic separation, or froth flotation are required. Heavy media separation is an appropriate process when liberation occurs at 16 mesh or coarser, and the phosphate mineral has a significantly higher density than the gangue (dolomite, calcite, quartz, shale). Low-intensity magnetic separation will remove highly magnetic minerals, such as magnetite,9 from phosphate. High-intensity magnetic separation will remove ankerite10 and other paramagnetic iron-bearing minerals11 from phosphate. Froth flotation is the most widely practiced operation for recovering phosphate rock from fines (-20 mesh). Variations of this process are used commercially to separate phosphate from barite, calcite, dolomite, feldspars, nepheline, phlogopite, and quartz. Flotation plants have been constructed and operated in Brazil, Canada, China, Finland, Jordan, Mexico, Russia, Senegal, South Africa and the United States. Dolomite flotation from phosphate is of increasing interest. In the United States, one commercial plant has a dolomite flotation circuit,12 and other dolomite flotation processes have been demonstrated by pilot plant testing of Florida low-grade pebble.13

Electrostatic removal of quartz from apatite is technically feasible although it is impractical and costly.14

A generic scheme for mining and beneficiating central Florida phosphate ore is presented in Fig. 10.5 as a block flow diagram. Liberation of phosphate from the gangue occurs during ore transport and washing. First, a low-cost product, called pebble, is recovered by screening the ore at about 16 mesh. Secondly, a low-grade product (intermediate pebble) is recovered by sizing the flotation feed at about 24 mesh. Clays are removed from the flotation feed by three or more stages of desliming with hydrocyclones. Finally, a more expensive but higher grade concentrate is obtained by a two-stage flotation process. A rougher phosphate concentrate is recovered by direct flotation with anionic reagents. After deoiling with sulfuric acid and rinsing with water, the rougher concentrate is conditioned with cationic reagents and subjected to inverse flotation. The phosphate rock product, comprised of pebble, intermediate pebble, and floatation concentrate, dispatched to a chemical plant for conversion to phosphoric acid.

Fig. 10.5 Unconsolidated sedimentary ore beneficiation flow diagram.



The initial beneficiation steps for consolidated phosphate ores generally differ from those of unconsolidated ores. Figure 10.6 depicts the flow diagram for the San Juan de la Costa phosphate mine in Mexico. The high-grade ore, slightly more than 1 m in thickness, is extracted by room-and-pillar mining. Continuous miners rip ore from the mining face and load shuttle cars, which transfer the ore to feeder-breakers and a belt conveyor systems. Outside the mine, ore is loaded into haul trucks and transported to the beneficiation plant. Liberation of the phosphate is accomplished by crushing to 9 mm followed by grinding to 0.7 mm. Following grinding, the ore is deslimed, attrition scrubbed, and deslimed a second time to remove clays and carbonate minerals from the flotation feed. The feed is conditioned with anionic reagents and subjected to rougher and cleaner direct flotation using sea water. The concentrate is washed with fresh water to remove sea salt prior to use in the chemical plant.

Fig. 10.6 Consolidated sedimentary ore beneficiation flow diagram.

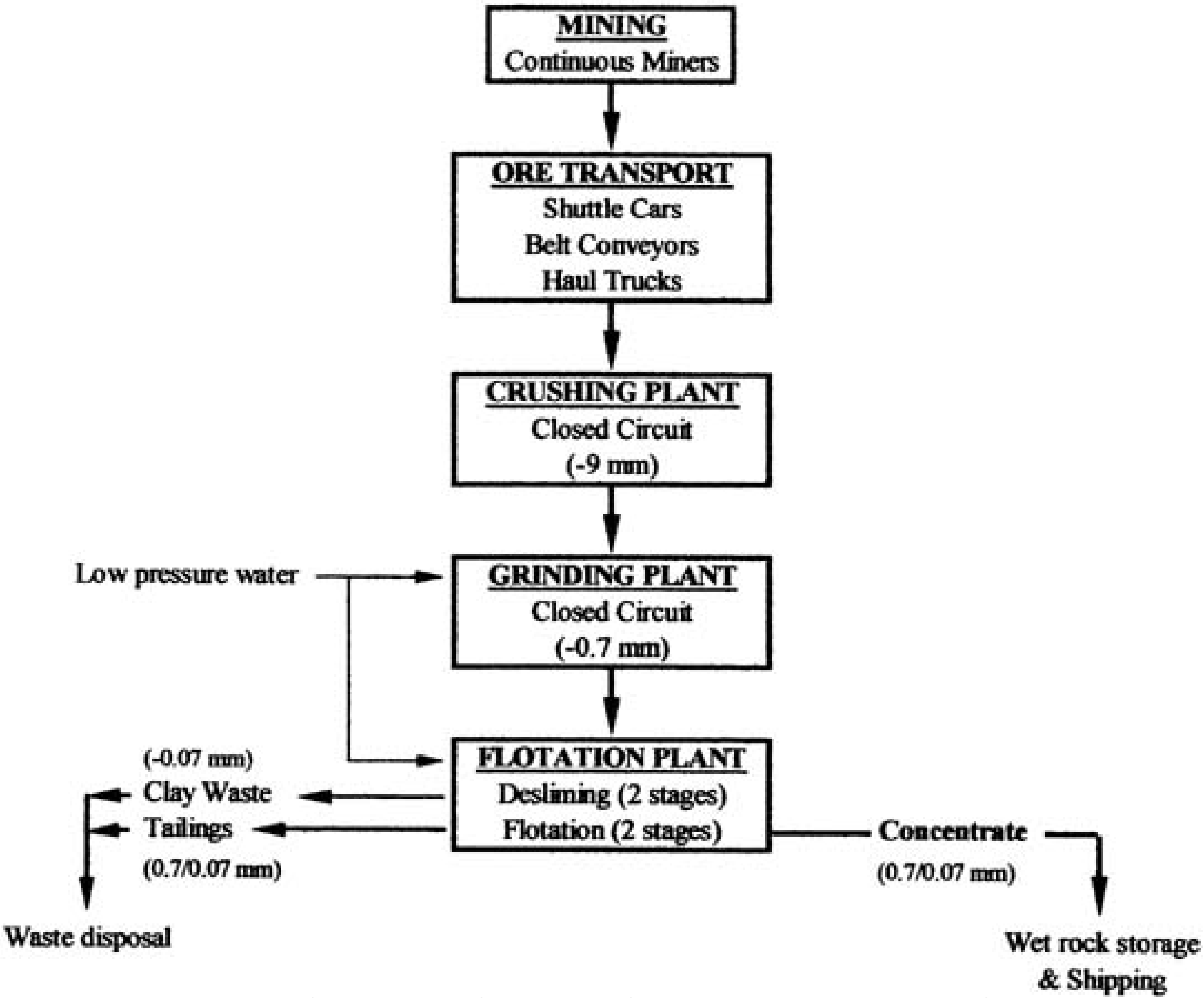
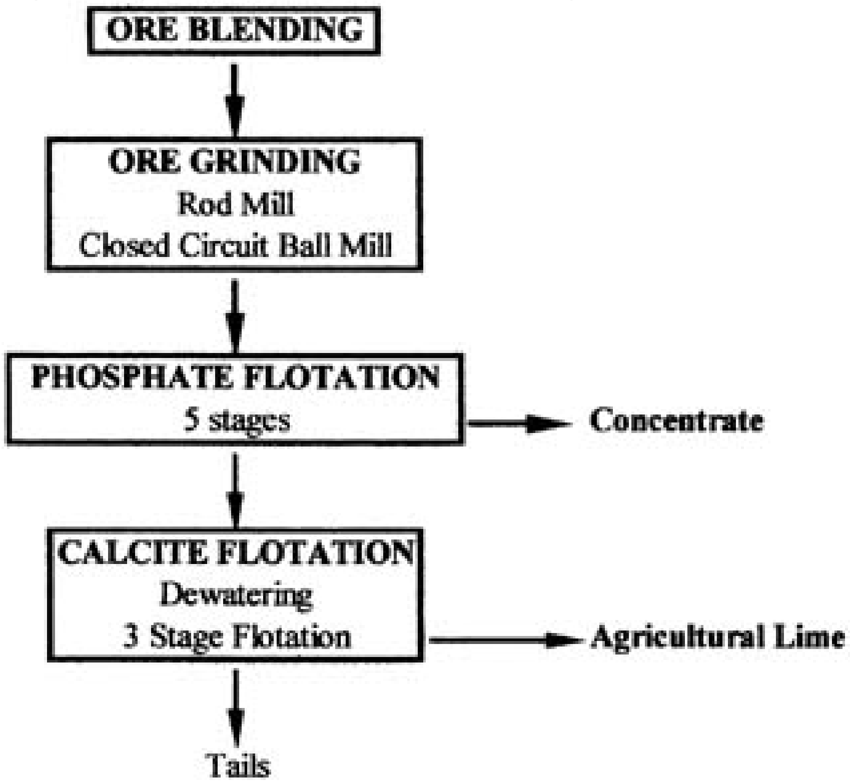


Figure 10.7 presents a simplified mid-1980s flow diagram for the Siilinjarvi phosphate beneficiation plant in Finland. The low-grade igneous ore to this plant is carefully blended to avoid changes in plant feed characteristics. Liberation of phosphate is achieved by rod milling followed by closed circuit ball milling.

The ground ore is conditioned without desliming, at pH 11, using an amphoteric flotation reagent. A phosphate concentrate is recovered by five stages of direct flotation. The tailings from phosphate flotation are dewatered, conditioned with anionic reagents, and subjected to three stages of direct flotation to recover calcite, which is used as agricultural lime.

Fig. 10.7 Igneous ore beneficiation flow diagram.



Production and Value

The U.S. marketable production of phosphate rock over the last three decades is summarized in Table 10.2. Significant increases in unit value during the 1970s stimulated increases in tonnage and total value. Over the last two decades U.S. marketable production of phosphate rock has declined in terms of tonnage and total value, while the unit value (in current dollars FOB mine) has fluctuated between \$21 and \$24 per metric ton. If the post-1975 data were adjusted for inflation, a considerably more pronounced reduction in total value would be evident, as would a decline in unit value (in constant dollars). Because the unit value of phosphate rock has not kept up with inflation, it is increasingly difficult to justify the capital investment for new mines. Consequently, as U.S. phosphate mines have been depleted, their production capacities have not been replaced on a one-for-one basis. Nevertheless, as shown on Table 10.3, the United States remains the leading producer of phosphate rock.

TABLE 10.2 Phosphate Rock Value (U.S. Marketable Production)

Year	Total Value ($billion)	Unit Value ($/metric ton)
1970	0.204	5.80
1975	1.107	25.00
1980	1.148	21.34
1985	1.236	24.31
1990	1.075	23.20
1995	0.947	21.75
2000	0.932	24.14

Source: Phosphate Rock, Annual Review, USBOM, and USGS Mineral Industry Surveys.

TABLE 10.3 World Production of Phosphate Rock (Million Metric Tons)

	1970	1980	1990	2000
United States	35.2	53.5	45.4	39.2
Morocco	11.5	18.9	21.2	21.6
China	1.2	10.7	17.3	19.4
USSR/Russia	20.5	24.7	36.9	11.7
Tunisia	3.0	4.6	6.6	8.3
Jordan	1.2	4.3	5.9	5.5
Other	12.8	13.1	22.8	26.1
Total World	85.3	129.8	156.0	131.7

Sources: Zellars-Williams Company (1979/1980) and the International Fertilizer Industry Association (1990/2000).

The global demand for food stimulated increased fertilizer usage and consequently increased phosphate rock consumption through 1990. Over the following decade, rock production has been somewhat flat. Although production capacity has declined in the United States, new mine capacity has been added elsewhere. From Table 10.3 it is evident that significant increased phosphate rock production capability has been added in Morocco, China, Tunisia, Jordan, and the countries comprising Other.

Many countries use indigenous phosphate rock as a source of phosphorous for industrial chemicals and fertilizers. Few countries are self-sufficient and supplemental sources of phosphate rock are essential. The cost of imported phosphate rock is markedly influenced by freight, and the low-cost producer is not necessarily the low-cost supplier.

The reduction in U.S. marketable production of phosphate rock has been accompanied by a reduction in exported phosphate rock. During the 1980s, U.S. phosphate rock exports declined from 14.3 to 7.8 million metric tons per year. As shown in Table 10.4, the decline in U.S. phosphate rock exports continued throughout the 1990s. Morocco has now replaced the United States as the major exporter of phosphate rock.

TABLE 10.4 Major Exporters of Phosphate Rock (Million Metric Tons Exported)

	1990	1995	2000
Morocco	11.7	9.4	10.5
USSR/Russia	2.6	1.7	4.5
China		1.0	3.4
Jordan	4.9	3.9	3.1
Togo	2.4	2.7	1.2
Senegal	1.4	0.9	0.4
United States	6.9	3.1	0.2

Source: International Fertilizer Industry Association.

10.3 Chemical Processing of Phosphate Rock

Phosphate rock is converted into usable chemicals by two methods. In the first, the rock is charged to an electric furnace with silica and coke to produce elemental phosphorus. The phosphorus then is converted into phosphoric acid and other compounds. In the second, the phosphate rock is reacted with sulfuric acid in a medium of phosphoric acid and calcium sulfate crystals to form dilute, impure phosphoric acid, The acid is separated and used to make fertilizers. This is known as the wet process method.15

The wet process is further divided into two subprocesses based on the type of calcium sulfate crystal produced. The dihydrate process wherein gypsum (CaSO4·2H2O) is produced has been the dominant process, but processes making hemihydrate (CaSO41/2H2O) have become more important over the past decade.

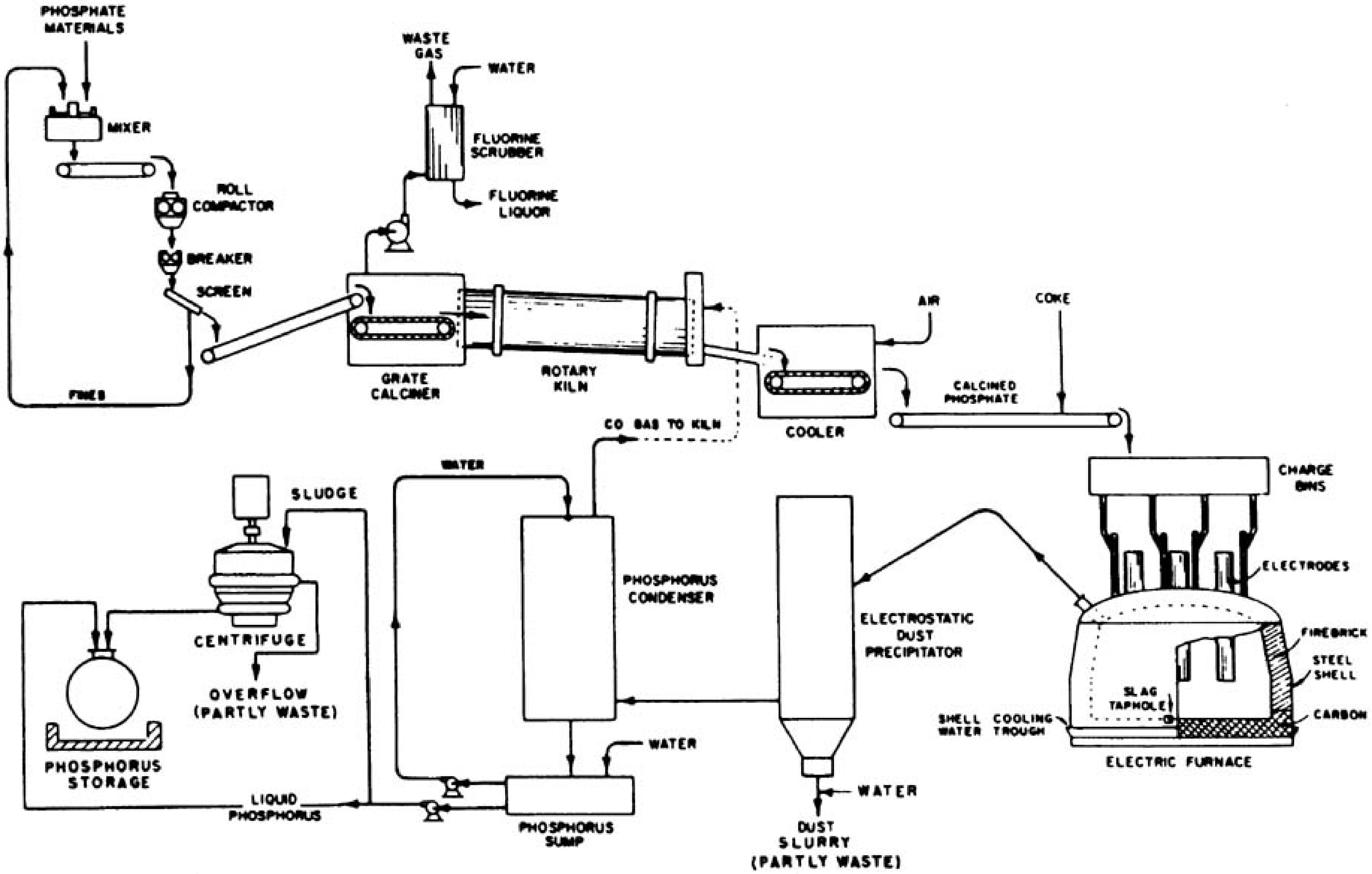
Thermal Process for Phosphorus and Phosphoric Acid

The furnace or thermal process is shown in Fig. 10.8. The approximate reaction is:

$$2\mathrm{Ca_3(PO_4)_2} + 6\mathrm{SiO_2} + 10\mathrm{C}$$

$$\rightarrow 6\mathrm{CaSiO_3} + \mathrm{P_4} + 10\mathrm{CO}$$

Fig. 10.8 Electric furnace process for production of elemental phosphorus.



The phosphorus leaves the furnace as a vapor and is condensed by direct contact with water. Phosphoric acid of high purity is made by burning phosphorus with air and hydrating the resulting P2O5 with water, according to the reaction:

$$\mathrm{P_4} + 5\mathrm{O_2} + 6\mathrm{H_2O} \rightarrow \mathrm{H_3PO_4}$$

If even less water is used for hydration, a product known as polyphosphoric acid results. Ordinary phosphoric acid is a solution of the monomer, H3PO4, in water, and is called orthophosphoric acid. If a molecule of water is removed between two orthophosphate molecules, the dimer, pyrophosphoric acid, H4P2O7, is formed. Similarly, the trimer and higher polymers can be made. Superphosphoric acid is a mixture of orthophosphoric acid and polyphosphoric acid and is now made from wet process acid as described later in this chapter.

In 1990, there were about eight plants in operation, some with multiple furnaces, in the United States. By 20002001, only one plant remained. New emission standards, high capital and operating costs, and competitive lower-cost wet acid purification technology have spelled doom for most of the furnace plants. A more thorough discussion of electric furnace processing is to be found in the ninth edition of this Handbook.

Industrial Phosphates

Furnace phosphorus currently produced in the United States is consumed to make compounds such as phosphorus pentoxide, phosphoric trichloride, and phosphorus pentasulfide, which find use in the preparation of drying agents, plasticizers, oil additives, fire retardants, and insecticides. These products are derived from phosphorus and, therefore, cannot be made from purified wet process acid so there is a continuing demand for a small amount of thermal product. However, for orthophosphoric acid use, the purified wet process acid is cheaper.16 This has now replaced furnace acid in soft drinks, candy, baked goods, and various other food products. It is also used for pickling metals.

The cheapest and most important salts of purified phosphoric acid are the sodium salts, made by reacting the acid with sodium carbonate or sodium hydroxide. Sodium phosphates may be classified in a general way as (1) orthophosphates, (2) crystalline condensed phosphates, and (3) glassy condensed phosphates.

Three sodium orthophosphates can be prepared, depending on whether one, two, or three hydrogen atoms are replaced by sodium. Monosodium phosphate is formed in the following reaction:

$$2\mathrm{H_3PO_4} + \mathrm{Na_2CO_3}$$

$$\rightarrow 2\mathrm{NaH_2PO_4} + \mathrm{H_2O} + \mathrm{CO_2}$$

Sodium carbonate also is used to make disodium phosphate, Na2HPO4, but sodium hydroxide must be used to replace the third hydrogen in trisodium phosphate, Na3HPO4. The orthophosphates have a wide range of uses in industry.

Condensed phosphates are made by eliminating water from orthophosphates. The most important compound is sodium tripolyphosphate (STPP), made according to the following reaction:

$$2\mathrm{Na_2HPO_4} + \mathrm{NaH_2PO_4} \xrightarrow{-\mathrm{H_2O}} \mathrm{Na_5P_3O_{10}}$$

The most important use of sodium tripolyphosphate is as a builder in detergents. However, legislative restrictions on the use of phosphorus compounds in household detergents have caused a worldwide flattening of consumption.

Glassy condensed phosphates are represented by sodium hexametaphosphate, in which the O:P ratio is 3:1. There can be considerable variation in the Na2O to P2O5 ratio.

The principal use of the condensed phosphates is to sequester metallic ions in water. They form water-soluble complexes with the metals and prevent metallic compounds from precipitating to cause discoloration, scale, and sludges.

Wet Process Phosphoric Acid

In the wet process, phosphate rock is reacted in a slurry of phosphoric acid and calcium sulfate crystals containing a controlled quantity of sulfuric acid. The simplified reactions for the dihydrate process is as follows:

$$\mathrm{Ca_3(PO_4)_2} + 3\mathrm{H_2SO_4}$$

$$\rightarrow 3\mathrm{CaSO_4} \cdot 2\mathrm{H_2O} + 2\mathrm{H_3PO_4}$$

Until 1969, virtually all wet process acid was made at maximum strengths of 2632 percent P2O5, separating the calcium sulfate as gypsum containing two molecules of water. Since 1969, a substantial number of commercial hemihydrate process plants have been built in various modes. In the hemihydrate process, reaction conditions are higher in temperature and phosphoric acid strength so that the stable solid phase is calcium sulfate with one-half molecule of water.

The hemihydrate process may be a single-stage process, known as HH, in which the hemihydrate solids are the waste product, or the process can have a second step in which the hemihydrate is re-crystallized to gypsum, known as hemi-dihydrate (HDH). In the hemihydrate process, acid strengths of 38 percent P2O5 to about 42 percent P2O5 normally are produced although strengths up to 50 percent can be produced under somewhat more difficult circumstances requiring more filter area.

The neutralization reaction of the above equation is conducted in one or more strongly agitated reaction vessels, whether in a gypsum or in a hemihydrate mode. The system is highly exothermic and the slurry is maintained at 8085°C for dihydrate processing, 95100°C for hemihydrate, by evaporative or air cooling. During the reaction of phosphate rock with sulfuric acid, fluorine is evolved and must be scrubbed from the vent gas.

Table 10.5 lists the production of phosphoric acid in the recent past, and shows the substantial changes in location of P2O5 production over the last decade. The East and West Europe fertilizer industry was in a steep decline by 1999. North Africa and the Middle East are continuing to expand based on their huge reserves of high-grade phosphate. China, with substantial deposits, has now begun to develop these. The years 2000 and 2001 show a decline in overall P2O5 over the figures shown for 1997, illustrating the cyclic nature of the world wide phosphate business.

TABLE 10.5 Annual Production of Phosphoric Acid, TPY P2O5

	1980	1988	1999
West Europe	3,669,700	3,322,735	1,588,100
East Europe	33,466,000	6,305,597	2,797,600
Africa	2,043,700	4,661,621	5,432,900

North America	10,016,800	10,411,968	11,439,000
Mexico	260,400	474,000	699,700
South America	303,100	762,806	979,600
Middle East	157,000	813,354	1,911,200
China	9,000	50,000	1,380,000
South and East Asia	1,414,300	1,637,380	1,616,300
Australia	180,000	107,650	
	21,520,000	28,547,111	27,788,400

Courtesy British Sulphur Corporation and The International Fertilizer Manufacturer's Association.

Dihydrate Process

The conventional dihydrate process, as operated over the last 70 years, was first demonstrated by the Dorr Company in the Cominco plant at Trail, British Columbia, in 1931.1 The principles discovered at that time for making an acid strength of up to 32 percent P_2O_5 and a highly filterable gypsum crystal are still used today. The salient features of the process included maintaining H_2SO_4 content in the digestion acid of about 2.03.0 percent. Reaction slurry was recycled at a ratio of 1015:1 relative to the volume of product slurry sent to the filter. The relatively higher recirculation of seed gypsum than had previously been practiced, together with control of the free sulfate, resulted in product acid well above the previous 22 percent P_2O_5 possible with batch and countercurrent decantation systems.

If the sulfate in the solution is allowed to exceed certain limits, depending on rock reactivity and particle size, the rock becomes coated with gypsum, and the reaction becomes uncontrollable. If the sulfate level is too low, the precipitated gypsum filters poorly and contains excessive quantities of crystallized (solid solution) P_2O_5, leading to excessive P_2O_5 losses. Over the years, the process has evolved to employ increased slurry recirculation and different reactor configurations, In addition, several different filter designs are available.

The principal dihydrate processes in use as of 2000 are shown in Table 10.6. During the decade 19902000, many of the smaller plants and even some larger than 450 tons per day P_2O_5 plants have been shut down due to environmental or market conditions. This has resulted in more production in Morocco, Jordan, and India, and less in Europe. Most of the world's phosphoric acid is produced by the dihydrate method, but there is likely to be increased production by the hemihydrate method because of advantages in some situations. The dihydrate processes listed have been modified over the years so that many slightly different configurations of the same process may exist.

TABLE 10.6 Phosphoric Acid Plants, Worldwide Dihydrate Process, 2000

Process	Number of Lines 450 MTPD or Larger
Prayon	2025
Rhone Poulenc	1015
Nissan H	7
Jacobs/DorrOliver	7
BadgerIsothermal	7

The unique feature of each process is the reactor system configuration. The reactor or reactors normally provide from 2 to 6 hr detention for the gypsum slurry. This is about 0.82.5 m3 of reactor volume per ton of P_2O_5 per day, meaning relatively large vessels totaling 1500 m3 to over 2000 m3 in size for large plants. The individual reactor systems are described briefly below.

The objective of the reaction system is to produce a highly filterable gypsum crystal that washes well, and that also contains a minimum amount of insoluble P_2O_5. The filterability of the gypsum slurry depends on the reactor configuration and on the rock source. Highly filterable gypsum slurries are produced from Florida, Togo, and Senegal phosphates. These rocks may provide more than double the P_2O_5 from a given filter when compared with gypsum produced from other sources. However, because phosphate rock is costly to transport, the use of local, less treatable phosphates can be quite economical. Table 10.7 lists the approximate filtration design rates for these groups of commercially available phosphate rock.

TABLE 10.7 Filtration Design Rates for Phosphoric Acid Produced from Various Rock Sources

Rock Source	Tons P_2O_5 Produced per m2 Active Area
Togo	
Florida	7.59.0
Senegal	
Morocco Khouribga	
Western U.S.	4.57.0
Kola (USSR)	
North Carolina	
Morocco Safi	
Algeria	2.55.0

The advantages of the dihydrate process vs. the various, newer hemihydrate configurations are as follows:

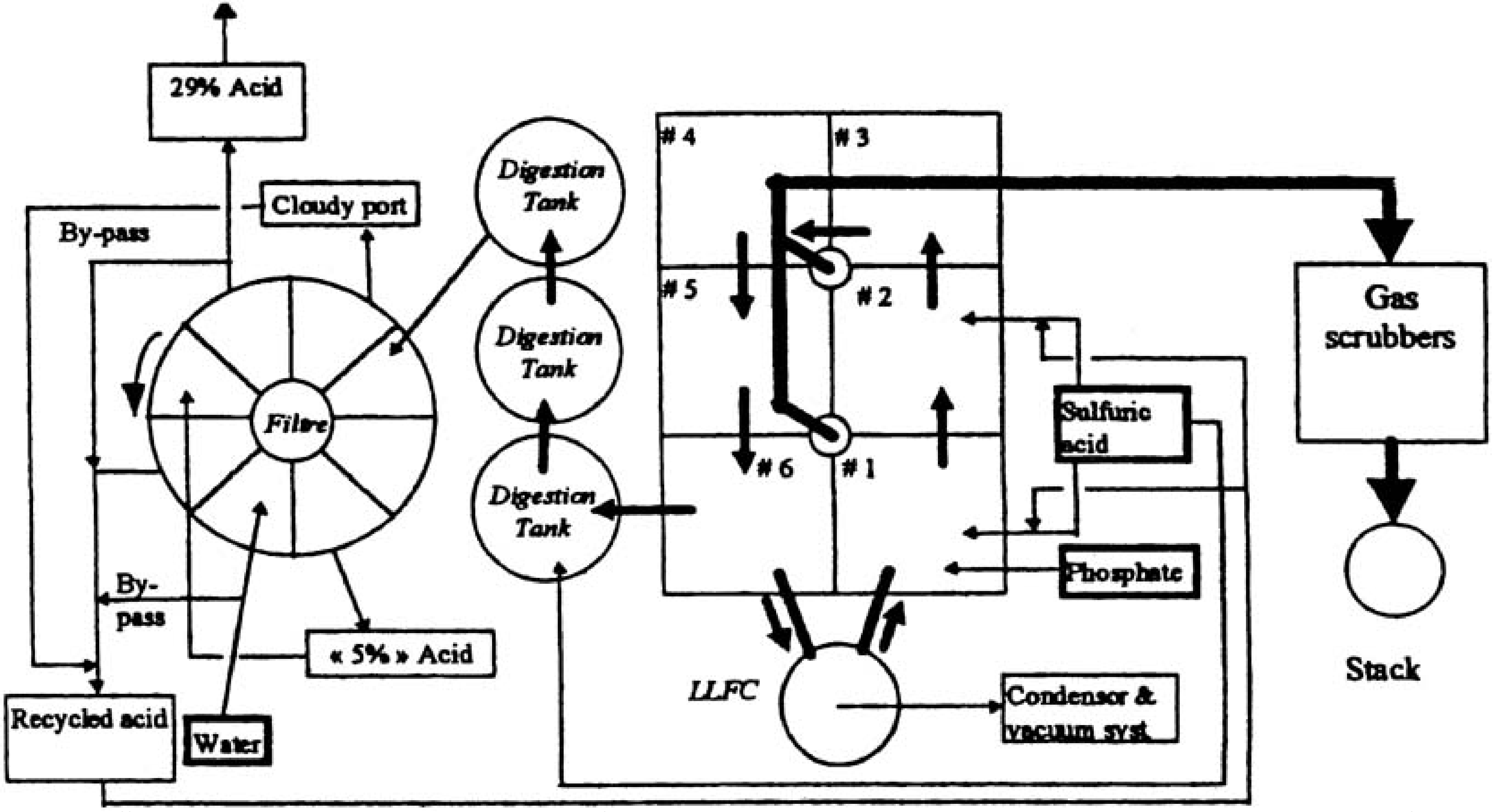
1. The water balance permits the use of wet rock slurry feed. This eliminates the cost of drying the rock and the dust nuisance. This is the principal benefit that has deterred most Florida producers from using the hemi routes. Because gypsum has two moles of water of hydration vs. one half mole for hemihydrate, and because product acid strength is 2628 percent P_2O_5 normally, the ground phosphate feed slurry can be as low as 65 percent solids for the dihydrate process. This still leaves enough makeup water for adequate washing of the gypsum. For hemihydrate processes the maximum moisture is generally considered to be about 1518 percent in the feed rock. Such a low moisture would call for a dewatering step for ground Florida pebble rock. Producers in Florida have thus far chosen not to go this route because of the difficulty and expense of dewatering.
2. Dihydrate gypsum, in the case of most phosphates, filters at relatively higher P_2O_5 throughput rates than hemihydrate, comparing dihydrate filter performance at 28 percent P_2O_5 and hemihydrate at 42 percent P_2O_5.
3. Dihydrate plants are proven at huge outputs, in excess of 2000 metric tons per day P_2O_5, thus offering economies of scale.
4. A vast amount of operating data and experience exists on a wide variety of phosphate rocks for the dihydrate mode.
5. Capital costs for the dihydrate system can be less than for the two-stage hemi processes even when evaporation and rock grinding are considered, because of the larger reaction vessels and two stages of filtration in the hemi process.
6. The yield across the filter for the dihydrate process, generally about 96 percent of the P_2O_5 fed, is about 34 percent above the single-stage hemi processes.
7. Dihydrate process maintenance costs are substantially less than those for hemi processes due to less severe process conditions. The on-stream factor is also higher for the average dihydrate facility.

Major Dihydrate Processes

Prayon Process

The Prayon process17 has evolved from the cascade system first used by the Dorr Company in the 1930s. Instead of round, steel vessels with rubber and brick lining, Prayon developed a multicompartmented reinforced concrete vessel, rectangular in shape, lined with a membrane and carbon brick in the early 1960s. The construction has proved to be exceedingly durable. A similar concrete construction is used by Jacobs and by Rhone Poulenc. A diagram of the Prayon Mark IV reactor configuration is shown in Fig. 10.9. The Prayon process uses vacuum cooling, a low-level vacuum chamber through which reaction slurry is circulated to maintain a reactor temperature of 8085°C. There are numerous installations of large Prayon plants in the United States, although the center of process technology is Belgium.

Fig. 10.9 Prayon Mark IV reaction and filtration system.



Prayon also offers the PrayonCentral Glass process, a name derived from its Japanese origins. In this process, gypsum is crystallized in a dihydrate mode in the first stage. After separation of most of the P_2O_5 from the slurry by centrifugation or filtration, the gypsum is recrystallized to hemihydrate employing the sulfuric acid ultimately required in the first stage. The recrystallized hemi is quite low in P_2O_5 and impurities and suitable for byproduct wallboard and plaster, and is readily washed in a second stage of filtration, followed by an agglomeration step that results in a semi-granular by-product. This process has allowed the Engis, Belgium, plant to continue operation by converting all of its gypsum to a saleable product. The Prayon Mark III and Mark IV dihydrate processes are used in Florida and Louisiana in some eight lines operating between 1350 and 2000 tons per day P_2O_5, and in one revamped plant in Jordan.

SpeichimRhone Poulenc Process

Most of Rhone Poulenc's existing plants are of a single stirred vessel configuration. However, a two-vessel arrangement, the Diplo system,18 has been offered and retrofitted into several of the previous single reactor plants.

The original Rhone Poulenc Single Tank system is an exceedingly simple plant. It is air-cooled by passing a flow of air over slurry splashers, and does not have a filter feed tank, the slurry simply overflowing the reactor to a vertical pipe on top of the filter feed pump. The newer Diplo system is said to offer a better yield, higher P_2O_5 product acid strength, and a more filterable gypsum by providing a better concentration gradient for gypsum growth.

Rhone Poulenc installations are located in France, Belgium, Morocco, Senegal, Brazil, and China.

Nissan H Process

In this process, the rock attack is done under conditions favoring the formation of an unstable hemihydrate. The slurry then is cooled and seeded to recrystallize to gypsum at high sulfate levels, producing a gypsum low in co-precipitated P2O5. Yields in the 9798 percent range are reported. The process, in operation since the 1960s, has been favored in many instances where the phosphogypsum can be used for wallboard or other building materials. Major installations are located in Japan and Morocco. A large plant at Pernis, Netherlands, has been closed and it appears that Nissan no longer offers phosphoric acid technology.

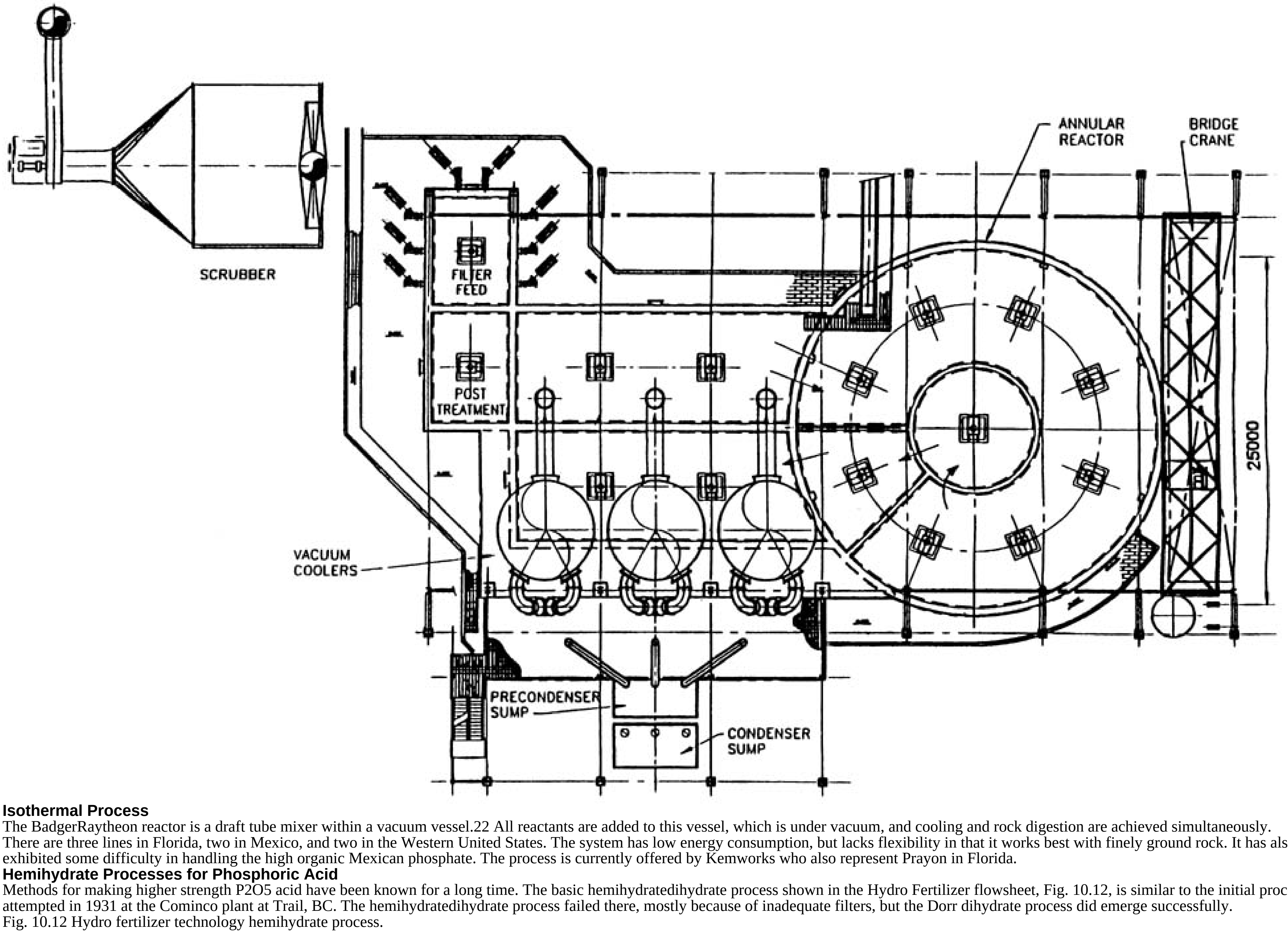
Jacobs Process

Jacobs Engineering acquired the Dorrco process technology in 1974 and has carried on the annular reactor design begun by Dorr in the early 1960s. The reactor configuration is a compromise between a multicompartment system, as used in the earlier DorrOliver cascade system and in the compartmented Prayon reactor, and the true single stirred vessel used by Rhone Poulenc and Badger.

In the Jacobs system, the annulus of a large concrete tank is fitted with a series of agitators. There is baffling, but there are no walls between the agitators. High slurry recirculation rates are achieved by a combination of back-mixing and slurry pumping. Vacuum cooling normally is used. The Jacobs plants include a 1500 tons per day plant in Tampa, Florida, a 900 tons per day facility at Paradeep, India, and several smaller units. Figure 10.10 shows the Oswal Fertilizers and Chemicals construction site of the world's largest phosphoric acid plant1921 2650 metric tons per day P2O5 also at Paradeep, Orissa, in mid-1999. Figure 10.11 is a diagram of the reactor itself. Starting in the center foreground of the photo and moving clockwise, the large building, partially roofed, holds five 110 m two-belt filters, next is the wet grinding and screening structure. The concrete reactor tank, having 4350 m3 of slurry volume, is shown prior to the installation of the agitators. The three small cylindrical vessels are slurry vacuum coolers. Six evaporators with fluosilicic acid recovery are pictured next. The large tanks are for acid storage and clarification, and finally, on the far right, the cooling tower foundations. Slurry flow is clockwise around the annulus, into the center compartment, then to the long narrow vacuum cooler feed compartment. The cooled slurry flows to the cooler seal compartment opposite the feed compartment. The bulk of the slurry recycles to the annular reactor with the net flow proceeding to the filter feed tank. In principal, the flow pattern is similar to the Prayon reactor system shown in Fig. 10.9. The ground phosphate rock is mixed with recycled cool slurry in the first position of the annulus followed by the addition of sulfuric acid diluted in a mixing tee with weak phosphoric acid from the filter.

Fig. 10.10 Oswal phosphoric acid plant during constructionMay 1999in Paradeep, Orissa State, India.

Fig. 10.11 Jacobson reaction system, 2500 tons per day P2O5.



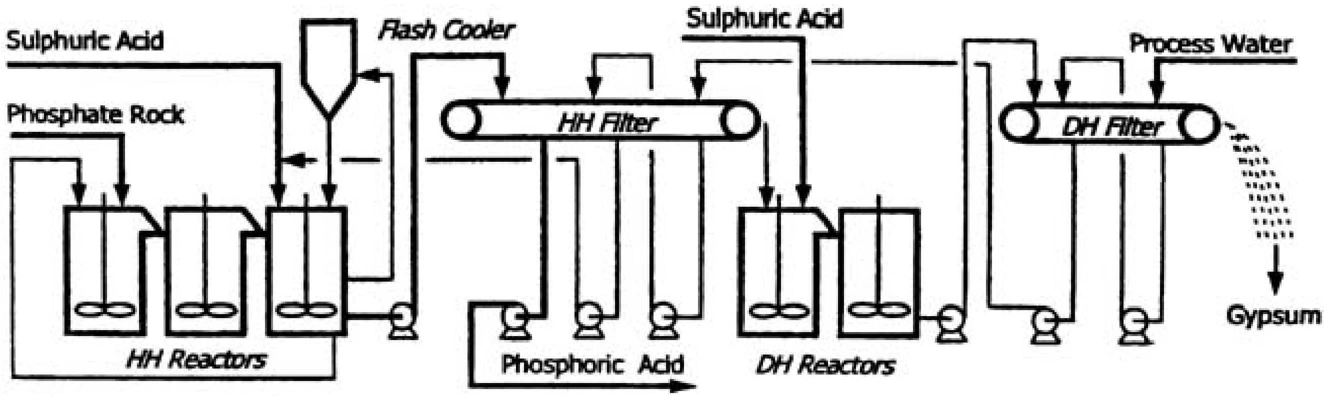
Isothermal Process

The BadgerRaytheon reactor is a draft tube mixer within a vacuum vessel.22 All reactants are added to this vessel, which is under vacuum, and cooling and rock digestion are achieved simultaneously. There are three lines in Florida, two in Mexico, and two in the Western United States. The system has low energy consumption, but lacks flexibility in that it works best with finely ground rock. It has also exhibited some difficulty in handling the high organic Mexican phosphate. The process is currently offered by Kemworks who also represent Prayon in Florida.

Hemihydrate Processes for Phosphoric Acid

Methods for making higher strength P2O5 acid have been known for a long time. The basic hemihydratedihydrate process shown in the Hydro Fertilizer flowsheet, Fig. 10.12, is similar to the initial process attempted in 1931 at the Cominco plant at Trail, BC. The hemihydratedihydrate process failed there, mostly because of inadequate filters, but the Dorr dihydrate process did emerge successfully.

Fig. 10.12 Hydro fertilizer technology hemihydrate process.



The first large hemi plant of more modern times was the Kemira hemi-dihydrate two-stage plant at Siilinjarvi, Finland, which started operation in 1969, using the Dorr HYS process. After about four years of operation at 250 metric tons per day P2O5, its rated design, the plant was expanded but operated in the dihydrate mode. In the hemihydrate mode, filtration and recrystallization problems had plagued the plant, which operated on Kola rock.

Hemihydrate processes are available for new facilities and also for the retrofitting of existing dihydrate plants. Several conversions to higher-strength acid have been made, where the steam saved in evaporation can replace fuel. Plants that make superphosphoric acid, or where sulfuric acid plant steam is not available, are likely targets for conversion.

The single-stage hemi process is similar to the front end of the hemi-dihydrate process (see Fig. 10.13). However, in this case, the hemi is sent to waste without deliberate recrystallization after washing. Yields for the single-stage process are generally below the yields of the dihydrate process, whereas yields for the two-stage hemi-dihydrate process are very high; see Table 10.8.

Fig. 10.13 Hydro fertilizer technology hemihydrate process.

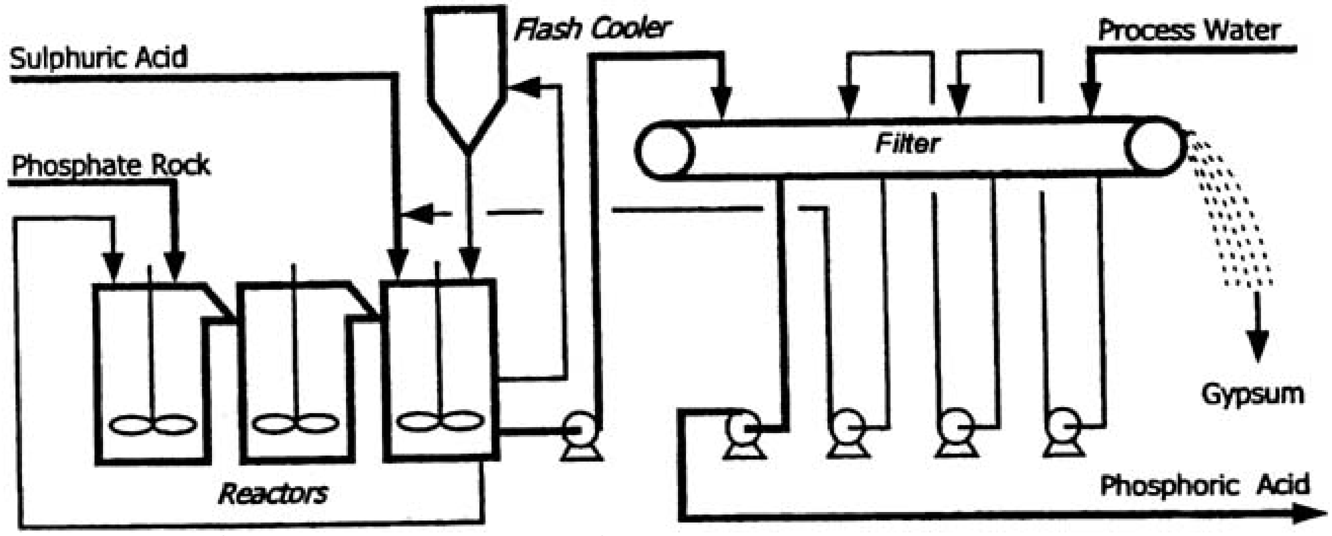


TABLE 10.8 Range of P2O5 Yield through Filtration as % of P2O5 Fed

Phosphate Acid Process	
Hemi-dihydrate	9899
Conventional dihydrate	9597
Single-stage hemi	9194

Table 10.9 lists the major hemihydrate plants operating in 2000. Many of the smaller facilities on the ninth edition list have been shut down. However, two major single-stage hemi plants have recently started operation. The Indo-Jordan facility has been particularly successful operating at full capacity from the initial startup, and up to 1000 tons per day P2O5 on two 80-m2 Eimco belt filters.

TABLE 10.9 Major Hemihydrate Installations (2000)

Owner	Location	Year in Operation	Process	Product Acid Strength	Rock	Capacity P2O5 (Million Tons per Day)
Chinhae	Korea	1990	N-H(C)(NDH)	45	FL	250
PCS (Arcadian)	U.S.	1980	N-H(C)(H)	40	Bou Cra	600
PCS (Occidental)	Florida	1980	Oxy(H)	38	FL	1400
Gresik	Indonesia	1984	Nissan C(HDH)	42	Jordan	550
Nam Hae	Korea	1988	Nissan C(C)(HDH)	42	FL/Jordan	1100
Yong Nam	Korea	1989	Nissan C(C)(HDH)	42		400
Copebras	Brazil	1987	Nissan C(C)(HDH)		Brazil	450
Coop Chem.	Japan	1987	Nissan C(HDH)			230
Yunnan	China	1992	N-H(HDH)	45	China	210
Inda-Jordan	Jordan	1998	N-H(H)	42	Jordan	750
WMC	Australia	2000	N-H(H)	42	Queensland	1500

N-H = Norsk Hydro; (C) = Conversion; (H) = Hemihydrate, single-stage; (HDH) = Hemihydratedihydrate.

The WMC plant23,24 extends proven hemi capacity to 1500 tons per day P2O5, but the plant has had to contend with a highly abrasive, low-grade rock, 23.5 percent P2O5 with over 35 percent SiO2. In addition, the PCS (Occidental) single-stage hemi plant, in operation since the early 1980s, has been termed by some as one of the best phosphoric acid plants in the world. Also, the conversion at Arcadian has met the test of time.

The advantages of the hemihydrate processes compared with the conventional dihydrate processes are:

1. Energy savings due to higher product acid strength.
2. Higher P2O5 recovery and lower H2SO4 consumption for the two-stage process.
3. Lower capital cost for the single-stage process.
4. Hemi-hydrate recrystallized gypsum that is relatively pure, 0.20.4 percent total P2O5. This makes it more suitable for a cement additive or in wallboard.
5. Less rock grinding required.

Energy Savings in Phosphoric Acid

As has just been noted, the hemi processes provide energy savings due to the reduced steam consumption for evaporation. This steam, even though low-pressure, 2.03.0 kg/cm2, has significant value for the co-generation of power. In some cases, the hemi process also saves rock-grinding power. On the other hand, dihydrate processes have the ability to use wet rock slurries and to absorb, in the process, more contaminated water resulting from rainfall collected off the waste gypsum stacks. The latter advantage may be important to maintaining a zero water balance and eliminating costly effluent treatment. The use of wet grinding and slurry feeding eliminates the fuel and electricity consumed in drying the rock.

Another energy-saving option is the use of hot water instead of steam for evaporation. This is an alternative, energy-wise, to the higher-strength hemi processes. Hot water normally is available from the heat of absorption of the sulfuric acid plant. Strengths of 4042 percent P2O5 can readily be achieved. A few commercial installations exist in Europe. A more expensive alternative is offered in the HRS sulfuric acid process by Monsanto, which converts the absorption heat to low-pressure steam.

Unit Operations

In addition to the reaction step discussed above, there are a number of other unit operations used in producing wet process phosphoric acid.

Calcination

Phosphate rock normally is used as a dry rock or in slurry form. However, in some cases, particularly where the raw phosphate is high in carbonaceous matter or it is desirable to have a clean acid, the rock is calcined. Also, in a few cases, the phosphate rock is calcined, the product slaked, and free lime separated as a beneficiation step. Calcination is energy intensive and produces a less reactive rock and, in some cases, a less filterable gypsum. Therefore, the use of calcination is diminishing, and is being replaced by a wet oxidation step to produce green acid.16 In separating calcium carbonate, flotation, where it is successful, is favored over calcination because of its lower cost.

Rock grinding

Until 1973, most phosphate was ground dry in roller or ball mills. In that year, Agrico, at South Pierce, Florida, converted one of its dry mills to wet slurry grinding and proved that the plant water balance could manage the rock at a 6568 percent solids slurry. Since that time, most U.S. installations have converted to wet grinding.

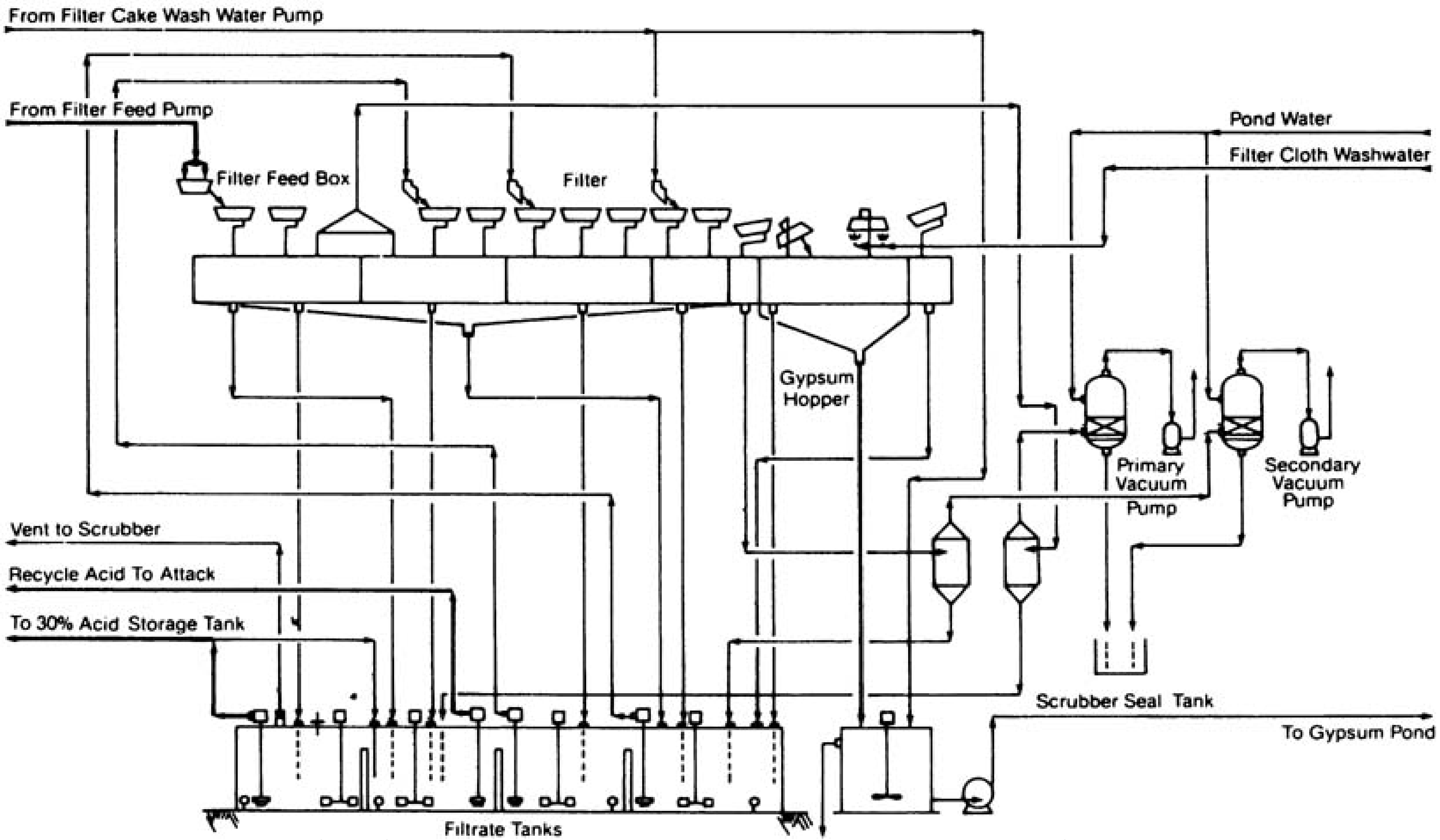
Relatively fine phosphates, such as Kola, North Florida, Senegal, and Togo, can be processed unground as dry concentrates or as dewatered beneficiated product with 1218 percent moisture. In the hemi processes, somewhat coarser feed, -20 mesh, may be tolerated. For dihydrate, it is desirable to feed from 1.5 percent +35 mesh (Tyler) to about 8 percent +35 mesh.

Filtration

The separation of phosphogypsum or hemihydrate from its mother liquor has always been a difficult operation. The process has been subject to the formation of calcium sulfate, sodium fluosilicate, and other types of scale that clog the cloth and necessitate periodic filter washing. Filter cloth wear is severe, requiring cloth changes as often, in some cases, as two or three weeks apart.

Three types of filters have predominated over the past 20 years, the most widely used being the BirdPrayon tilting pan filter shown in Fig. 10.14. The Ucego, a table filter with a peripheral side wall belt that leaves the filter to permit cake sluicing, has been popular worldwide since the late 1960s. In the late 1970s and the 1980s, belt filters became more readily accepted. The belt filter has been used on phosphoric acid since the 1940s and 1950s, but in the past it was plagued by mechanical problems and materials failures.

Fig. 10.14 Flow diagram of filtration section of a wet process phosphoric acid plant. *(Prayon process, courtesy Davy Mckee Corp.)*



During the last decade, many successful belt filter installations have been made by Eimco, Filtres Philippe, Delkor, and Gaudfrin. Even in hemi service at Namhae, Indo-Jordan, and WMC, operating at 95°C or above, the belt filters appear to be successful. Because of their long narrow configuration, the belt filter is well suited as a supplemental filter and three have been installed in Florida for this purpose.

The filtration step is a countercurrent washing using two or three washes. Usually the final wash is a contaminated pond water or a cooling loop water, thus providing for, in most cases, a zero effluent plant.

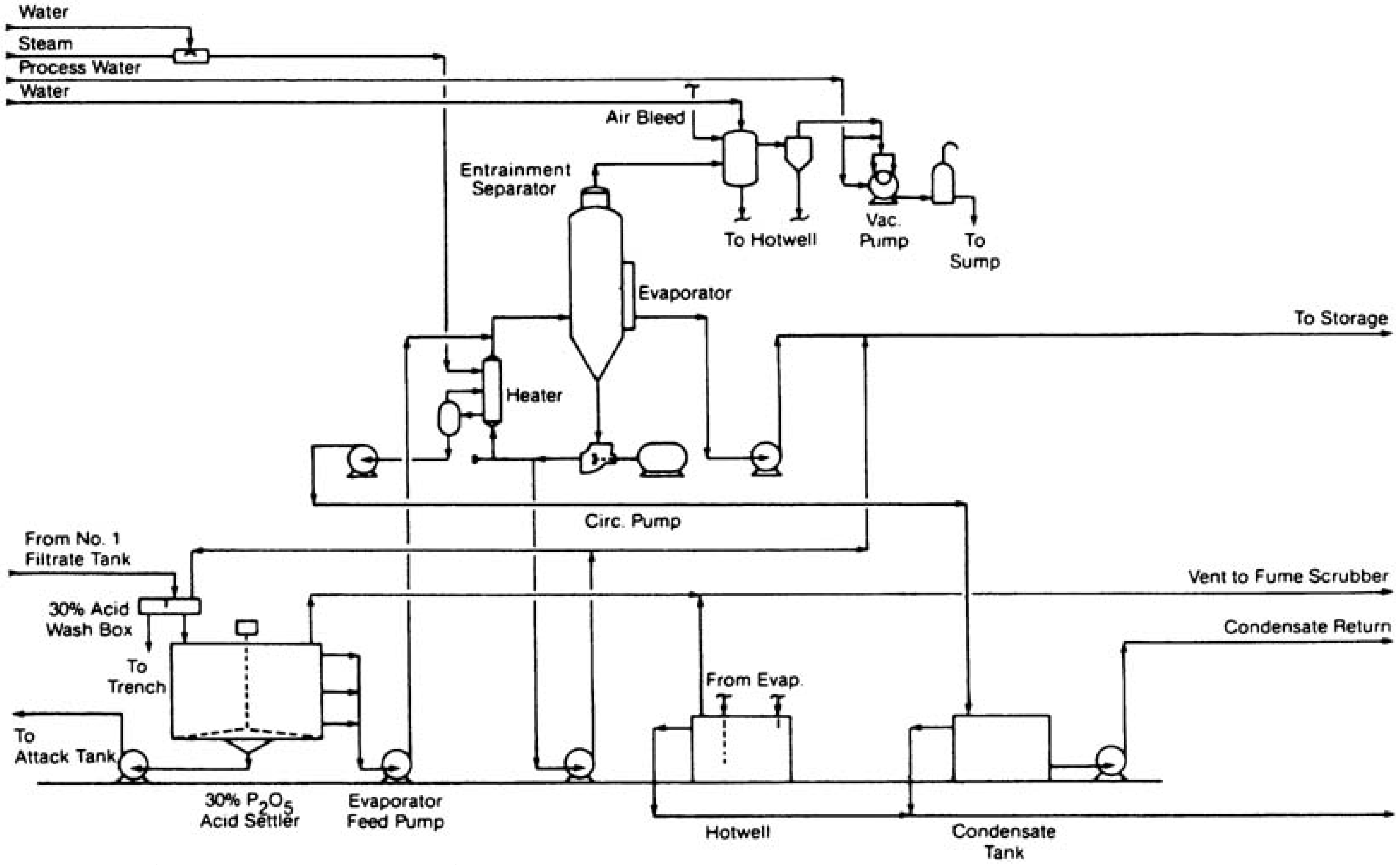
Sizes of the BirdPrayon and Ucego filter can be very large, over 200 m2 of active surface area, allowing rates up to 16002000 metric tons per day P_2O_5. Belt filters are, so far, 110 m2 or less; however, they are relatively inexpensive because little alloy steel is used, and normally two belt filters would be less in first cost than one large tilting pan or table filter.

Evaporation

Phosphoric acid is used for downstream products mostly at 28 percent P_2O_5, 40 percent P_2O_5, and 54 percent P_2O_5. Many plants also make clarified merchant grade acid (MGA), which, at 5254 percent P_2O_5, is a world traded product.

Evaporation normally is done under vacuum using forced circulation via an axial flow pump. Vacuum can be accomplished by steam ejectors, vacuum pumps, or with condenser water using an entraining condenser downleg. Heat exchangers normally have been shell and a tube, with graphite tubes in the United States. French practice has been to use carbon block exchangers, but these tend to scale more and are no longer in favor in many areas. Some newer shell and tube exchangers have been Sanicro 28 or Hastelloy G3 or G30. A flow sheet is shown in Fig. 10.15.

Fig. 10.15 Flow diagram of evaporation section of a wet process phosphoric acid plant. *(Courtesy Davy Mckee Corp.)*



Typical analyses of wet process phosphoric acid made from North Carolina calcined concentrate are shown in Table 10.10.

TABLE 10.10 Typical Analyses, Wet Process Phosphoric Acid Made from North Carolina Calcined Concentrate

	Weight Percentage	
Component	Concentrated Acid	Superphosphoric Acid
P_2O_5, total	53.0	69.5[a]
Solids	0.2	
Free water	22.0	
Fe_2O_3	1.5	2.0
Al_2O_3	0.7	1.0
F	0.6	0.3
MgO	1.1	1.3
SO_4	2.7	3.7
CaO	0.1	0.2
Sp.g, at 75°F	1.68	2.0

[a] About 36% of the total P_2O_5 is present as polyphosphates. (*Courtesy Texasgulf, Inc.*)

Superphosphoric Acid

Solution fertilizers have become very popular in the United States. The principal source of P_2O_5 for these fertilizers is wet process Superphosphoric acid containing about 70 percent P_2O_5, where 35 percent or more of the P_2O_5 is present in the polyphosphoric form. When this acid is ammoniated and diluted, the iron, aluminum, and magnesium compounds naturally present remain in solution, sequestered by the polyphosphates. Clear solutions result, and there is no clogging of the sprays used for applying the fertilizer. When ordinary phosphoric acid is neutralized with ammonia, heavy sludges form, and the resulting solution is difficult to store and apply.

Superphosphoric acid is made by the additional concentration of clarified phosphoric acid in vacuum evaporators of the falling-film or forced-circulation type. High-pressure steam or Dowtherm vapor is used for heating. Corrosion is a problem, so the equipment is made from high alloy stainless steels. The acid is shipped in special insulated tank cars to the solution fertilizer plants, which are located close to the farm areas they serve. Organic matter contributes to sludge problems in making solution fertilizers; therefore, calcination of the phosphate rock used for making the acid is advantageous. Organics also may be removed by oxidation with nitric acid or ammonium nitrate,[16] and several SPA producers have gone to such treatment to improve their product quality.

10.4 Wet Process Acid By-Products

Phosphogypsum

About 5 tons of gypsum on a dry basis are made for each ton of P2O5 produced in a wet process phosphoric plant. This material usually is disposed of as a waste, by impounding it in old mine pits, stacking it in huge piles, or, in some cases, discharging it into very large rivers or river mouths. Phosphogypsum is sold to farmers in California for control of salt buildup in irrigated soils; a small quantity is sold to peanut farmers in the southeastern United States. However, there has been concern about the utilization of gypsum because of its low-level radioactivity in some instances. In addition, it contains fluosilicates and P2O5 so that utilization in building products, such as wallboard and blocks, has been limited to Europe and Japan, where natural gypsum is more costly than manufactured gypsum. In these cases, the phosphogypsum generally comes from a hemihydrate process producing a slightly purer form than natural gypsum.

Regeneration of SO2 from gypsum has been done via the KruppOSW process, the most important facility formerly being the Fedmis plant in South Africa, which made about 300 tons per day H2SO4 and a similar amount of cement. Apparently, because of the high cost of cement and sulfur in this remote location, the plant was not economically viable and shut down in the late 1980s. Because of the present low cost of sulfur and its future abundance as a result of the removal of sulfur from sour natural gas and because of SO2 removal processes in power generation, interest in sulfur recovery from phosphogypsum has waned.

The Florida Institute of Phosphate Research (FIPR) has studied phosphogypsum utilization at length, but the U.S. Environmental Protection Agency has prohibited its movement from its stacks (piles) because of its low level radiation. The FIPR hopes to show that its use in road building subsurfaces and as an additive to enhance digestion of municipal waste in landfills can be accomplished without harm to the public now or in the future.23,24

Fluorine Recovery

Phosphate rock contains about 3.5 percent fluorine, some of which is recovered as a by-product in manufacturing wet process phosphoric acid. During acidulation, the fluorine is released as hydrofluoric acid, HF, which reacts with the silica present as an impurity in the rock to form fluosilicic acid, H2SiF6. Some of the fluorine is lost with the gypsum as sodium or potassium fluosilicates, and some remains dissolved in the filter acid. When the acid is concentrated, much of the fluorine in the feed is boiled off, appearing as HF and silicon tetrafluoride, SiF4, in the vapors.

Fluorine is recovered at the evaporator station by scrubbing the vapors leaving the flash chamber. The vapors pass through an entrainment separator to remove fine droplets of phosphoric acid and then into a spray tower where they are scrubbed with a weak solution of fluosilicic acid according to the reaction:

$$2\mathrm{HF} + \mathrm{SiF}_4 \rightarrow \mathrm{H}_2\mathrm{SiF}_6$$

Part of the circulating solution is continuously withdrawn as a 2025 percent aqueous solution of H2SiF6. The solution is shipped in rubber-lined tank cars and is used for fluoridation of drinking water, the preparation of fluosilicates, and production of AlF3. These salts find use in ceramics, pesticides, wood preservatives, concrete hardeners, and aluminum production.

Uranium Recovery from Wet Process Phosphoric Acid

Uranium recovery was briefly described in the ninth edition. Since 1990, all uranium recovery contracts have expired in the United States and the recovery facilities moth-balled or scrapped. There is no indication the situation will change in the near future, since the reduced price of uranium no longer makes its recovery from phosphoric acid economical.

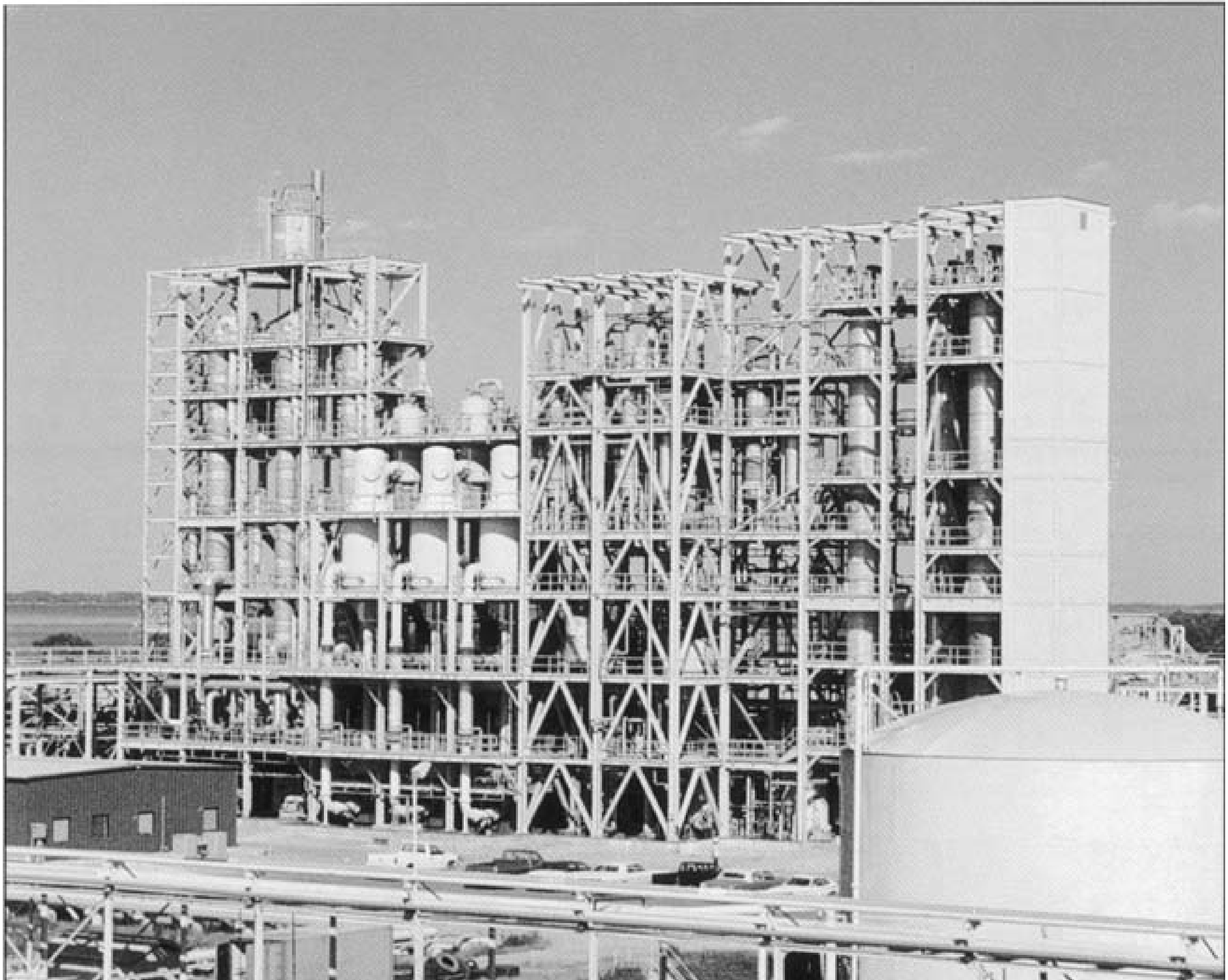
Animal Feed Supplements

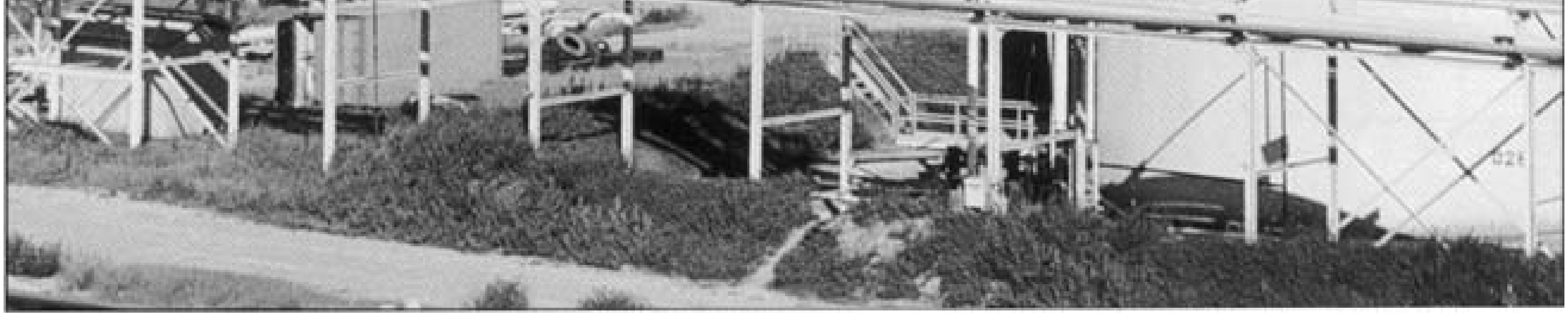
Calcium phosphates for use in animal and poultry feeds are made from both furnace and wet process phosphoric acids. Dicalcium phosphate, CaHPO4, containing 18.5 percent P, and mono calcium phosphate, Ca(H2PO4)2·H2O, containing 21.0 percent P, are made in large tonnages. Both grades are prepared by reacting phosphoric acid with pulverized limestone in a pug mixer. The limestone must be quite pure, and the phosphoric acid must have a low fluorine content, below 100 P to 1.0 F. If 54 percent P2O5 wet process phosphoric acid is used, it is defluorinated first by adding diatomaceous earth and then sparging the acid with steam. An alternate method is to use wet process superphosphoric acid, which has a low fluorine content. The superphosphoric acid is hydrolyzed by diluting it with water and heating. The pug mixer product is a fine granule, minus 12 mesh, which is dried and shipped in bulk to feed-mixing plants.

Purified Phosphoric Acid

Currently only one furnace acid plant remains in operation in the United States to supply elemental phosphorus and a few uses that cannot be satisfied by purified wet acid. The plant pictured in Fig. 10.16 is being doubled in capacity and a new purified acid plant in the Western United States started operation in 2001. The RhonePoulenc purification plant at Geismar, LA, continues in operation.

Fig. 10.16 Purified phosphoric acid plant for P.A. Partnership, Aurora, NC. (*Courtesy P.A. Partnership and Jacobs Engineering*.)





Wet process phosphoric acid made from calcined rock is preferred because it is devoid of the soluble organics and sludges present in acid made from sedimentary phosphates. However, calcination is expensive so that some plant operators are willing to go through laborious clarification to avoid it.

Clear acid is fed to a column or a battery of mixer-settlers and extracted with a solvent such a butyl alcohol or tributyl phosphate. Generally, about three-fourths of the phosphoric acid transfers to the organic phase, leaving the impurities in the raffinate, which is sent to a fertilizer unit to recover its P2O5. The yield of cleaned acid can be increased by adding another mineral acid such as sulfuric acid or hydrochloric acid to the extraction step.

After washing, the phosphoric acid is stripped from the solvent with water, and the solvent is returned to the extraction section. The phosphoric acid now is quite dilute and still contains small amounts of impurities. The acid then is concentrated, and the impurities are removed by steam stripping and the addition of reagents and adsorbents followed by filtration. The exact details of the process vary, depending upon the process technology, which is proprietary, the impurities present in the feed acid, and the solvent used. Purified wet phosphoric acid is suitable for both industrial and food-grade use, although food-grade requires another level of purification over industrial or technical grade acid. There are now several purified acid plants in the United States. Figure 10.16 is a photo of the PCS plant at Aurora, NC. There are also plants in Morocco, Belgium, Japan, and Israel.

Environmental Aspects

As implied earlier in this chapter, gypsum disposal is a problem that generally has defied an inexpensive solution. As a waste material, it is relatively benign, but P2O5, sulfate, fluorine, low-level radioactivity, and other contaminants, including some heavy metals in small quantities, can leach from waste stacks into the nearby groundwater. In Florida, the underlying layers of limestone afford some protection, but the EPA will require future stacks be lined with a membrane to prevent seepage. The "gypsum dilemma" has loomed as an increasing problem over the last decade.2527

Although dumping of waste gypsum slurries into the ocean still is practiced in some locations outside the United States, and harmful effects are generally difficult to quantify, there is continuing pressure from environmental groups to cease dumping into rivers and the seas.

Gaseous emission from phosphoric acid plants can be scrubbed with cool contaminated recycle cooling water to relatively low emission levels. However, to minimize cooling tower or cooling pond emission, fluorine recovery often is necessary on those streams that have a significant fluorine content.

10.5 References

Leyshon, D. W., The Origin of the Modern Dihydrate Phosphoric Acid ProcessCominco 1931, A.I.Ch.E., Orlando, FL, 1990.
McClellan, G. H., "Mineralogy of Carbonate Fluorapatites," J. Geol. Soc. Lond., 137, 675681 (1980).
Van Kauwenbergh, S. J., Cathcart, J. B., and McClellan, G. H., "Mineralogy and Alteration of the Phosphate Deposits of Florida," U.S. Geol. Surv. Bull., 1914 (1990).
Bernardi, I. P., and Hall, R. B., "Comparative Analysis of the Central Florida Phosphate District to its Southern Extension," Mining Eng., 12561261 (Aug. 1980).
McKelvey, V. E., "Phosphate Deposits," Geo. Surv. Bull., 1252-D (1967).
Northolt, A. J. G., Shelton, R. P., and Davidson, D. F., Phosphate Deposits of the World, Vol. 1, Phosphate Rock Resources, Cambridge University Press, 1990.
Northolt, A. J. G., Shelton, R. P., and Davidson, D. F., Phosphate Deposits of the World, Vol. 2, Phosphate Rock Resources, Cambridge University Press, 1990.
ZellarsWilliams, Inc., Evaluation of the Phosphate Deposits of Florida Using The Minerals Availability System, Final Report submitted to the U.S. Department of the Interior Bureau of Mines, June 1978
Busnardo, C. A., and Olivario, R. N., "Optimization of the Grinding Circuit of the Jacupiranga Carbonatite Ore in Jacupiranga, Brazil", SME/AIME Annual Meeting, New York, pp. 8598 (1985).
Nofal, A. M., "Egyptian Phosphate Rocks: Important Factors Affecting Thek Economic and Technical Evaluation," in Beneficiation of Phosphate: Theory and Practice, Society for Mining, Metallurgy, and Exploration, Inc., 1993.
Pressacco, R., "Overview of the Agrium Kapuskasing Phosphate Operation," CIM Bull., 94(1049) (April 2001).
Allen, M. P., "The Vernal Phosphate Rock Mill," in Beneficiation of Phosphate: Theory and Practice, Society for Mining, Metallurgy, and Exploration, Inc., 1993.
Gruber, G. A., Guan, C. Y., and Kelahan, M. E., "Dolomite FlotationPilot Plant Studies," in Beneficiation of Phosphates, Vol. III, St Pete Beach, Dec. 2001.
Lawver, J. E., "General Principals and Types of Electrostatic Separators," in SME Mineral Processing Handbook, Society of Mining Engineers, 1985.
Becker, P., Phosphates and Phosphoric Acid, 2nd ed., Marcel Dekker, New York, 1989.
Leavith, Kranz, Gorman, and Stewart, U.S. Patent 4808,391.
Theys, T., and Smith, P., IMACID, a 1000 Ton Phos Acid Unit in Morocco, A.I.Ch.E., Clearwater, FL, June 2002.
Satier, B., and Apostoleris, G., "Speicbjrn/Phone Poulenc Process," IFA Technical Conference, The Hague, Netherlands, October 1992.
Blythe, B. M., Leyshon, D. W., and Jaggi, T. N., "Inception of the World's Largest Phosphoric Acid Plant," IFA Technical Conference, Marakesh, Morocco, October 1998.
Leyshon, D. W., "Phosphoric Acid Technology at Large, Part II," Phosphorus & Potassium, November/December, 1999.
Davies, L., "Strength in Diversification," Phosphorus & Potassium, May/June, 2000.
Felice, C., Martinez, J., and Hilakos, S., Raytheon's Isothermal Reactor Process, A.I.Ch.E., Clearwater, FL, May 1998.
Collen, D., and Duckworth, G., "Ammonium Phosphate Plant Makes Its Debut," Phosphorus & Potassium, March/April, 2000.
Gobitt, J., A New Chapter in Hemihydrate Technology, AI.Ch.E., Clearwater, FL, June 2000.
Leyshon, D. W., "The Gypsum Dilemma," Phosphorus & Potassium, March/April, 1996.
Leyshon, D. W., "The Gypsum Dilemma, New Concerns," Phosphorus & Potassium, March/April, 2001.
Wissa, A. E. Z., and Fuleihan, N. F., "Phosphogypsum Stacks and Ground Water Protection," Phosphorus & Potassium,* May/June, 2000.

*Phosphorus & Potassium *is a bi-monthly publication of British Sulphur Publishing, London, England*.

11.1 Introduction

Fertilizers provide plants with the nutrients they need for their growth and development. Plants live, grow, and reproduce by taking up water and nutrients, carbon dioxide from the air, and energy from the sun. Apart from carbon, hydrogen, and oxygen, which collectively make up 9095 percent of the dry matter of all plants, other nutrients needed by plants come essentially from the media in which they growessentially in the soil. The other nutrients are subdivided into primary nutrients (nitrogen, phosphorus, and potassium) and secondary nutrients (calcium, magnesium, and sulfur). In addition, plants also need other nutrients in much smaller amounts, and they are referred to as micro-nutrients (boron, chlorine, copper, iron, manganese, molybdenum, and zinc).

To maintain soil fertility and productivity and prevent land degradation, nutrients taken up by crops must be replenished through the application of fertilizers. The use of fertilizer results in many benefits to producers, consumers, and the environment, starting with increased agricultural outputs (mainly food and fiber) to contributing to soil organic matter maintenance, water-holding capacity, biological nitrogen fixation, soil erosion control, other physical and chemical properties, and less extensive land use. These benefits contribute to increased agricultural growth and agribusiness activities, which are catalysts for broadly based economic growth and development in most developed and developing economies; agriculture's links to the nonfarm economy generate considerable employment, income, and growth in the rest of the economy.

A commercial fertilizer is a material that contains at least one of the plant nutrients in chemical form that, when applied to the soil, is soluble in the soil solution phase and assimilable or "available" by plant roots. Most often, this implies chemical forms that are water soluble. However, in the case of phosphorous, solubility in special reagent solutions (citric acid, neutral ammonium citrate, or alkaline ammonium citrate) often is used as a guide for availability to plants. In the case of nitrogen, slow solubility in water may be more desirable from an environmental and efficiency standpoint than easy solubility.

Fertilizer products are customarily designated by a series of numbers separated by dashes. This set of numbers is called "grade" of the fertilizer product. Each of the numbers indicates the amount of a nutrient that the manufacturers guarantee is contained in the fertilizer product. This number includes only the amount of nutrient found by accepted analytical procedures, thereby excluding any nutrient present in a form that is deemed to be unavailable for plant nutrition. The content of each nutrient, expressed as a percentage of total weight, is the guaranteed minimum rather than actual, which is usually slightly higher.

Usually, three numbers are used when giving the grade of a fertilizer product, and these three numbers always refer in order to the content of the primary nutrients: nitrogen, phosphorus, and potassium. If other nutrients are present, their content can also be indicated in the grade of the fertilizer product; each extra number is followed by the chemical symbol of the nutrient it represents. Many countries indicate the content of phosphorus and potassium not in the elemental form but in the oxide form, P2O5 and K2O. Thus, a fertilizer product with a grade of 126222MgO is guaranteed by the manufacturer to contain: 12 percent N, 6 percent P2O5, 22 percent K2O, and 2 percent MgO.

The beginning of our dependence on inorganic fertilizer can be traced back to the nineteenth century when Justus von Liebig articulated the theoretical foundations of crop production and when John Bennett Lawes began producing fertilizers containing phosphorus.1 However, only since the 1960s when global starvation became a real possibility have fertilizers assumed a predominant role in increasing agricultural productivity. Fertilizer was an integral part of technological trinityseed, water, and fertilizerresponsible for bringing about the "Green Revolution" that helped many densely populated countries, including India, China, and Indonesia, achieve food self-sufficiency in a short span of 2025 years. Since the 1960s, global cereal production has more than doubled, increasing from 877 million metric tons (mt) in 1961 to 2100 million mt in 2000, with developing countries accounting for nearly 70 percent of the increase (Fig. 11.1). It is estimated that fertilizers accounted for 56 percent of the rise in average yields per hectare and about 30 percent of the total increase in production. Consequently, cereal production closely parallels fertilizer use (Fig. 11.2). Today there is adequate food for the world's 6 billion inhabitants, yet there remains a segment of the population that is underfed and undernourished.

Fig. 11.1 Cereal production in developed and developing countries, 19612000.

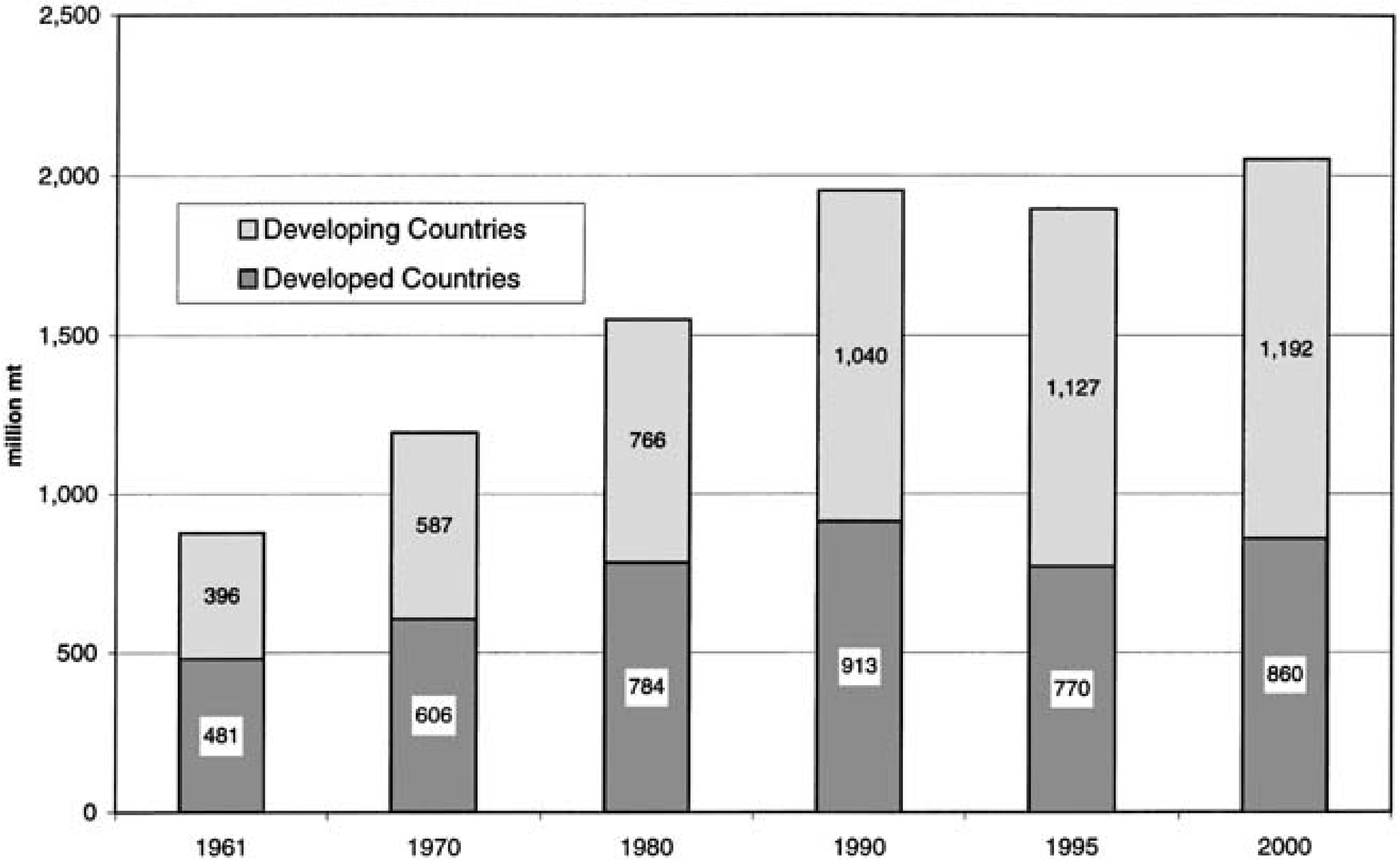
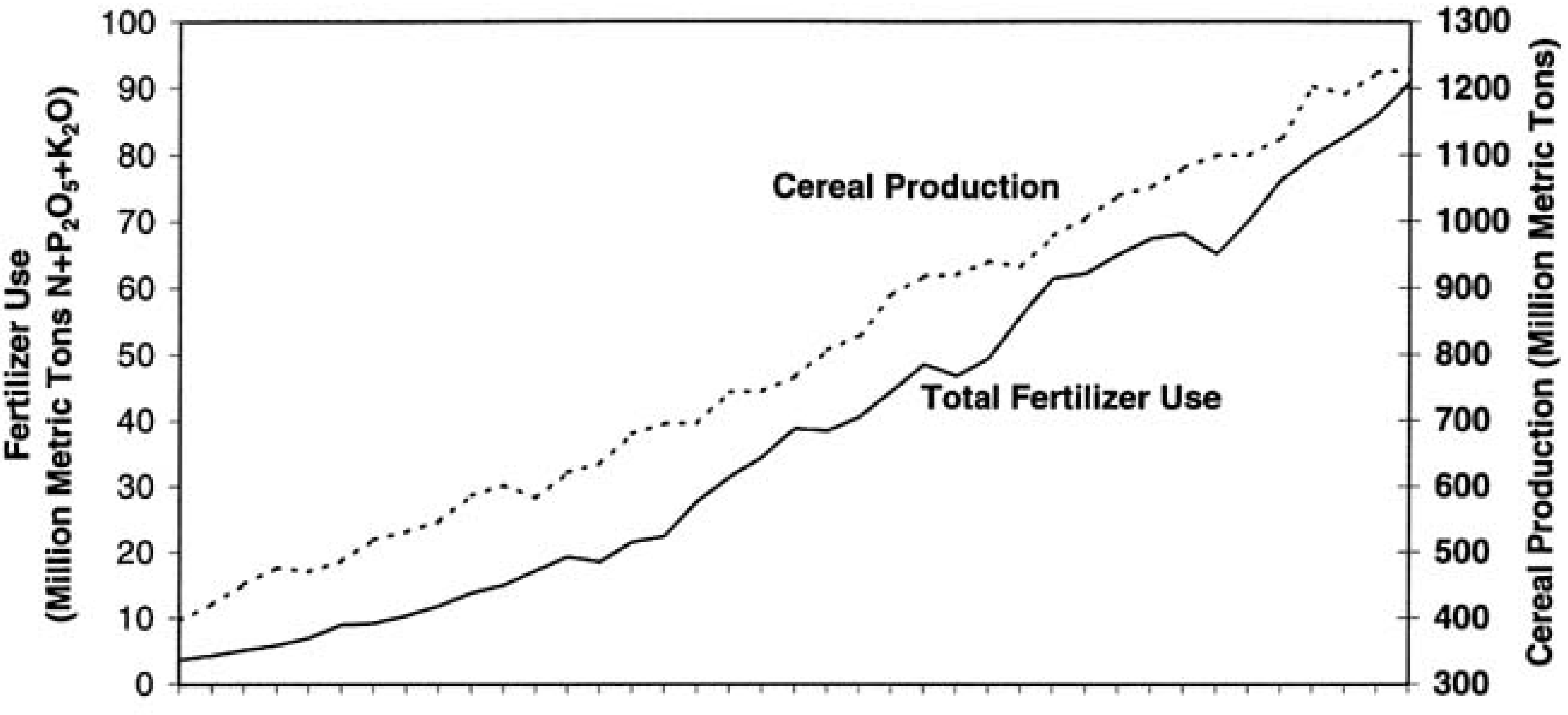
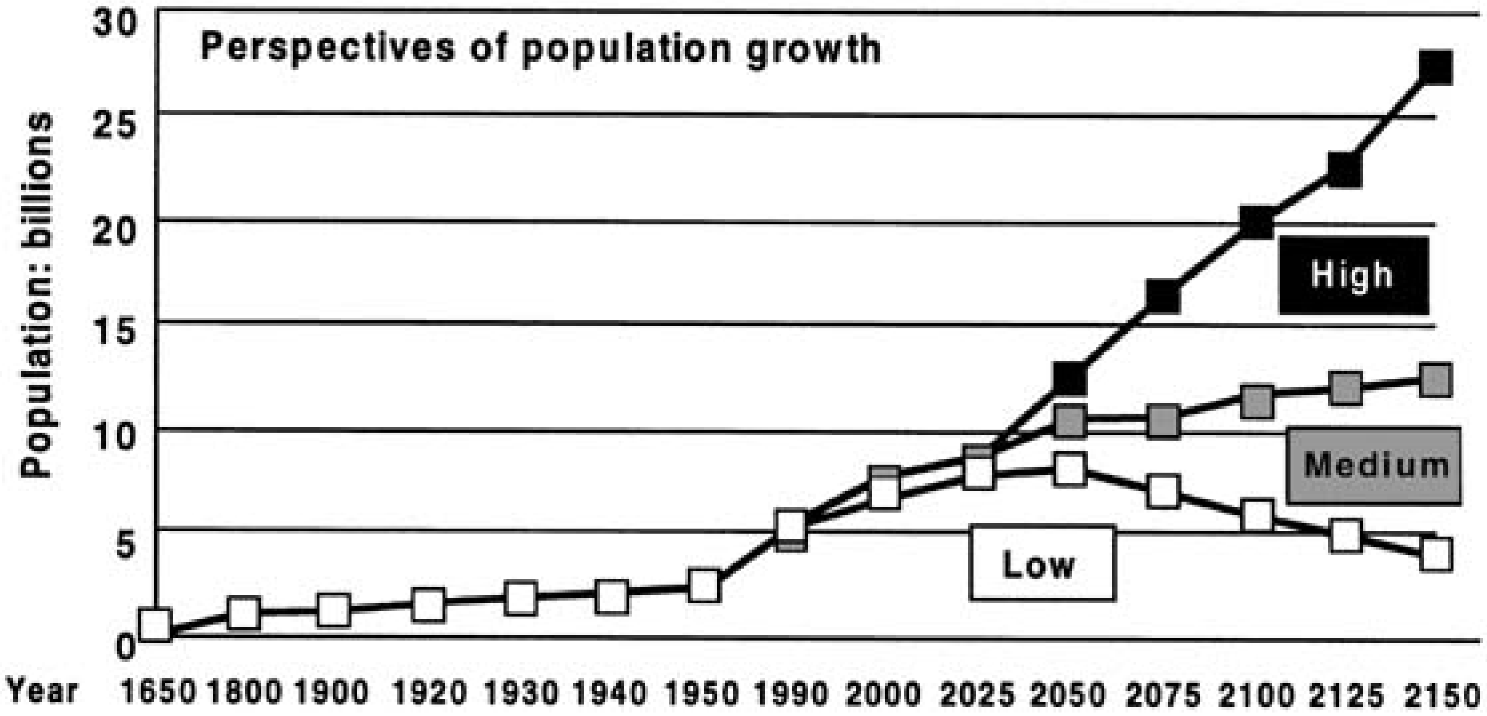


Fig. 11.2 Fig. 11.2. Developing countries: total cereal production and total fertilizer use, 1961/621999/2000.



Growth in the future demand for food will be influenced by changes in a number of forces, but mainly by growth in population, income levels and economic development, and changes in the food preferences of consumers. Although the United Nations recently adjusted downward its estimates of population growth, still during the next 20 years approximately 75 million people will be added to the world's population each year, increasing it from about 6 billion in 2000 to over 8 billion by 2030 (Fig. 11.3). Much of this increase will occur in the cities of the developing world, which will account for most of the increased demand for food, including meat products. These increased demands will require the world's farmers to produce 50 percent more cereals in 2030. Because of the need to preserve the forests and wildlife habitat, most of the increase in cereal production will have to come from the existing land under cultivation. This can be achieved only through improvements in crop yields, which would require a 30 percent increase in fertilizer use. This increased fertilizer use has to be balanced against the environmental and human health concern stemming from intensive fertilizer applications, particularly in industrialized countries.

Fig. 11.3 World population projection.



11.2 Overview of the Fertilizer Industry

Since medieval times, farmers have realized the need to maintain the productivity of soil to achieve improved crop yields. Until the last 200300 years, however, the approach was highly empirical; only by accident or by trial and error was it found that applications to the soil of various organic wastes or naturally occurring mineral substances such as manure, compost, fish, ashes, saltpeter, and other substances would sometimes increase yields or apparently restore productivity to fields that were considered "worn out."

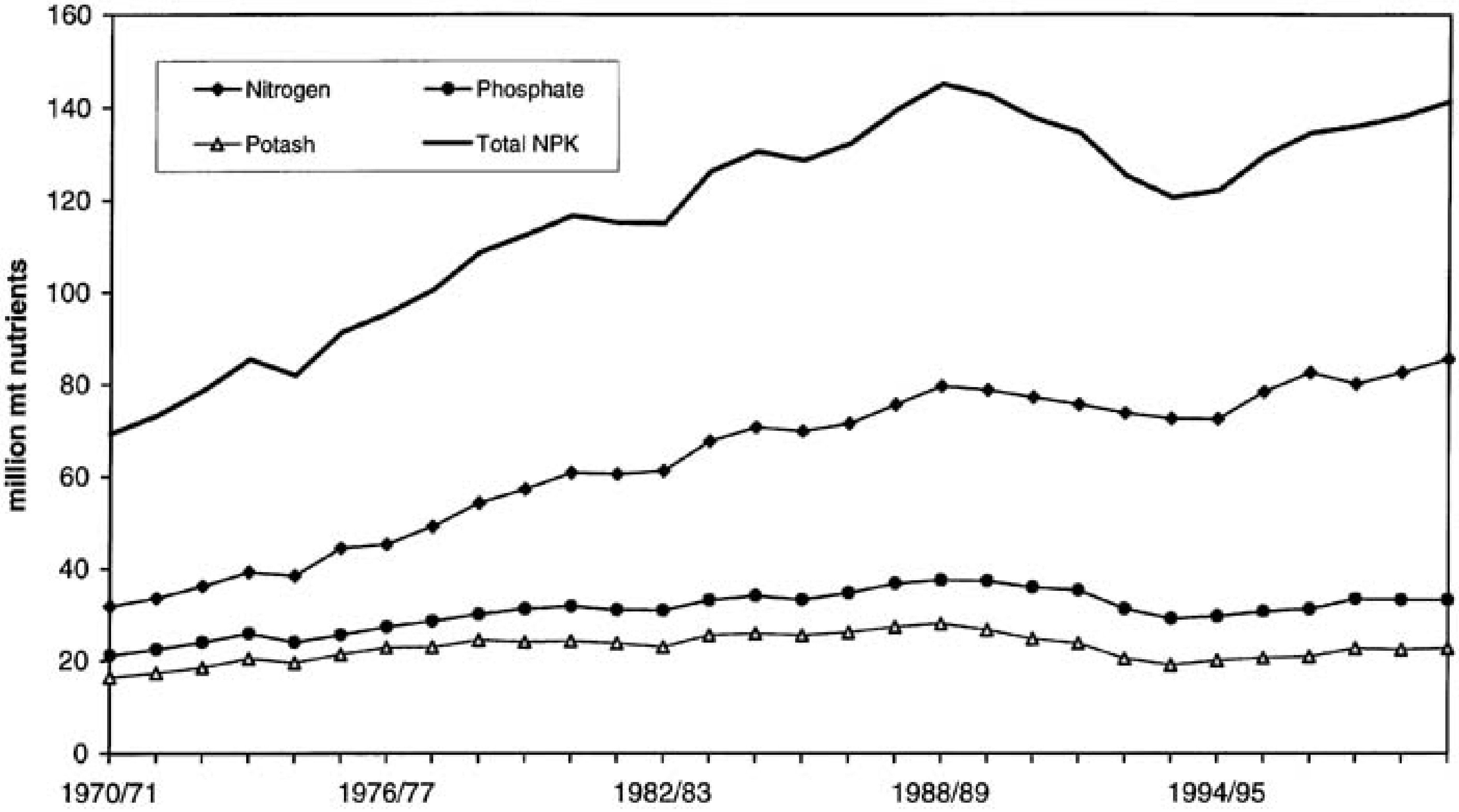
As more and more chemical elements were identified, scientists became interested in determining the amounts and relative importance of various elements in plants. The German Scientist Liebig clarified the value of elements derived from the soil in plant nutrition and stressed the necessity of replacing those elements to maintain soil fertility. He usually is credited with initiating the fertilizer industry. Liebig recognized the value of nitrogen but believed that all plants could get nitrogen from the air, a concept that unfortunately is not true. He envisioned a fertilizer industry with nutrients such as phosphate, lime, magnesia, and potash prepared in chemical factories. In 1840 Liebig published a recommendation that pulverized animal bones be treated with sulfuric acid to make the phosphate more readily available to plants. This practice was accepted, and the production of fertilizers by chemical processing thus began.

Natural organic materials and various chemical byproducts represented a large proportion of the total world fertilizer supply until about the middle of the twentieth century; in the later years of the century, however, the dependence shifted almost entirely to synthesized or chemically processed materials. Only by this means has it been possible to keep up with increasing populations, increased farm acreage, and increased plant food needs of new and improved crop varieties. Today, the fertilizer industry utilizes many facets of highly sophisticated chemistry and engineering in the manufacture of fertilizers, and in pollution control, including disposal of waste products. The procuring and handling of raw materials and the distribution and marketing of products also involve the latest technology and many innovations. In addition, the agronomic aspects of fertilizer usage engage the efforts of many individuals and organizations worldwide. The development and introduction of "high-yielding" varieties of seed required more concentrated and efficient fertilizers. Many organizations and private companies worldwide have carried out extensive research and development on improving fertilizer production, developing new products, and increasing efficiency of usage. The most notable among them include the U.S. Tennessee Valley Authority, the International Fertilizer Development Center, and the Rothemstadt Experiment Station.

The fertilizer industry is a large volumeone that falls within the basic industrial structure of a country. The magnitude of this industry is apparent from Fig. 11.4, which shows the yearly global consumption of primary plant nutrients (N, P_2O_5, and K_2O) since 1970/71. For 1999/2000 the total consumption of N, P_2O_5, and K_2O was 141.4 million metric tons (mt). Of this amount, China, the United States, and India consumed 36.7 million mt, 19.9 million mt, and 18.9 million mt, respectively. Collectively, these three countries accounted for about 53 percent of the world consumption. Assuming an average nutrient content of all fertilizers for that period of about 40 percent shows that some 354 million mt of fertilizers was produced and handled by the industry during that year.2

In general, the industry includes a number of "basic producers" each of which typically concentrates on producing large amounts of single-nutrient or high-nutrient fertilizer products, usually at locations near raw materials. Numerous other components of the industry concentrate on locating close to local farmers and distributing the basic products either in the form received from basic producers or after various secondary-processing operations such as granulation, blending, or conversion to fluids. In the United States, fluid mixtures account for more than 20 percent of the total fertilizer sales. To a lesser extent, fluid mixtures have been introduced into Europe, North America, and South America. Fluid mixtures are used on high-value crops in some middle eastern countries, particularly Israel, where efficient use of water is as important as the efficient use of plant nutrients.

Fig. 11.4 World: nitrogen, phosphate, potash, and total NPK consumption, 1970/711999/2000.



11.3 Raw Materials for Fertilizer Production

The primary raw materials for nitrogen fertilizers are natural gas, naphtha, fuel oil, and coal. The manufacturing of phosphate fertilizers most often requires phosphate rock. Naturally occurring potassium salts form the basis of the production of most potash fertilizers.

Natural gas, naphtha, fuel oil, and sulfur have fairly definable specifications. In contrast, phosphate rock and coal are products that can vary significantly in composition and other characteristics. These variations can affect the processes used to upgrade the "as-mined" ores or the processes for manufacture of fertilizers from beneficiated products. Potash ores also vary greatly in composition depending on origin; however, the end products of mining, beneficiation, and processing generally have relatively constant compositions.

Adequacy of requisite raw materials is the most obvious concern when facing a substantial increase in future demand. Two separate yet intertwined issues in the case of fertilizers are: sufficiency of raw materials and availability of energy to convert them into final products. Potassium is of least concern among the three primary nutrients. Not only is this element abundantly present in the earth's crust, but also it can be found in conveniently concentrated deposits in both deeply buried and near-surface sediments. Potassium deposits in descending order of known reserves are in North America (Canada and the United States), Germany, Russia, Belarus, Brazil, Israel, and Jordan. Even the most conservative reserve base estimates indicate a reserve on the order of 500 years at the level of the late 1990s production.3 Mining, crushing, and beneficiation of potassium usually amount to about 10 gigajoules (GJ)/t of the nutrient, energy demand comparable with that required in making cement.

Phosphate deposits are not as abundant as those of potassium minerals; recently published totals range between 1.5 and 3.5 billion tons of P reserves, 4.79.5 billion tons of potential reserves, and about 13 billion tons of resources.4 Phosphate deposits in the descending order of known reserves are in Morocco and Western Sahara, United States, South Africa, Jordan, and Tunisia. At the current rate of extraction, global reserves would last about 80 years, and the estimates are that resources could support the recent rate of application for nearly 250 years. This time horizon can be extended by tapping less accessible (but plentiful) deposits at a higher price. Depending on the final product, energy costs of phosphatic fertilizers range from about 18 to 32 GJ/t P.

Nitrogen fertilizers via ammonia synthesis account for more than 90 percent of the world's nitrogen fertilizers. Nitrogen supply for ammonia synthesis is truly inexhaustible since the atmosphere contains 3.8 quadrillion tons of the element. Various feedstocks can be used to obtain hydrogen, and during the last several decades the focus has been to improve the energy efficiency of ammonia synthesis. Natural gas is the preferred feedstock and the best natural gas-based plants now use less than 33 GJ/t N. The global mean, which is affected by more energy-intensive reforming of heavier hydrocarbons (naphtha and fuel oil) and coal, is now between 49 and 55 GJ/t N, roughly half of the level prevailing during the early 1950s.

Even if all of the energy needed to fix the fertilizer nitrogen were to come from natural gas, it would still be less than 7 percent of the recent annual global consumption of the fuel and less than 2 percent of all energy derived from fossil fuels.5 Clearly, there is little reason to be concerned about either the current needs or the future supplies of energy for producing nitrogenous fertilizers via the ammonia synthesis route. Moreover, there is no doubt that higher absolute energy needs for nitrogen fertilizers will be partially offset by improved efficiency of ammonia synthesis and by higher efficiencies of fertilizer use. Because today's low-income countries will experience much faster growth of energy needs in sectors other than the fertilizer industry, the share of global fossil fuel consumption for ammonia synthesis by the middle of the twenty-first century may be only marginally higher than it is today.

The global natural gas resources are considered abundant, and they are conservatively estimated to be between 1.1 and 1.7 times as large as all the natural gas that has already been extracted or found in proven reserves.6 In the absence of natural gas, naphtha, or fuel oil, ammonia synthesis could proceed easily (albeit more costly) by tapping the world's enormous coal deposits or using a variety of biomass feedstocks.

11.4 Nitrogen Fertilizers

Nitrogen is a component of amino acids that make up proteins; chlorophyll (the molecule that captures the sun's energy); enzymes; and the genetic material, nucleic acids. Therefore, this nutrient is required in large amounts by all plants and forms one of three primary nutrients. Although nitrogen is available in abundance (79% by volume) in the earth's atmosphere, only a very limited number of plant varieties, chiefly legumes, can utilize this nitrogen directly from the air. For most plants, including such important cereal crops as rice, wheat, and corn, the nitrogen must be in a chemical form dissolved in the soil solution. Atmospheric phenomena, including lightning discharge, convert nitrogen to nitrogen oxides, which then is absorbed in rain water and enters the soil. However, this contribution is quite small and the resultant nitrogen content of soils is normally very low (less than 0.3%). Continuous cropping without replenishment quickly depletes the soil of this important nutrient.

Until about the early 1950s, the more widely accepted method of supplying nitrogen was through the application of manures/ organic waste and the use of crop rotation. Crop rotation entailed growth of a nitrogen-fixing legume crop such as peas, clover, or alfalfa, which then was plowed into the soil to provide nitrogen for a subsequent nonlegume crop, for example wheat and corn. This method gives the added bonus of improving soil organic matter status. Such a crop rotation is relatively inefficient with regards to land usage and labor requirement. Modern, high-yield agriculture cannot, under most conditions, be sustained by such a system. Nevertheless, the use of crop rotation is being promoted in many countries of sub-Saharan Africa, particularly land-locked countries where natural soil fertility is very low and the price of fertilizers relatively high.

Natural Organics

Animal and human wastes have long been used as fertilizers, especially in Europe and Asia, particularly China. Even some American Indians are said to have planted a dead fish in each corn hill to increase yields. Such materials contain small percentages of nitrogen and other plant nutrients that are assimilable by plants. Today, the use of raw sewage on crops persists in Asia and Europe, but volumewise is not of great significance. In the United States and most European countries, the use of raw sewage is not considered acceptable, but some municipal sewage plants employ special sterilization, heavy metals removal, and drying processes that yield acceptable fertilizers of low nitrogen content. The volume of such output is small, however. Other organic materials traditionally employed as fertilizers of usable nitrogen content include guano (deposits of accumulated bird droppings), fish meal, and packinghouse wastes including bone meal and dried blood. However, the cumulative importance of all such natural nitrogen sources in modern agriculture is minor. Less than 1 percent of the total fertilizer nitrogen now used comes from such sources. These products for the most part are not chemically altered prior to use. Processing is mainly physical in nature to improve handling and distribution properties.

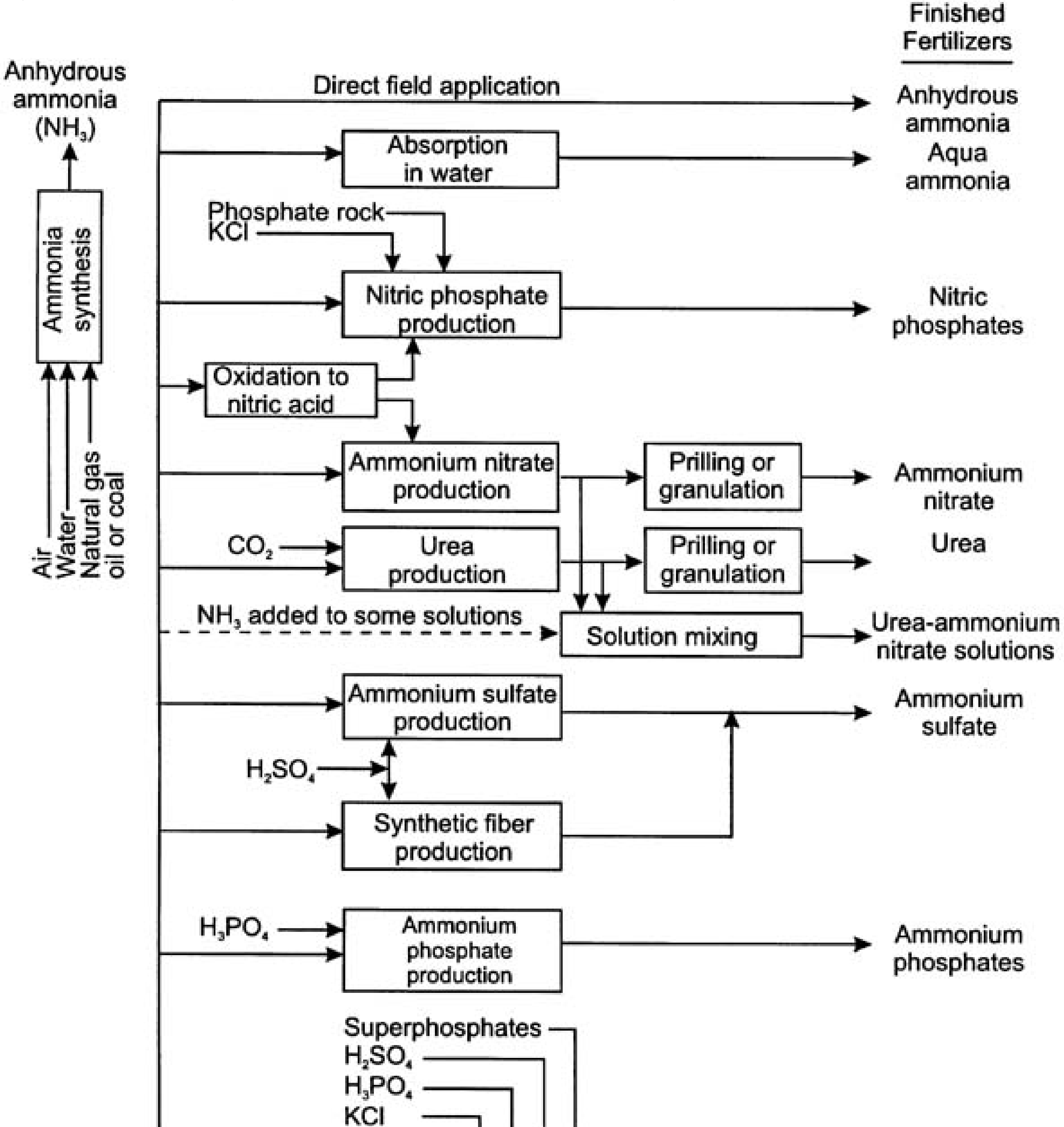
Nitrogen Fertilizers from Synthetic Ammonia

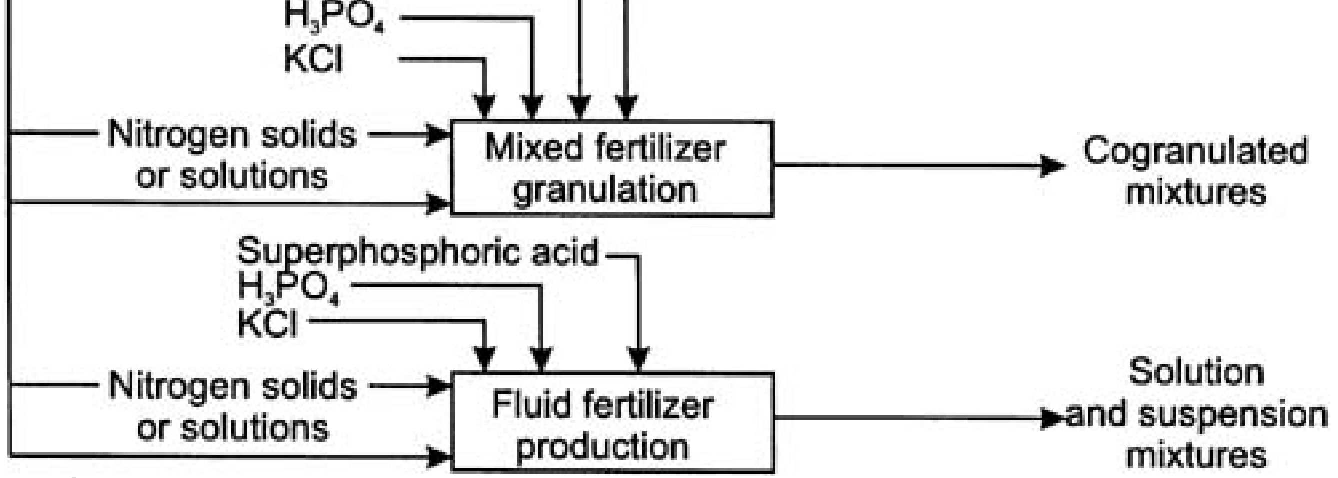
The development of a practical ammonia synthesis process in the early years of the twentieth century was a profound scientific achievement of great social significance in view of the subsequent dependence of the world on fertilizer for support of its growing population.

Both the historical background and detailed technical descriptions of modern ammonia synthesis processes are covered in Chapter 29. These processes are all refinements of the original one conceived and developed by Fritz Haber and Carl Bosch in Germany during the period 190413. Basically, a gaseous mixture of nitrogen and hydrogen, in proper proportions to form ammonia, is compressed to very high pressure in the presence of an activated iron catalyst. Ammonia forms and is removed by cooling and condensation. The nitrogen feed is obtained from air by any of several routes that remove the oxygen. Likewise, there are several routes by which hydrogen feed is obtained, most of which involve decomposition of water by reaction with a carbon source such as natural gas, naphtha, or coke. The preparation and the purification of the synthesis gas and the synthesis itself are highly sophisticated modern processes that require great skill and know-how in design, construction, and operation of plants.

There are many processing routes by which synthetic ammonia subsequently finds its way into finished fertilizers. The major routes are outlined graphically in Fig. 11.5 and are discussed in some detail below.

Fig. 11.5 Major routes of synthetic ammonia into finished fertilizers. *(Courtesy TVA.)*





Anhydrous Ammonia

Anhydrous ammonia, without further processing, is an excellent nitrogen fertilizer when properly injected in gaseous form below the surface of many soils. Such direct injection first came into farm use in the early 1940s and rapidly became accepted. Anhydrous ammonia is an important nitrogen fertilizer in the United States, Canada, Denmark, and Mexico, but in some other countries, for various reasons it is a relatively minor nitrogen source.3

Anhydrous ammonia injection is popular for economic reasons. Ammonia is naturally the cheapest form of fixed nitrogen at the point of production because no further processing is required. Also, the nitrogen content of anhydrous ammonia (82.4% N) is almost twice that of the next most concentrated nitrogen fertilizer (urea: 46% N); this high nitrogen content reduces shipping and handling weight proportionally.

From production to field application, ammonia is handled in liquefied form. Because ammonia is a gas at normal atmospheric pressure and temperatures, its maintenance in liquid form requires either pressurization or refrigeration. Although such handling requires specialized equipment and considerable safety precautions, there are advantages and economies inherent in the handling of a fluid rather than a solid.3

In order to maintain ammonia in liquid form at atmospheric pressure, it is necessary to refrigerate it to at least -33°C. In storage and handling without refrigeration, rather high pressures must be contained when atmospheric temperatures reach high levels. For example, a pressure of l9 kg/cm2, which for safety reasons has been set as the upper limit allowable for storage vessels at retail locations, will maintain the liquid form only up to a temperature of 46°C. Safety pop-off valves are set to vent ammonia at higher pressures.

Facilities for handling liquid ammonia from points of production to points of use are highly developed in the United States. A considerable amount of ammonia is moved hundreds of miles through interstate pipelines. Major pipelines run from points of production in Texas, Oklahoma, and Louisiana oil fields all the way to the intensively farmed mid-western area of the United States. Storage at the points of production and at large terminals sometimes is in pressurized vessels (spheres), but more often it is at atmospheric pressure in large, insulated, refrigerated vessels. Refrigeration is effected simply by allowing vaporization of some of the contained ammonia, compressing the resulting vapor, and recondensing it in a water-cooled condenser. Barges with similar self-sustaining refrigeration systems commonly move large amounts of anhydrous ammonia on rivers and on the high seas. Also, specially designed railroad tank cars ranging in capacity from 24 to 73 tons are used widely for ammonia transport. These cars typically are nonrefrigerated, instead depending on pressure retention to maintain the liquid form. Large pressurized tank-trucks also are used. At retail depots in farming areas, storage normally is in pressure vessels, often with some compressor capability to recondense vapor and to facilitate transfers. Movement of ammonia to farm fields is via wheeled pressure tanks known as "nurse" tanks. Application equipment, usually tractor-drawn, is equipped with pressure tanks that are filled by transfer from nurse tanks.

Because anhydrous ammonia is gaseous at atmospheric conditions, for fertilizer use it must be injected below the surface of the soil to ensure absorption by the soil solution. Most applicators are equipped with a metering system and a series of soil-penetrating "knives"; the metered ammonia is piped through tubing down the trailing edge of each knife to a single opening at the deepest extremity. A penetration of 1525 cm is typical. Retention of ammonia in the soil is best when abundant soil moisture is present.

Aqua Ammonia

Ammonia in a simple water solution is popular as a nitrogen fertilizer in some locations, but overall it claims less than 1 percent of the total nitrogen fertilizer market. This product is made from anhydrous ammonia in units referred to as "converters." These units simply provide cooling as ammonia and water are combined in metered proportions. Aqua ammonia normally contains 20 percent nitrogen (24% NH_3). Its chief advantage over anhydrous ammonia is its low vapor pressure; this allows handling in nonpressure equipment, which simplifies design and reduces safety hazards. The chief disadvantage of aqua ammonia as a fertilizer is its relatively low nitrogen content, which limits its shipping range and increases field handling requirements. As with anhydrous ammonia, the preferred application method is knifing below the soil surface to limit volatilization loss; however, a relatively shallow injection depth of only 713 cm can be used.

Ammonium Nitrate and Urea

Almost half of the total global fertilizer nitrogen application is as ammonium nitrate or urea, in either solid or solution form. The history and the technology of ammonium nitrate and urea production from synthetic ammonia are covered in detail in Chapter 29. The production of nitric acid from synthetic ammonia, which is a prerequisite to the production of ammonium nitrate, also is covered completely in Chapter 29. The production plants for all these materials are, like synthetic ammonia plants, highly sophisticated, involving all facets of modern engineering.

The production processes for both ammonium nitrate and urea yield their products first in the form of saturated solutions. Typical ammonium nitrate solution concentration is 7585 percent (2630% N) at 4075°C, and typical urea solution concentration is about 75 percent (34% N) at 93°C. To produce the solid forms of these products, the solutions first must be concentrated to extremely low water contents (12%) and then fed to prilling or granulation equipment. The resultant solid products that range in size between 2 and 4 mm are highly acceptable to farmers as high-nitrogen fertilizers and to "bulk blenders" as nitrogen sources in the formulation of dry-blend mixed fertilizers, as will be described later. The nitrogen content of solid ammonium nitrate is 34 percent, and that of urea is 46 percent.

The concentration and the solidification of ammonium nitrate and urea solutions to form solid fertilizers require considerable expenditure of energy and processing expense; so it is attractive to use the solutions in the formulation of liquid fertilizers, without processing them into solid form. However, neither ammonium nitrate nor urea alone is sufficiently soluble at ambient temperatures to be attractive for direct use as liquid fertilizer. With either compound alone, a solution with a salting-out temperature of 0°C would contain only about 18 percent nitrogen. Fortunately, however, the joint solubility of the two compounds in water is much more favorable than this; so the liquid nitrogen solutions marketed for fertilizer use almost invariably are joint solutions of ammonium nitrate and urea. Typical commercial solutions contain 2832 percent nitrogen, with compositions and characteristics as shown in Table 11.1.

TABLE 11.1 Physical and Chemical Characteristics of UreaAmmonium Nitrate Nonpressure Nitrogen Solutions

Grade, % N Composition by weight	*28*	*30*	*32*
Ammonium nitrate, %	40.1	42.2	43.3
Urea, %	30.0	32.7	35.4
Water, %	29.9	25.1	20.3
Specific gravity, 15.6°C	1.283	1.303	1.32

Source: *Fertilizer Manual,* United Nations Industrial Development Organization and International Fertilizer Development Center, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

The high-volume use of these solutions is a result of their economy of production and the safety and convenience of their handling and application. The use of nonpressure equipment and surface application (in contrast to knifing) is normal. Nitrogen solutions of other compositions also are prepared commercially, but are used primarily as intermediates in the preparation of solid fertilizers. Such solutions usually consist of various combinations of water, ammonia, ammonium nitrate, and/or urea. Because of the content of free ammonia, most of these solutions require pressurized handling.

Ammonium Sulfate

Although an excellent nitrogensulfur fertilizer with good physical properties, ammonium sulfate is now outranked economically by ammonium nitrate and urea. The chief reasons for the unpopularity of ammonium sulfate are its relatively low nitrogen content (21%) and its relatively low solubility in liquid fertilizers. At most U.S. locations, it is uneconomical to produce ammonium sulfate from synthetic ammonia and virgin sulfuric acid. As a result, most of the ammonium sulfate available for fertilizer use is a by-product of other processes. A larger source now is by-product ammonium sulfate solution from the production of caprolactum and acrylonitrile, which are synthetic fiber intermediates. This by-product solution normally is fed to steam-heated evaporator crystallizers to yield the solid ammonium sulfate. Additional production of by-product ammonium sulfate is obtained from "spent" sulfuric acid, by reaction with synthetic ammonia. In this case, the production of ammonium sulfate is a method of disposing of acid that already has been partially spent in other processes. Some ammonium sulfate is produced by scrubbing ammonia-containing coke-oven effluent gas with sulfuric acid. This production is discussed later.

By-product ammonium sulfate produced by any of the above-mentioned methods is usually in the form of crystals that are 90 percent or less than 1 mm in size, and thus are much smaller than the size preferred for present-day finished fertilizers. This small size is not undesirable if the material is to be used as feed to a fertilizer granulation plant, but it is undesirable for bulk blending (dry mixing) or direct field application. It is possible to produce suitably large crystals, but only at the expense of greatly reducing crystallizer capacity. Both compaction and granulation processes have been employed with some success on a limited scale to convert the small-crystal material to a larger-size product.7

Ammonia in Multi-Nutrient Fertilizers

All the ammonia-based fertilizers discussed thus far contain only one of the three major plant nutrients, namely nitrogen. Referring again to Fig. 11.5, one may see that ammonia is the source of fertilizer nitrogen also, either directly or via nitrogen solids or solutions, in multi-nutrient fertilizers. These are fertilizers that contain two or three of the major plant nutrientsnitrogen, phosphorus, and potassium. Ammonium phosphates, both mono- and di-, are phosphate fertilizer products that also carry nitrogen. (These will be discussed later as major suppliers of phosphate.) As suppliers of nitrogen, the ammonium phosphates and other multi-nutrient solid and fluid fertilizers collectively furnish about 20 percent of global fertilizer nitrogen. (Production of these fertilizers also will be described later.) The source of their nitrogen is, in all cases, ammonia and/or ammonia-based solids or solutions. Nitric phosphates are suppliers of nitrogen through their content of nitric acid and ammonia. They are not presently produced or used to any significant extent in the United States but are produced in Europe and some other locations. (Their production too will be discussed briefly later.)

Miscellaneous Low-Volume Nitrogen Fertilizers

Mineral Nitrogen

The only nitrogencontaining mineral that is mined commercially for fertilizer use is saltpeter (sodium nitrate, $NaNO_3$). An impure form of sodium nitrate known as "coliche" is found in large, rocklike deposits in a relatively dry desert area of Chile, and these deposits have long been mined to provide natural sodium nitrate for the fertilizer market. The ore is blasted from open mines, sized, and leached with water in vats. Evaporation of the resultant solution yields the commercial crystalline fertilizer product known widely as "Chilean nitrate," "nitrate of soda," or among some farmers simply "sody." Its nitrogen content is 16 percent. This product first was imported and marketed in the United States in 1830, so it is one of the oldest fertilizers on the American market. Sodium nitrate now, however, is a relatively insignificant component of the world's nitrogen fertilizer market, having been supplanted by synthetic nitrogen compounds.

By-product Ammonia

Coke, an essential ingredient in the smelting of iron ore and in some other metallurgical processes, is produced by the heating of coal in an essentially oxygen-free atmosphere. This heating volatilizes numerous hydrocarbon components of the coal and results in decomposition of some of these compounds. The nitrogen content of coals normally is about 1 percent, and approximately half of this is volatilized as ammonia gas in the coking process. Recovery of this ammonia has long been effected by passing the coke-oven effluent gas through scrubbercrystallizers containing sulfuric acid. The resultant crystalline by-product, ammonium sulfate, is an excellent fertilizer source of both nitrogen and the secondary plant nutrient sulfur. The nominal nitrogen content of the product is 20 percent, and the sulfur content is 24 percent. Prior to the introduction of synthetic ammonia production processes (191325), coke-oven ammonia was a major source of fertilizer nitrogen; however, its production was very small when compared with today's demand for fertilizer nitrogen. Today, ammonia production from coke ovens, although still carried out, is insignificant in volume compared with that from ammonia synthesis processes.

11.5 Phosphate Fertilizers

Phosphorus intake is essential to the metabolism of both plants and animals. In the case of plants, the entire uptake is from the soil solution via root absorption. Absorption occurs only when the phosphorus is present in the soil solution in the form of $H_2PO^-_4$, HPO^{2-}_4, or PO^{3-}_4 ions.8 Furthermore, a large proportion of the soil phosphorus is present in chemical forms that are not "available" (assimilable), or are only very slowly available, to crops. Cropping of native soils without phosphate fertilization soon depletes the supply of phosphorus and renders the soil barren. In times before fertilizers became readily available, it was not uncommon to "wear out" a farm by repeated cropping. Crop rotation, which was mentioned earlier as a useful method of converting atmospheric nitrogen to available soil nitrogen, is ineffective as a means of phosphorus fertilization. In fact, the nitrogenconverting legumes generally are voracious consumers of phosphate.

Modern, high-yield agriculture is dependent on regular fertilization with phosphorus compounds that are either immediately soluble in the soil solution or become soluble at a rate sufficient to supply the crop. A factor to contend with in phosphate fertilization is soil "fixation" of phosphorus; that is, the characteristic of many soils to convert at least a portion of applied phosphate to chemical forms that are unavailable to plants. The fixation tendency varies with soil type, and generally is highest in soils of high clay content. The overall result of fixation is that considerably more phosphate must be applied as fertilizer than is removed with crops.

Natural Organic Phosphate Fertilizers

In animal metabolism, phosphorus concentrates in the bones and, in the case of Crustacea, in the shells. Manures and human wastes, previously mentioned as effective but low-volume sources of fertilizer nitrogen, are much less efficient sources of phosphorus than of nitrogen. The phosphorus content of such products is normally only 1.33.0 percent, which is only 2550 percent as great as their nitrogen content. Bird droppings, including chicken manure and guano deposits, are somewhat higher in phosphorus content (7% P), but are available in only relatively insignificant amounts overall.

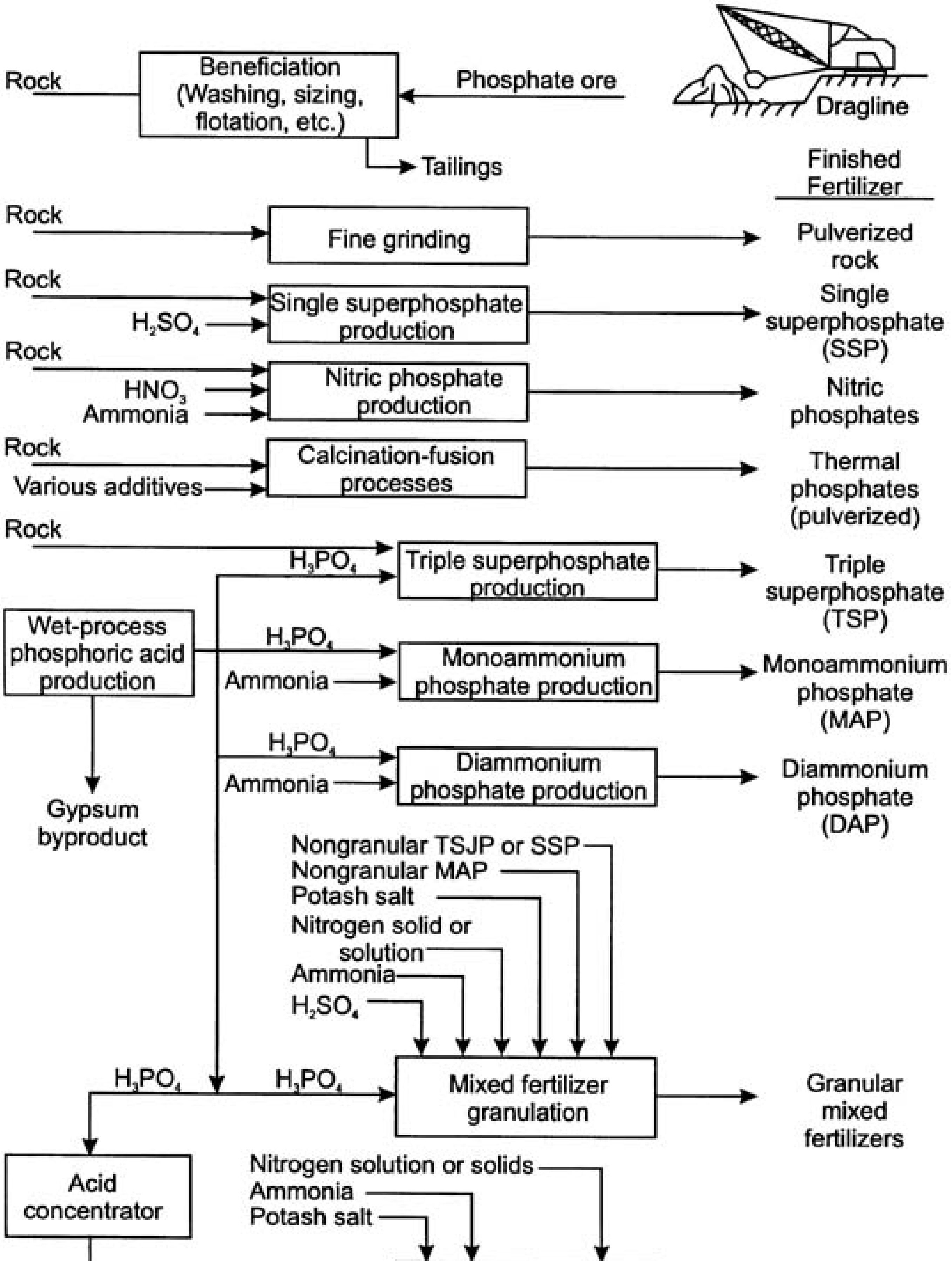
Raw animal bones normally contain 810 percent phosphorus (2025% P_2O_5); thus they are a relatively rich source. The commercial grinding of bones for fertilizer use began in Europe in the early nineteenth century.9 The practice persists today in the production of bone meal, but only in very low volume, chiefly to furnish slowly available nutrients for greenhouse and house plants. In about 1830, it was found that pretreatment of bone meal with dilute sulfuric acid greatly enhanced the "availability" of the phosphorus to plants, and the practice became common. The initiation of this practice now is regarded to have had considerable historic significance as the apparent beginning of the chemical fertilizer industry. The supply of bones being very limited, it soon was found (about 1840) that a similar acid treatment of mined phosphate ore rendered the phosphorus "available" as a fertilizer. Thus was born the still important "superphosphate" industry, and the supplanting of the very limited organic phosphorus sources by widely available mineral sources was begun. Today, organic sources of fertilizer phosphorus account for less than 1 percent of total worldwide consumption.

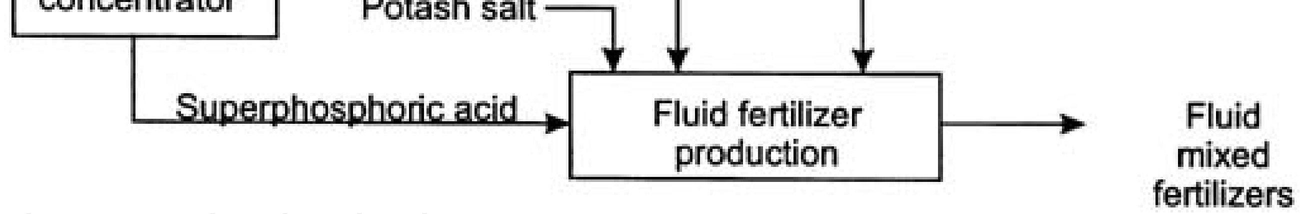
Fertilizers from Mineral Phosphates

Essentially all fertilizer phosphorus now is derived from mined ores. (The occurrence, mineral characteristics, mining, and beneficiation of major phosphate ores were described in some detail in Chapter 10.) Worldwide, about 85 percent of the mined phosphate eventually finds its way into fertilizer.3 As mentioned earlier, the most conservative estimates indicate a sufficiency for hundreds of years at expected consumption levels. Supply problems of the immediate future will relate chiefly to exhaustion of the better ores, with the result that ores of lower grades and higher impurity contents will have to be processed.

Major routes of mineral phosphate (phosphate rock) into finished fertilizers are outlined in Fig. 11.6 and are discussed below. It is obvious from the figure that although phosphate rock is used directly in several major fertilizer production processes (single superphosphate, nitric phosphates), most important processes require that the rock phosphate first be converted to phosphoric acid (H_3PO_4). Phosphoric acid production, then, is a very significant component of the phosphate fertilizer industry. The processes used for the production of fertilizer-grade phosphoric acid are known collectively as "wet" processes, and involve, initially, dissolution of the rock in sulfuric acid. (The highly specialized technology of wet-process acid production was discussed in Chapter 10.)

Fig. 11.6 Major routes of mineral phosphate into finished fertilizers. *(Courtesy TVA)*





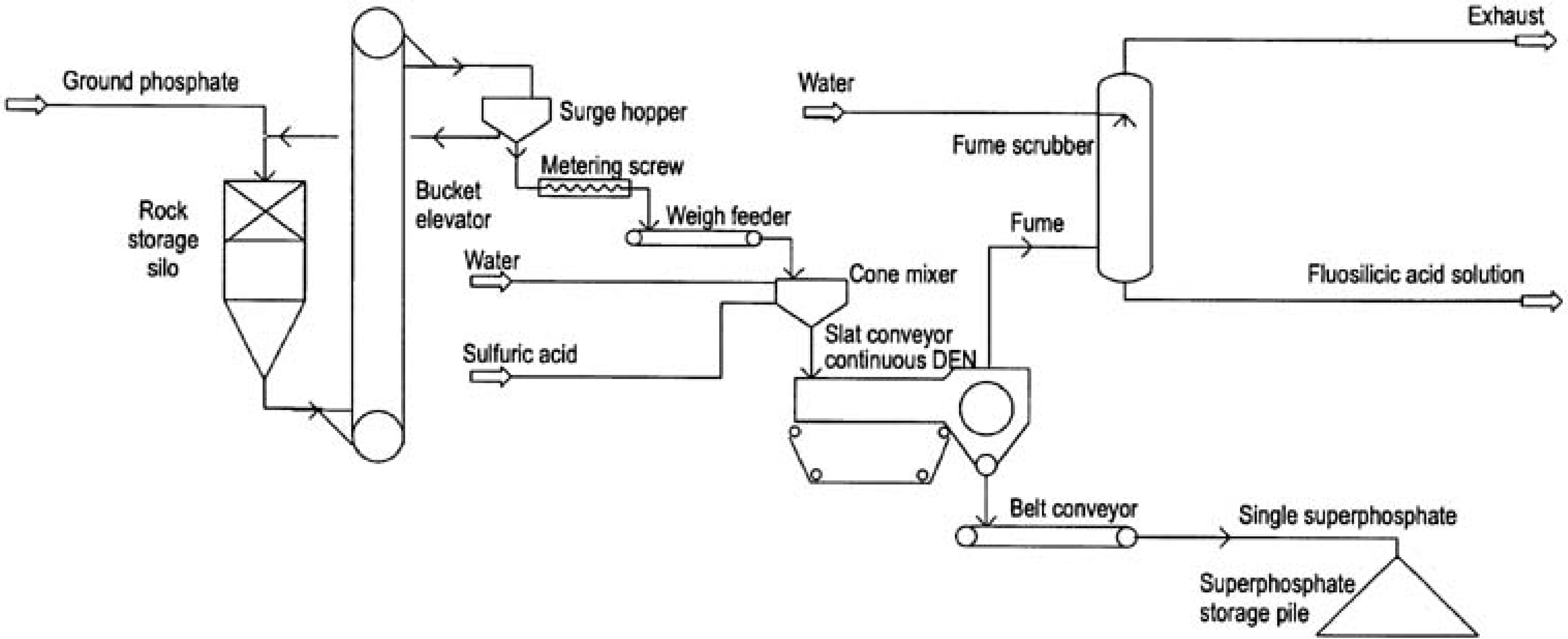
Single Superphosphate (SSP)

In 1842, Lawes in England followed up on the pioneering work of Liebig and received a patent on the use of sulfuric acid in reaction with raw phosphate material to produce "superphosphate" fertilizer. The industry grew slowly, and in 1862 about 150,000 tons of what later would be referred to as "ordinary" (OSP) or "normal" (NSP) or "single" (SSP) superphosphate (1618% P_2O_5) was produced in England. By 1870, there were 70 single superphosphate plants in the United Kingdom and 7 in the Charleston, South Carolina, area of the United States.10 In 1888, commercial shipment of phosphate rock from Florida was initiated, and a major mining industry followed in that state, which continues today. As late as 1955, SSP supplied over 60 percent of the world's phosphate fertilization and was the major phosphate fertilizer in the United States. Rock from Florida or South Carolina was shipped throughout the country to local SSP production facilities. These facilities usually included the capability of blending the SSP with nitrogen and potash materials and bagging the resultant mixtures.

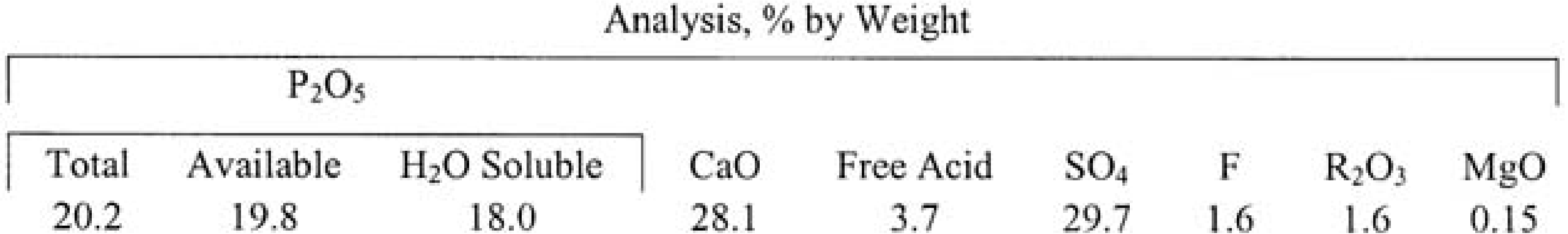
In recent years the production of SSP, to a large extent, has given way to the production of the higher-analysis "triple superphosphate" and ammonium phosphates. SSP retains greater importance in some countries. For example, SSP remains popular in Australia, at least in part because its sulfur content is particularly needed there. SSP is the simplest, as well as the oldest, of manufactured phosphate fertilizers. Its manufacture consists of reacting pulverized phosphate rock with sulfuric acid in comparatively simple equipment, followed by sufficient aging (curing) to complete the reaction and improve the physical condition. The acid rock reaction converts the water-insoluble apatite structure of the rock to soluble monocalcium phosphate. When produced from high-grade rock of 3032 percent P_2O_5 content, SSP contains about 20 percent available P_2O_5; the product made in earlier years from lower-grade rocks contained only 1418 percent P_2O_5. SSP can be used for direct application, for bulk blending, or in the production of granular NPK fertilizers. If the product is to be used for direct application or bulk blending, it usually is granulated to improve handling properties.

The main equipment for manufacturing SSP consists of a mixer to bring together sulfuric acid and phosphate rock, followed by a den, and if granulation is used, some suitable type of equipment for granulation with steam or water. Early plants used batch-type mixers for acidrock mixing, but these mixers mostly have been replaced by continuous types. Several types of continuous mixers have been employed; of these, the lowest-cost, simplest, and apparently most satisfactory one is the TVA cone mixer, which has no moving parts. Mixing is accomplished by the swirling action of the acid. Short, single-shaft or double-shaft pug mills also are used for continuous mixing. The mixers discharge into a den, the purpose of which is to allow time for the acidrock reaction to progress and the physical condition to reach a dry, friable state. The dens can be of either the batch or the continuous type. A slat-conveyor continuous den, commonly referred to as a Broadfield den, is used in many continuous systems.11 A diagram of a typical SSP plant employing a cone mixer and a continuous den (slat-conveyor) is shown in Fig. 11.7. The reaction of phosphate rock with sulfuric acid to produce SSP can be approximated in chemical stoichiometric terms. However, in general practice the proportioning usually is based on a simpler relationship of about 0.6 kg of sulfuric acid (100% H_2SO_4 basis) per kilogram of phosphate rock (3032% P_2O_5). The phosphate rock usually is pulverized to about 90 percent smaller than 0.15 mm and 70 percent smaller than 0.075 mm.

Fig. 11.7 Continuous process for manufacture of single superphosphate. *(Source: Fertilizer Manual.)*



Gases that are released while the superphosphate is solidifying (setting) cause the mass in the den to become porous and friable so that it can be "cut" (disintegrated) and handled readily. SSP made from typical rock will "set" in 4050 min in a continuous den, whereas the set time in a batch den is 1.52 hr. The superphosphate usually is held in storage piles (cured) for 46 weeks in order to obtain better handling properties and to allow the chemical reactions to continue. The usual grade of SSP made from Florida rock is 20 percent available P_2O_5. A typical analysis is shown below:


Analysis, % by Weight

P_2O_5								
Total	Available	H_2O Soluble	CaO	Free Acid	SO_4	F	R_2O_3	MgO
20.2	19.8	18.0	28.1	3.7	29.7	1.6	1.6	0.15

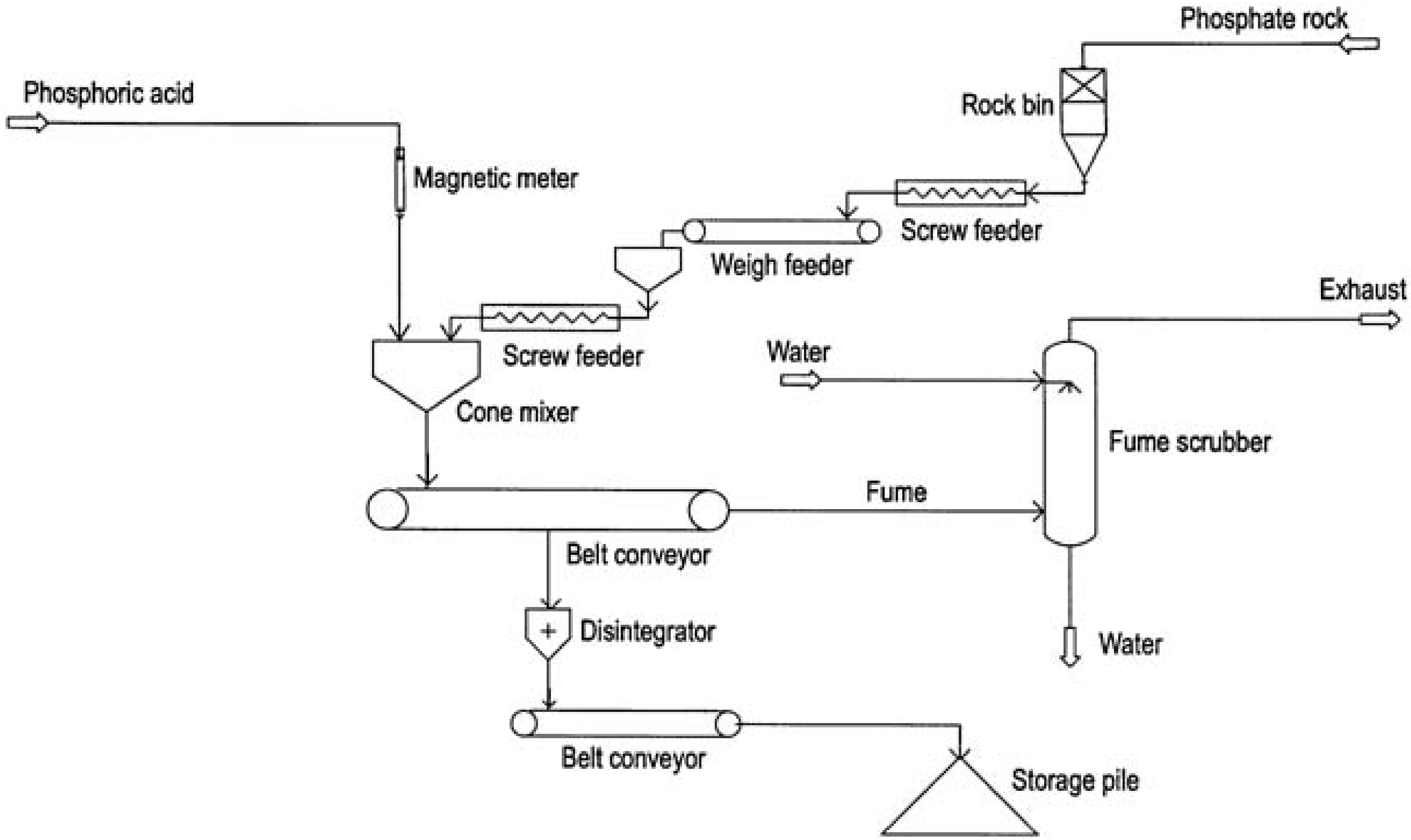
Because of the low phosphate analysis of SSP (20% P_2O_5), economics favor shipping the rock (32% P_2O_5) to local plants and producing the superphosphate there. A majority of the producers use the locally produced SSP in formulations for granular NPK fertilizers made in an ammoniationgranulation process. The maximum amount of SSP possible is used in the NPK formulations because usually it is the lowest-cost form of P_2O_5; however, because of the relatively low P_2O_5 content of SSP, not much can be used in high-analysis NPK formulations. SSP will readily absorb about 6 kg of ammonia per 20 kg of P_2O_5 during the ammoniationgranulation process. Production of granular SSP can be carried out by starting with either cured or ex-den SSP. The SSP is fed to a rotary drum or pan granulator along with a water spray and/or steam. The granulator is followed by a dryer and screening (sizing) system. Fines and crushed oversize material are fed back to the granulator.

Triple Superphosphate

Triple superphosphate (TSP) is made by acidulation of phosphate rock with phosphoric acid, using equipment and processes similar to those for SSP. As in SSP production, the acidrock reaction converts the water-insoluble apatite structure of the rock to water-soluble monocalcium phosphate. TSP, with its relatively high P_2O_5 content of 4546 percent, did not appear on the scene in any appreciable quantity until wet-process phosphoric acid (see Chapter 10) was produced commercially. TVA carried out demonstration-scale production of triple superphosphate in the late 1930s by using electric-furnace phosphoric acid. Extensive agronomic testing and market development, through the use of large tonnages of this TSP by farmers and fertilizer manufacturers in TVA demonstration programs, led to the rapid acceptance of TSP. Producers of phosphate rock in the United States and other countries moved into production of wet-process acid and TSP. Logistics favored production of the higher-analysis TSP (46% P_2O_5 vs. 32% P_2O_5 in phosphate rock) near the source of the phosphate rock, and shipment of the TSP to distribution facilities or blending plants near the markets. Today, most producers of TSP are located adjacent to phosphate mining operations in Florida, Morocco, and Tunisia. Since its introduction, TSP has steadily displaced SSP in the marketplace.

The cone mixer has been used almost universally in the production of nongranular TSP. Because the "set time" for TSP is only 1420 min, as compared with 4050 min for SSP, a simple, cupped conveyor belt usually is used to hold the acidulate until the TSP solidifies; this is instead of the slat-type den used for SSP. With a belt about 1.5 m wide and 30 m long, the production rate usually is 4050 tons/hr. A flow diagram for TSP production by the cone mixer- "wet-belt" system is shown in Fig. 11.8.

Fig. 11.8 Continuous system for manufacture of nongranular triple superphosphate. *(Source: Fertilizer Manual.)*



Proportioning for TSP typically is 2.42.5 kg of P2O5 from acid for each kg of P2O5 from rock. Nongranular TSP usually is cured 46 weeks prior to shipment or used at the site. Typical chemical analysis of TSP made with Florida rock is shown below:

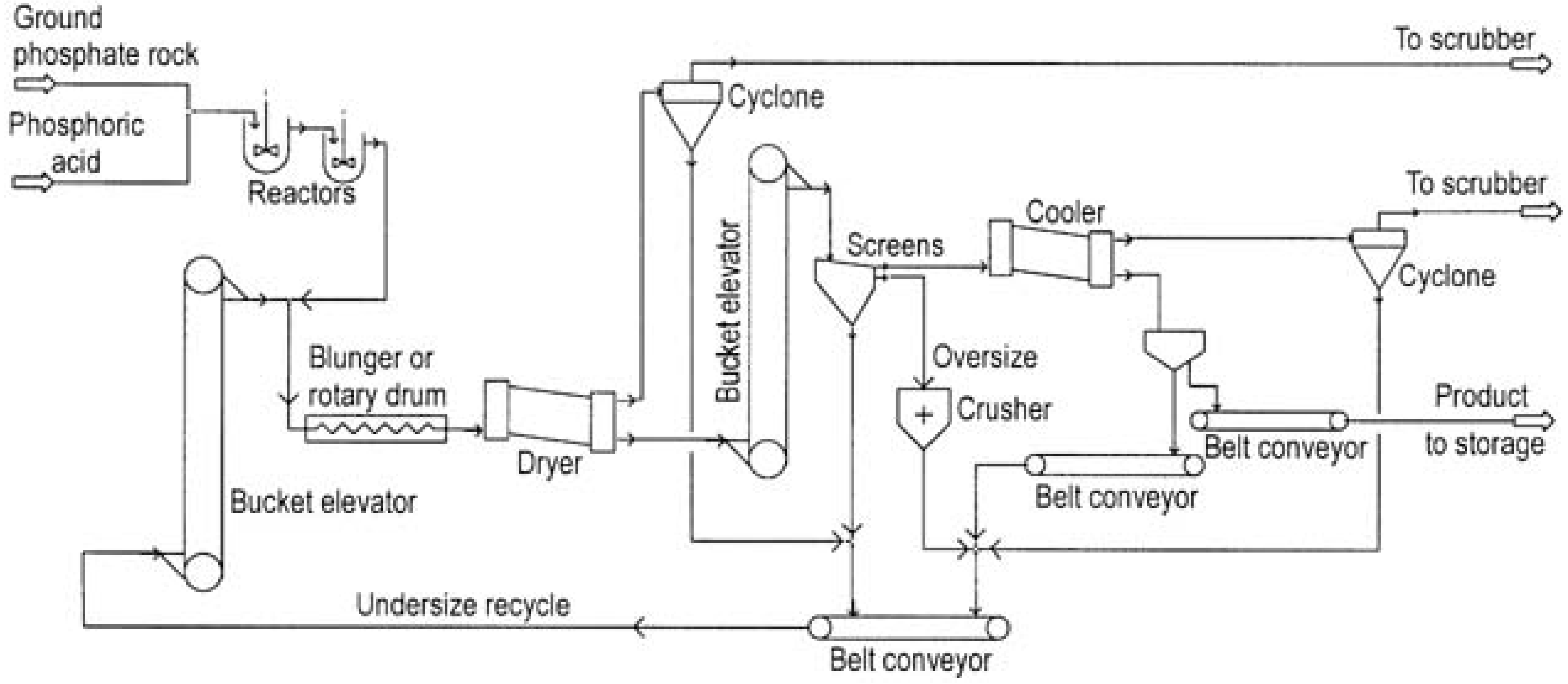
Analysis, % by Weight

P_2O_5								
Total	Available	H_2O Soluble	Free Acid	CaO	R_2O_3	MgO	F	H_2O
46.9	46.3	42.0	3.4	19.3	3.1	0.5	2.7	4.5

The use of TSP in granular NPK fertilizer formulations, together with or in place of SSP, has allowed the production of higher-analysis grades of granular NPK fertilizers, such as 131313 instead of 101010. The TSP can readily be ammoniated to about 3.5 kg of ammonia per 20 kg of P2O5 during the granulation process.

TSP is produced in large quantities in granular form for use in direct application and for use in bulk blends. In some processes cured or ex-den TSP, like SSP, is granulated in a rotary drum or pan granulator using steam and water to promote granulation.3 A slurry-type granulation process, as outlined in Fig. 11.9, is now the process of choice of major producers of TSP. Pulverized phosphate rock is treated with wet-process phosphoric acid in a two-stage reaction system, and the slurry is sprayed into a pug mill or rotary drum where it is layered on recycled undersize and crushed oversize at a ratio of 1 kg per 512 kg of recycle. Product granules (45% P2O5) are quite spherical and dense. The lower grade of the granular product, as compared with that of cured nongranular TSP, results from the need to decrease the acidulation ratio from the usual 2.45 kg of acid P2O5 per kilogram of rock P2O5 to about 2.25 in order to control the free-acid content and to minimize stickiness. For production rates higher than about 25 tons/hr, a rotary drum usually is used instead of a pug mill.

Fig. 11.9 Slurry process for the manufacture of granular triple superphosphate. *(Source: Fertilizer Manual.)*



Nitric Phosphate

Fertilizers that are referred to as nitric phosphate or nitrophosphate are produced by acidulation of phosphate rock with nitric acid or with mixtures of nitric and sulfuric or phosphoric acids. The primary advantage of nitric phosphate processes is that no sulfur or less sulfur is required as compared with superphosphates or ammonium phosphates; this is particularly important during a shortage of sulfur, or in locations where sulfur must be shipped long distances. A variety of processes and equipment have been used in Europe since the late 1930s.3,12 Also there are a number of plants in Central and South America and in Asia. The production of nitric phosphates is complex. Simple substitution of nitric acid in a superphosphate-type acidrock reaction is not feasible because (1) decomposition of the nitric acid would occur and cause noxious fumes and loss of nitrogen and (2) the product would be extremely hygroscopic and unstable.

One process modification that overcomes these difficulties uses only nitric acid for acidulation; the extraction slurry is cooled to crystallize calcium nitrate, which then is removed by centrifugation. This process is referred to as the Odda process. The calcium nitrate by-product is either sold as a fertilizer or converted to co-product ammonium nitrate. The extraction slurry, with its lowered calcium content, then is ammoniated and granulated. Various process modifications utilize a rotary drum, pug mill, or spray drum (Spherodizer) for granulation of the slurry. NPK nitric phosphate grades also have been produced by a prilling process in a few European plants. In earlier years, a main disadvantage of nitric phosphate processes was the low water solubility of phosphate in the products, as the main phosphatic constituent was dicalcium phosphate. The use of supplemental phosphoric acid, or "deep cooling" by refrigeration to remove a higher proportion of calcium nitrate in the Odda-type processes, now allows water solubility of 60 percent or higher. The remaining 40 percent of the phosphate, although water-insoluble, is almost completely citrate-soluble, and this is "available" as plant nutrient. Popular grades of nitric phosphates include 141414, 221111, 20200, and 16130.

Diammonium Phosphate (DAP)

Ammonium phosphates did not come on the fertilizer scene in significant quantities until the early 1960s; however, they have rapidly become the leading form of phosphate fertilizer in the world. Almost all new phosphate fertilizer complexes built in recent years, and those planned, are for the production of ammonium phosphate as the major product. In 1960, world production of ammonium phosphates was equivalent to only 0.5 million mt of P2O5. By 1990 it was up to 15.4 million mt and by 1999 it has reached 16.7 million mt. Ammonium phosphates now provide two-thirds of total world phosphate fertilization.

Prior to the late 1940s, diammonium phosphate (NH4)2HPO4 was considered to be too unstable for use as a fertilizer. However, at about that time, exploratory work by TVA indicated otherwise. The first demonstration-scale diammonium phosphate made by TVA for fertilizer use was a monocrystal product made by ammoniation of pure electric-furnace acid in a vacuum crystallizer.13 Demonstration programs showed this material to be stable and to be an excellent high-analysis fertilizer. Small amounts of similar product were produced by others using electric-furnace acid and by-product ammonia. However, general acceptance of diammonium phosphate by the fertilizer industry was not possible until procedures could be developed to produce it from wet-process phosphoric acid instead of the relatively expensive electric-furnace acid. This occurred in 196061 with the development and patenting of the "TVA process" for production of granular 18460 grade DAP from wet-process acid.14,15 This relatively simple process was rapidly adopted by the industry, and remains the basic process for production of this major phosphate fertilizer. Many granular DAP plants have production capacities of about 50 tons/hr, with some going as high as 70100 tons/hr.

The flow diagram of a typical granular DAP production unit of the TVA type that has become standard in the industries is shown in Fig. 11.10. Wet-process phosphoric acid of about 40 percent P_2O_5 content (often a mixture of 54% P_2O_5 acid and acid from the scrubbing circuit of 2830% P_2O_5 content) is fed to a preneutralizer vessel. Anhydrous ammonia is sparged into this preneutralizer through open-end pipes that project through the walls. A key feature of the process is that this neutralization of the acid is carried to a closely controlled NH_3:H_3PO_4 mole ratio of about 1.4. Ammoniation to this degree provides maximum solubility of ammonium phosphate, as is shown by the solubility curve of Fig. 11.11. The heat of the ammoniation reaction evaporates considerable water, with the result that the water content of the slurry is reduced to only 1620 percent. The slurry still is fluid and pumpable because of the high solubility at a 1.4 mole ratio.

Fig. 11.10 TVA process for production of granular diammonium phosphate. *(Source: Fertilizer Manual.)*

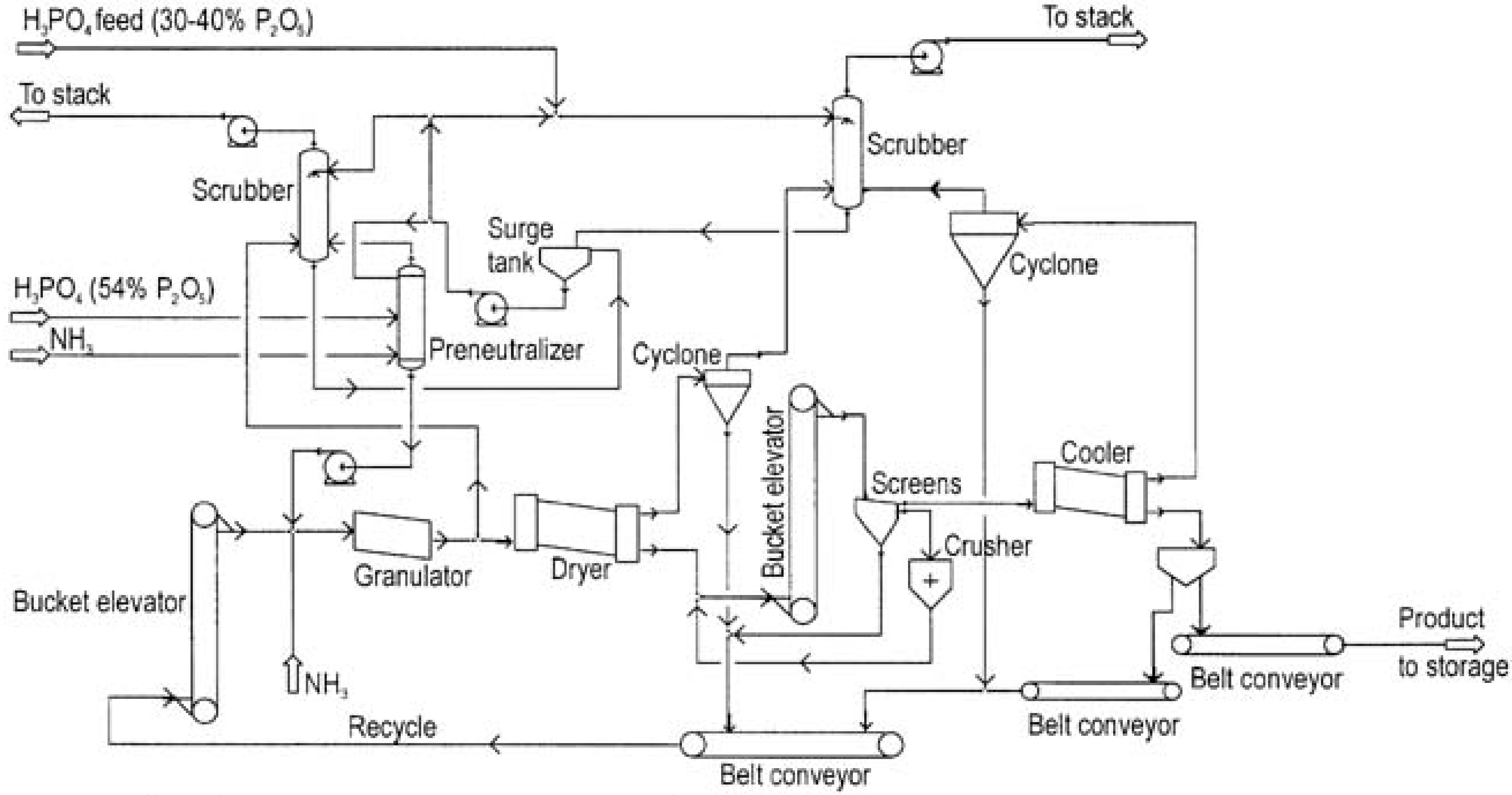
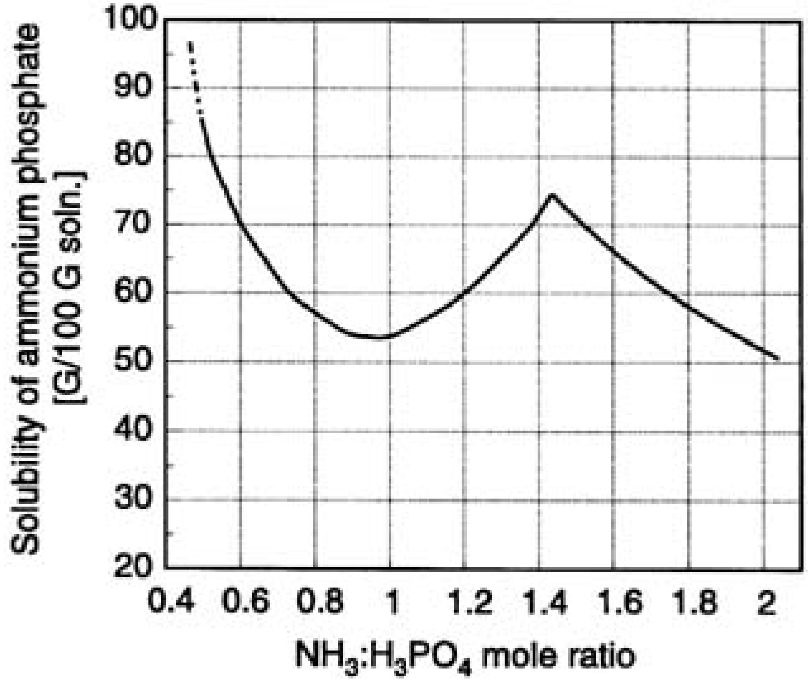


Fig. 11.11 Effect of NH_3: H_3PO_4 mole ratio on solubility of ammonium phosphate at 75°C. *(Source: Fertilizer Manual.)*



The preneutralizer slurry is pumped at a controlled rate to a rotary-drum, ammoniator granulator; there it is distributed onto a rolling bed of recycled solids. The most commonly used metering system for the preneutralized slurry is a variable-speed centrifugal pump with automatic control from a magnetic flowmeter. Some plants have had success with a magnetic flowmeter and automatic control valve of a special ball type.

Ammonia is sparged beneath the bed in the rotary drum to ammoniate the slurry further to near the theoretical DAP ratio of 2.0; the usual finishing NH_3: H_3PO_4 mole ratio is 1.851.94. Metallic phosphate impurities such as iron, aluminum, and calcium phosphates account for most of the deviation from theoretical. This ammoniation in the granulator drastically decreases ammonium phosphate solubility (Fig. 11.11), and thus promotes solidification and granulation. Ammonia absorption is not 100 percent complete in either the preneutralizer or the granulator. However, the unabsorbed ammonia is recovered in a scrubbing circuit containing phosphoric acid of about 30 percent P_2O_5 content. Solid recycle to the granulator drum consists of undersize from screening of dried product plus some crushed, dry product and oversize. These combined materials are recycled at a rate of 57 kg per kilogram of product, and variation of this recycle rate is the primary method of controlling granulation efficiency. Product discharged from the granulator is dried with moderate heat to a 180190°F product temperature. Higher drying temperatures must be avoided because of a relatively low melting point for DAP. Most plants screen hot and cool only the product fraction because the material is not too sticky for recycle without cooling. Rotary coolers or fluidized-bed-type coolers are used. The product, with a moisture content of 1.52 percent, does not require a conditioner. It has excellent storage and handling properties in bags or in bulk. It has been shown that the iron and aluminum phosphate impurities in the product are present in a gel-like structure that favorably hardens the granules and increases their resistance to moisture.16 The critical relative humidity of diammo-nium phosphate made from wet-process acid is about 70 percent; thus it is relatively nonhygroscopic.

Construction materials other than mild steel are required only for the acid lines, the preneutralizer, the slurry handling system, and the scrubbing circuits. Type 316L stainless steel or rubber- and brick-lined mild steel is used for the preneutralizer. Type 316L stainless is used also for the slurry pumps and piping. Fiberglass-reinforced polyester plastic and high-density polyvinyl chloride are sometimes used for wet-process acid pipes and for scrubbers. Scrubbers also may be constructed of rubber-lined mild steel.

Monoammonium Phosphate (MAP)

"Diammonium Phosphate," "DAP," and the grade designation "18460" have become extremely familiar terms in the world fertilizer industry. However, substantial interest has developed also in fertilizer use of monoammonium phosphate, $NH_4H_2PO_4$. Agronomically, MAP is favored where soils are mainly alkaline, for example in Canada, Egypt, and Pakistan. From an economic standpoint, the relatively high P_2O_5 contents of typical MAP grades (11520 to 10540) make them attractive where the primary interest is in producing and shipping phosphate; the MAPs provide a higher P_2O_5 payload than does DAP.

TVA developed two comparatively minor modifications of the granular DAP process to allow production of granular MAP.17 In one method, the acid is ammoniated to an NH_3:H_3PO_4, mole ratio of only about 0.6 in the preneutralizer and then to about 1.0 in the granulator drum. The 0.6 ratio is, like the 1.4 used in DAP production, a high-solubility point in the ammonium phosphate system (Fig. 11.11); the subsequent ammoniation to 1.0 in the granulator decreases solubility and thus promotes granulation. In the other procedure, which has been preferred by the industry, acid in the preneutralizer is ammoniated to the high-solubility NH_3: H_3PO_4 ratio of about 1.4, as in DAP production; then additional wet-process acid is distributed onto the bed in the granulator to adjust back to the low-solubility MAP mole ratio of about 1.0. With either modification, the remainder of the process is the same as for DAP, except that, because of the higher temperature stability of MAP, a higher drying temperature can be used to increase the production rate for MAP.

Starting in about 1968, simple processes were developed for the production of nongranular (sometimes called powdered) MAP. Chief developers were Fisons and Scottish Agricultural Industries (SAI) in the United Kingdom, Swift in the United States, and Nissan in Japan. These processes involve variations in the simple reaction of wet-process acid with ammonia followed by spray-type drying with air. On the basis of this work, a number of comparatively low-cost units for production of nongranular MAP have been built commercially, including plants in the United Kingdom, the Netherlands, Japan, Australia, Spain, the United States, Brazil, and Iran. This intermediate usually is shipped to other plants, where it then is used in the production of NPK solid or fluid fertilizers. Thus far, however, the nongranular MAP has not attained the popularity that was predicted for it in the mid-1970s.

Miscellaneous Low-Volume Phosphate Fertilizers

Phosphate Rock for Direct Application

Application of finely pulverized raw phosphate rock directly to the soil has been utilized almost from the beginning of fertilization practice. Although such direct use eliminates the need for most processing, there are major drawbacks. The ores are of low phosphorus content in comparison with most manufactured fertilizers, which increases shipping cost. Also, the very fine grinds that are required to encourage solubility are difficult to handle and to apply. The major drawback, however, is agronomic. The agronomic effectiveness of raw rock depends upon the origin and the chemical and mineralogical nature of the particular rock, the pH and other characteristics of the soil, and the crop produced. Researchers at TVA and IFDC have characterized phosphate rocks from a large number of sources as to their reactivity.3,18 In general, phosphate rock of sedimentary origin is more reactive than igneous source. Phosphate rock from North Carolina (U.S.), Sechura (Peru), and Gafsa (Tunisia) are at the top, and are about equal in reactivity and suitability for direct application. Other phosphate rocks that are reasonably reactive and are marketed for this purpose include some types from Morocco, Israel, Algeria, and a few other locations. In 1998, a world total of about 2.0 million mt of phosphate rock was used for direct application. This amount represented about 1.7 percent of the total phosphate consumption for 1998.

Defluorinated Phosphate Rock

There is substantial production of defluorinated phosphate rock for fertilizer use in Japan (about 100,000 mt/year). Ground, high-grade rock is mixed with small proportions of sodium carbonate or sulfate and wet-process acid. The mixture is calcined at a temperature of 1350°C in an oil-fired rotary kiln 45.0 m in length and 2.7 m in diameter. The product contains 3842 percent P_2O_5 of which more than 90 percent is soluble in neutral ammonium nitrate solution and is an effective fertilizer on acid soils. During the production of defluorinated phosphate rock, substantially all fluorine is driven off. Sodium bifluoride ($NaHF_2$) is recovered as a by-product. A similar product is made in the United States, but it is mainly used for animal feed supplement.

Calcined Aluminum Phosphate

Ores consisting mainly of hydrated aluminum phosphate minerals are useful for direct application after thermal treatment (calcination) at a controlled temperature. Calcination removes the combined water and destroys the crystallinity of the phosphate minerals, thereby improving the reactivity of the phosphate.3

At present, only two aluminum phosphate ores are known to be used for direct application after thermal treatment. One such ore is found near Thies, in Senegal. The "as-mined" ore contains about 29 percent P_2O_5 on a dry basis; after calcination it contains over 34 percent P_2O_5. The calcined product is ground to about 95 percent smaller than 0.15 mm and marketed for direct application under the trade name "Phosphal." About 75 percent of the P_2O_5 is soluble in alkaline ammonium citrate, and the product is effective as superphosphate in calcareous, neutral, and acid soils. The product is used in France, Senegal, and some nearby countries.

A similar ore occurs in the upper stratum ("C-grade") of the phosphate deposit on Christmas Island. The C-grade ore contains about 25 percent P_2O_5 dry basis. Calcination is carried out in a fluidized bed, operated in the range of 400600°C. The maximum solubility in alkaline ammonium citrate solution is about 70 percent.

Partially Acidulated Phosphate Rock

Not all phosphate rock is suitable for direct application and may require the addition of a more soluble form of phosphate. An alternative is partial acidulation to render its P_2O_5 more available. The partially acidulated phosphate rock (PAPR) process depends on treating ground phosphate rock with only a portion of the stoichiometric value of acid, for example 50 percent PAPR.19 The amount of watersoluble phosphate in PAPR varies according to the degree of acidulation. The crop response to the product is often quite similar to fully acidulated products such as SSP or TSP.

Sulfuric and phosphoric acids are the most common acids used for partial acidulation. Possibilities exist for using three processes: a conventional run-of-pile process followed by granulation, IFDC-developed singlestep acidulation and granulation, and a slurry granulation triple superphosphatetype process.

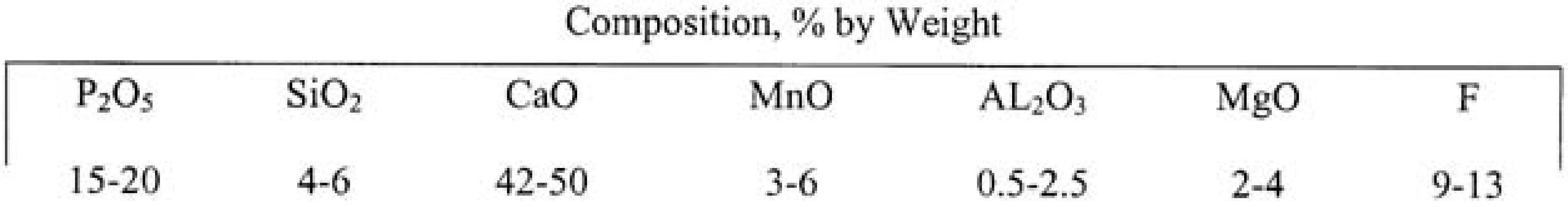
Enriched Superphosphate

This product, which is also referred to as "double" superphosphate, is essentially a mixture of SSP and TSP, usually made by acidulation of phosphate rock and a mixture of sulfuric and phosphoric acids. The desired concentration of mixed acid can be obtained by mixing concentrated sulfuric acid (93 or 98% H_2SO_4) with dilute phosphoric acid (30%) thus avoiding the need for concentrating the latter. Production processes and equipment are about the same as for SSP.20

Theoretically, any grade between SSP and TSP can be produced, but the usual range is 2535 percent P_2O_5. The product is useful in sulfur-deficient areas where SSP would supply more sulfur than necessary. Only a small amount of enriched superphosphate is produced and used in Australia.

Basic Slag

This is also known as Thomas slag and is a by-product of the steel industry. Iron made from high-phosphorus ore is converted to steel in a Thomas converter by oxidation in contact with a basic (high CaO) slag. Sometimes phosphate rock is deliberately added to the blast furnace charge to increase the phosphorus content of the iron thereby increasing the P_2O_5 content of the slag. A typical range of composition of high-grade basic slag is:



Composition, % by Weight

P_2O_5	SiO_2	CaO	MnO	AL_2O_3	MgO	F
15-20	4-6	42-50	3-6	0.5-2.5	2-4	9-13

While the present market for basic slag is quite small, it is primarily produced and used in Western Europe where high-phosphorus iron ore occurs.

Fused Calcium Magnesium Phosphate

In a process developed by TVA, a mixture of phosphate rock and olivine or serpentine (magnesium silicate) is fused in an electric furnace.21 The molten product is quenched with water and used in a finely divided state as a fertilizer. The product, a calcium magnesium phosphate (CMP) glass, contains about 20 percent P_2O_5 and 15 percent MgO. Over 90 percent of the product is soluble in citric acid.

The theoretical compositions of some magnesium-containing minerals that can be used to produce CMP are:

Olivine	$(Mg, Fe)_2SiO_4$
Serpentine	$Mg_3H_4Si_2O_9$
Garnierite	$(Mg, Ni)H_2SiO_4$
Magnesite	$MgCO_3$

Although small compared with soluble phosphates, CMP is used in Japan, China, Korea, Taiwan, and South Africa. The reason for the decline in production and use is the relatively high (850 kWhr/ton of product) energy requirements.

The CMP is found to be more effective than superphosphate when used on acid soils. The product has a liming value equivalent to 0.50.7 ton of calcium carbonate per ton of material. The magnesium oxide and soil-soluble silica is available to growing plants.

Rhenania Phosphate

Rhenania phosphate is another thermally produced phosphate fertilizer. It is made by calcining a mixture of phosphate rock, sodium carbonate, and silica in a rotary kiln at 1250°C.3 Enough sodium carbonate is used to form the compound $CaNaPO_4$ and enough silica to form Ca_2SiO_4 with excess calcium. Typical charge proportions are one part sodium carbonate to three parts of phosphate rock and enough silica to raise the SiO_2 content of the product to about 10 percent. The overall reaction in producing Rhenania phosphate is assumed to be:

$$\mathrm{Ca_{10}F_2(PO_4)_6 + 4Na_2CO_3 + 2SiO_2}$$
$$\mathrm{\rightarrow 6CaNaPO_4 + 2Ca_2SiO_4 + 2NaF + 4CO_2}$$

Any grade of phosphate rock can be used, but since the grade of the product is determined by the grade of the rock, a high grade is preferred.

A typical product contains 2830 percent P_2O_5, nearly all soluble in neutral or alkaline ammonium citrate solution. The product is applied in a powdered form or granulated with potash salts. Some grades are produced containing magnesium or boron, which are added during granulation as kieserite or borax, respectively.

11.6 Potassium Salts

The element potassium has been recognized to be beneficial to plant growth since J. R. Glauber in the Netherlands first proposed, in the middle of the seventeenth century, that saltpeter (KNO3) was the "principle" of vegetation. This salt, derived from the leaching of coral soils, was thus the first chemical compound intentionally applied to crops to satisfy their nutrient needs. Justus von Leibig had established the essentiality of the element for plant growth in his pioneering work published in 1840. The potassium or potash chemical industry dates from the year 1861, following the discovery by German chemists of a process for recovery of KCl (muriate of potash) from rubbish salts, the residues remaining from the extraction of sodium chloride (NaCl).

The term "potash" generally refers to potassium chloride, but it is also used to designate other potassium compounds that may be qualified by the words muriate of potash, sulfate of potash, etc.

The functions of potassium in the plant are manifold. This element serves to activate or catalyze a host of enzyme actions, to facili-tate the transport of nutrients and assimilates in the xylem and phloem, to maintain the structural integrity of the plant cell, to regulate turgor pressure, to mediate the fixation of nitrogen in leguminous plant species, and to protect plants to some degree from certain plant diseases.

Potassium is a relatively abundant element in the earth's crust, ranking seventh in concentration. It is widely distributed geographically and is commonly found in association with sodium compounds. The feldspars, muscovite (white mica), granite, and gneiss are rich sources; but because they are siliceous and refractory, they are difficult and costly to convert to forms suitable for use as fertilizers. The common potassium minerals are listed in Table 11.2.

TABLE 11.2 Common K Minerals

Mineral	Formula	K Content (g/kg)	K2O Content (g/kg)
Sylvite	KCl	524.4	631.7
Carnalite	KCl · MgCl2 · 6H2O	140.7	169.5
Kainite	KCl · MgSO4 · 3H2O	157.1	189.2
Langbeinite	K2SO4 · 2MgSO4	188.5	227.0
Leonite	K2SO4 · MgSO4 · 4H2O	213.3	256.9
Schoenite	K2SO4 · MgSO4 · 6H2O	194.2	233.9
Polyhalite	K2SO4 · MgSO4 · 2CaSO4 · 2H2O	129.7	156.2

Source: Stewart, J. A., in *Potassium in Agriculture*, R. D. Munson (Ed.), p. 83, ASA-CSSA-SSSA, 1985.

The name potash derives from an early production method in which potassium carbonate, leached from wood ashes, was crystallized by evaporating the leachate in large iron pots. The salt potassium chloride (muriate of potash or KCl) is now the major source of the element (95%); other important salts are potassium sulfate (sulfate of potash, K2SO4), potassium magnesium sulfates of varying K/Mg ratios, and potassium nitrate (KNO3).

The potash industry is based on very large deposits of water-soluble potassium minerals resulting from the evaporation of shallow seas or natural brine lakes over a geological time span. These evaporates normally are located at depth in the earth's mantle. Ironically, the best deposits are found in areas quite remote from the more productive agricultural regions, which thus are the areas most in need of this element. Typically, the deposits are to be found in horizontal tabular bodies or beds and may occur at depths up to 2100 m or more. The beds may be only a few centimeters to a few meters thick, but commercial production is limited to strata that are at least 1 m thick. As with hard-rock mining, the potash-bearing ores are extracted or harvested with continuous mining machines and brought to the surface through vertical shafts. Where the ores occur below a depth of about 1100 m, or where the beds exhibit geological anomalies (e.g. folding), the potash is dissolved in a brine solution and pumped to the surface for recovery using solution-mining techniques.

Potassium Minerals

The ore zone or stratum typically contains potassium or potassiummagnesium minerals along with halite (sodium chloride). Muriate of potash is refined from sylvinite ore, a mechanical mixture of potassium chloride (KCl) and sodium chloride (NaCl). Because the latter salt is injurious to most crop plants, the KCl (sylvite) must be separated from the NaCl (halite).

For coarse-grained ores, physical methods employing froth flotation and/or heavy-media methods are employed to achieve separation. The heavy medium may consist of, for example, pulverized magnetite in a brine solution. The specific gravity of the medium is adjusted so that it falls between that of KCl (1.99 g/cm3) and that of NaCl (2.17 g/cm3).

When the crushed ore is placed in this medium, the potassium values are floated off, and the contaminating sodium values sink and are drawn off and rejected from the bottom of the flotation vessel. Entrapped magnetite from both the product and waste streams is easily recovered magnetically and returned to the flotation vessel.

In froth flotation the crushed ore, which has been scrubbed to remove contaminating clay particles, first is treated with an aliphatic amine acetate and a froth-promoting alcohol. The water-repellant amine acetate is attracted preferentially to the sylvite particles. In a subsequent step, the ore or pulp is maintained in suspension, and the coated sylvite particles are attracted to and entrained on the rising air bubbles generated by the flotation agitator. They rise to the surface and are collected by hydraulic and mechanical means.

The potassium-depleted pulp is then discarded. The sylvite or muriate is collected, centrifuged to remove the hydraulic medium, dried, screened, into various particle size ranges, and sent to storage.

PotassiumMagnesium Minerals

Other important sources of potassium for fertilizer use are the double salts kainite (KCl · MgSO4 · 3H2O) and langbeinite (K2SO · 2MgSO4). The former is recovered from potash mines in Germany and Sicily; the latter is recovered and processed from mines in New Mexico (U.S.). These minerals supply important quantities of both magnesium and sulfur as well as potassium, and all are essential nutrients for plant growth. Kainite has a theoretical composition of 15.99 percent K (19.26% K2O), 9.94 percent Mg, and 13.11 percent S, and langbeinite contains 18.85 percent K (22.7% K2O), 11.71 percent Mg, and 23.18 percent S.

In the processing of langbeinite, the contaminating chlorides (NaCl, KCl) are removed by extracting the crushed ore with water. Centrifugation, drying, and screening follow to prepare the product in various size grades for the market. Where the mixed ore contains recoverable quantities as KCl as well as langbeinite, froth flotation and heavy-media methods may be employed to recover both potassium-containing salts.

Potassium Sulfate

Sulfate of potash (K2SO4), unlike the earlier-discussed potash salts, does not occur as natural deposits. It can be recovered by fractional crystallization from such natural brines as those of the Great Salt Lake in Utah and Searles Lake in California. Here separation and recovery are achieved by solar evaporation in shallow ponds. These processes can be utilized only where a suitable brine source is available, and where solar evaporation rates are high.

A much more important source of K2SO4 is the product resulting from either a controlled decomposition reaction of a complex sulfate salt, reaction of the salt with KCl, or both. Thus, kainite (KCl · MgSO4 · 3H2O) may be reacted with water in a two-step reaction, first to form schoenite (K2SO4 · MgSO4 · 6H2O), with the schoenite then reacted with KCl to form K2SO4.

Langbeinite likewise is reacted with KCl to form sulfate of potash, according to the following reaction:

$$\underset{\text{langbeinite}}{K_2SO_4 \bullet MgSO_4} \quad + \quad \underset{\substack{\text{potassium}\\\text{chloride}}}{4KCl} \longrightarrow \underset{\substack{\text{potassium}\\\text{sulfate}}}{3K_2SO_4} \quad + \quad \underset{\substack{\text{magnesium}\\\text{chloride}}}{2MgCl_2\ \text{(solution)}}$$

Approximately 50 percent of world K2SO4 production is derived from the reaction of KCl with either kainite or langbeinite.

For those producers who do not have access to such suitable sulfate salts to serve as the sulfate donor, the so-called Mannheim process, in which sulfuric acid is reacted with KCl, may be utilized. The following reaction unfortunately generates the corrosive substance hydrochloric acid as a co-product:

$$\underset{\substack{\text{potassium}\\\text{chloride}}}{2KCl} \quad + \quad \underset{\substack{\text{sulfuric}\\\text{acid}}}{H_2SO_4} \longrightarrow \underset{\substack{\text{potassium}\\\text{sulfate}}}{K_2SO_4} \quad + \quad \underset{\substack{\text{hydrochloric}\\\text{acid}}}{2HCl}$$

Potassium sulfate, though relatively costly to produce, is a preferred source of potassium for some vegetable and fruit crops and tobaccocrops that are sensitive to the chloride forms.

Potassium Nitrate

The third most widely used potassium salt in agriculture is potassium nitrate (KNO3). An ore containing sodium nitrate, potassium nitrate, some chlorides, and sulfates is mined in Chile. The ore is leached and part of the sodium nitrate is crystallized and removed.

The salt is produced in small amounts commercially in the United States, Israel, and Norway by the reaction of KCl with nitric acid. Though the salt is an attractive fertilizer material containing the two essential nutrients, nitrogen and potassium, in plant-available form, it is expensive to produce and thus enjoys only a limited market.

Potassium Phosphates

Potassium phosphates are excellent fertilizers, and their very high analysis is an advantage that has stimulated much research in an effort to find an economical production process. However, no process has been developed that is economical enough to result in widespread production; therefore, present use is limited to special purposes for which the high cost can be justified.

At present, most of the potassium phosphates used in fertilizers are produced from potassium hydroxide or carbonate and phosphoric acid and are used in liquids for foliar application or other specialty uses.

Some of the potassium phosphates are:

		Approximate grade
Monopotassium phosphate	KH2PO4	05235
Dipotassium phosphate	K2HPO4	04054
Tetrapotassium pyrophosphate	K4P2O7	04357
Potassium metaphosphate	KPO3	06040

In addition, a potassium polyphosphate solution of 02627 grade has been produced from superphosphoric acid and potassium hydroxide; it contains a mixture of ortho, pyro, and higher polyphosphates.

Another process that received attention involves the use of concentrated sulfuric acid to treat potassium chloride to produce potassium bisulfate:22,23

$$KCl + H_2SO_4 \longrightarrow KHSO_4 + HCl$$

The hydrogen chloride is recovered as a dry gas and can be used for certain oxychlorination reactions or other industrial purposes for which elemental chlorine is now used. The potassium bisulfate is used, with more sulfuric acid, to make monopotassium phosphate and phosphoric acid. A simplified equation, representing phosphate rock as tricalcium phosphate, is:

$$\begin{aligned}&KHSO_4 + 2H_2SO_4 + Ca_3(PO_4)_2 + 6H_2O\\&\quad\longrightarrow 3CaSO_4 \bullet 2H_2O + KH_2PO_4 + H_3PO_4\end{aligned}$$

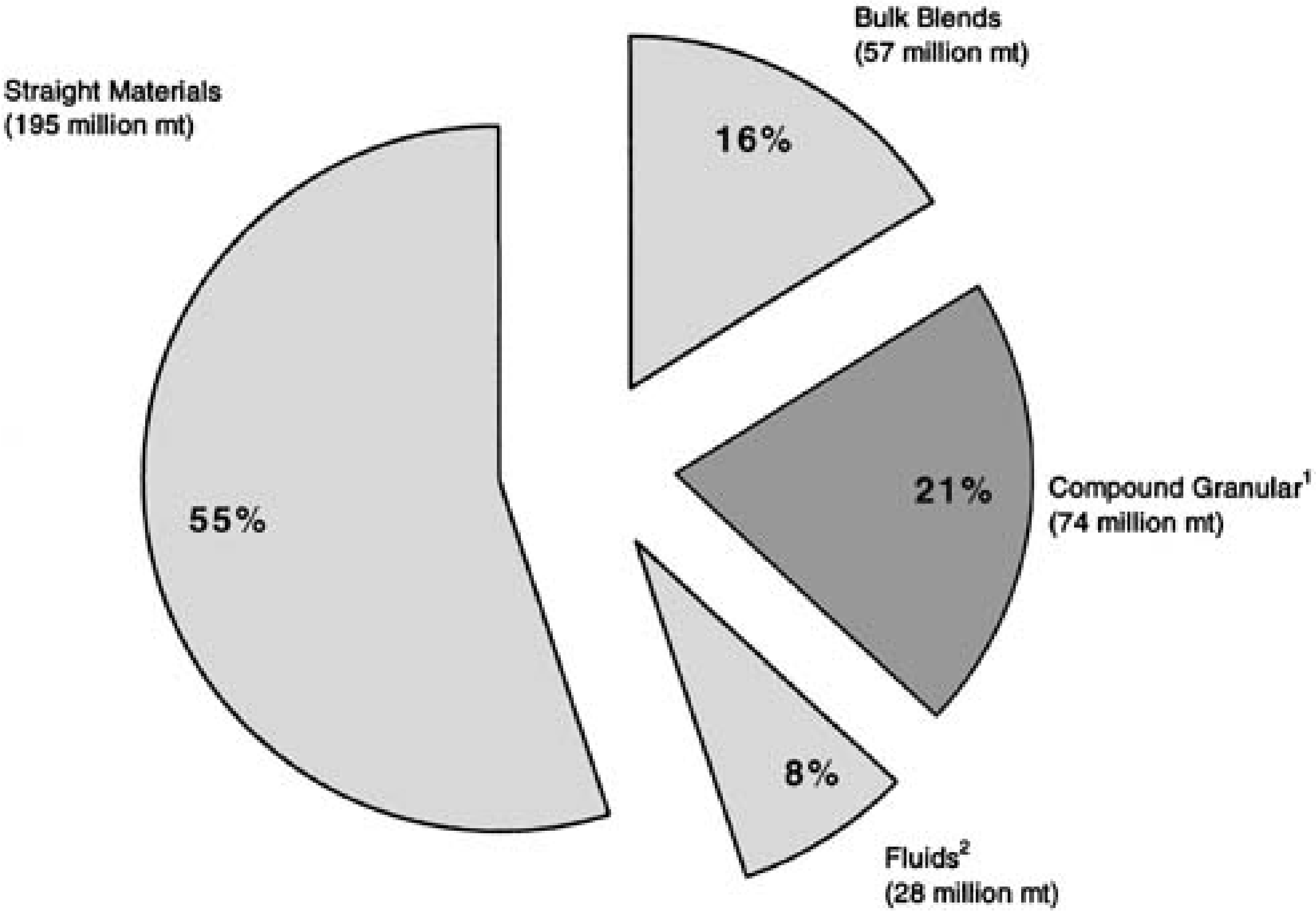
After removal of the gypsum by filtration, the mixture of phosphoric acid and monopotassium phosphate can be ammoniated to produce a chloride-free NPK fertilizer such as 84816, or monopotassium phosphate can be separated from the phosphoric acid by precipitation in methanol solution and marketed separately. If desired, the monopotassium phosphate can be converted to a pyro- or polyphosphate by pyrolysis.

The obvious advantage of the process is the recovery of HCl in a useful form. Another advantage is that a standard wet-process phosphoric acid plant can be adapted with a minor change. Numerous other processes for making potassium phosphate from KCl have been proposed, patented, or developed on a commercial scale, but none is known to be in commercial use for fertilizer purposes.

11.7 Compound Fertilizers

The preceding sections of this chapter have described the production of nitrogen fertilizers, phosphate fertilizers, and potash fertilizers from basic raw materials. These fertilizers, without further processing, are referred to as "straight materials." Worldwide, in 1999/2000 (Fig. 11.12), straight materials accounted for about 195 million mtabout 55 percent of the total production. The remaining 45 percent was in the form of "mixed" fertilizers, that is, fertilizers that contain more than one of the major plant nutrients nitrogen, phosphorus, and potassium. Application of nitrogen fertilizer as straight material is much more prevalent than phosphate fertilizer. The reasons for these patterns are both agronomic and economic. Agronomically, large applications of nitrogen alonefor example, in the rice production and "side dressing" of cornoften are of considerable importance. For phosphate, there usually is no similar agronomic need for single-nutrient application; so application in mixtures, with the resultant elimination of multiple handling, is preferred. In the United States and Europe, with highly mechanized farming practices, the present trend is toward "precision" agriculture, which involves variable rates of application of nutrients at different locations in the same field based on the nutrient status. This method of application requires straight fertilizers rather than compounds.24

Fig. 11.12 Estimated world production of fertilizers by product type. 1. Steam/water, chemical (including ammonium phosphate), and compacted granular products. 2. Single- and multiple-nutrient fluids, including anhydrous ammonia. 3. Nutrient production (1999/2000 reporting year) amounted to 141.4 million mt (N + P2O5 + K2O).



Mixed fertilizers can be divided, by physical characteristics and production methods, into the following categories, all of which will be discussed in some detail:

1. nongranular mixtures
2. compound granulars
3. bulk blends
4. fluid mixtures

Nongranular Mixtures

Until about 1950, essentially all the mixed fertilizers available to farmers were of the nongranular type. Production was by simple batch weighing and blending of available nitrogen, phosphorus, and potassium ingredients, usually in locations that were convenient with respect to rail or water receipt of raw materials and out-shipment of bagged mixtures to dealers in farm areas. The source of phosphate almost invariably was nongranular SSP produced at the mixing location from in-shipped phosphate rock and locally produced sulfuric acid. In-shipment of elemental sulfur was required to produce the acid, usually by the now obsolescent "chamber" process. The major nitrogen ingredient used in the mixtures was fine-crystal by-product ammonium sulfate, and the potash source was nongranular potassium chloride. In some plants the superphosphate was "ammoniated" by reaction with ammonia solution, either in a separate operation before mixing with the other ingredients or during the mixing operation. Such ammoniation was beneficial in reducing the corrosiveness (acidity) of the superphosphate and in improving its physical condition. The ammonia also provided nitrogen at relatively low cost, but the amount that could be absorbed by the superphosphate was limited. Organic materials often were included in the mixtures as nutrient sources, bulking agents, or physical conditioners. Materials used included fish meal, packinghouse wastes, dried blood, tobacco stems, cottonseed hulls, and others. Mineral materials such as kaolin clay, diatomaceous earth, and vermiculite sometimes were included as conditioners. Batch-type drum mixers of 1- to 3-ton capacity were used with the ammonia solution piped directly into the mixer. Mixture grades were low by present-day standards, and storage and handling properties of these nongranular mixtures were generally poor. Caking problems, in particular, were considerable. Production of this type of fertilizer now persists in only a very few locations.

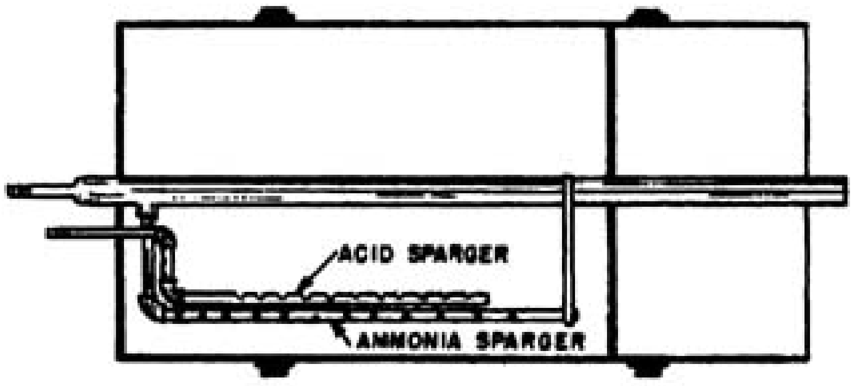
Compound Granulars

Continuous AmmoniatorGranulator Process

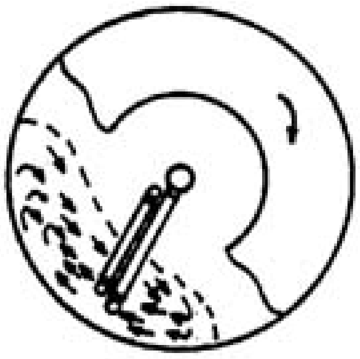
Following the close of World War II in 1945, farmer demands developed for much increased quantities of fertilizer and for products with better handling properties. Granulation, which provides an increase in particle size and the elimination of fines, was a promising avenue for the physical improvement that was pursued by researchers in both the United States and Europe. A milestone in this movement to granulation was the development, patenting, and demonstration by TVA, in 1954, of a continuous rotary-drum granulation process for mixed fertilizers.25,26

The rotary drum granulator in this process is essentially the same as was pictured for the diammonium phosphate granulation process. In early versions of the process, feed to the drum SSP consisted of (1) crushed oversize product and fines, (2) nongranular SSP, (3) nongranular potassium chloride, (4) ammonium sulfate crystals, (5) sulfuric acid, and (6) anhydrous ammonia or an ammonia-containing nitrogen solution. The acid and the ammonia or ammoniating solution were sparged under the rolling bed of solids in the drum with a sparger arrangement, as shown in Fig. 11.13. With the proper balance of liquid and solid feeds, granulation occurs as a result of the rolling action in the action in the drum, and it is completed, with hardening, in a rotary-drum dryer.

Fig. 11.13 Sparger location under rolling bed in TVA-type ammoniator-granulator. *(Courtesy TVA)*



SIDE VIEW



FEED END VIEW

This process was rapidly and widely adopted by mixed-fertilizer producers throughout the United States and later adopted by producers in Europe, Latin America, and Asia.

The process was particularly attractive because: (1) the raw materials were essentially the same as materials that were already available at most batch-type nongranular plants, (2) higher proportions of relatively cheap anhydrous ammonia or nitrogen solution could be used, (3) continuous-type processing allowed higher production rates, and (4) product physical properties were much improved as a result of granulation. Today, the same basic process, but with numerous modifications, is used in most mixed-fertilizer granulation plants. However, the number of plants has dwindled in favor of bulk-blending and fluid fertilizers. One early modification to the process was the addition of phosphoric acid as a feed material, either along with or in place of sulfuric acid. Other feed materials used in some variations include TSP, monoammonium phosphate, ammonium nitrate, and urea. The use of urea is becoming more attractive in compound fertilizer production in Asia because higher grades can be achieved. The molten urea or concentrated urea solution is sprayed onto the bed of materials in the granulator and the wet product requires gentle drying. Most urea-based compound fertilizers are more hygroscopic, requiring dehumidified bulk storage facilities before the product is bagged.27

Use of Preneutralizer (Slurry Granulation)

A major improvement to the basic TVA granulation process was the incorporation in most plants of a preneutralizer vessel for prereaction of phosphoric acid with ammonia. The procedure, as applied to the production of DAP, was described earlier and was pictured in Fig. 11.10. As in the DAP process, the preneutralizer used in the production of mixed fertilizers is operated at an NH_3:H_3PO_4 mole ratio of about 1.4 to give maximum ammonium phosphate solubility and thus maximum fluidity of the slurry pumped to the granulator. Subsequent further addition of ammonia in the granulator sharply decreases solubility and thus promotes solidification and hardening of granules. As in ammonium phosphate production, the iron and aluminum impurities normally present in wet-process phosphoric acid contribute significantly to the hardening of granules during ammoniation.17 Sulfuric acid, when used, is fed directly to the granulation drum, not to the preneutralizer. The use of the preneutralizer allows the inclusion of higher proportions of acid in formulations. Also, better use is made of the heat of the ammoniaacid reaction for the evaporation of water, with resultant savings in dryer fuel.

Use of a Pipe Reactor (Melt Granulation)

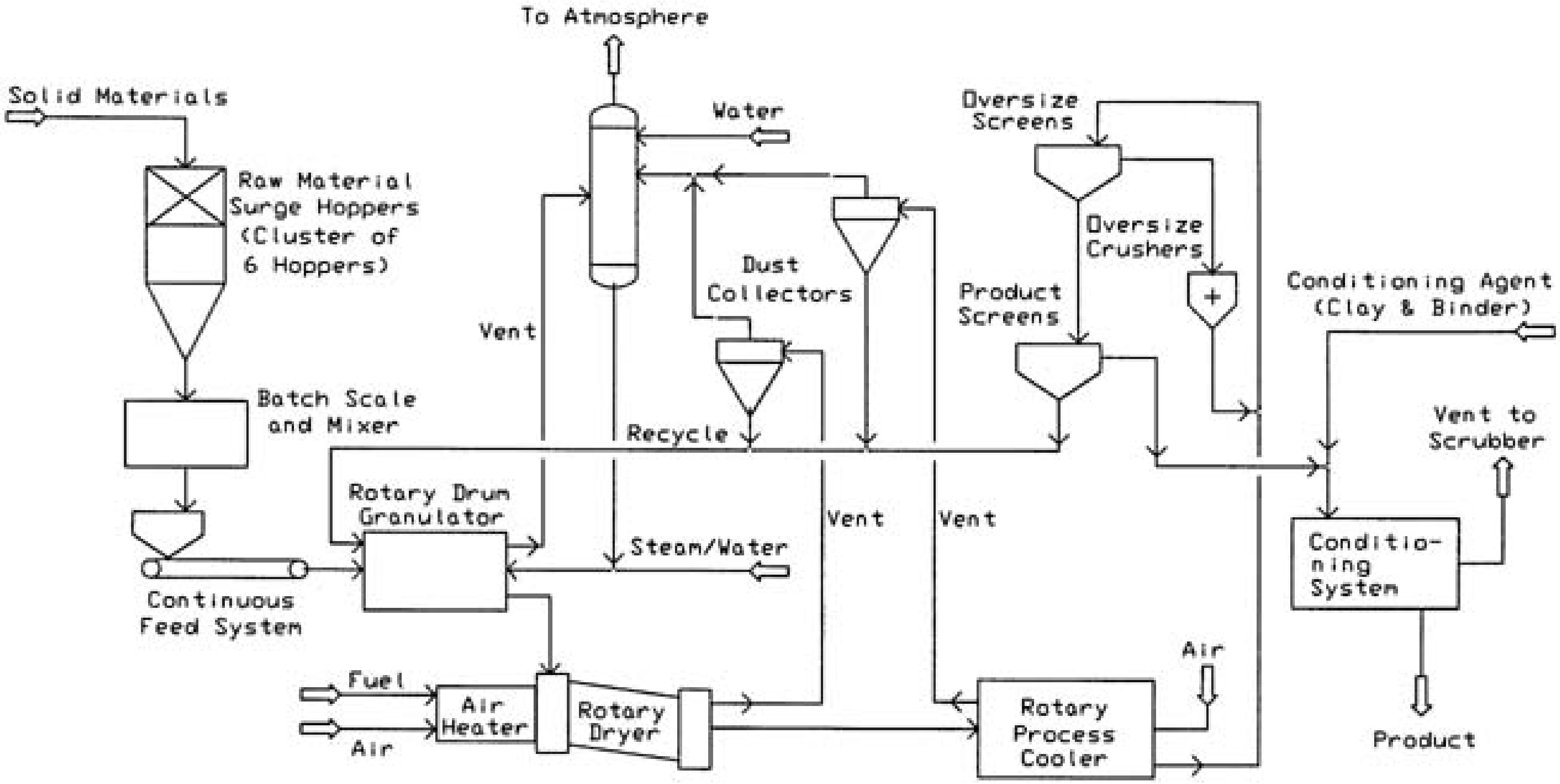
A more recent development that has been adopted in most U.S. granulation plants involves the substitution of a pipe reactor or a pipe-cross reactor for the preneutralizer vessel as the locale for the ammoniaacid reaction.2832 A simple pipe reactor differs from the pipe-cross reactor in that only one acid inlet port, usually for phosphoric acid, is provided.

The salient feature of the pipe reactor and pipe-cross reactor systems is that the heat of the ammonia-acid reaction is confined to the pipe and is efficiently utilized there to vaporize essentially all the moisture in the feed acid. The resultant steam exits the pipe discharge end in the granulator and is swept away in an air stream. The ammonium phosphate melt, which is of very low moisture content, also exits the pipe and is discharged onto the rolling bed of dry solids in the granulator. This melt serves as the binder for granule formation. The reaction temperature in the pipe reactor usually is held below 150°C to avoid a buildup of troublesome scale in the pipe. In view of this temperature limitation, there is little or no conversion of the phosphorus to polyphosphate forms. By allowing the pipe temperature to rise to about 200°C or higher, polyphosphate can be formed without scaling, but this mode of operation seldom is used in mixed-fertilizer granulation. TVA has used such high pipe temperatures to produce 11570 and 28280 grade ammonium polyphosphate granules in which 1525 percent of the P_2O_5 is in polyphosphate form.28 The physical properties of these products are exceptionally good, and the polyphosphate content makes them particularly suitable for dissolution to form fluid fertilizers. Because of the very low moisture content of the pipe reactor melts, even at the lower operating temperatures, the normally expensive drying of finished granular product usually is not needed: passage of the product through a cooler with an air sweep is sufficient for final drying. This results in a major fuel saving, which is a very significant economic advantage.

Steam Granulation

A simple method for granulating dry mixtures of fertilizer materials was developed in Europe in the early 1950s and still is employed there to some extent. In this procedure, which is pictured in Fig. 11.14, the feed materials are finely ground to ensure their incorporation in the granules. A dry mixture is made and then is moistened in a rotary drum or pan-type granulator with either water spray or steam, or both. This added moisture forms a salt solution by partial dissolution of the mixture ingredients. The granules form because of the plasticity and rolling action in the granulator. Initially, these granules are very soft; but, with further rolling and drying in a subsequent drying operation, salt bridging imparts the required granule hardness. In contrast to the TVA ammoniationgranulation process, the steam granulation process usually causes no significant reaction between mixture ingredients. In Europe, this feature is of considerable significance when superphosphate is used as a feed material. Ammoniation of superphosphate, which occurs in the TVA ammoniationgranulation process, converts part of the water-soluble P_2O_5 content of the superphosphate to water-insoluble but citrate-soluble forms that are not accepted as marketable phosphate fertilizer in some European countries. The citrate-soluble forms are accepted as effective, marketable phosphate forms in the United States. The European Union is currently debating whether to allow the citrate-soluble form as a part of saleable phosphates.

Fig. 11.14 Typical NPK fertilizer granulation plant using steam and/or water. *(Source: Fertilizer Manual.)*



Compaction Granulation

This process involves the use of mechanical force to form granules (dense particles) from finely divided solid materials.7 In this process (Fig. 11.15) the fine materials are weighed and, depending on the desired grade, thoroughly mixed. The mixture is then continuously fed to a pressure-roll compaction machine that forms the powdered mixture into a dense, hard sheet of material usually about 23 cm thick and 60100 cm wide. The compaction machine consists of two horizontally opposed rollers that turn inward toward each other. The shaft of one roller is fixed while the other is movable. The pressure from a hydraulic system is applied to the movable shaft. The pressure is dependent on the fertilizer materials. The compacted sheet of material is crushed in a controlled manner to form smaller particles. The crusher discharge is screened to obtain the desired product-size fraction. The oversize material is crushed and recycled to the screens and the undersize material is returned to the compaction machine after thorough mixing with fresh feed. This is a dry process and does not involve chemical reaction. Some typical pressing force for fertilizer materials is shown in Table 11.3.

TABLE 11.3 Typical Pressing Forces for Fertilizer Compactiona

Feed Material	Pressing Forcea (kN/cm)b	Water Content (%)	Feed Particle Size (mm)
Ammonium sulfate	100120	0.51.0	<1.0
Potassium chloride			
Feed temperature >120°C	4550	nil	1.0, with max.
Feed temperature 20°C	70	nil	of 3% <0.06c
Potassium sulfate			
Feed temperature >70°C	70	1.0	<0.5
Potassium nitrate	100	0.51.0	<1.0
Calcium nitrate	60	nil	<1.0
Urea	3040	nil	13
Compound fertilizer containing			
· No raw phosphate rock	3080	<1.0	<1.0
· Raw phosphate rock	>80	<1.0	<1.0
· Urea	3040	<1.0	<1.0

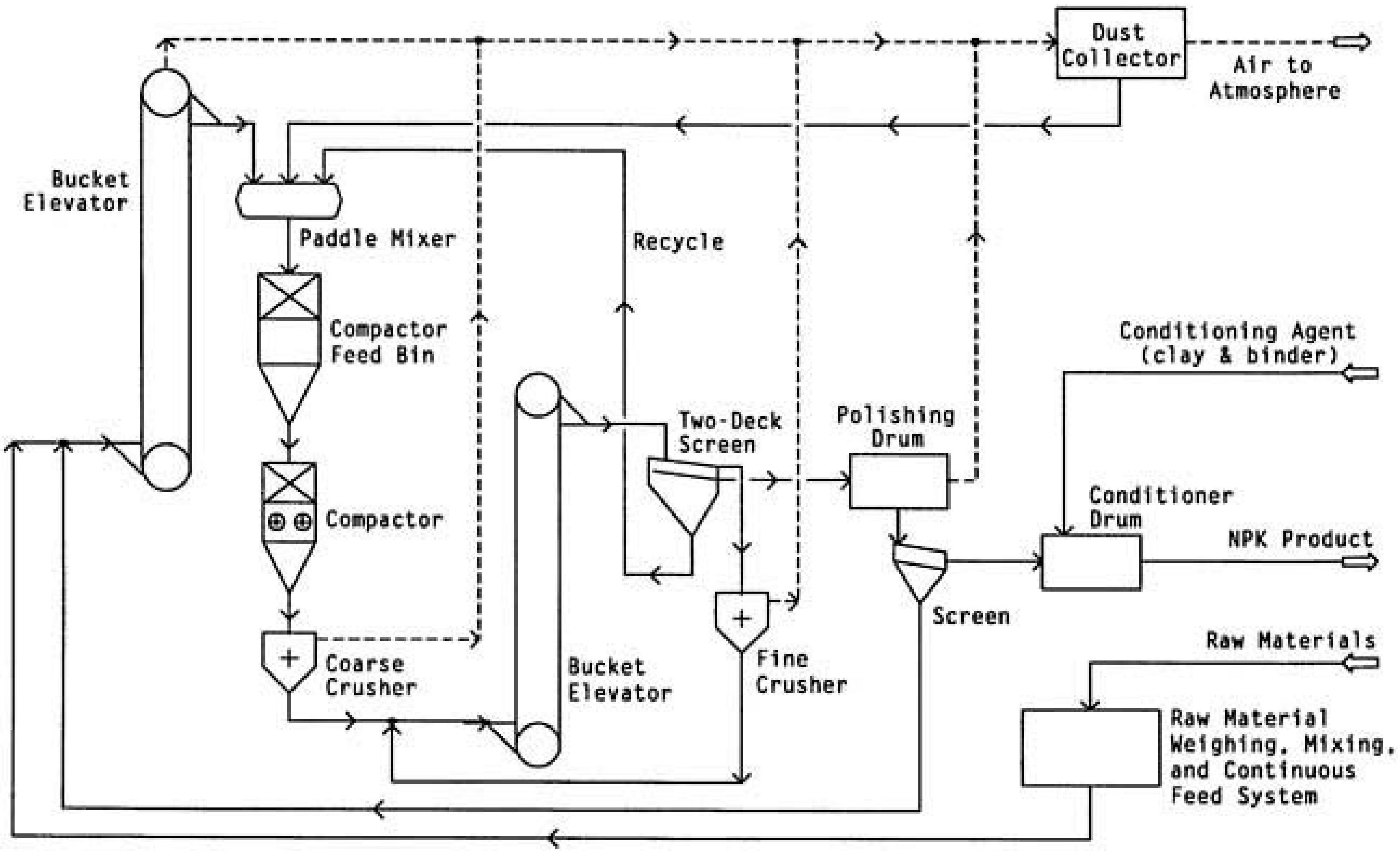
Source: *Fertilizer Manual*, United Nations Industrial Development Organization and International Fertilizer Development Center, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

aIndicated pressing force is for machine having 1.0-m diameter rollers.

b1 kN/cm = approximately 0.1 tonne/cm.

cSize criteria apply to material feed temperature as indicated.

Fig. 11.15 Typical NPK Fertilizer Compaction Granulation Plant. *(Source: Fertilizer Manual.)*



Compacted fertilizers have a more irregular shape as compared with the more smooth and round granules produced by other granulation processes. This irregular shape may cause some resistance among users who are accustomed to the more traditional spherical fertilizer granules. However, compacted fertilizers are produced and used in Europe.

Bulk Blends

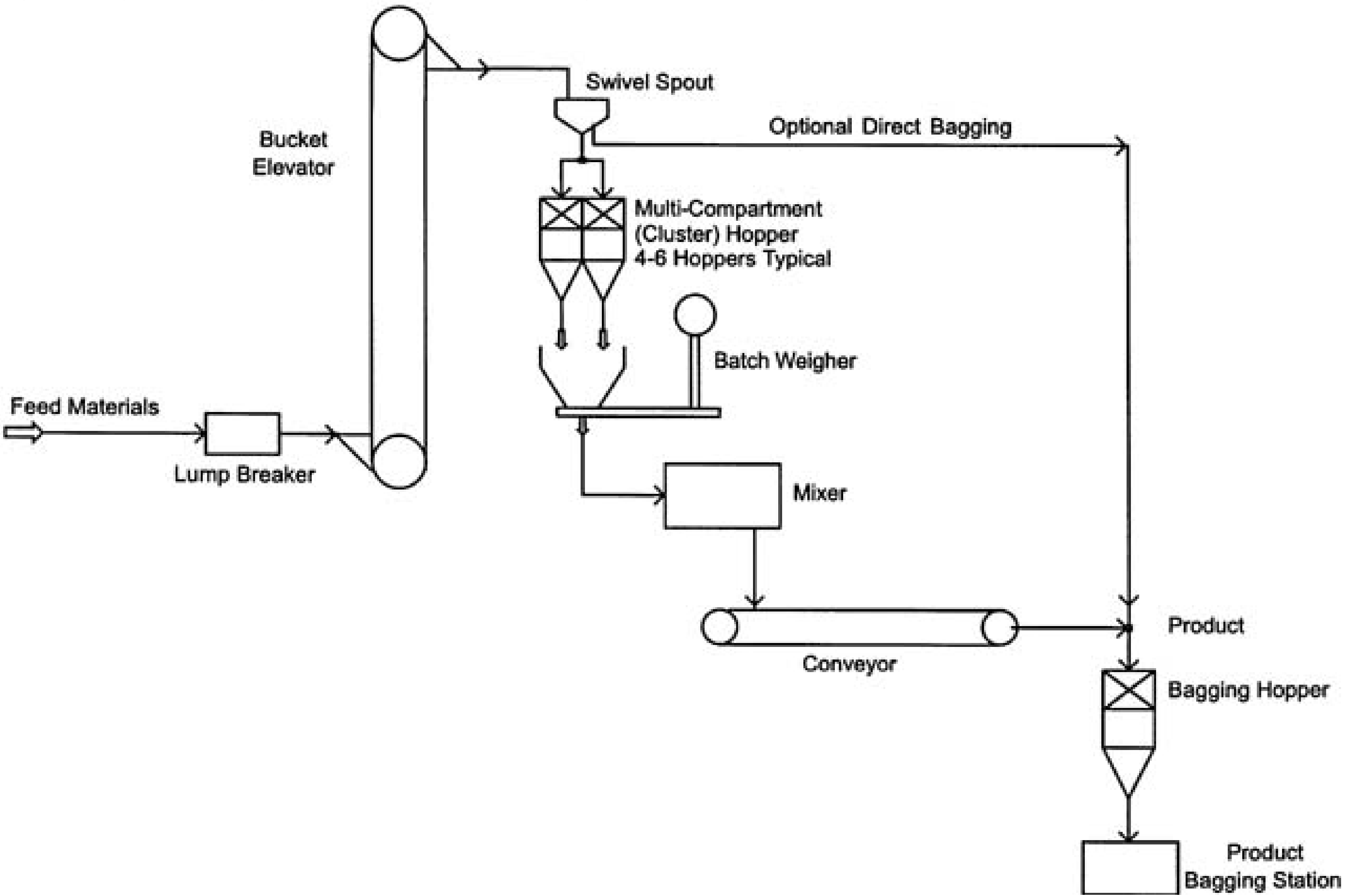
History and Growth

In the "compound" (mixed) fertilizer granulation processes just described, the starting materials are nitrogen, phosphorus, and potassium in the form of (1) nongranular solids, (2) liquids, or (3) gas (ammonia). When nitrogen, phosphorus, and potassium materials are all available from basic producers in granular form, as they now are, granular mixed fertilizers of almost any nutrient proportion (grade) can be made by simple proportioning and dry blending of granular nitrogen, phosphorus, and potassium materials; further granulation or other processing is not required. This is the basis of the presently popular "bulk-blending" system of mixed fertilizer preparation and distribution.

In 1999/2000 about 57 million mt of bulk-blended fertilizers was produced. This corresponded to about 16 percent of the global fertilizer production.

The simplicity of the bulk-blending procedure is illustrated in Fig. 11.16. The operations consist of (1) in-shipment of granular fertilizer materials, (2) temporary storage, (3) proportioning, (4) blending, and (5) movement of the finished blend to the field for spreading.

Fig. 11.16 Typical batch-type NPK fertilizer bulk-blending plant. *(Source: Fertilizer Manual.)*



Key factors in the rapid growth of bulk blending have been: (1) convenience in providing the farmer with a wide choice of plant-food ratios or grades on a "prescription" basis in relatively simple, small mixing plants located in the immediate farm area; (2) economy due to handling in bulk rather than bags; and (3) provision of spreading and other farmers services by the local blender.

Typical blending plants produce and market only 20006000 tons of blend per year. However, the present trend is for larger capacity plants covering a 30-mile radius of farming area designed to serve relatively large areas. The capital investment in blending plants is comparatively low. Blending plants operate on a quite seasonal basis, and often the operators have supplemental business enterprises.

Intermediates Used in Bulk Blending

The bulk-blending system came into being only after suitable nitrogen, phosphorus, and potassium intermediates became available in granular form. In particular, the introduction of granular DAP of 18460 grade in about 1960, as described in a previous section, was a major catalyst to bulk blending. Characteristics of that product that make it an ideal material for shipping and subsequent blending are: (1) a high plant food content of 64 percent, (2) a high phosphorus content of 46 percent, (3) excellent storage and handling properties, (4) compatibility with essentially all other fertilizer materials, and (5) economy of production near phosphate deposits. Most NPK bulk blends are prepared with DAP as the phosphate source. Granular TSP is used for no-nitrogen (PK) grades, but for NPK blends TSP has the disadvantage of being incompatible (reactive) with urea.

In the preparation of bulk blends, it is of great importance to ensure that all ingredients of the blend are reasonably matched in particle size.33 The use of unmatched materials results in poor mixing, segregation during handling, and uneven distribution on the field. Nitrogen materials used with the greatest success in bulk blends are (1) granular urea, (2) granular ammonium nitrate, and (3) large-crystal or granulated ammonium sulfate. Prilled ammonium nitrate can also be made of large-enough particle size for good blending, but prilled urea normally is of smaller size than desirable. The chief potash material used in blends is granular-grade potassium chloride, which is sized specially by potash producers for use in blends. Although granular-size potash usually consists of irregular crusher fragments rather than well-rounded granules, studies have shown that this particle shape difference does not promote segregation or otherwise impair blend quality.34 A somewhat smaller-size, coarse-grade potassium chloride sometimes is used in blends because it is cheaper, but segregation problems are invited. Potassium sulfate and potassium magnesium sulfate are made in granular form for the blending of specialized formulations.

The addition of micronutrients to bulk blends poses special problems because of the small quantities usually added. Although many micronutrients are available in granular form, intergranule spacing on the field can be excessive when the granular micronutrient constitutes only a very small proportion of the blend. Procedures have been devised and used for adding finely powdered micronutrients to blends together with a small amount of liquid binder.35 The result is that each blend granule is coated with micronutrient, and thus the field distribution is better.

Operating experience and TVA studies have identified several combinations of materials that are incompatible in blends and should be avoided: (1) unammoniated superphosphates react with urea, and the result is wetting and caking and (2) urea in contact with ammonium nitrate forms an extremely soluble hygroscopic mixture that liquefies in the presence of only minute amounts of moisture. Therefore, these two materials can never be used together in the same blend. Essentially all other combinations of available fertilizer materials are compatible in blends.

Equipment and Operating Technique

There are different arrangements of equipment in the blending plants throughout the world. The operations in such plants generally consist of: (1) receipt of the intermediate granular components in bulk by truck or railroad car, (2) storage of the intermediates separately in bins, (3) removal of the granular blend components from storage and weighing in desired proportions, (4) mixing to obtain uniformity, and (5) discharge of the mixed batch directly into a spreader truck or into a holding bin. The most popular type of mixer is the batch-type rotary drum.

Such mixing plants can be highly automated so that the output with a mixer of only 1-ton capacity can be as much as 1520 tons/hr. Mixers of several types are available with capacities ranging between 5 to 8 tons per batch; in some cases batch mixer capacities can be 10 tons per batch.36

Problems in Bulk Blending

Bulk blending is a simple practice that can provide uniform mixtures in almost any desired proportions. However, unless proper materials are used and certain handling precautions are taken, segregation can occur, and nonuniform blends will result. The factors involved and the precautions that are necessary have been described in several publications.3 The main requirements are the use of granular materials with a well-matched range of particle size and the handling of the mixture after blending in ways that will minimize segregation. The mixing step is seldom a problem; it is quite easy to obtain a uniform blend by the use of proper materials and a few minutes of mixing. However, maintaining this uniformity during handling on the way to the farm and onto the soil requires precautions. Handling procedures that may cause segregation include coning (as occurs if the blend is allowed to drop from a discharge pipe onto a pile), vibration in hauling vehicles, and ballistic action imparted by some types of spreaders.

The problem of adding small quantities of micronutrient materials to bulk blends was mentioned earlier. This problem is being met by either using micronutrients in granular form or coating the blend particles with finely ground micronutrients and a liquid binder.35

Fluid Mixtures

Mention was made earlier of the significant quantity of nitrogen applied to farmlands in the form of ureaammonium nitrate solution and aqua ammonia. Since the late 1950s, there has also been a growing practice of applying mixed fertilizers (NP and NPK) in fluid form. In fact, the growth of the fluids market paralleled bulk-blended products. In 1999/2000 about 28 million mt of fluids was produced and corresponded to 8 percent of the world production. North America, and Europe are major markets for fluids. Fluids are also used in high-value crops in some countries of the Middle East and Latin America.

Numerous advantages have been cited to explain the growth of fluid mixed fertilizers.3 Of these, the most significant probably is the ease and the precision with which fluids can be applied to the soil. When placement of the fertilizer is important, as in row fertilization, fluids often have an advantage. Also important is the adaptability of fluids to the homogeneous incorporation of micronutrients, herbicides, and insecticides. Another factor that encouraged the adoption of fluid mixtures was the prior existence of equipment and facilities for handling and applying nitrogen fluids; it was only logical for this practice to be expanded to include the preparation and application of mixtures. There are some claims of agronomic superiority for fluid fertilizers over similarly applied solids, but such claims are debatable and are not widely accepted.

Solution Fertilizers

Fluid fertilizers in use today can be divided into two major categories: (1) solutions, in which the plant food content is entirely in solution, and (2) suspensions, in which the plant food content exceeds its solubility, and the excess is held in suspension, usually by the addition of a suspending agent. The solution-type mixtures were the first to achieve commercial importance. Beginning about 1950, the practice of making solution mixtures was begun with pure electric-furnace orthophosphoric acid (54% P_2O_5) as the source of phosphorus. Ammoniation of this acid yielded an 8240 grade ammonium orthophosphate solution that could be shipped to mixing plants in farm areas and used as a base for blending with other fertilizer materials. Blending usually was with ureaammonium nitrate solutions (2832% N) and a pure grade (solution grade) potassium chloride.

With the system that employed these raw materials, numerous advantages of fluid mixtures were established, but a need to improve the economics of the procedure was obvious. Pure electric-furnace acid is much more expensive than wet-process acid, so ways were sought to use the cheaper but impure wet-process acids. Direct ammoniation of wet-process acid causes precipitation of numerous impurities as an intolerable, gelatinous sludge. Also, plant-food solubilities are undesirably limited in the orthophosphate system, so the concentration (grade) of the product made with electric-furnace orthophosphoric acid was limited to 8240. A series of developments by TVA led the way to solving both of these problems.

The problem of limited plant-food solubility was much alleviated by TVA's introduc-tion of superphosphoric acid in 1957.37 Superphosphoric acids are acids of high P_2O_5 content (up to 80% P_2O_5) that contain significant proportions of polyphosphates (nonortho, polymerized phosphates). The superphosphoric acid first produced by TVA was a pure electric-furnace product of 76 percent P_2O_5 content, of which about 50 percent was in polyphosphate form. With ammoniation of this acid, it was possible to produce an 11370 grade base solution instead of the 8240 grade maximum made from ortho acid. Demonstration use of 11370 base in a number of mixing plants showed that solution mixtures of relatively high concentration could be made. Also, it was found that the polyphosphate content enhanced the solubility of micronutrient additives by a sequestering action.

The unfavorable economics of the electric-furnace acid process, however, discouraged the commercial production of electric-furnace superphosphoric acid; so attention was directed toward concentration of the cheaper wet-process phosphoric acid to give wet-process superphosphoric acid. Satisfactory concentration methods were developed,38 but the effects of acid impurities on the handling properties of the concentrated acids limited the final P_2O_5 concentration and hence the polyphosphate content. The best handling properties were obtained with acids made from calcined rocks and with P_2O_5 concentration limited to 6870 percent P_2O_5. In acid of this concentration, from 20 to 35 percent of the P_2O_5 is in the polyphosphate form, and the acid is referred to as low-conversion superphosphoric acid. Such acid today is the chief source of P_2O_5 for solution-type fertilizers.

The most popular scheme by which solution-type fertilizers are made from low-conversion, wet-process superphosphoric acid is outlined in Fig. 11.17. The first step is ammoniation of the acid to produce ammonium polyphosphate base solution. By carrying out this ammoniation in a pipe reactor, a temperature of over 200°C is developed, which promotes an increase in polyphosphate content to 6570 percent of total P_2O_5 in the ammoniated acid.39,40 The grade of the resultant base is usually 10340; an 11370 grade can be made, but its storage properties are not as good, especially in cold weather. A sketch of a typical pipe-reactor system for the production of high-polyphosphate base solution from low-conversion, wet-process superphosphoric acid is shown in Fig. 11.18.

Fig. 11.17 Production methods for solution-type and suspension-type mixed fertilizers.

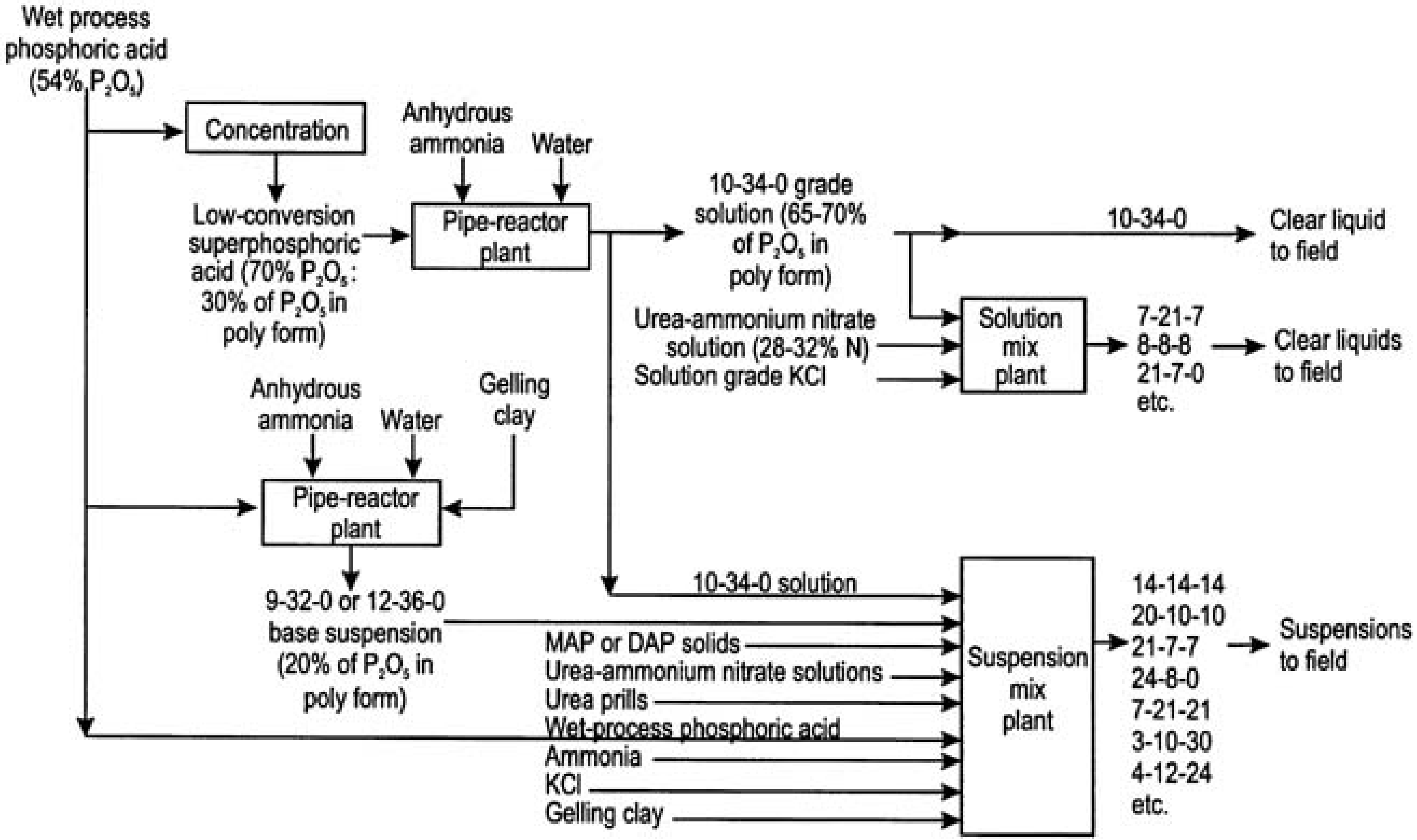
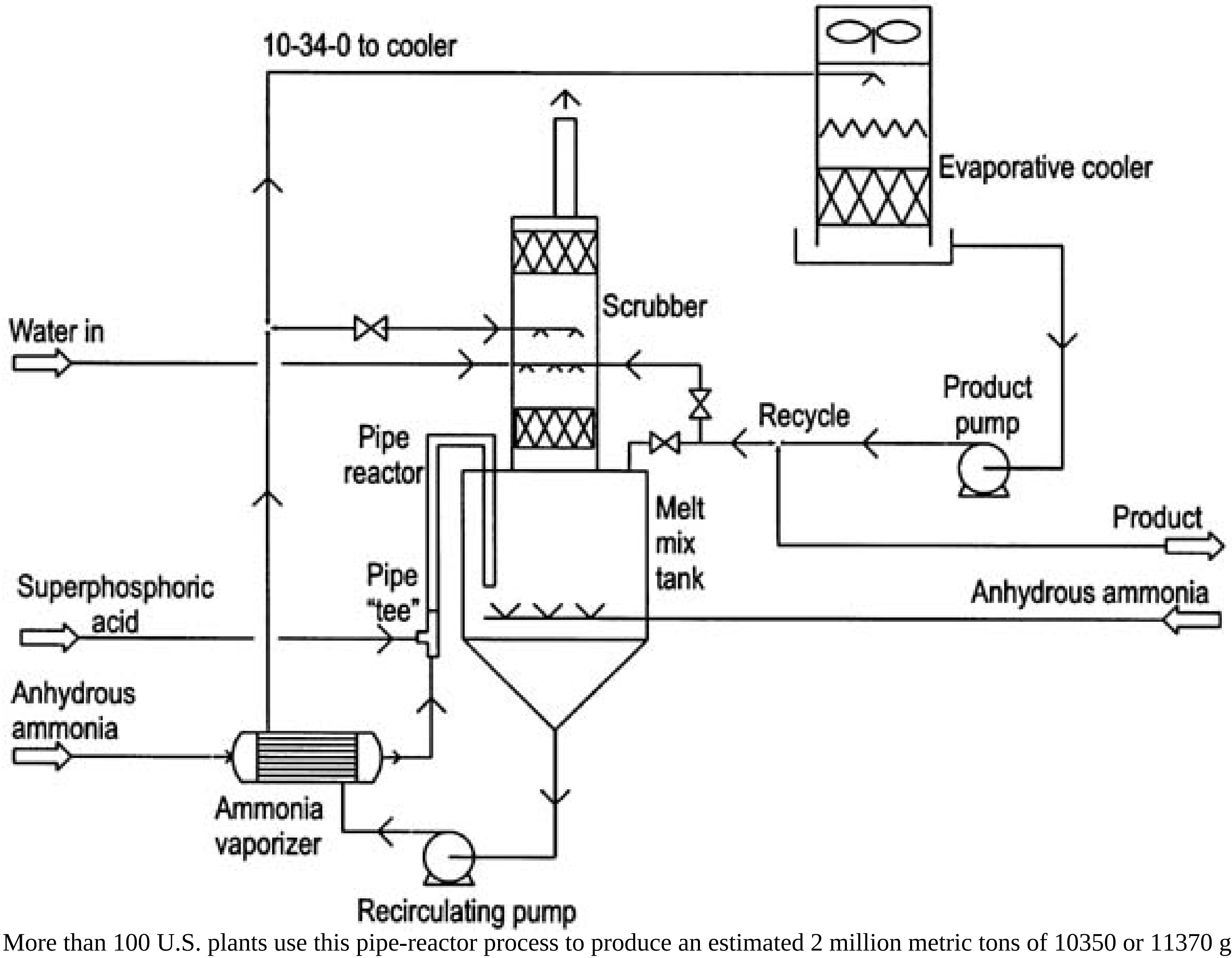
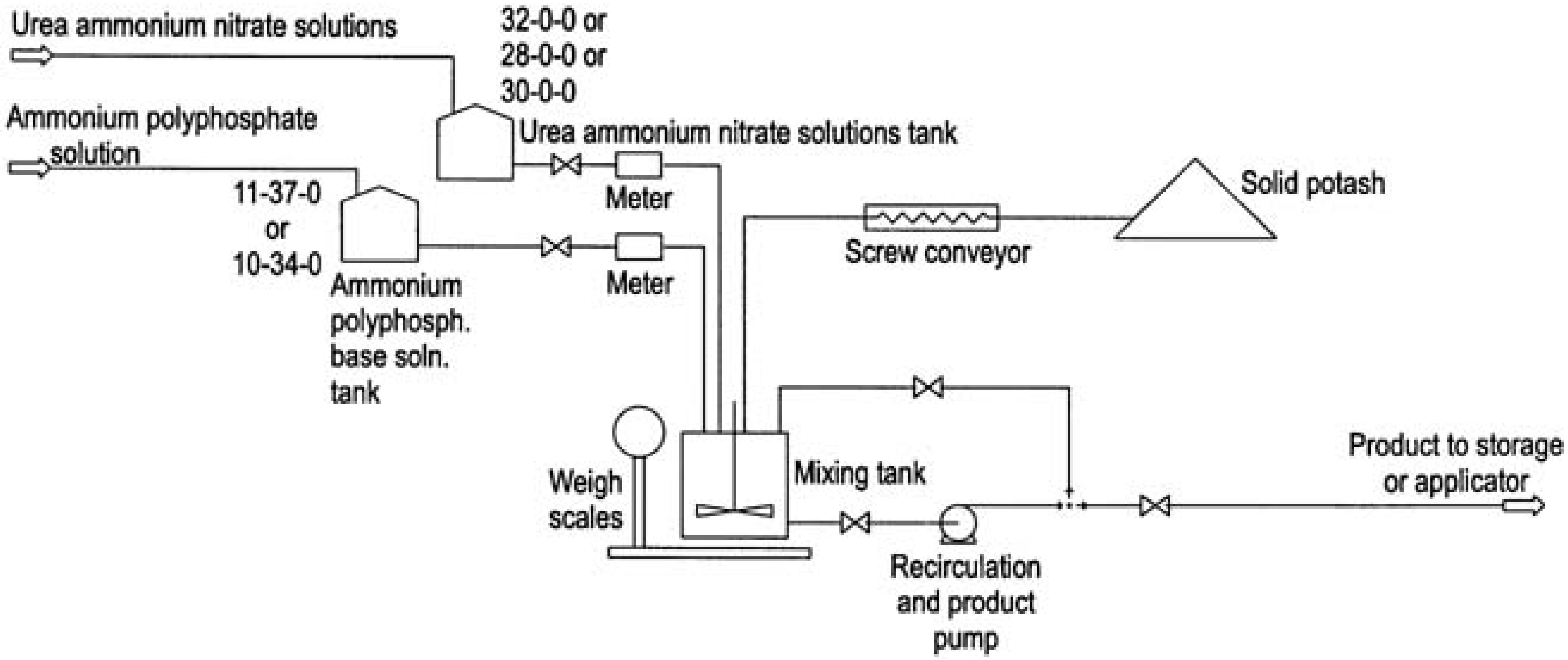


Fig. 11.18 Typical pipe-reactor plant for production of ammonium polyphosphate solution. *(Source: Fertilizer Manual.)*



More than 100 U.S. plants use this pipe-reactor process to produce an estimated 2 million metric tons of 10350 or 11370 grade product per year. Other countries such as Belgium, France, and the Former Soviet Union (FSU) also use this process to produce ammonium polyphosphate solution. A considerable amount of the solution is used for direct application in the U.S. wheat belt and other areas in which potassium is not deficient. The rest is used in small mix plants to produce mixtures. The ammonium polyphosphate base solution is mixed with nitrogen solutions containing 2832 percent N (ureaammonium nitrate solutions) and sometimes also with potash to produce such clear liquid grades as 2170, 7217, and 888. As with bulk blending, the solution-mixing operation usually is carried out in small mixing plants located in farm areas (Fig. 11.19).

Fig. 11.19 Liquid fertilizer mix plant. *(Source: Fertilizer Manual.)*



Suspension Fertilizers

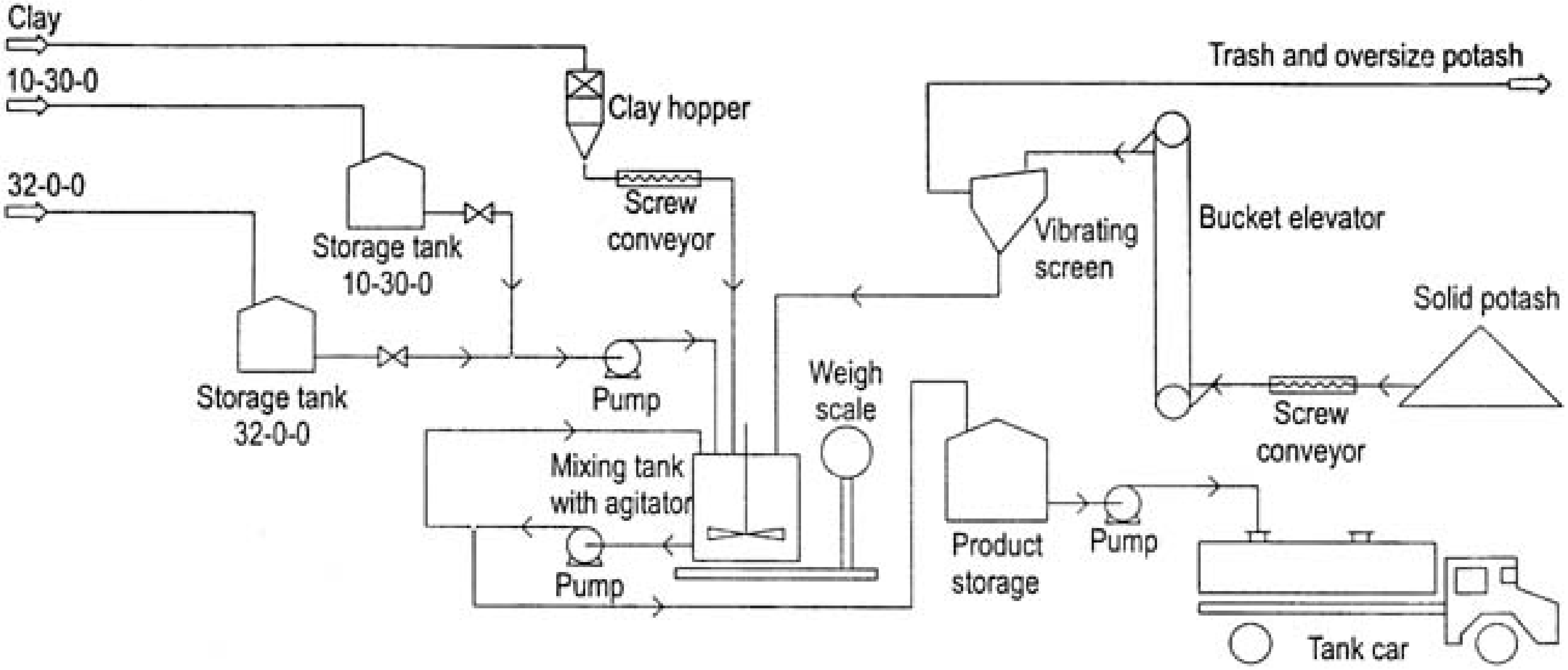
Suspension fertilizers were developed chiefly as a means of breaking the grade barriers imposed on solution fertilizers by solubility limitations. The low solubility of potassium chloride was particularly troublesome, in that high-potash solution grades were not possible.

The feasibility of preparing and handling suspension-type fertilizers was demonstrated by TVA and others in 1959.41,42 Two essential requirements for the preparation of a successful suspension were shown to be: (1) small particle size of the suspended solids and (2) the addition of a gelling-type clay as a suspending agent along with sufficient high-shear agitation to properly disperse the clay. The most satisfactory gelling-type clay is attapulgite.43 Attapulgite is a hydrated magnesium aluminum silicate, which is composed of needle-shaped crystals so small that they can be seen only with the use of high magnification. In the dry state, these crystals bond together in bundles as larger particles. These bundles must be disrupted and the individual crystals freed before they can assume the structural arrangement that is characteristic of a gel. When this disruption of the crystal bundles is carried out in the presence of an electrolyte, such as any of the common fertilizer salts present in fluid fertilizers, the individual crystals immediately assume the regular structural orientation of a weak gel, and thus provide the desired suspending action.

Wetting of the clay by the fluid fertilizer causes some of the required disruption of crystal bundles and freeing of individual crystals but not enough to provide good gelling without high-shear agitation. The addition of 23 percent of attapulgite is sufficient for most fertilizer suspensions. Instead of providing high-shear agitation for the entire batch of suspension, some operators carry out a pre-gelling operation in which high-shear agitation is applied only to the clay plus a limited amount of water or fertilizer solution. The pre-gelled mixture then is added to the major body of fertilizer with only mild mixing. Pre-gelled clay is also marked as liquid clay and is transported in tank cars. The provision of a high-shear mixer is essential not only for gelling of the suspending clay but also for disintegration of any granular products such as DAP or MAP that might be used in the suspension. A high-shear recirculation pump is also an aid to gelling of the clay.

Figure 11.20 shows the layout of a typical suspension mix plant in which all the major suspension ingredients are received as solids. This is presently a popular mode of operation. Raw material costs for such a plant often are less than for bulk blending because nongranular materials can be used. Piping for the optional addition of phosphoric acid and ammonia is shown. Such addition develops heat, which hastens the disintegration of some solids.

Fig. 11.20 Typical mix plant for production of suspension mixtures. *(Source: Fertilizer Manual.)*



The storage and handling properties of suspensions can be enhanced by the inclusion of some polyphosphate in the suspension. Low-temperature storage properties, in particular, are enhanced. One popular method of providing polyphosphate is to supply all or part of the P2O5 as 10340 ammonium polyphosphate base solution, such as is used in solution fertilizers. A second method is the use of a pipe reactor to prepare 9320 or 12360 grade ammonium polyphosphate base suspension from relatively inexpensive wet-process orthophosphoric acid (54% P2O5), ammonia, and gelling clay.44 About 20 percent of the P2O5 in the 9320 form or 12360 product is in polyphosphate form. The clay content is 2 percent. When these base solutions are used for mixing with other suspension ingredients, their clay content often is sufficient to provide all the clay required in the final suspension. Commercial acceptance of 9320 or 12360 base suspension preparation has not yet occurred. At present, apparently it is most economical to furnish polyphosphate as 10340 ammonium polyphosphate solution.

Controlled-Release Fertilizers

The term "controlled-release fertilizer," as used in the following discussion, refers to fertilizers that release, either by design or naturally, their nutrient content over an extended period of time. Potential advantages claimed for controlled-release fertilizers are increased efficiency of uptake by plants; minimization of losses by leaching, fixation, or decomposition; a reduction in the release of nutrients and by-products to air and water; and avoidance of burning of vegetation or damage to seedlings. In spite of this impressive list of potential advantages, the controlled-release fertilizer market remains small and primarily restricted to high-value crops and/or ornamental plants and golf courses.

Controlled-Release Phosphate Fertilizers

Several controlled-release phosphates have been produced commercially. Important among them are ground phosphate rock, basic slag, and fused calcium magnesium phosphate. These products have been discussed earlier.

Most of the insoluble or slightly soluble phosphate materials are used because they are less expensive than soluble phosphates or because they supply other elements. These products are rarely agronomically superior to soluble phosphates. However, some of them are valued for specialty uses because they may be safely placed in contact with seeds or roots in some situations. Slightly soluble phosphates are found to be useful in tropical agriculture where conditions are quite different from those in temperate zones.

Controlled-Release Nitrogen Fertilizers

Because of economics and environmental considerations, the slow-release nitrogen fertilizers are much more important than phosphate (or potash), particularly under certain soil and climatic conditions. In most cases unutilized phosphate and potash remain available for subsequent crops. In contrast, there is seldom much carryover of fertilizer nitrogen from one crop to the next. Nitrogen applied as fertilizer is lost from the soil through: (1) leaching or surface drainage, (2) volatilization as ammonia, and (3) decomposition. Because of these losses, nitrogen use efficiency by crops seldom exceeds 50 percent and may be as low as 20 percent under flooded rice conditions when grown under conditions of intermittent flooding or poor water control. Thus, improved utilization is particularly important because it would favorably impact the cost of production and simultaneously reduce the release of nitrogen compounds to air and water.45

Slightly Soluble Materials

One group of controlled-release nitrogen fertilizers comprises chemical compounds that are inherently only slightly soluble in water or soil solution. Ureaaldehyde compounds are the principal representatives of this group that are produced commercially. Isobutylidene diurea (IBDU) is produced in Japan through the reaction of urea with isobutyraldehyde in a 2:1 mole ratio. When pure, it contains 32.18 percent N.46 Crotonylidene diurea (CDU), also called cyclodiurea, is produced in Japan and Germany through the reaction of urea with crotonaldehyde or acetaldehyde. The pure compound contains about 32 percent nitrogen.

Ureaformaldehyde reaction products, usually called "ureaform," are produced by about six manufacturers in the United States and several other countries. Unlike IBDU and CDU, ureaform is not a definite chemical compound. It contains methylene ureas of different chain lengths; the solubility increases with decrease in chain length. It usually contains about 38 percent nitrogen.

Perhaps the most useful component of ureaform as a slow-release nitrogen material is trimethylene tetraurea; shorter chain lengths are too rapidly decomposed, and longer chains are highly resistant to decomposition. Unfortunately, it is very difficult to produce commercially a pure trimethylene tetraurea; ureaform contains both shorter and longer chains.

A typical ureaform may contain 30 percent of its nitrogen in forms that are soluble in cold water (25°C). The quality of the remaining 70 percent is judged by the percentage that is soluble in hot (boiling) water as determined by prescribed analytical procedures. At least 40 percent of the nitrogen insoluble in cold water should be soluble in hot water for acceptable agronomic response; typical values are 5070 percent.

Ureaform is produced by reaction of urea with formaldehyde in aqueous solution in the presence of an acid that catalyzes the reaction. The degree of polymerization depends on the mole ratio of urea to formaldehyde and on the pH, time, and temperature of reaction. The reaction is stopped by neutralizing the acid. Production processes have been described by Clark et al.47 and by Kolterman and Rennie.48

In Japan, ureaform is also produced as a component of compound fertilizers by the addition of formaldehyde solution to the urea solution fed to the granulator.49 In one process a methylene urea reactor is used to react urea and formaldehyde solution to produce a slurry that feeds directly to a granulation plant with other materials to make compound fertilizers.50 Some slow-release nitrogen is formed, and the physical properties of the products are improved.

In the United States, ammoniating solutions containing urea, formaldehyde, and ammonia are used to prepare compound fertilizer containing water-insoluble nitrogen. The composition of one such solution is: urea35.3 percent, formaldehyde8.0 percent, CO28.4 percent, NH325.0 percent, and H2O23.3 percent. Other solutions with high formaldehyde content (up to 60%) are available for use with additional urea.51

The addition of sulfuric or phosphoric acid (or superphosphate) neutralizes the ammonia and catalyzes the formation of methylene urea. Products containing about 2030 percent of their nitrogen in a water-insoluble form are produced. Good control of conditions in the granulator and dryer is needed to produce a satisfactory quality of water-insoluble nitrogen. In some cases, a prereactor is used prior to the granulator.50

The consumption of ureaform in the United States is approximately 50,000 tons/year, not including that formed in compound fertilizers from solutions. Most of the production goes into compound fertilizer for specialty uses, such as for lawns, flower gardens, and golf courses. The wholesale price of ureaform per unit of nitrogen is about three times that of urea, ammonium nitrate, or ammonium sulfate.

In tests made by TVA and IFDC of many other organic nitrogen-containing compounds as fertilizers, some proved readily available, some slowly available, some inert, and some toxic. Among those that showed promise as slow-release fertilizers were oxamide, glycouril, cyanuric acid, ammeline, and ammelide. For lack of an economical process, none of these materials has been manufactured commercially for fertilizer use.

Magnesium ammonium phosphate (MgNH4PO4) is a slowly soluble source of nitrogen although it contains more phosphate than nitrogen. The commercial product, called "Mag Amp," contains a variable amount of water of crystallization and some impurities; a typical grade is 8400. A similar product containing potash is marketed under the trade name "En Mag" in the United Kingdom. The grade is 52410. The rate of solution of these slightly soluble materials is controlled by the size and hardness of the granules.

Coated Soluble Materials

A wide range of materials and techniques has been explored with the object of making controlled-release fertilizers by coating soluble fertilizer materials with plastic films, resins, waxes, asphaltic materials, or other barriers. The first commercial production of coated controlled-release fertilizer utilizes a process developed by Archer Daniels Midland Company (ADM). The main component of the coating is a copolymer of dicyclopentadiene with a gly-cerol ester.52 Applied in several layers that vary in composition, the coating releases fertilizer solution by osmotic exchange with moisture from the soil. The coated granular fertilizer was manufactured by ADM under the trade name Osmocote starting in 1964. It now is produced and marketed by Sierra Chemical Company under exclusive license from ADM. Three grades of Osmocote are available: 141414, 1899, and 1648. The weight of the coating ranges from 10 to 15 percent of the gross weight. Osmocote products are recommended for turf, floriculture, nursery stock, and high-value row crops.

Sulfur-coated urea (SCU) is a controlled-release material developed by TVA during the 1960s and 1970s.53 Sulfur was selected as the coating material on the basis of economy and efficiency after many coating materials were tried. Urea was selected as the material to be coated because of its high nitrogen content, low cost, and commercial availability.

Initial studies indicated that sulfur alone was not adequately resistant to moisture penetration. However, it was discovered that the addition of an oil wax sealant over the sulfur coating provided a successful coating. The process has been commercialized and the product is mainly used for turf and golf courses. In this process granular urea is preheated in the first rotating drum to 77°C with electric radiant heaters to prevent the sulfur from freezing too rapidly on the granules. The molten sulfur is atomized and sprayed onto the rolling bed of granules in the second drum. Wax is applied in the third drum. The coated granules are then cooled in a fluidized bed cooler, and an absorbent powder is applied in a fourth drum to eliminate the tacky condition of the wax and to prevent the granules from floating when applied in water (as in a rice paddy).

The rate of dissolution of coated urea can be varied by varying the thickness of the sulfur coating and, to some extent, by varying the amount of sealant. In addition to the controlled-release characteristics, coated urea has excellent storage and handling characteristics and might even be handled in bulk in humid climates. Also, it may be blended with TSPs or SSPs, which are incompatible with uncoated urea. Also the sulfur coating is a slowly available source of sulfur, which would be an advantage in sulfur-deficient areas.

The sulfur-coating technique may be applied to fertilizers other than urea. Experimental work has included a coating of DAP, potassium chloride, potassium sulfate, and various compound fertilizers, including a pan-granulated urea-potassium chloride product of 32016 grade.

Nitrification Inhibitors

Nitrogen fertilizers in the ammonium form are immobilized by sorption on clay particles and hence are resistant to leaching. However, the ammonium form is converted to the nitrate form under aerobic conditions by microbiological processes known as nitrification. The rate of conversion depends on the temperature; below 10°C it is very slow, but at 20°C and higher it is rapid. The nitrate form is more readily available to most plants although some plants (e. g. rice) can readily use the ammonium form. Thus, some of the purposes of controlled release (resistance to leaching, prevention of denitrification losses, and delayed availability) may be attained by delaying nitrification of ammonium nitrogen.

Numerous organic chemicals have been identified as nitrification inhibitors. The inhibition is due to the toxicity of these chemicals to organisms that convert ammonium nitrogen to nitrite, which is the first step of the nitrification process.

One of the nitrification inhibitors that has received attention in the United States is 2-chloro-6-(trichloromethyl) pyridine which is promoted under the name N-Serve. The minimum concentration of N-Serve in the soil for delaying nitrification at least 6 weeks is said to range from 0.5 to 10 ppm.

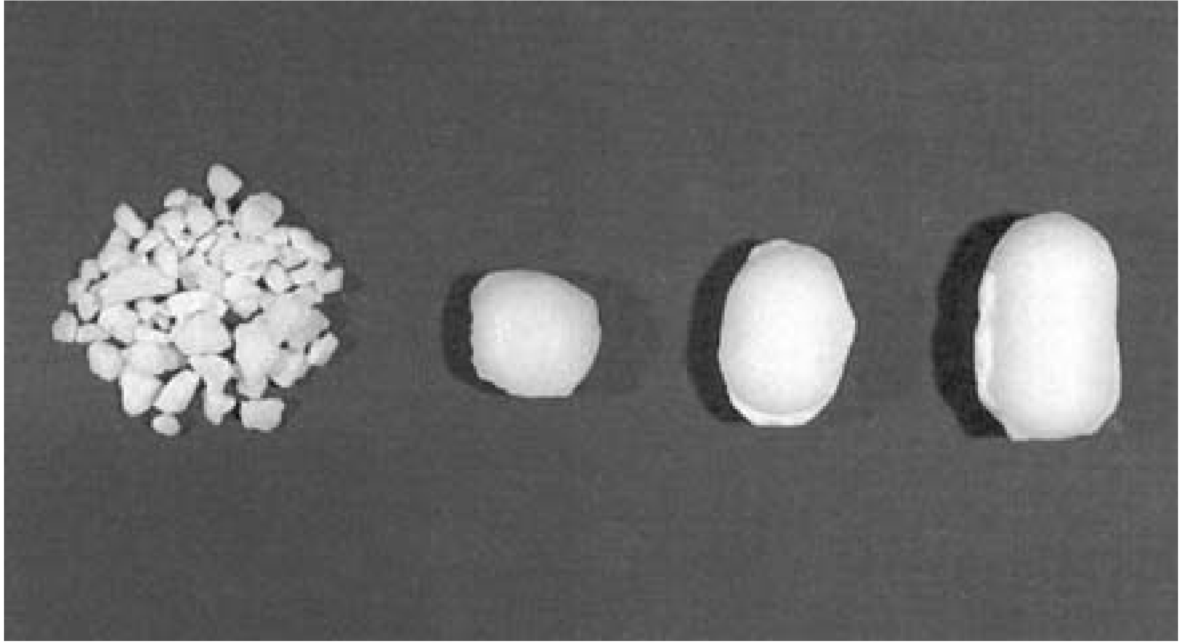
The N-serve literature points out that a nitrification inhibitor is helpful only when conditions favor high nitrogen loss from the soil. Such conditions are heavy rainfall or heavy irrigation, coarse-textured soil, and soil in the pH range where nitrification readily occurs. Nitrate thus formed will leach down into anaerobic soil layers.

Potassium azide (26% KN3) dissolved in anhydrous ammonia was effective in lowering the rate of nitrification in tests conducted by the U.S. Department of Agriculture in cooperation with agricultural experiment stations in the states of Louisiana and Washington. PPG Industries, Inc., is exploring this and other agricultural uses (e.g. as a herbicide in rice culture) for the azide.

Another nitrification inhibitor that has received much attention in Japan is 2-amino-4-chloro-6-methyl pyrimidine. It is manufactured by Mitsui Toatsu Chemicals, Inc. (formerly Toyo Koatsu Industries, Inc.) under the trade name AM. Other nitrification inhibitors that have been used in Japan in compound fertilizers are sulfathiazole, dicyandiamide, thiourea, N-2, 5-dichorphenyl succinamide, 4-amino-1, 2, 4-triazole hydrochloride, and guanylthiourea.50

In flooded or waterlogged soils, as for example, in a continuously flooded rice paddy, anaerobic (reducing) conditions exist in the soil below the surface, and hence nitrification of ammoniacal nitrogen does not occur in this layer. Therefore, when ammonia, ammonium salts, or urea (which hydrolyzes in the soil to ammonia and carbon dioxide) is placed in the reducing zone of the soil, it remains in the ammonium form and is more or less immobilized by sorption on the clay. Some of the advantages of controlled release may be attained by placement of ammoniacal nitrogen in the reducing zone of the rice paddy soil. However, this placement is difficult since applicators quickly become clogged with mud and wet fertilizer. To facilitate subsurface placement, large granules or briquettes of urea have been produced experimentally that can be pressed into the soil by hand (or foot) (Fig. 11.21). Experiments with this placement have shown substantially increased efficiency of nitrogen utilization by rice as compared with the usual broadcast placement.

Fig. 11.21 Compacted urea product and briquetted urea particles of three particle sizes (from left).



IFDC has designed a village-level briquetting machine that can produce urea briquettes ranging in weight from 1 g to 2.7 g (Fig. 11.22). The capacity of these machines varies from 20 kg/hr to 1 ton/hr and can be operated intermittently. As of 2001 more than 650 of these machines are operating in Bangladesh, providing briquettes for nearly 400,000 ha of rice paddy.

Fig. 11.22 A village-level briquetting machine.

11.8 References

Smil, V., Cycles of Life, Scientific American Library, New York, NY, 1997.

Food and Agriculture Organization of the United Nations (FAO), FAOSTAT database <http://faostat.fao.org>, Rome, Italy (2002).

United Nations Industrial Development Organization (UNIDO) and International Fertilizer Development Center (IFDC), Fertilizer Manual, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

Herring, J. R., and Fantel, J. R., "Phosphate Rock Demand into the Next Century: Impact on World Food Supply," Nonrenewable Resources, 2(3), 226241 (1993).

British PetroleumAmoco, *Statistical Review of World Energy*, 1998.

Odell, P., Fossil Fuel Resources in the 21st Century, International Atomic Energy Agency, Vienna, Austria, 1999.

International Fertilizer Development Center (IFDC), *CompactionAlternative Approach for Granular Fertilizers*, IFDC Technical Bulletin T-25, 1983.

McVicar, M. H., Using Commercial Fertilizers, 3rd ed., The Interstate Printers and Publishers, Inc., 1970.

Hignett, T. P., "History of World Fertilisers and Manufacturing Processes," Ind. Chem. Manufacturer, 9(4), 1317 (1971).

Young, R. D., and Achorn, F. P., "Trends in U.S. Fertilizer Technology," TVA Bull. Y-133, pp. 57, Tennessee Valley Authority, Muscle Shoals, AL (Aug. 1978).

The Role of Phosphorus in Agriculture, Am. Soc. Agron., Crop Science Soc. Am., and Soil Science Soc. Am., pp. 198199, Madison, WI, 1980.

Hignett, T. P., "Nitrophosphate Process Advantages and Disadvantages" in Proc. 15th Annual Meeting Fert. Ind. Round Table, pp. 9295, 1965.

Getsinger, J. G., Houston, E. C., and Achorn, F. P., "Production of Diammonium Phosphate by Continuous Vacuum Crystallization," J. Agr. Food Chem., 5, 433436 (June 1957).

Young, R. D., Hicks, G. C., and Davis, C. H., "TVA Process for Production of Granular Diammonium Phosphate," J. Agr. Food Chem., 10, 442447, (Nov. 1962).

U.S. Patent 3,153,574.

"New Developments in Fertilizer Technology," 8th Demonstration, TVA Bull. Y-12, Tennessee Valley Authority (1970).

Young, R. D., and Hicks, G. C., "Production of Monoammonium Phosphate in a TVA-Type Ammonium Phosphate Granulation System," Comm. Fert., 114 (2), 2627 (Feb. 1967).

Lehr, J. R., and McClellan, G. H., "Phosphate Rocks; Important Factors in Their Economic and Technical Evaluation," CENTO Symposium on the Mining and Beneficiation of Fertilizer Minerals, pp. 192242, November 1924, 1973.

Pelovski, Y., and Garrett, M. K., *Partially Acidulated Phosphate Rock FertilisersAgronomic, Environmental and Production Aspects*, Proceedings No. 364, The Fertiliser Society.

Yates, L. D., Nielsson, F. T., Fox, E. J., and Magness, R. M., "Enriched and Concentrated Superphosphate," Industrial and Engineering Chemistry, 45(3), 681690 (1953).

Walthall, J. H., and G. L. Bridger, "Fertilizer by Fusion of Rock Phosphate with Olivine," Ind. Eng. Chem., 35(7), 744777 (1943).

Thompson, W. H., "New Route Cuts Costs for Potassium Orthophosphates," Chem. Eng., 78(8), 8385 (1971).

Drechsel, E. K., "Potassium Phosphates: The New Generation of SUPER Phosphates," Paper presented to the American Chemical Society, Division of Fertilizer and Soil Chemistry, Chicago, Illinois, August 28, 1973.

Giese, A., "Environmental Responsibility on the FarmThe Impact on the Fertilizer and Agriculture Industry," in Environmental Challenges of Fertilizer ProductionAn Examination of Progress and Pitfalls, Proceedings of an International Workshop, J. J. Schultz and E. N. Roth (Eds.), IFDC, Muscle Shoals, AL, September 1719, 1998.

Yates, L. D., Nielson, F. T., and Hicks, G. C., "TVA Continuous Ammoniator for Superphosphates and Fertilizer Mixtures," Part I and Part II of *Farm Chemicals* (August 1954).

U.S. Patent 2,741,545.

Environmental Challenges of Fertilizer ProductionAn Examination of Progress and Pitfalls, Proceedings of an International Workshop, J. J. Schultz and E. N. Roth (Eds.), Muscle Shoals, AL, September 1719, 1998.

"New Developments in Fertilizer Technology," 9th Demonstration, TVA Bull. Y-50, Tennessee Valley Authority (1972).

"New Developments in Fertilizer Technology," 11th Demonstration, TVA Bull. Y-107, Tennessee Valley Authority (1976).

Medbery, J., "Experience with Pipe-Cross Reactor," Proc. Fert. Ind. Round Table, pp. 5255, 1971.

Baggett, K. J., and Brunner, D. J., "MFA's Experience with TVA Cross-Pipe Reactor," Proc. Fert. Ind. Round Table, pp. 6470, 1977.

Fittell, R. S., Hollingworth, L. A., and Forney, J. G., "Manufacture of Granular Ammonium Phosphates Using a Pipe Reactor Process," Proc. Fert. Ind. Round Table, pp. 7081, 1977.

Hoffmeister, G., Watkins, S. C., and Silverberg, J., "Bulk Blending of Fertilizer Material: Effect of Size, Shape, and Density on Segregation," J. Agr. Food Chem., 12, 6469 (Jan./Feb. 1964).

Hoffmeister, G., "Compatibility of Raw Materials in Blended FertilizersSegregation of Raw Materials," Proc. 12th Annual Meeting Fert. Ind. Round Table, pp. 8388, 1962.

Young, R. D., "Providing Micronutrients in Bulk-Blended, Granular, Liquid and Suspension Fertilizer," Comm. Fert., 118(1), 2124 (Jan. 1969).

Young, R. D., "Production of Compound Fertilizers from Intermediates in Local Plants," TVA Bull. Z-30, pp. 1220, Tennessee Valley Authority (1971).

McKnight, D., and Striplin, M. M., "Phosphoric Acid of High Concentration," J. Agr. Chem., 13, 3334 (Aug. 1958).

Phillips, A. B., "Concentration Wet-Processes Acid to Superphosphoric Acid," Farm Chem., 126, 36, 38, 62 (June 1963).

Meline, R. S., Lee, R. G., and Scott, W. C., "Use of Pipe Reactor in Production of Liquid Fertilizers with Very High Polyphosphate Content," Fert. Soln., 16 (2), 3245 (Mar./Apr. 1972).

U.S. Patent 3,775,534.

Walters, H. K., Jr., "Salt Suspension Fertilizers," Comm. Fert., 2526 (Sept. 1959).

Newsom, W. S., Jr., "Suspension Fertilizers," Solutions, 3035 (Jan.Feb. 1960).

Sawyer, E. W., Polon, J. A., and Smith, H. A., "The Stabilization of Suspension Fertilizers with Colloidal Attapulgite," Solutions, 3643 (Jan.Feb. 1960).

Mann, C., II, McGill, K. E., and Jones, T. M., I&EC Product Research and Development, pp. 488495 (Sept. 1982).

Trenkel, M. E., Improving Fertilizer Use Efficiency: Controlled-Release and Stabilized Fertilizers in Agriculture, International Fertilizer Industry Association, Paris, France, 1997.

Hamamoto, M., "Isobutylidine Diurea as a Slow-Acting Nitrogen Fertilizer and Studies in this Field in Japan," Proceedings No. 90, The Fertiliser Society, London, 1966.

Clark, K. G., Yee, J. Y., and Love, K. S., "New Synthetic Nitrogen Fertilizers, Preparation and Properties of Urea-Form," Ind. Eng. Chem., 40(7), 11781183 (1948).

Kolterman, D. W., and Rennie, W. W., "Ureaform Fertilizers," in The Chemistry and Technology of Fertilizers, pp. 4854, V. Sauchelli (Ed.), Rheinhold Publishing Co., New York, NY, 1960.

Ando, J., "Developments in Granulation of Mixed Fertilizers in Japan," Proceedings of the 20th Annual Meeting Fertilizer Industry Round Table, pp. 8593, Memphis, TN, 1970.

Kuwabara, M., Hayamizu, S., and Hatakeyama, A., "Trends in Urea-Based Compound Fertilizer Technology," in Granular Fertilizers and Their Production, pp. 125147, British Sulphur Corporation, London, England, 1977.

Nielsson, F. T., "IMC Experience in Specialty Fertilizer Manufacture," Proceedings of the 23rd Annual Meeting of the Fertilizer Industry Round Table, pp. 3342, Washington, DC, 1973.

Powell, R., Controlled Release Fertilizers, Noyes Development Corporation, Park Ridge, NJ, 1968.

Young, R. D., "TVA'S Development of Sulfur-Coated Urea," TVA Bull. Y-79, Tennessee Valley Authority, Muscle Shoals, AL (1974).

12.1 Sodium Chloride

Common salt, sodium chloride, occurs in nature in almost unlimited quantities. It is a direct source of chlorine, caustic soda (sodium hydroxide), sodium chlorate, synthetic soda ash (sodium carbonate), sodium metal, and sodium sulfate. Indirectly, it is also the source of hydrochloric acid and a host of sodium salts. It has an imposing list of uses, placing it among the more important substances in the economic world. It is used to preserve meat, fish, and hides; it is a necessary component of the animal diet; it is used in refrigeration systems; and large quantities are used for ice control on highways in colder climates. Salt is used by the soap maker to separate soap from glycerine and lye, and by the dye manufacturer to precipitate dye products. In addition, salt is used extensively for the regeneration of water-softening resins (see Table 12.1).

TABLE 12.1 1996 Uses of Salt in the United States (Millions of Metric Tons)
3

Market	Consumption
Chlor-alkali	21.3
Highway deicing	17.7
General industrial	3.5
Agricultural	1.6
Food	1.6
Other chemicals	1.1
Other uses	6.0
Total	52.8

Source: *Chemical Economics Handbook*, Sodium Chloride, SRI International, Menlo Park, CA, 770.8000 U, 1998.

Salt is mined as a solid in shaft mines with depths of 5002000 ft, in Michigan, Ohio, New York, Kansas, Louisiana, and Texas. Run-of-the-mine salt contains 9899 percent NaCl. Over 25 percent of the world's salt is produced in the United States.

More frequently, salt is obtained by solution mining. Typically, this is achieved by drilling a well into a salt formation, installing concentric piping into the well, pumping water in one pipe, and retrieving nearly saturated brine from the other pipe. The brine concentration is controlled by the rate of pumping and is kept slightly undersaturated to avoid salting up the brine lines at the wellhead. The brine is purified by chemical treatment, settling, and filtration. Such artificial brines permit a cheaper operating cost and are well adapted to the manufacture of synthetic soda ash, chlor-alkali products, and table salt.

Very extensive production of artificial brines is carried out along the U.S. Gulf Coast, principally in Texas and Louisiana, where huge domelike deposits of rock salt exist in readily accessible areas. Some of these domes are 4 miles in diameter and over 8 miles deep,1 containing over 100 billion tons of salt each. In order to mine salt efficiently and safely and to utilize the subsequent cavern for storage, tests are performed concerning the mechanical behavior of salt under the stresses of unequal and changing loads. Crude oil and processed hydrocarbons are conveniently stored in the caverns produced in these domes by solution mining of the salt. The hydrocarbons float on the brine within the cavern and are readily recovered by simply pumping brine back into the well. The Gulf Coast, with its salt and hydrocarbon deposits, deep water ports, and ready sources of energy, is ideally suited for the chemical manufacture of salt derivatives, especially chlorinated hydrocarbons.2

Artificial brines also are produced extensively from stratified salt deposits in Michigan, Ohio, New York, West Virginia, Ontario, and Western Canada. Much of the early chemical industry in North America was concentrated in these areas because of the availability of these brines and the need for salt in the production of soda ash, chlorine, and caustic. The stratified salt exists in layers 20200 ft thick, associated with anhydrite ($CaSO_4$). The nature of these deposits is such that it leads to higher calcium levels in the brine when recovered by solution mining compared with the brine obtained from salt domes. In the vicinity of many of these stratified salt deposits, there also exist naturally occurring brines, which consist mainly of $CaCl_2$, with lesser amounts of $MgCl_2$, KCl, and NaCl, a few thousand parts per million of bromide and up to a few hundred parts per million of iodide. These natural brines are mined principally for their bromine content.

Salt also is derived from seawater in those places in the world, such as California, where solar evaporation is sufficient to concentrate the seawater in large basins until some of the salt deposits. The salt is harvested by scrapers or front end loaders. A recent improvement in harvesting technology uses a global positioning satellite to guide the harvester uniformly about the final brine pond.5 The salt-based chemical industry in Asia and the West Coast of the United States derives its salt from solar evaporation, which accounts for about half the total world production (see Table 12.2).

TABLE 12.2 U.S. Salt Production in 1996 (Millions of Metric Tons)4

Evaporated salt	7.1
Rock salt	13.5
Brine	21.5
Total	42.2

Source: *Chemical Economics Handbook*, Sodium Chloride, SRI International, Menlo Park, CA, 770.8000 P, 1998.

Salt and other chemicals also are derived from naturally occurring brines, such as brine from Searles Lake in California, which typically contains 16.5 percent NaCl, 6.8 percent Na_2SO_4, 4.8 percent KCl, 4.8 percent Na_2CO_3, and 1.5 percent $Na_2B_4O_7$. Another source is the Great Salt Lake, with 27 percent salinity, of which 80 percent is NaCl.

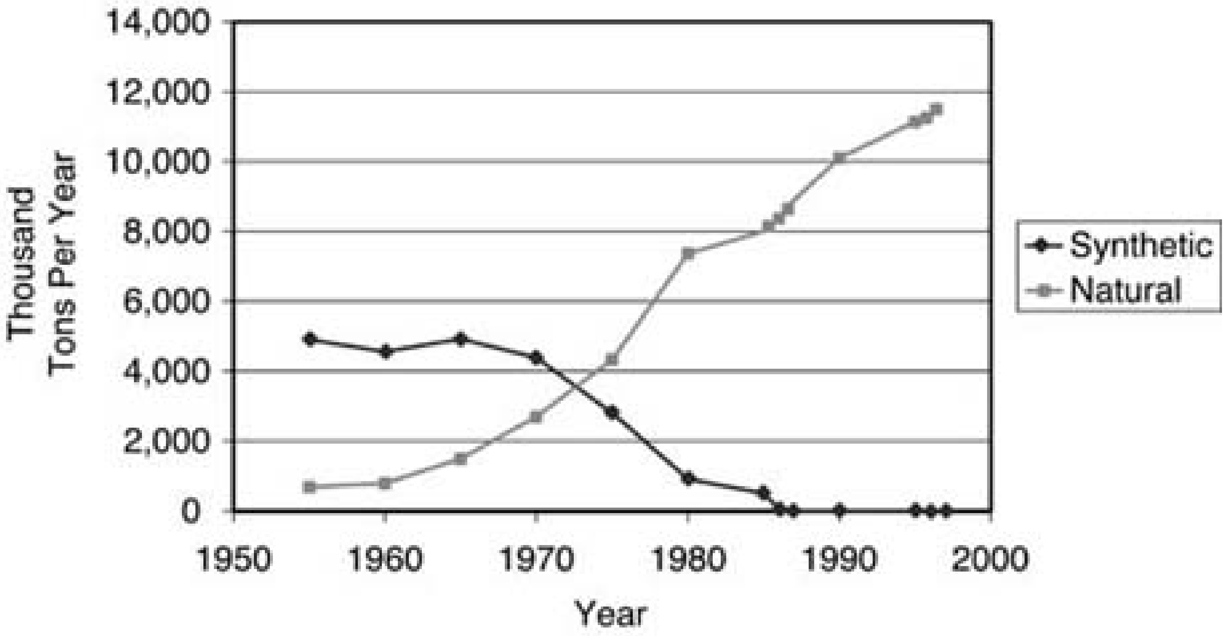
At Searles Lake, the brine is concentrated in a multi-effect evaporator, from which NaCl is recovered. The process liquor is fractionated further to yield a KClborax concentrate and burkeite, a double salt of Na_2CO_3 and Na_2SO_4, which is processed still further to yield soda ash and salt cake (Na_2SO_4) as products.

The use of salt for ice control on highways is second only to the salt usage for caustic chlorine manufacture. In order to minimize damage to the environment, Germany has practiced the application of damp salt on highways. In this process, solid NaCl is dampened with $CaCl_2$ solution, which prevents the salt from being blown from the road and initiates the thawing process more quickly.

12.2 Soda Ash

Soda ash, sodium carbonate, is derived commercially from three sources, which are, in order of decreasing importance: (1) naturally occurring trona ore; (2) the Solvay ammonia-soda process; and (3) naturally occurring alkaline brines. There has been a dramatic shift away from the synthetic or Solvay process ash in recent years, as is shown in Fig. 12.1.6 No synthetic sodium carbonate has been produced in the United States since 1986. The Solvay process is still a major factor in ash production in Europe, but may decline there as well.

Fig. 12.1 U.S. production of synthetic vs. natural soda ash.6



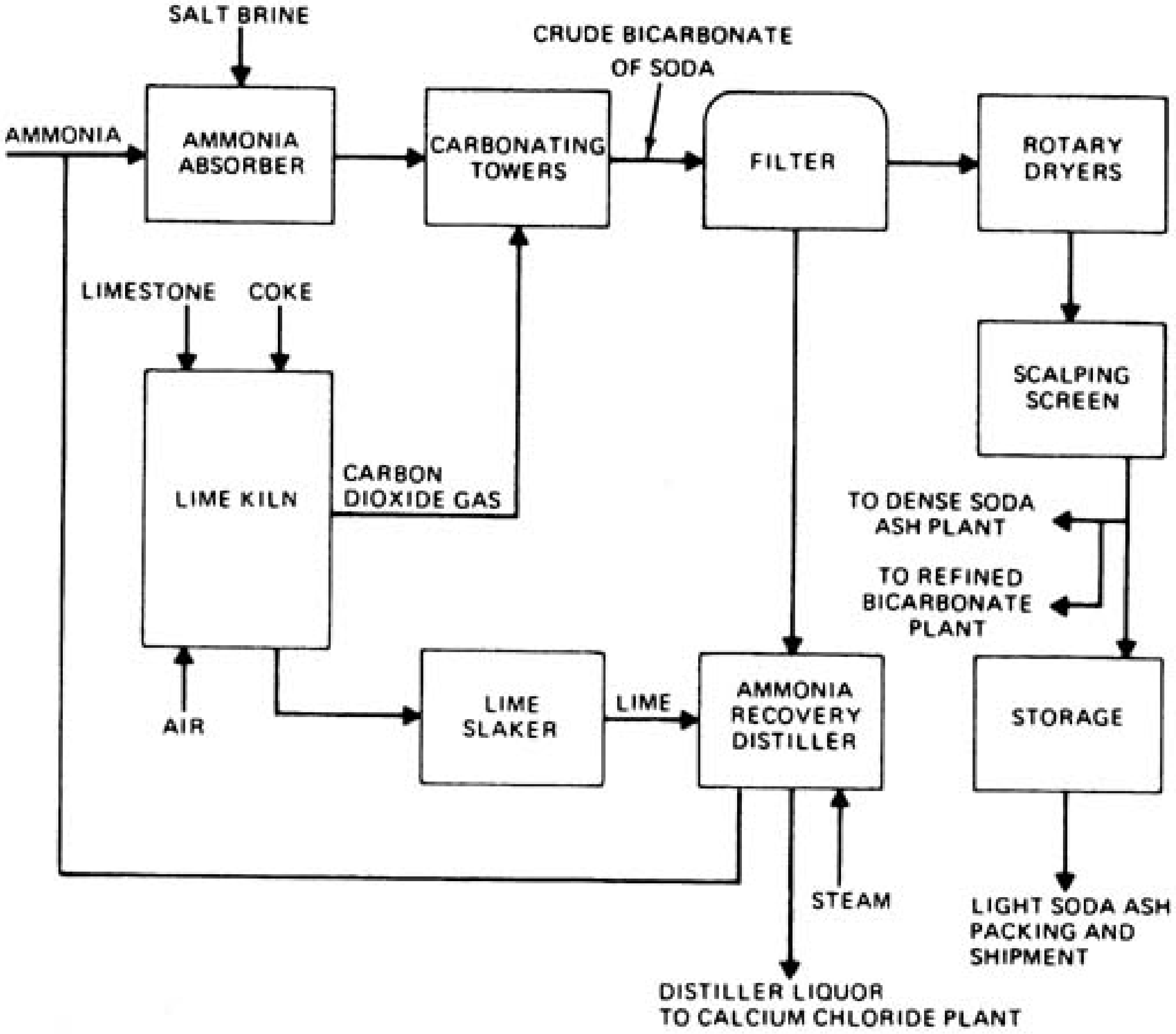
Like its predecessor, the Leblanc process, the Solvay process is on the decline for economic reasons. Increasing costs of production and environmental factors are the key issues in the many closings of synthetic ash plants in the last several years. However, a complete worldwide takeover by natural ash is doubtful because of two factors: (1) the limited amount of natural ash compared with the widespread availability of salt and limestone (the essential ingredients of the synthetic ash process); and (2) the locations of natural ash deposits relative to the locations of the ash consumers. Most of the U.S. natural ash is derived from the area of Green River, Wyoming. The Green River area of Wyoming contains over 1011 metric tons of trona, enough to satisfy global demand for over 2000 years.7 Significant amounts also are recovered from the alkaline brines of Searles Lake, California.

The synthetic ash process, or ammonia soda process, was perfected by Ernest Solvay in 1865. The process is based on the precipitation of $NaHCO3$ when an ammoniated solution of salt is carbonated with $CO2$ from a coke-fired lime kiln. The $NaHCO3$ is filtered, dried, and calcined to $Na2CO3$. The filtered ammonium chloride process liquor is made alkaline with slaked lime, and the ammonia is distilled out for recycle to the front end of the process. The resultant calcium chloride is a waste or by-product stream. The net overall chemical change in the process (shown schematically in Fig. 12.2) is represented by the following stoichiometry:

$$\mathrm{CaCO_3(limestone) + C(coke) + O_2(air) + 2NaCl(brine)}$$

$$\mathrm{\rightarrow Na_2CO_3 + CO_2 + CaCl_2(brine)}$$

Fig. 12.2 Simplified diagrammatic flow sheet for the Solvay ammoniasoda process.



The process requires a large amount of fuel to calcine both the limestone and sodium bicarbonate and to generate steam for ammonia recovery. For the reaction proper, no fuel is required. In fact, large volumes of cooling water are required to remove the heat generated by the absorption and reaction of ammonia and carbon dioxide. The process has an imperfection in that an undesirable solution of calcium chloride also is produced.

Prior to being fed to the process, the NaCl brine must be purified to remove calcium and magnesium ions so that they will not be precipitated when the brine is carbonated, producing objectionable scale on equipment surfaces, as well as contaminating the product. A solution of soda ash and caustic is added to the brine to precipitate the calcium as $CaCO3$ and the magnesium as $Mg(OH)2$. These impurities are flocculated, removed in a settler, and discarded. The purified brine is fed to the ammonia absorber in the process.

Lime for recovering ammonia in the process and $CO2$ for reacting with the ammoniated brine are produced by calcining the best available limestone with foundry coke in vertical shaft kilns. This type of equipment is preferred for producing the maximum yield of active lime and the maximum concentration of CO2 in the kiln gas. The dry lime, drawn from the vertical kiln, is cooled by the entering air, which is thereby preheated. Likewise, for maximum fuel economy, the exiting gas preheats the incoming limestone and fuel. The gas composition should exceed 40 percent CO2 with only factional percentages of CO and O2.

The dry lime and hot water are fed to a rotating cylindrical slaker to produce milk of lime, which is pumped to the distillers for the recovery of ammonia.

The Solvay process recycles large quantities of NH3, and it is necessary to minimize losses in scrubbing the various NH3-containing gas streams. Thus, the incoming purified brine is used to wash the ammonia-bearing air, which is pulled through the bicarbonate cake on the vacuum filters. This is accomplished in a packed absorber. The brine then passes through a second packed section where it absorbs the NH3 in the gases from the carbonating towers. The brine then flows to the main NH3-absorber, which is also a packed tower. The brine is circulated through water-cooled heat exchangers to remove the heat of absorption. Ammonia, CO2, and small amounts of H2S released in the distiller are absorbed in the packed tower. In addition, a small amount of makeup ammonia is added. A typical analysis of the cooled, ammoniated brine is:

Temperature	38°C
NH3	90 g/L
CO2	40 g/L
NaCl	260 g/L
H2S	0.1 g/L

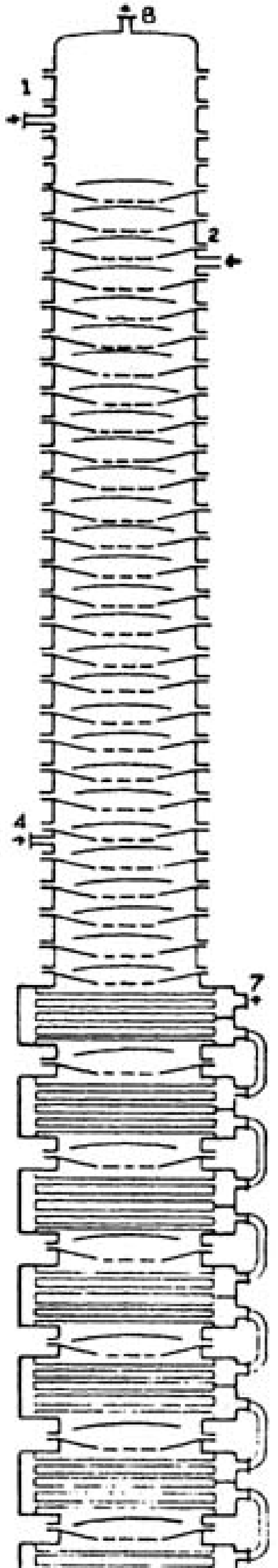
The total heat removed from the ammonia absorption is 1.25 million BTU per net ton of soda ash produced.

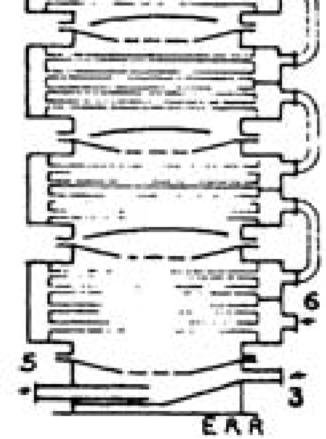
Next, the ammoniated brine is carbonated to a point just short of crystallization. Then the brine is given a final carbonation and cooled to produce the crude NaHCO3 crystals. The equipment used in the carbonation step consists of groups of five identical towers having alternate rings and discs in the upper section to assure mixing of the falling liquor with the rising gas stream without plugging by the crystallizing solid phase. The lower section of each tower consists of a series of heat-exchanger bundles alternating with rings and discs.

The ammoniated brine is passed downward through one of the group of five columns, which has become fouled with sodium bicarbonate after operating as a crystallizing unit for several days. The 40 percent CO2 gas from the kilns is pumped into the bottom of the column to provide agitation and heat in order to dissolve the crystalline scale and bring the liquor to a composition just short of crystallization. The liquor is adjusted to the desired temperature by passage through a heat exchanger in preparation for feeding to the crystallizing towers, and then it is fed into the top of each of the other four columns in the group.

A mixed gas of 6075 percent CO2 derived from mixing the 40 percent CO2 from the kilns and the 90 percent CO2 from the calcination of bicarbonate, is fed to the bottom of these crystallizing units. Absorption of CO2 in the highly alkaline ammoniated brine results in crystallization of crude sodium bicarbonate. Because of the heat evolved in the absorption and neutralization of the carbonic acid gas and from crystallization of the sodium bicarbonate, the temperature of the liquor in the column rises from 38°C to a maximum of about 6264°C. In normal operation, the temperature of the discharge slurry is maintained at about 27°C by automatic adjustment of the water flow through the cooling tubes. A tower such as that shown in Fig. 12.3 has a capacity for producing 50 tons per day of finished soda.

Fig. 12.3 A carbonating tower in the ammoniasoda process, 69 feet high and 6 feet in diameter. (1) Entry for ammoniated brine, used when the tower is being cleaned; (2) entry for the ammoniated brine for the regular bicarbonate precipitation; (3) and (4) carbon dioxide entries; (5) outlet for bicarbonate slurries; (6) cooling water inlet; (7) cooling water outlet; (8) escape for uncondensed gases. (*Modeled after Kirschner*.)





The heat removed from the carbonator is about 260,000 BTU per net ton of ash produced, whereas that removed from the crystallizing units is 1.25 million BTU per ton of product.

It is noteworthy that the presence of sulfide in the feed liquor to the towers serves to maintain a protective film of iron sulfide on the cast iron equipment, which minimizes the contamination of product crystals from iron corrosion products.

Slurry drawn from the crystallizing columns is filtered in rotary vacuum filters where the $NaHCO_3$ crystals are water-washed. The filtered liquor then flows to the distillation unit for recovery of ammonia. The crude bicarbonate filter cake contains 3 percent NH_4HCO_3, 1215 percent H_2O, and 80 percent $NaHCO_3$.

Next, the liquor is pumped to the NH_3-still preheater. At this point, the sulfide solution required for corrosion protection is added. The preheated process stream enters a stripper where excess CO_2 is removed prior to treatment with lime and distillation removal of the NH_3. The hot liquor from the stripper flows to an agitated vessel, to which is added milk of lime to release NH_3:

$$\mathrm{Ca(OH)_2 + 2NH_4Cl}$$

$$\mathrm{\rightarrow CaCl_2 + 2NH_3 + 2H_2O}$$

The lime-treated solution then is fed to the top of the bubble-cap distillation unit. Steam is injected at the bottom, stripping out the ammonia down to a residual level of only 0.001 percent.

The crude $NaHCO_3$ is calcinated in dryers constructed with rotating seals and gas-tight feed and discharge mechanisms, to ensure the production of CO_2 that is undiluted with air:

$$\mathrm{2NaHCO_3 \rightarrow Na_2CO_3 + H_2O + CO_2}$$

The heat requirement is about 2 million BTU per ton of soda ash produced. Product from the dryers is cooled for shipment, or converted to other products. A typical analysis of good commercial light soda ash made by this process is:

Na_2CO_3	99.70%
NaCl	0.12%
H_2O	0.12%
Fe	15 ppm
Ca and Mg	75 ppm

The waste liquor from the process may be evaporated to produce by-product $CaCl_2$ and NaCl. However, the major portion is clarified and pumped into water courses whose natural flow is sufficient to provide the dilution needed for disposal.

Production of soda ash from trona ore mined in the Green River area of Wyoming is rapidly expanding. Two basic processes are used, involving dry mining of the ore from a depth of about 1500 ft. A third process based on solution mining of the ore has been developed. In the older dry-mining process, the ore is crushed and dissolved at the surface. Then the solution is purified by settling and filtration, followed by evaporative crystallization to form sesquicarbonate: $Na_2CO_3 \cdot NaHCO_3 \cdot 2H_2O$. The crystals are centrifuged off and calcined to ash in steam-tube dryers. In a more recent process the crushed ore is calcined immediately, then dissolved, purified, and evaporatively crystallized to sodium carbonate monohydrate, which is centrifuged and dried to product ash.

Relatively small quantities of soda ash are produced from alkaline brines at Searles Lake, California, by a process of fractional crystallization that also produces other sodium and potassium salts. Table 12.3 shows the current distribution of soda ash in the United States.

TABLE 12.3 1999 U.S. Consumption of Sodium Carbonate (Thousands of Short Tons)8

Total exports	3,550
Glass	3,240
Chemicals	1,780
Soaps and detergents	730
Flue gas desulfurization	200
Pulp and paper	130
Water treatment	130
Miscellaneous	400
Total	10,160

In 2000 the list price of bulk dense soda ash was $105/ton f.o.b. Green River, WY. Market prices were typically $7275/ton for large orders.8

12.3 Sodium Bicarbonate

Sodium bicarbonate is produced from a saturated solution of sodium carbonate by carbonation:

$$Na_2CO_3 + CO_2 + H_2O \rightarrow 2NaHCO_3$$

The precipitated bicarb crystals are centrifuged and then dried at low temperature to avoid reversion to carbonate.

Sodium bicarbonate also can be obtained by solution mining of natural sodium bicarbonate (nahcolite). In the mining procedure, pairs of well are sunk, warm water is pumped down one well, and dissolved sodium bicarbonate solution is pumped up from the other well. Solid sodium bicarbonate subsequently is crystallized from the solution.

In 2000, there was 570,000 short tons per year of sodium bicarbonate capacity based on the soda ash route and 110,000 tons per year based on mining nahcolite. Two additional nahcolite-based plants were planned for an additional 200,000 tons per year capacity. All nahcolite-based plants are located in Colorado.9

The major use of sodium bicarbonate in North America is in foods as baking soda and as feed supplement for dairy cattle. Table 12.4 gives a breakdown of uses of $NaHCO_3$. The 2000 list price of $NaHCO_3$ was $22.80 per hundredweight.9

TABLE 12.4 1999 U.S. Consumption of Sodium Bicarbonate (Thousands of Short Tons)9

Food	168
Animal feed	126
Cleaning products	47
Pharmaceuticals and personal care	47
Chemicals	42
Water treatment	32
Fire extinguishers	11
Paint blast media	11
Miscellaneous	42
Total	526

12.4 Sodium Sulfate

Sodium sulfate is produced from natural sources and as by-product in a variety of processes including: ascorbic acid, battery recycling, resorcinol, silica pigment, viscose rayon, and sodium bichromate. Other salt cake production processes are based on the reaction of sulfuric acid (Mannheim process) or SO_2 and air (Hargreaves process) with sodium chloride. The Mannheim process is no longer in use in the United States. U.S. Hargreaves capacity is very limited. However, both processes are used widely in the remainder of the world. Table 12.5 lists the sources of salt cake in the North American Free Trade Zone.10

TABLE 12.5 2001 Sodium Sulfate Production By Process10

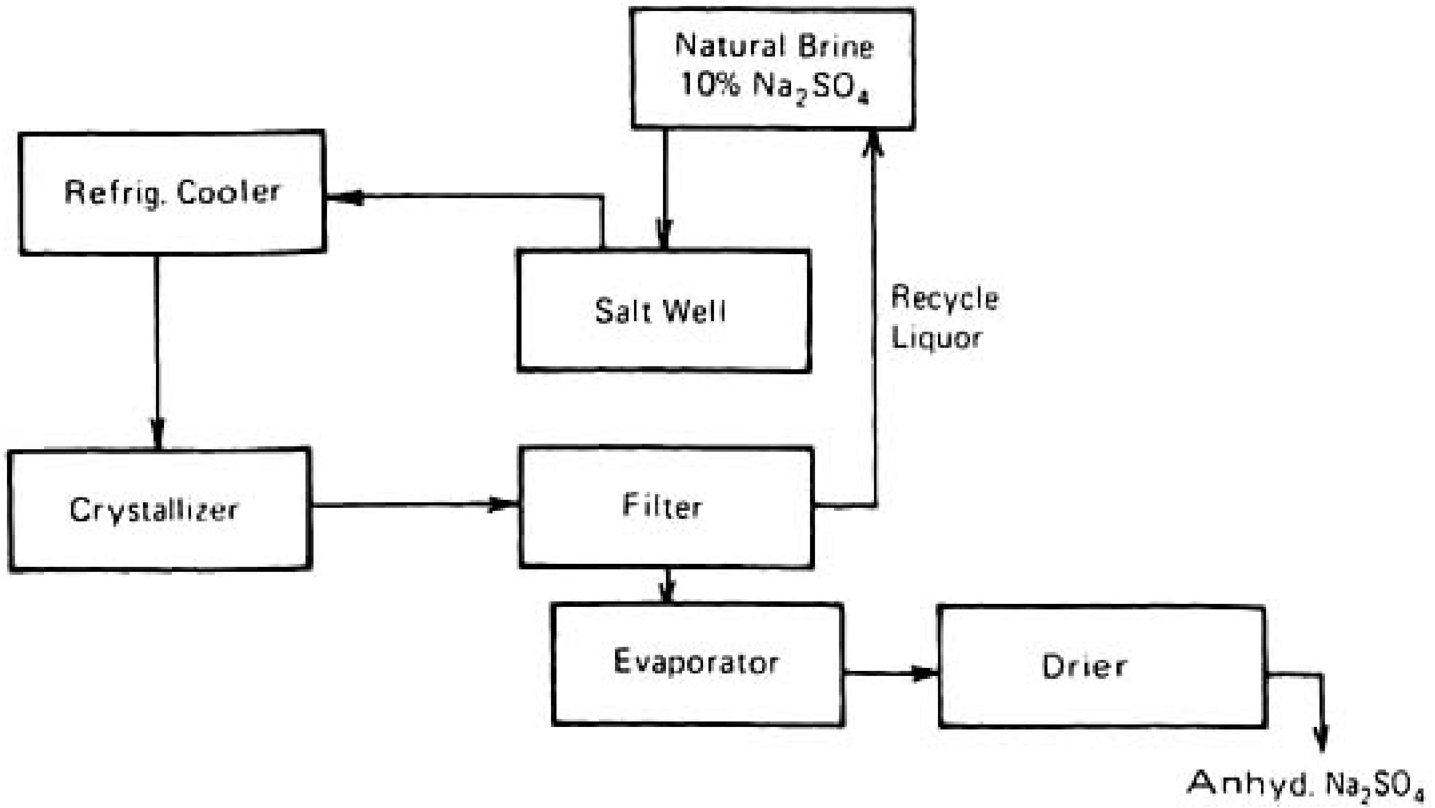
Process	Thousands of Tons per Year
Natural	1402
Sodium bichromate	120
Rayon	95
Battery recycling	75
Ascorbic acid	26
Resorcinol	35
Silica pigment	55
Unidentified	344

One large source of natural sodium sulfate in Coahuila, Mexico, with a capacity of 760,000 tons per year can supply all U.S. demand605,000 tons in 2000. This is not reflected in imports which amounted to 38,000 tons in 2000. Exports of sodium sulfate were 163,000 tons in that same year.10 In addition, a small amount of Glauber's salt ($Na_2SO_4 \cdot 10H_2O$) still is produced for sale, but it has very little effect in the world of sodium sulfate.

The term "salt cake" now refers generally to sodium sulfate, but originally meant sulfate produced in the Mannheim furnace, which yields a product containing less than 99 percent Na_2SO_4, compared with the much purer anhydrous Na_2SO_4 derived from natural brines.

In the latter process (shown schematically in Fig. 12.4) the natural brine, containing about 10 percent Na_2SO_4, is saturated with NaCl by pumping it into a salt deposit. The concentrated brine then is pumped from the salt well through a refrigerated heat-exchanger, where it is cooled to around -10°C, and then into a crystallizer where quite pure crystals of Glauber's salt are formed. This solid hydrate of Na_2SO_4 then is melted and the water removed by evaporation using a submerged burner as the heat source. The wet Na_2SO_4 is dried further in a rotary kiln, producing anhydrous Na_2SO_4. Over 10 tons of natural brine and 1.6 million BTU of heat energy are required to produce 1 ton of product. The most significant production location is in Texas, where the sulfate-containing brine is conveniently located near deposits of both domed salt and natural gas.

Fig. 12.4 Simplified flow sheet for the production of Na_2SO_4 from natural brine.



A similar process is practiced at Searles Lake. There the brine first is carbonated and chilled to remove sodium carbonate and borax. Further chilling crystallizes Na_2SO_4 as the hydrate, Glauber's salt, and some remaining borax. The coarse crystals of Glauber's salt are separated from the fine crystals of borax in a hydraulic classifier. The sulfate fraction then is filtered, washed, dried, and evaporated to produce anhydrous Na_2SO_4.

In another process at Searles Lake, the brine first is evaporated to produce NaCl and a double salt of Na_2CO_3 and Na_2SO_4. The two solids are separated in a hydraulic classifier, and the sulfate fraction is redissolved and recrystallized to Glauber's salt. Anhydrous Na_2SO_4 is recovered by mixing the Glauber's salt with sodium chloride brine. This dissolves the Glauber's salt and allows anhydrous Na_2SO_4 to precipitate because of its lowered solubility in the presence of sodium chloride.

In the Mannheim process, NaCl and 100 percent H_2SO_4 in amounts equivalent to complete conversion to Na_2SO_4 are fed to a circular muffle furnace made of cast iron. The furnace is equipped with a shaft, which penetrates from below and carries four arms, each of which is fitted with two cast iron plows. The shaft rotates at 0.5 rpm, slowly plowing the heated mixture to the circumference where the burned cake discharges through a chute. The furnace charge is heated to about 840°C, just below the fusion temperature of the salt cake. HCl is liberated, cooled, and absorbed in water to produce 32 percent HCl for sale. The salt cake is really a by-product of the operation.

A substantial amount of by-product Na_2SO_4 also is produced in a variety of processes involving the use of H_2SO_4 to react with a sodium salt, or to neutralize free caustic soda. Chief among these are the viscose rayon process and the manufacture of sodium dichromate. In the viscous process, Na_2SO_4 is formed by the reaction between H_2SO_4 and cellulose xanthate plus free caustic in the rayon fiber spinning bath. Spin-bath liquor is evaporated to crystallize Glauber's salt, which is centrifuged off, melted, and then evaporated to anhydrous Na_2SO_4. This is a significant source of sulfate, as 1.1 lb of Na_2SO_4 is produced per pound of rayon fiber spun.

In the manufacture of sodium dichromate, by-product Na_2SO_4 crystallizes directly as the anhydrous material when H_2SO_4 is added to a boiling solution of sodium chromate:

$$2Na_2CrO_4 + H_2SO_4 + H_2O$$
$$\rightarrow Na_2Cr_2O_7 \cdot 2H_2O + Na_2SO_4$$

Decreased use of chrome by automobile and small-appliance manufacturers and mandated recycling of chromium wastes in the electro-plating industry have led to a decline in Na_2SO_4 produced from this source.

Table 12.6 lists the consumption of sodium sulfate by market application. U.S. demand has shrunk over the last decade and this is expected to continue. World demand continues to grow slightly. The 2001 list price is approximately \$113/ton.10

TABLE 12.6 2000 U.S. Consumption of Sodium Sulfate10

Industry	Thousands of Tons	Percent
Detergents	278	46
Kraft pulping	79	13
Textiles	73	12
Glass	67	11

Carpet freshener	42	7
Other	85	14

In the kraft pulping process, sodium sulfate (salt cake) is added as makeup to the recycle black liquor just prior to entering the recovery furnace. In the recovery furnace the sodium sulfate is reduced to sodium sulfide. A decade ago, sulfur losses accounted for a sodium sulfate makeup of about 80 lb per ton of pulp. Since then, antipollution regulations have encouraged technical innovations to decrease sulfur losses in the pulping process to a level where a makeup of only 40 lb salt cake per ton of pulp is required. Also, in recent years the use of Na2SO4 in the manufacture of kraft paper has been giving way somewhat to the use of sodium sulfide, sodium hydrosulfide, and emulsified sulfur, which are more readily available as by-products of the petroleum industry, derived from the caustic scrubbing of sulfide-containing hydrocarbon gases. The sulfides often are preferred by the paper industry because the sulfate must be reduced to sulfide in the pulping process.

In dry detergents, sodium sulfate is used largely as a low-cost inert filler and diluent. In recent years detergents have become more concentrated, and sodium sulfate requirements have been reduced.

In glass manufacture, sodium sulfate is used in small amounts for its sulfur and sodium oxide values as a melting aid. In order to minimize the deleterious effects of volatilized sodium sulfate from glass manufacture, the amount of sodium sulfate consumed per ton of glass has been decreasing in recent years.

12.5 Sodium Sulfides

Sodium sulfide (Na2S) and sodium hydrosulfide, sometimes referred to as sodium sulfhydrate (NaHS), share the same derivation: caustic soda and H2S. Earlier processes based on reduction of salt cake with coal have largely been supplanted. In 2000, the U.S. installed capacity for NaHS was 184,000 tons/year. Additionally, there is more than 260,000 tons/ year of lower-purity, lower-strength material recovered as by-product from natural gas and small refinery operations.11 The source of H2S used in the manufacture of NaHS is essentially all by-product. Controlled caustic scrubbing of H2S forms a solution of NaHS:

$$\mathrm{NaOH} + \mathrm{H_2S} \rightarrow \mathrm{NaHS} + \mathrm{H_2O}$$

The NaHS solution is filtered to remove the sulfides of heavy metals, such as Fe, Hg, Ni, Mn, and Cu. The clear filtrate may be sold as a 4446 percent solution of NaHS or evaporated in stainless steel equipment to crystallize a solid hydrate containing 7072 percent NaHS, which is sold as a flake product. NaHS is easily converted to Na2S by further reaction with caustic:

$$\mathrm{NaHS} + \mathrm{NaOH} \rightarrow \mathrm{Na_2S} + \mathrm{H_2O}$$

By using a NaHS solution of the proper concentration with a flake caustic, a hydrate product containing 6062 percent Na2S is obtained. This is sold directly as a flake product, or fused solid in drums. When high-quality H2S and NaOH are used, the product Na2S is suitable for use in dyes, photography, rayon, and leather manufacture.

Lower-quality sulfides are obtained by using H2S produced as a by-product in the manufacture of CS2 from methane (or other low molecular weight hydrocarbons) and sulfur:

$$\mathrm{CH_4} + 4\mathrm{S} \rightarrow \mathrm{CS_2} + 2\mathrm{H_2S}$$

The gas mixture from this catalytic reaction is cooled and scrubbed with caustic. The sulfides so produced contain small amounts of mercaptans (e.g. CH3SH), which lend a very objectionable odor to the product.

Sulfides also are obtained as by-products in the manufacture of BaCO3 from the barite ore, BaSO4, by roasting with coal, water leaching, and treatment with soda ash:

$$\mathrm{BaSO_4} + 4\mathrm{C} \rightarrow \mathrm{BaS} + 4\mathrm{CO}$$

$$\mathrm{BaS} + \mathrm{Na_2CO_3} \rightarrow \mathrm{Na_2S} + \mathrm{BaCO_3}$$

Table 12.7 shows the distribution for consumption of Na2S and NaHS by industry use. They are both used extensively in a variety of industries. Use in kraft pulp mills continues to grow but much of the material is the lower grade variety. The 2000 carload price for sodium hydrosulfide 45 percent liquid was $585/ ton on a 100 percent basis. NaHS supply far exceeds U.S. demand of 114,000 tons/year of the high-purity, high-strength material. Prices are expected to remain steady.11

TABLE 12.7 Uses of Sodium Sulfides (Percent of Total Market)1214

Industry	Na2S	NaHS
Pulp and paper	3	40
Dyestuffs and miscellaneous chemicals	25	15
Rayon and film	3	5
Metals and minerals	3	28
Leather	65	12

12.6 Sodium Thiosulfate

Over 75 percent of crystal sodium thiosulfate (hypo) ($Na_2S_2O_3$) is used in photography (Table 12.8) because of its ability to dissolve water-insoluble silver salts. U.S. production of sodium thiosulfate was 37,000 tons/year in 1955. It fell to less than half that amount by the mid-1970s and has recently recovered to 31,000 tons/year in 1998. Photography remains the largest single use of the material.15 The introduction of new technologiesfirst videotape followed by digital imagingconsiderably limits the future growth in this market.

TABLE 12.8 1998 Major Uses of Crystalline Sodium Thiosulfate15

Industry	Percent
Photography	76
Water treatment	5
Oil recovery	6
Leather processing	3
Others	10

Note: About one-third of domestic production is sold as an aqueous solution. Seventy-five percent of the soluble product is used in water treating; dechlorinating industrial and municipal wastewater. Although this market has considerable upside potential with added emphasis on reducing man's environmental impact, the largest installations use sulfur dioxide for this purpose.15

Most current production is derived as by-product from the manufacture of sulfur dyes and Na_2S:

$$Na_2S + SO_2 + H_2O \rightarrow Na_2SO_3 + H_2S$$

$$Na_2SO_3 + S \rightarrow Na_2S_2O_3$$

$$2Na_2S + Na_2CO_3 + 4SO_2 \rightarrow 3Na_2S_2O_3 + CO_2$$

In producing sulfur dyes, by-product leach liquor contains sodium thiosulfate, which is simply derived by evaporation and crystallization. Much of the thiosulfate is sold as the pentahydrate, $Na_2S_2O_3 \cdot 5H_2O$. Stainless steel equipment is used in processing thiosulfate.

Of less importance today is the older soda ashsulfur process. Ash is dissolved in hot water, and the solution is pumped to SO_2 absorption towers. The resulting $NaHSO_3$ liquor is then heated with powdered sulfur in an agitated stainless steel digestion tank at elevated temperature. The product solution is filtered, evaporated, and crystallized to yield $Na_2S_2O_3 \cdot 5H_2O$ product, which is centrifuged, washed, dried, and packaged in airtight containers (to discourage efflorescence). In 1999, sodium thiosulfate pentahydrate sold for 36.5 cents/lb.15

12.7 Sodium Sulfite

Sodium sulfite (Na_2SO_3) is produced by reacting sulfur dioxide with a solution of sodium carbonate or sodium hydroxide. Sodium sulfite also is recovered as a by-product of resorcinol production. About half of the sodium sulfite produced in North America is used by the pulp and paper industry (i.e. chemithermomechanical process). It also is used in photography, textile bleaching, and food preservation, and as a chemical intermediate. Some producers can switch from sodium sulfite to sodium bisulfite or metabisulfite in the same facility. In 1998, the market demand for sodium sulfite was 102,000 tons in the United States, excluding that produced and used captively by some paper mills. The demand has decreased substantially over the last decade as pulp mills have reduced sulfur discharges. The 1999 selling price for anhydrous 95100 percent material was approximately $30.50 per hundredweight.16

12.8 Sodium Bisulfite

Most of the sodium bisulfite of commerce is really the anhydride, $Na_2S_2O_5$, sodium metabisulfite or sodium pyrosulfite. Sodium bisulfite is not stable and is generally transported in the metabisulfite form, which is a solid containing 98 percent $Na_2S_2O_5$, 1.5 percent Na_2SO_3, and 0.5 percent Na_2SO_4. Methods of manufacture of sodium metabisulfite are all variations of the same theme:

$$\mathrm{Na_2CO_3 + 2SO_2 \rightarrow Na_2S_2O_5 + CO_2}$$

In one variation, SO_2 is sparged into a stainless steel absorber through which is passed a solution of Na_2CO_3. Formation of product crystals from the saturated solution is achieved by lowering the temperature. The crystals are centrifuged off and dried rapidly in a flash dryer in order to avoid air oxidation to sulfate. The markets for sodium metabisulfite include chemical intermediates, pharmaceuticals, food preservatives, and dye fixing. Additionally, sodium metabisulfite is used in photography, as an antichlor to remove excess chlorine from bleached kraft pulps, and for the manufacture of sodium hydrosulfite. It is estimated that U.S. production is more than 45,000 tons/year. The 1995 list price was approximately 29 cents/lb.[17]

12.9 Sodium Hypo Sulfite (Hydrosulfite)

Not to be confused with "hypo" (the term used in photography to refer to thiosulfate), sodium hyposulfite (or hydrosulfite) ($Na_2S_2O_4$) also is referred to as sodium dithionite. $Na_2S_2O_4$ is a powerful reducing agent used principally for the reduction of vat dyes and brightening of mechanical wood pulps and clays (Table 12.9). There are numerous methods for production of sodium hyposulfite, all of which involve reduction of SO_2 or sodium bisulfite. Sulfur dioxide or $NaHSO_3$ is reduced by Zn, sodium amalgam, sodium formate, sodium borohydride, or cathodic reduction to produce $Na_2S_2O_4$.

TABLE 12.9 2000 U.S. Markets for Merchant Sodium Hydrosulfite18

Industry	Percent
Pulp bleaching	55
Textile dying	24
Clay bleaching	16
Miscellaneous	5

There are two variations of the zinc method. In the first, a solution of sodium hydrogen sulfite is reduced with zinc dust in the presence of excess SO_2:

$$2NaHSO_3 + Zn + SO_2 \longrightarrow Na_2S_2O_4 + ZnSO_3 + H_2O$$

Milk of lime is added to neutralize remaining sulfurous acid, and $CaSO_3$ and $ZnSO_3$ are removed by filtration. Common salt is added to the remaining solution to salt out $Na_2S_2O_4 \cdot 2H_2O$. The suspension of crystals is heated to 60°C to dehydrate the product, water is extracted with alcohol, and the product crystals are vacuum-dried. The crystals are stable only when completely dry.

The second method based on zinc consists of treating an aqueous suspension of zinc dust with SO_2 at 80°C:

$$Zn + 2SO_2 \longrightarrow ZnS_2O_4$$

Treatment of this solution with soda ash precipitates a basic zinc carbonate, which is filtered off. The dry crystalline product $Na_2S_2O_4$ is obtained by salting out and drying, as in the first method. The zinc methods have become obsolete because of the environmental concerns associated with zinc discharge.

The amalgam process uses a dilute sodium amalgam (0.5% Na) derived from the cathode reaction of a mercury chlorine cell.19 The amalgam is admixed with a solution of $NaHSO_3$ whose pH is controlled in the range of 57:

$$4NaHSO_3 + 2NaHg \longrightarrow Na_2S_2O_4 + 2Na_2SO_3 + 2H_2O + Hg$$

The mercury is returned to the mercury cell to form more amalgam, and the $Na_2S_2O_3$ is allowed to react with aqueous SO_2 to form more $NaHSO_3$.

One common method for sodium hyposulfite production involves the use of sodium formate:

$$HCOONa + NaOH + 2SO_2 \longrightarrow Na_2S_2O_4 + H_2O + CO_2$$

The above aqueous reaction requires the use of methanol under acidic conditions. In this process, the anhydrous sodium hydrosulfite is directly precipitated and is filtered off.

Sodium hyposulfite (57% $Na_2S_2O_4$) also is produced on-site at pulp and paper mills by reduction of SO_2 with sodium borohydride solution under alkaline conditions (Borol process):

$$\underset{\text{Borol solution}}{(NaBH_4 + 3.4NaOH)} + 5.6NaOH + 9SO_2 \longrightarrow 4Na_2S_2O_4 + NaBO_2 + NaHSO_3 + 6H_2O$$

US Patent 4,793,906 describes an electrochemical process for producing hydrosulfite solutions. The cathode reaction is:

$$2HSO_3^- + 2H^+ + 2e^- \longrightarrow S_2O_4^{2--} + 2H_2O$$

It is anticipated that such a process would be viable for on-site production of sodium hyposulfite.

In 1999, U.S. demand was 101,000 tons of $Na_2S_2O_4$ including 20,000 tons of imports and 4,000 tons of exports. During the mid-1990s the United States became a net importer of sodium hydrosulfite with most of the material coming from Canada. In addition, approximately 13,000 tons per year of $Na_2S_2O_4$ are produced captively in the United States by the Borol process. The on-site process has the advantage of minimum hyposulfite degradation to thiosulfate, which causes corrosion in pulp mills. In 2000, the list price for $Na_2S_2O_4$ was 76 cents/lb. U.S. capacity of 109,000 tons per year is nearly in balance with the demand. With no significant new uses on the horizon, prices are expected to remain stable.18

12.10 Sodium Phosphates

In 1997, 37 percent of the phosphorous produced in the United States was used in the manufacture of the family of sodium phosphate products, chief among them is sodium tripolyphosphate, $Na_5P_3O_{10}$. Sodium tripolyphosphate (STPP) is approximately 57 percent P_2O_5 and 42 percent Na_2O. In 1997, 242,300 tons of sodium phosphates were sold at 4142 cents/lb.[20]

It is made by the reaction between soda ash and phosphoric acid.[21] By far, the major source of phosphoric acid is furnacegrade acid, because of its high purity as compared with the less pure wet-process acid. A hot solution of Na_2CO_3 is reacted with 60 percent H_3PO_4, with a molar ratio of sodium to phosphorus of 1.67 maintained in order to achieve the desired stoichiometry. The solution is purged with steam to evolve all the CO_2, and then it is filtered, spray-dried, and granulated for shipment. By changing the ash-to-acid ratio, a variety of products can be produced.

This process forms the basis for an entire family of products, fundamental to which, in addition to the tripolyphosphate, is the production of the di- and trisodium salts. They may be produced in the same process by crystallizing out and recovering $Na_2HPO_4 \cdot 12H_2O$.

The major use of STPP is in home laundry, industrial, and institutional detergents, as indicated in Table 12.10.

TABLE 12.10 1997 STPP Uses (Thousands of Tons)[20]

Detergents	128.5
Water treatment	19.8
Food and beverage	66.1
Dentifrices	0.9
Metal and plastic finishing	2.4
Other	24.7
Total	242.3

Phosphate use in detergents in the United States has declined from 385,000 tons in 1984 to 128,500 tons in 1997, as a result of the ban on phosphates in home laundry detergents. The switchover to other detergent builders has been completed. It appears that the decline in sodium phosphate demand has bottomed out. Flat future demand is expected.[20]

The mono-, di-, and pyrophosphates are used in foods, the latter two in baking powders. The monosodium salt is produced from Na_2HPO_4 by further processing with H_3PO_4, evaporation, and crystallization. Prolonged heating of the monosodium derivative at 230°C produces the pyrophosphate, $Na_2H_2P_2O_7$. Further heating of the monosalt to the point of fusion produces hexametaphosphate, a glassy substance of uncertain stoichiometry, which is used in water treatment.

12.11 Sodium Silicate

The sodium silicate family is analogous to the phosphates in that a whole series of derivatives are produced in a reaction between Na_2CO_3 and silica by varying the ratio of soda ash to silica. The products often are referred to by the ratio of SiO_2 to Na_2O in a given composition. U.S. demand for sodium silicates was 1.27 million tons in 1998. Production capacity is approximately 1.42 million tons. The market applications for this group of materials are presented in Table 12.11.[22]

TABLE 12.11 1998 Uses for Sodium Silicates[22]

Industry	Percent
Detergents	38
Catalysts	15
Pulp and paper	12
Elastomers	7
Food and health care	5
Coatings	3
Agriculture	3
Other	17

Sodium silicates are made in batches by fusing sand and soda ash in a furnace at about 1300°C, in whatever ratio is necessary to achieve the desired properties of the final compound. In general:

$$\mathrm{Na_2CO_3} + n\mathrm{SiO_2} \rightarrow \mathrm{Na_2O(SiO_2)}_n + \mathrm{CO_2}$$

$$2\mathrm{NaOH} + n\mathrm{SiO_2} \rightarrow \mathrm{Na_2O(SiO_2)}_n + \mathrm{H_2O}$$

The product is called water glass, because, when solid, it actually is a glass; but, unlike lime-soda glass (ordinary window glass), it is soluble in water. The process is carried out in large tank furnaces similar to window glass furnaces. The materials are introduced in batches at intervals, but the products may be drawn off continuously if desired. A mixture of salt cake and coal may replace a portion of the soda ash.

As the melt leaves the furnace, a stream of cold water shatters it into fragments. These are dissolved with superheated steam in tall, narrow steel cylinders with false bottoms,[23] and the product liquor is clarified.[24] Sodium silicates are sold as solutions that vary from the most viscous, 69°Be, to the thinner solutions, down to 22°Be, suitable for use in paints. The composition of typical commercial grades of sodium silicate solutions is shown in Table 12.12. A dry powder also is marketed (Table 12.13); it is made by forcing the thick liquor through a very fine opening into a chamber swept with a rapid current of cold air, which carries off the moisture.[25] Because the solid silicates are hygroscopic, they often are blended with Na_2SO_4 to prevent caking. Silicate solutions also are processed further with mineral acids to manufacture hydrated silica, which yields silica gel upon being dried. Silica gel is used extensively in air-drying equipment and in the manufacture of catalysts.

TABLE 12.12 Composition of Typical Commercial Sodium Silicate Solutions[26]

Weight Ratio SiO_2: Na_2O	% SiO_2	% Na_2O	Solids Content (%)	Specific Gravity (°Be)	Viscosity at 20°C (Centipoises)	Density (Pounds Per Gal)
3.22	28.7	8.9	38.0	41.2	290	11.7
3.00	30.9	10.3	41.2	44.6	412	12.0
2.84	31.9	11.2	43.1	47.0	690	12.3
2.58	32.1	12.5	34.6	NA	780	12.6
2.54	26.9	10.6	37.5	42.0	63	11.7
2.40	33.4	13.9	47.0	52.2	1,700	13.0
2.00	29.4	14.7	44.1	50.0	335	12.7
1.80	24.1	13.4	37.5	54.6	60	13.4

TABLE 12.13 Commercially Available Alkaline Sodium Silicate Solids[26]

Name	Chemical Formula	Molecular Ratio Na_2O:SiO_2
Sodium metasilicate, anhydrous	$Na_2O \cdot SiO_2$	1.0
Sodium metasilicate, pentahydrate	$Na_2O \cdot SiO_2 \cdot 5H_2O$	1.0
Sodium sesquisilicate	$3Na_2O \cdot 2SiO_2$	1.5
Sodium orthosilicate, anhydrous	$2Na_2O \cdot SiO_2$	2.0
Sodium orthosilicate, hydrate	$2Na_2O \cdot SiO_2 \cdot 5H_2O$	2.0

Though there has been some rationalization in the industry, there continues to be excess production capacity for sodium silicates in the United States. The growth in the sodium silicate market is being driven by increasing demand for custom silicas and zeolites, such as precipitated and colloidal silicas and zeolite catalysts.[22] The sharply growing use of peroxide bleaching and waste paper deinking is boosting sodium silicate demand in the pulp and paper industry, particularly in Canada. The silicate sequesters iron and other metal ions that reduce the peroxide efficiency. It also keeps the pH buffered in the alkaline range required for peroxide bleaching of pulp. Sodium silicate also has become an important ingredient of the chemical formulation for the deinking of printed waste paper. In this process, in addition to being a buffering and sequestering agent, silicate is known to provide detergency and anti-redeposition properties during the flotation deinking process.

In 1999, a 41°Be solution with a 3.223.25 SiO_2:Na_2O weight ratio sold for about 10 cents/lb. The price for a similar ratio solid product was about 25 cents/lb.[22]

12.12 Chlor-Alkali (Chlorine and Caustic Soda)

Ranking among the higher-volume chemicals produced in the world are chlorine and caustic (sodium hydroxide). In 1998 North America produced over 30 million tons of chlorine and caustic soda with an estimated value approaching $8 billion.2729 Table 12.14 contains global production data for chlorine and caustic soda.27

TABLE 12.14 1998 Global Chlor-Alkali Production (Thousands of Tons)27

Region	Caustic	Chlorine
United States	14,183	13,547
Canada	1,117	1,089
Mexico	512	462
Brazil	1,577	1,366
Western Europe	11,031	10,219
Eastern Europe	3,637	3,245
Middle East	1,700	1,546
Japan	4,677	4,224
Greater Asia	8,237	7,216
Rest of World	1,470	759
Total	48,141	43,673

Historically, the electrolysis of brine accounts for approximately 96 percent of all chlorine production. About 2 percent of the total electrical power generated in the United States is used for chlor-alkali manufacture. Approximately 30 percent of the total world chlor-alkali capacity is located in the United States in about 50 plants. The greatest concentration of U.S. capacity is along the Gulf Coast of Texas and Louisiana.2830 This concentration derives from the combined availability in that region of salt, hydrocarbons (fuel and organic products based on chlorine and caustic), and deep-water ports for transportation. The intercoastal canal, the Mississippi and Ohio rivers, and their connecting waterways are also well suited for heavy barge traffic from the Gulf to much of the heartland.

Table 12.15 gives the distribution of U.S. chlorine by industry use. Table 12.16 similarly lists the market distribution for caustic.

TABLE 12.15 1998 U.S. Chlorine Markets28

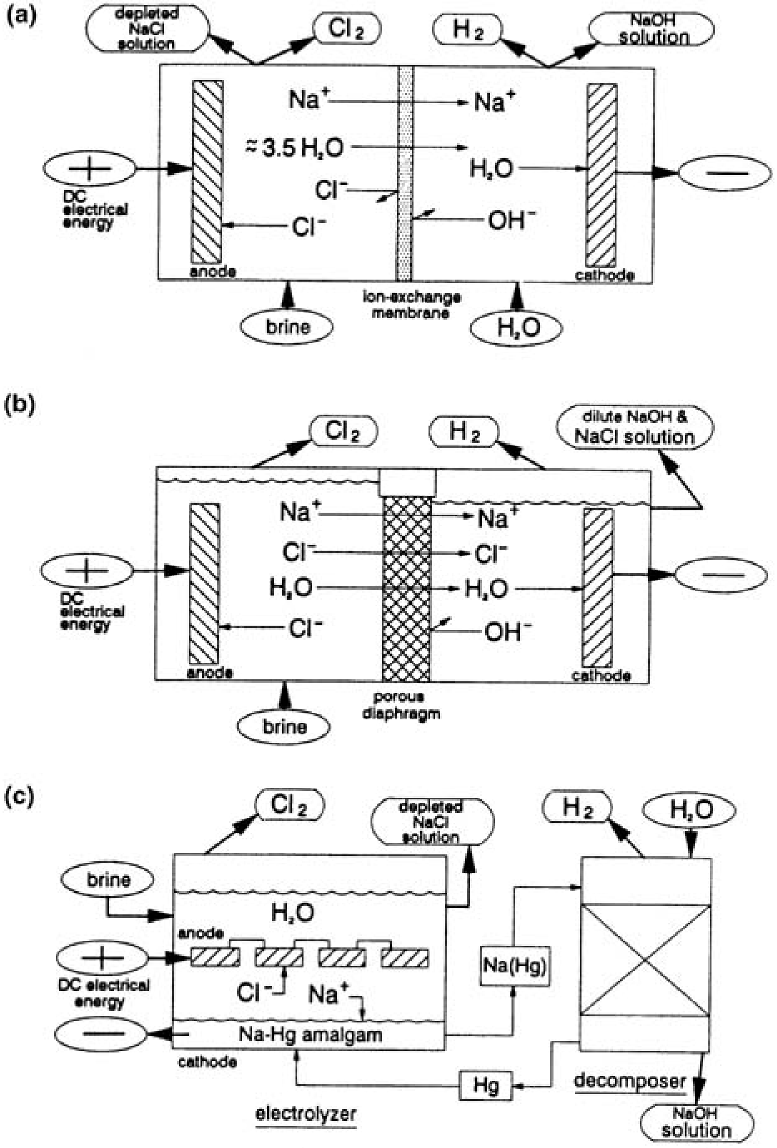
Market	Percent
Ethylene dichloride/vinyl chloride monomer	37
Other organic chemicals	25
Inorganic chemicals	17
Water treatment	6
Pulp and paper	5
Miscellaneous	10

TABLE 12.16 1998 U.S. Caustic Soda Market29

Market	Percent
Pulp and paper	13
Soaps and detergents	5
Alumina	3
Petroleum	4
Water treatment	3
Textiles	3
Organic chemicals	35
Inorganic chemicals	11
Miscellaneous	23

The heart of the chlor-alkali process is the electrolytic cell in which saturated, purified NaCl is electrolytically decomposed. There are basically three types of chlor-alkali cells currently in use, which are shown schematically in Fig. 12.5.3133 These cells are designed for continuous electrolysis of brine to produce chlorine gas, hydrogen gas, and aqueous caustic soda.

Fig. 12.5 Chlor-alkali electrolysis cells: (a) diaphragm cell; (b) membrane cell; (c) mercury cell. (*Courtesy the Dow Chemical Company*)31



The diaphragm cell derives its name from the separator, or diaphragm, which is a mat of inorganic fibers vacuum-deposited directly on a woven-wire, mild steel member. The steel mesh serves doubly as a support for the diaphragm and as the cathode, or negative electrode, of the cell. The diaphragm separates the anode and cathode compartments of the cell to discourage direct mixing of the chlorine and caustic, and also to prevent chlorine and hydrogen from mixing, which can be a dangerous situation because they are capable of reacting violently with one another under certain conditions. A head of brine is maintained higher on the anode (the positive electrode, at which chlorine is produced) side of the diaphragm than the cell liquor level on the cathodic side. Thus brine slowly percolates through the diaphragm and into the cathode chamber of the cell, where water is decomposed to form hydrogen gas and hydroxide ion. The effluent cell liquor contains about 12 percent NaOH, 14 percent NaCl, and small amounts of chlorate, hypochlorite, sulfate, and carbonatewith the remainder being water. The usual caustic produced for sale is a 50 percent NaOH solution; so this cell effluent requires evaporation and separation of the salt.

The mercury cell differs from the diaphragm cell in that the cathode is a thin film of mercury, with no separator between it and the anode. Although chlorine is produced at the anode in both types of cell, the cathode reactions differ markedly. Instead of water decomposing at the cathode of the mercury cell, as it does in the diaphragm cell, sodium metal is deposited in the mercury to form a dilute (0.5%) amalgam. This amalgam flows out of the cell into a decomposer, in which water and the amalgam react directly to form a 50 percent NaOH solution of high purity, there being no evaporation required. The denuded mercury is returned to the cell.

Another difference between the two cell types is the brine flow. In the diaphragm cell about 50 percent of the NaCl is decomposed, and the remainder flows out of the cell in the weak caustic effluent. In the mercury cell, only a small fraction of the salt is decomposed per pass through the cell. The brine is immediately dechlorinated, resaturated with salt, and returned to the cell.

At first glance, it would appear that the mercury cell process is far superior to the diaphragm cell process. In many ways it is. However, the mercury cell requires much more electrical energy than does the diaphragm cell. In certain areas, especially where electrical energy is cheap and where small-capacity, easily run plants are required, the mercury cell is quite economical. Also, the mercury cell may be preferred if fuel costs are high, as that discourages the plant designer from installing evaporatorswhich require burning fuel just to make steam to run them. In large installations where it makes more sense to install a powerhouse at the site, the generation of power cogenerates lower-pressure steam, which may be used quite economically to drive a multiple-effect evaporator.

The membrane cell has some operating features reminiscent of both the diaphragm and the mercury cell. This new cell is somewhat like a diaphragm cell, in that it has a separator. However, the similarity ends there. The separator in the newer type cell is a perfluorinated ion-exchange membrane, through which essentially only hydrated sodium ions migrate under the influence of the electrical field.34,35 Because there is no fluid flow through the membrane, the brine in the anode compartment becomes depleted in salt. As it does so, it is removed, dechlorinated, resaturated, and returned to the cell, just as in the case of the mercury cell. Also, very high-quality caustic is made in this cell, in concentrations ranging from 3035 wt percent NaOH with less than 50 ppm NaCl. These solutions are more economically evaporated than is the diaphragm cell effluent, because they are more concentrated and contain no salt, sodium sulfate, or sodium chlorate impurity to speak of.36 This technology has emerged in the last 20 years and in time will largely displace other cells.

As mentioned earlier, the ion-exchange membrane allows the transport of primarily sodium ions and water from the anolyte to the catholyte compartments, whereas the diaphragm is designed to allow percolation of all the anolyte through the separator. The cation-conducting, ion-exchange membrane is designed to reject anions, as indicated in Fig. 12.6. The chlor-alkali membranes in use today consist of one or more perfluorinated, ion-exchange polymeric materials, as illustrated in Fig. 12.7. The short-chain polymers shown in Fig. 12.7 have been developed by the Dow, Asahi Glass, and Asahi Chemical companies.31 Long-chain polymers are manufactured by Du Pont and sold under the trade name Nafion.*

Fig. 12.6 Chlor-alkali membrane cross-section.31

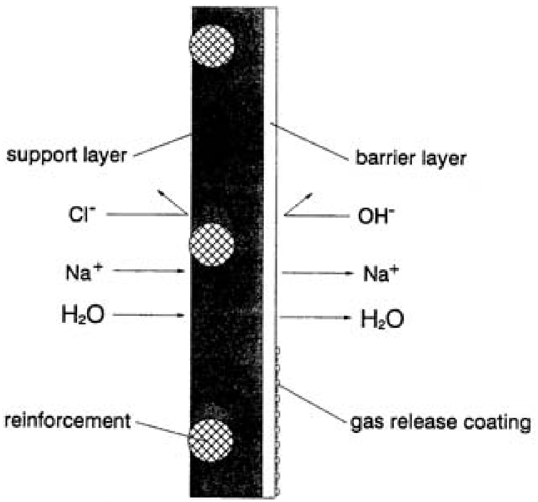
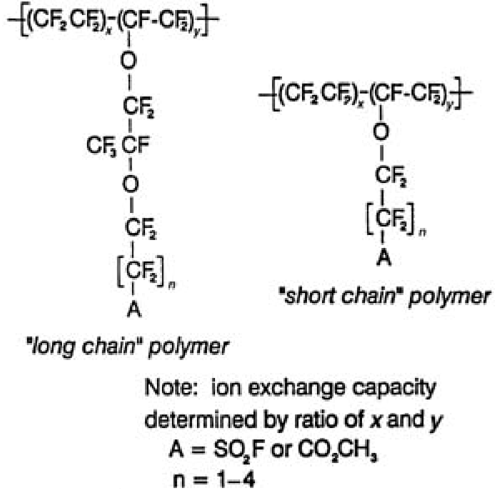


Fig. 12.7 Perfluorinated polymers for chlor-alkali membranes.



*Trademark of E. I. du Pont de Nemours & Company.

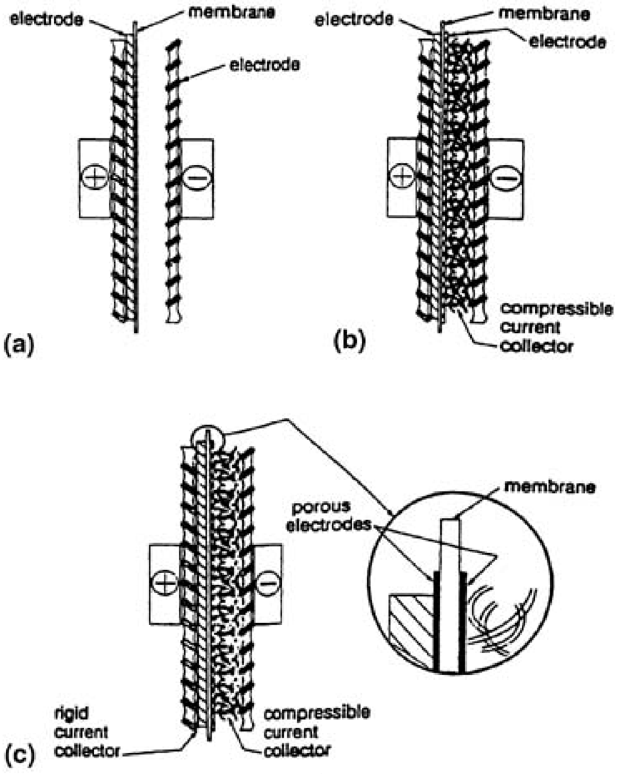
A chlor-alkali membrane, designed to produce 3035 percent NaOH, consists of at least two distinctly different polymer layers, as shown in Fig. 12.6. The anode side of the membrane is about a 0.1 mm film of sulfonic acid polymer, whereas the cathode side is about a 0.05 mm film of carboxylic acid polymer. A Teflon* fabric net is embedded in the polymer film to provide mechanical strength to the membrane. In some cases, the membrane surface is modified to facilitate gas bubble release.

A most significant advance in cell anode technology is being used to great advantage in all three cell types to conserve energy, reduce impurities, and lengthen cell life.37,38 Prior to this development the cell anodes were made out of graphite. The new anodes are made out of titanium, which has excellent corrosion resistance in hot, chlorinated brine. However, the very oxide film that lends corrosion resistance is not an electrical conductor. By applying electroactive coatings, largely composed of mixed oxides of ruthenium and titanium, to the surface of the titanium, it remains conductive. These anodes give a vast improvement in performance over graphite for two reasons: (1) they do not wear away like graphite, which, as it becomes increasingly thinner, widens the electrode gap in the cell, requiring higher and higher voltages to operate the cell; (2) the electroactive coating is actually a catalyst that allows chlorine to be formed on the anode surface with essentially zero activation energy. Both properties contribute much to reduced power consumption in running the cells. In the case of the mercury cell, there is the added advantage that the dimensional stability of the new anode precludes the continual adjustment of electrode spacing, which is constantly necessary with graphite electrodes.

The ion-exchange membrane and dimensionally stable anode developments have impacted greatly on chlor-alkali operations. Their impact has been accelerated by both environmental and energy cost considerations. The cell capital cost is skyrocketing because of the expense of these new components; however, their usage does lower operating costs. In practice, a balance between capital costs and operating costs is sought to achieve an economic optimum for a given plant site.

Both membrane and diaphragm cells use punched plate, expanded mesh, or woven wire cathodes of nickel, nickel-plated steel, or stainless steel. In most modern membrane cells some type of activated or catalytic cathode is used. Usually the gap between the anode and cathode is minimized to reduce the electrical resistance of the cell.31 Various cell configurations are shown in Fig. 12.8. A cell designed for zero gap operation usually has a compressible mesh to conduct the current. The SPE (Solid Polymer Electrolyte) concept shown in Fig. 12.8 is not commercial yet, but the concept is borrowed from fuel cell technology where the porous electrode is embedded into the membrane.

Fig. 12.8 Chlor-alkali cell configurations: (a) gap cell; (b) zero gap cell; (c) SPE® cell. ® Oronzio Benova co.



The chemistry at the anode is identical for all three cells:

$$2\mathrm{Cl}^- \rightarrow 2\mathrm{e}^- + \mathrm{Cl}_2$$

The cathode reactions also are the same for the diaphragm and membrane cells:

$$2\mathrm{H}_2\mathrm{O} + 2\mathrm{e}^- \rightarrow \mathrm{H}_2 + 2\mathrm{OH}^-$$

In the mercury cell, however, the cathode reaction is quite different, there being no hydrogen formed in the cell:

$$\mathrm{Na}^+ + \mathrm{e}^- \xrightarrow{(\mathrm{Hg})} \mathrm{Na}\ (\text{amalgam})$$

The amalgam may be used as a chemical reductant (i.e., to reduce NaHSO3 to the hydrosulfite19 or to hydrodimerize acrylonitrile to adiponitrile39), or decomposed with water to make caustic, as is generally done.

Environmental pressures on mercury discharges have dealt a blow to the practice of mercury cell technology. As a result, there have been many plant closings. An excellent job has been done in achieving substantial reductions of Hg losses, and this technology will continue to be practiced. However, all future capacity expansions will use the membrane process because it has the lowest investment and operating costs with product quality approaching that of the mercury process. Japan has switched over 95 percent of its chlorine capacity to the membrane process.27

The theoretical decomposition potential, $E°$, of a cell (i.e., the voltage below which it is impossible to form any products at all) is strictly dependent on the nature of the electrode reactions. It is related to the Gibbs free energy of the net overall chemical change occurring in the cell: $G = nFE°$, where n is the number of electrons involved in the electrode reaction, and F is the Faraday constant.40 $E°$ is the thermodynamic potential, to which the IR drops (voltage necessary to drive the current through the electrical resistances in the cell), and the overpotential (voltage associated with the kinetics of the reactions) must be added to get the actual cell voltage:

$$\begin{aligned}\text{cell voltage} \quad &= \quad \text{thermodynamic potential} \\ &\quad + IR \text{ drop} + \text{overvoltage}\end{aligned}$$

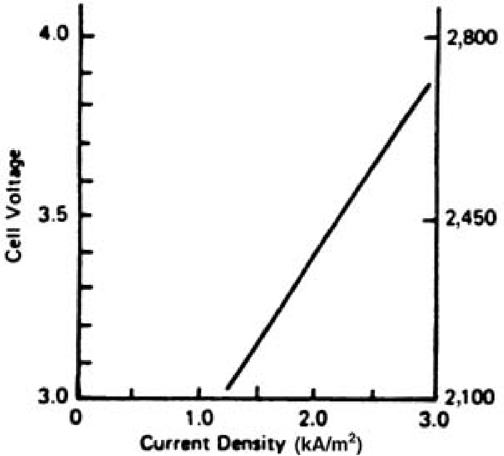
Typical voltage components for a membrane cell are given in Table 12.17.

TABLE 12.17 Typical Voltage Components for a Membrane Cell at 3.1 kA/m2, 85°C, 34 Weight % NaOH, DSA Anode, 200 g/L NaCl Anolyte and Activated Cathode41

Component	Voltage	% of Total
Reversible potential of cell reactions	2.20	72
Membrane *IR* + concentration polarization	0.50	16
Electrode overpotentials	0.15	5
Solution *IR*	0.15	5
Structural *IR*	0.06	2
	3.06	100

Using a modern diaphragm cell running at 1500 A/m2 as an example, the percentages of the total cell voltage attributable to these three components are: 65 percent thermodynamic, 25 percent *IR*, and 10 percent overvoltage. Cell design and development programs continue to search for better ways to reduce energy consumption. Thus, higher cell capital becomes justifiable on the basis of performance. But as cell capital increases, it tends to force operation at increased current density in order to produce more pounds per cell per unit of time. In a practical sense, this translates to a compromise between cell cost and current density in order to minimize production cost. Figure 12.9 is a typical currentvoltage curve for a diaphragm cell. The current density simply refers to total current divided by the nominal electrode area of either the anode or the cathode, often expressed as thousands of amperes (kA) per square meter. Figure 12.9 also shows the electrical energy consumption in kilowatt hours per 2000 lb of chlorine (kWhr/T Cl2).

Fig. 12.9 Currentvoltage curve for a typical diaphragm chlor-alkali plant.



The minimum energy requirement, unattainable in practice, is 1543 kWhr/T Cl2. This corresponds to 100 percent current efficiency (685.8 kAh/ton Cl2) and 100 percent voltage efficiency (i.e., no *IR* drops and zero overvoltage at both electrodes, so that the voltage is the thermodynamic potential, or 2.17 V). The energy efficiency of a typical diaphragm cell is easily expressed as:

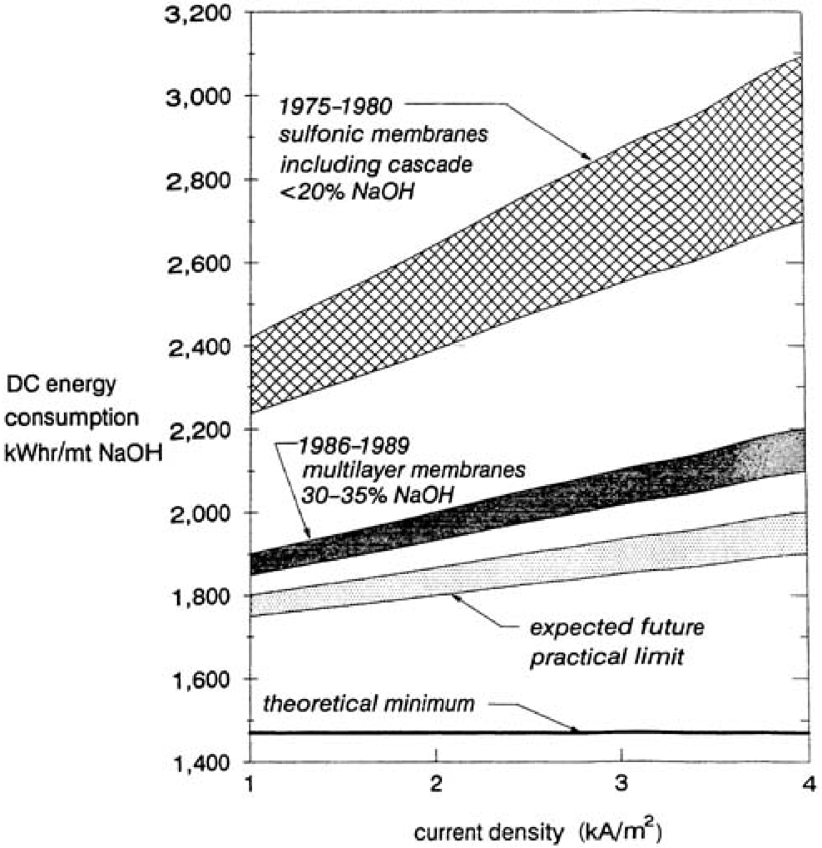
$$\frac{(2.17) \times (\%\ \text{current efficiency})}{(\text{cell voltage})}$$

For example, if a cell running at 80,000 A at a current density of 2.2 kA/m2 and a current efficiency of 96 percent requires a cell voltage of 3.40, the overall energy efficiency is:

$$\frac{2.17 \times 96}{3.40} = 61.3\%$$

Figure 12.10 shows the dramatic improvements made in the electrical energy consumption of membrane cells since the mid-1970s.31 Formerly, it was believed that the membrane performance limited stable operation to current densities below 4 kA/m2. Recent work has shown that today's membranes can work efficiently up to 10 kA/m2.42 However, the membrane is much more sensitive to its environment at ultra-high current density. Extremely meticulous attention to brine quality is one key operating variable allowing long-term stable operation at 10 kA/m2. The other is a hardware design with good electrolyte circulation to keep the concentration and temperature uniform throughout the cell.43 Stable power costs and the high capital investment for membrane cells make increased current density an attractive option.

Fig. 12.10 Electrical energy consumption improvements in membrane cells.31



Much of the inefficiency results in heat generation; a cell will heat itself almost to boiling when fed brine at about 40°C. The self-heating tends to lower the cell voltage because the electrical resistance of the cell decreases with increasing temperature. As modern rectification equipment (transformers, controllers, and rectifiers) is very efficient, the conversion of AC electricity into electrochemical products is essentially dependent on the cell efficiency. Typically, AC-kWhr/lb product = 1.03 × (DC-kWhr/lb product).

Figure 12.11 is a schematic of a commercial diaphragm cell, typically operated in the range of 75,000150,000 A. Many such cells are connected in series, as shown in Fig. 12.12, so that the current leaving at the anode of one cell enters the cathode of the next cell, with the anode terminal cell connected to the positive bus of the rectifier, and the cathode terminal cell connected to the negative bus of the rectifier. Bipolar and monopolar cell circuit layouts are shown in Fig. 12.13.31

Fig. 12.11 Cut view of ELTECH's H-4 diaphragm cell operating at a nominal load of 150 kA. (*Courtesy Eltech Systems Corporation*)

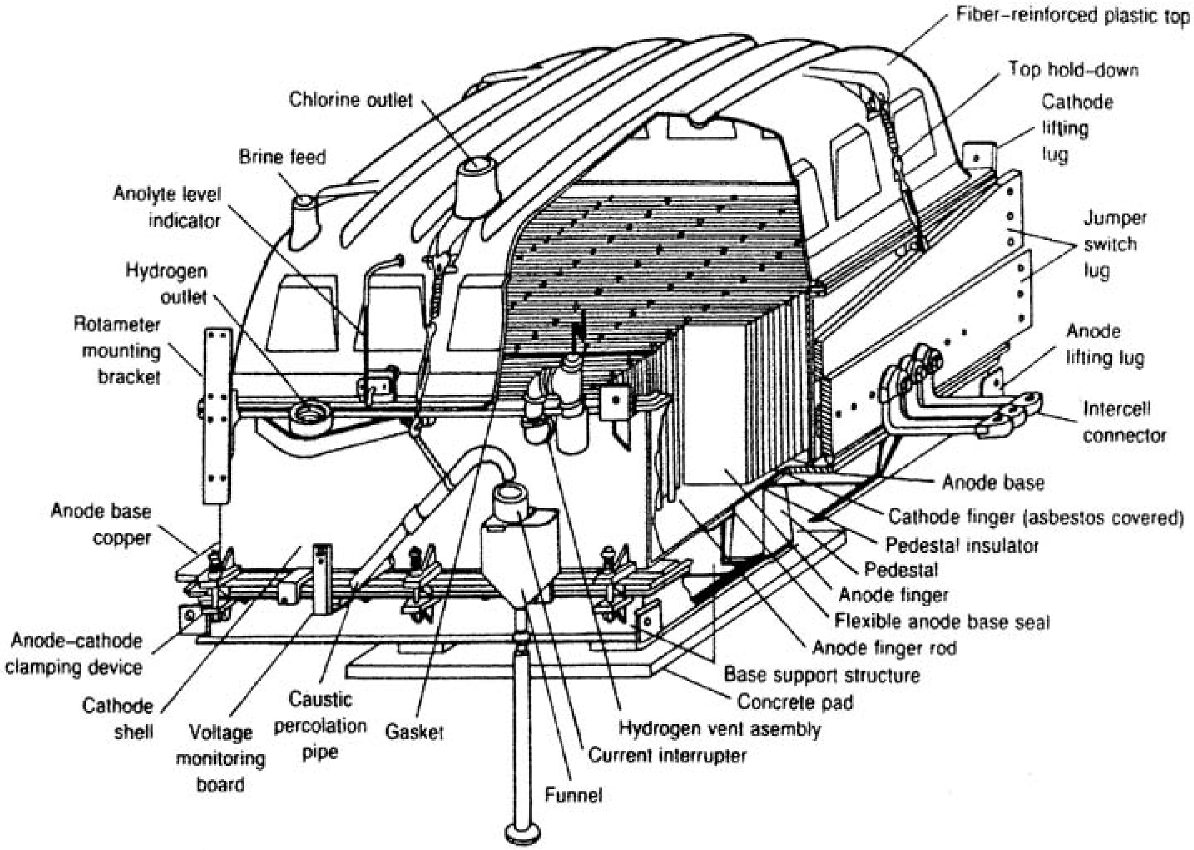


Fig. 12.12 An example of a membrane chlor-alkali cell commercial installation. (*Photo courtesy Eltech Systems Corporation.*)

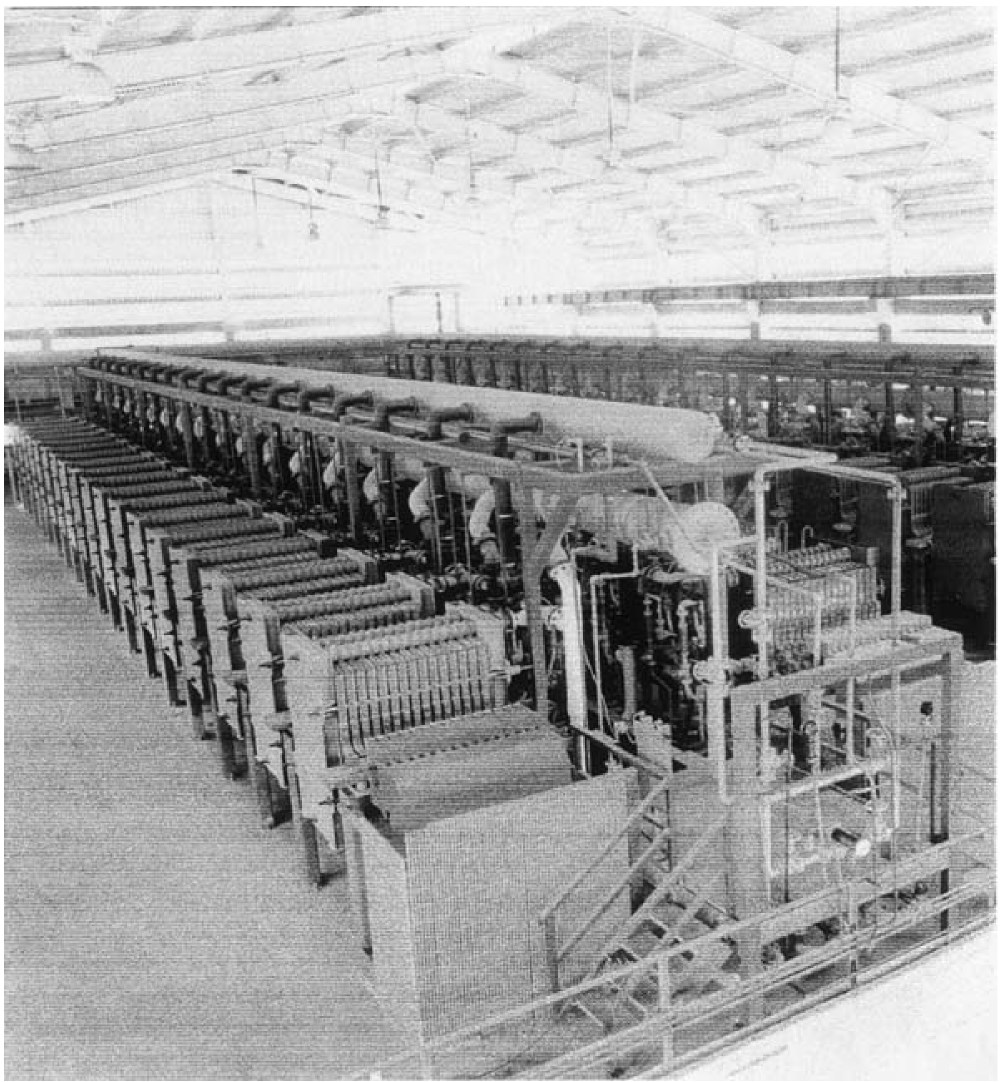
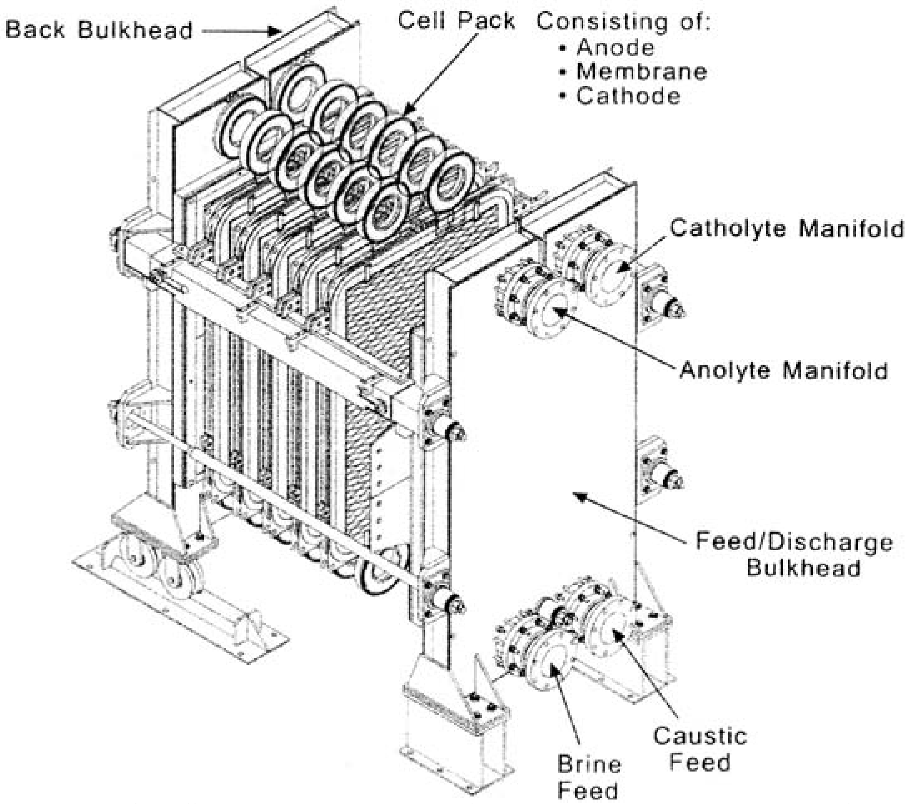


Fig. 12.13 Expanded view of Eltech ExL bipolar membrane cell electrolyzer. (*Courtesy Eltech Systems Corp. ExL is a trademark of Eltech Systems Corp.*)

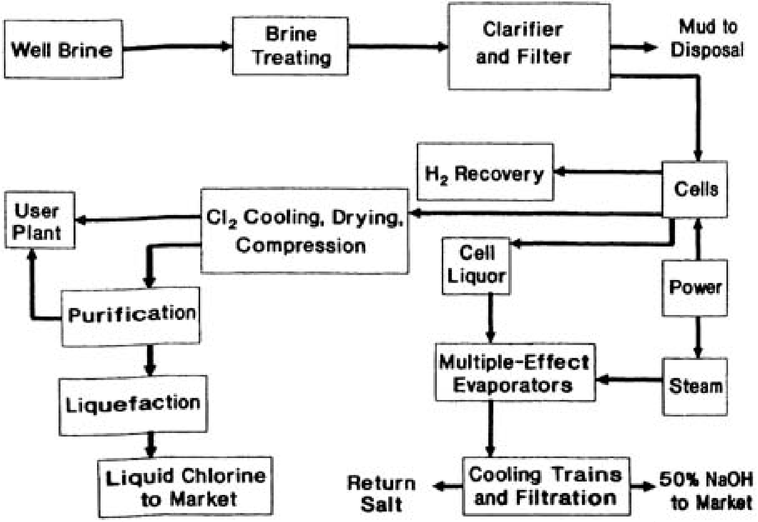


With allowance for inefficiencies, each 30 kA increment of current that passes through the series circuit of cells produces 2000 lb of chlorine, 2100 lb of caustic, and 57 lb of hydrogen, per cell, in 24 hr of operation.

If a production plant is required to produce 900 tons of chlorine per day using rectification equipment capable of sustained operation at 90,000 A, the plant must operate a minimum of 300 cells. If the cell voltage is 3.4 V at 90 kA, rectification equivalent to 1020 V is required. In practice, the logistics of production requirements, plant reliability, and maintenance outages will dictate the size (i.e., number of cells and nominal amperage) of each circuit. In the example above, the plant would use a minimum of two circuits, with two nominal 90-kA, 600 V rectifiers.

Figure 12.14 is a simplified flow sheet for a typical diaphragm cell chlor-alkali plant. The process begins with brine treating, in which the well brine is saturated with return salt from the caustic evaporators, then treated with hypochlorite to destroy ammonia and amines, and also treated with carb-caustic liquor to remove calcium and magnesium impurities. Membrane cells require a secondary purification of the brine feed to provide parts per billion levels of Ca, Mg, and other heavy metals. Alkaline earth and heavy metal ions deposit in the membrane, raise its electrical resistance and decrease its current efficiency. This process is not completely reversible even with severe acid treatments. A single exposure to excess alkaline earth metal ions will affect a membrane's performance for the remainder of its operating life.

Fig. 12.14 Simplified flow sheet for chlor-alkali plant.



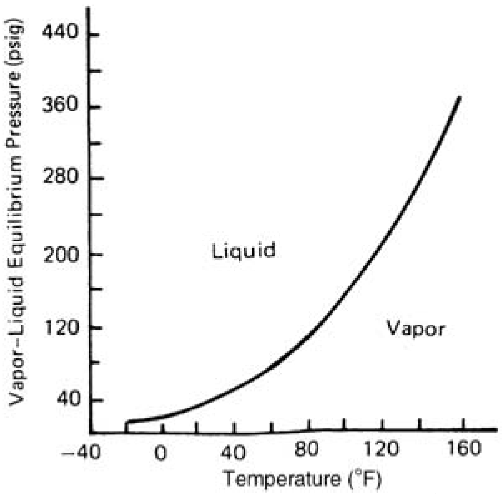
The removal of trace ammonia-nitrogen impurities is essential to avoid buildup of NCl_3 in the chlorine purification and liquefaction stages. The NCl_3 is unstable in concentrations above a few thousand parts per million, and can decompose violently. In the diaphragm cell process, the Ca and Mg impurities are removed to reduce plugging of the diaphragm, and to lengthen the time between diaphragm washing or diaphragm replacement. This was typically done every few months, but has now been extended to over a year, principally by replacement of the graphite anodes with the newer metal anodes. In the older cells, the slowly oxidizing graphite formed a sludge that tended to plug the diaphragm.

The hydrogen from the cells formerly was discarded (vented to the atmosphere) unless it was used in some of the larger plants to make NH_3 and other chemicals. It takes a large plant to produce a useful amount of H_2 because the cell produces only 0.028 lb H_2 per pound Cl_2. Also, the fuel value of H_2 is only a third that of methane. But since the energy crisis of the 1970s, the fuel value of the hydrogen has become sufficient to recover it for boiler fuel. This requires cooling to remove water, scrubbing to remove traces of chlorine, caustic, and salt, and then compressing sufficiently to satisfy the feed requirements of the burner. Of increasing importance is the burning of the hydrogen with the dilute chlorine stream known as tail gas from the liquid chlorine plant. Some chlorine producers sell their hydrogen to merchant hydrogen producers who further purify and liquefy the chlor-alkali electrolytic hydrogen for metallurgical use and rocket fuel.

The chlorine from the cell is saturated with water at the cell temperature, and contains lesser amounts of NCl_3, organic chlorides, and salt from brine entrainment. The gas is cooled to knock out the water, and this chlorinated water is returned to the brine plant. Then the cooled gas is dried by passage through towers of sulfuric acid. Strong (98%) acid is fed to one end of the drying train at a rate equivalent to the removal of weak (65%) acid from the opposite end. The waste acid often is used to neutralize alkaline wastes, and the reconcentration and recycling of this acid presents a challenge to a chlor-alkali manufacturer.

The moisture content of the chlorine at this stage usually is low enough to allow the use of steel in the remainder of the process. The gas is compressed, usually to whatever pressure is required by the user plant, or to sufficient pressure to liquefy it, the exact pressure depending upon the refrigeration capacity and the liquid chlorine storage requirements of a given plant. Figure 12.15 is the liquefaction curve for chlorine.[44]

Fig. 12.15 Liquefaction curve for chlorine.



In nothern climates, where ambient temperatures can fall below those required to liquefy chlorine at rather modest pressures, it is necessary to provide means for maintaining the temperatures in chlorine-gas distribution lines above the liquefaction point. This can be done by tracing or wrapping the lines with heating tape, using steam lines, and insulating the pipes. These precautions are especially important in organic process plants, which can run without incident while using gaseous chlorine but could explode if fed liquid chlorine.

Organic impurities and trace amounts of NCl_3 may be removed from gaseous chlorine in a simple extraction in which liquid chlorine is used as the extractant. The liquid bottoms flow from the extractor to a still in which chlorine is flashed off to recycle and the organic fraction (largely low molecular weight organic chlorides) is recovered for disposal by combustion to CO_2 and HCl, which is recovered for use.

There are many options in the liquefaction of chlorine, depending upon the storage requirements for the liquid and the system used to handle the tail gas. Low-pressure storage is recommended because of the reduced hazard in the event of a spill. However, this requires lower storage temperatures. Low temperatures also reduce the amount of chlorine in the tail gas from liquefaction. Yet lower temperature may not be necessary if a plant also deliberately produces bleach or HCl from the tail gas.

The components in the tail gas originate in the cells. At the anode, some water is oxidized to produce small amounts of oxygen. To this amount is added more oxygen from air, which may leak into the chlorine header from outside because the header is operated under a slight vacuum to prevent gas from escaping to the atmosphere. The cell gas typically contains about 97 percent Cl_2. Because the remaining 3 percent is noncondensable under the conditions in the liquid chlorine plant, it remains gaseous at the tail end of the liquefaction process, saturated with chlorine in an amount determined by the vapor pressure in that stage of the process. When air is deliberately added to the system to reduce the hydrogen content of tail gas to below 5 percent, the tail gas often increases in volume. With hydrogen above that level, there is a great risk of explosion. The hydrogen comes from the cathode compartment of the cell, which is operated under a slight pressure to avoid sucking air into the hydrogen collection system. Ordinarily, this leads to quite small amounts of hydrogen in the cell chlorine gas. But when diaphragm problems arise or pinholes develop in the membrane, the level can rise to undesirable values. In good plant practice, this is monitored continuously.

The cell liquor flowing from the cathode compartment of each cell is collected for evaporation to 50 percent product caustic. The diaphragm cell liquor contains 1012 percent NaOH, 1416 percent NaCl, 0.10.5 percent Na_2SO_4, 0.020.04 percent Na_2CO_3, and 0.010.2 percent $NaClO_3$. The evaporation is carried out in reverse feed nickel-lined multi-effect systems (i.e. feed liquor enters the opposite end of the train with respect to the primary steam entry). With rapidly rising fuel costs, it has become economically justifiable in the construction of new plants to install quadruple-effect evaporators. Previously triple-effect systems were installed. The salt content of membrane cell effluent is so low that only a simple flash evaporator is required. This represents a considerable reduction in investment versus diaphragm caustic plants.

Because of the high salt content of the feed liquor originating from diaphragm cells, the evaporation system is really an evaporative-crystallizer. Salt is removed from each effect in the evaporaation train. Hydroclones, centrifuges, settlers, and filters are used extensively to remove the salt as efficiently as possible and return it to the brine plant with as little caustic loss as possible. In this regard, good plant practice strikes an economic balance between caustic loss and the addition of wash water to the system, which requires additional heat input.

The sulfate content of caustic evaporator feed can cause a problem during evaporation. When the caustic concentration reaches about the 32 percent level, the triple salt $NaCl \cdot Na_2SO_4 \cdot NaOH$ crystallizes out along with NaCl crystals. To avoid sulfate buildup in the system, which would lead to excessive caustic loss, the sulfate must be purged from the system. If it is recycled with the salt, its concentration will rise to unacceptable levels. Various schemes are used, but all plants purge sulfate at some point.

The hot 50 percent caustic from the evaporator train is cooled and filtered to remove salt down to the solubility limit. Typical analysis of diaphragm cell product caustic, on a 50 percent NaOH basis (actual caustic content varies from 48.5 to 50.5%) is: 1.0 percent NaCl, 0.15 percent Na_2CO_3, 0.025 percent Na_2SO_4, 0.15 percent $NaClO_3$, 5 ppm Fe, and a maximum of 30 ppm combined heavy metals (Sb, As, Bi, Cd, Cu, Pb, Hg, Ag, and Sn). Most often the billing weight of a shipment of caustic is based on its Na_2O content, equivalent to 76 percent Na_2O:

$$\frac{(\text{pounds solution}) \times (\%\,\mathrm{Na_2O})}{76}$$
$$= \text{net pounds NaOH } (\mathrm{Na_2O}\text{ basis})$$

Actually, pure NaOH contains 77.5 percent Na_2O. The value 76 percent is derived from tradition, based on the fact that many decades ago the purest form of commercially available caustic contained 76 percent Na_2O. In isolated cases, where the market is far removed from the source, the caustic is further concentrated and shipped hot as 73 percent NaOH. This saves on freight costs, but the solution must be kept hot because it freezes at about 160°F. Similarly, 50 percent caustic must be maintained above 60° F.

In selected markets, such as the rayon industry, low salt levels are required (typically 100 ppm). Mercury cell and membrane cell caustic are ideally suited for these markets. Diaphragm cell caustic is purified further in a few installations, but primarily to remove $NaClO_3$, which is corrosive to the equipment used to manufacture anhydrous caustic. Chlorate alone may be removed by reaction with a reducing agent, such as sugar, which is injected into the hot caustic feed to the anhydrous concentrator system. This has the disadvantage of increasing both the salt and the carbonate levels in the caustic.

To remove chlorate and other salts, the 50 percent diaphragm cell caustic may be extracted with liquid ammonia in a pressurized system.45 The ammonia fraction then is processed through a stripper to remove and recycle the ammonia. The alkaline stripper bottoms are useful in neutralizing acidic waste streams. The purified caustic is evaporated further, and then fed to an anhydrous concentrator, typically an Inconel falling film evaporator, heated with molten salt. The anhydrous caustic, containing at least 97.5 percent NaOH, is marketed as one solid mass in drums, as flake caustic, or more desirably as beads or prills. The last are marketed in bulk or bags.

Since the binary system H_2ONaOH contains several hydrates that may be conveniently crystallized, another possibility exists for purifying caustic. Of special interest are the dihydrate, $NaOH \cdot 2H_2O$, which melts at about 55°F, and $NaOH \cdot 3.5H_2O$, which melts at 60°F. The impurities concentrate in the mother liquor; purified caustic then is produced by melting the crystals. The most recent attention to this method of purification has been in Japan.46 These processes are more difficult than they appear, because of the high viscosity of concentrated caustic solutions near their freezing points and the resultant challenges of both adequately transferring heat and separating the crystals from the mother liquor. But with the advent of the membrane cell, there may be more development activity in this area.

In the handling and storage of 50 percent caustic, it is important to limit the amount of iron pickup. The iron spec is 5 ppm on a solution basis. Stainless steel offers little advantage over mild steel with respect to iron contamination. Furthermore, hot, salt-containing caustic leads to severe stress corrosion cracking of stainless steel. Carbon steel storage tanks often are lined with an epoxy-phenolic or neoprene latex coating. Iron pickup from transfer piping increases with increasing temperature and increasing linear velocity of pumping. Unnecessary recirculation or pumping through steel lines should be avoided.47 Iron, copper, and nickel impurities are especially undesirable for the bleach manufacture.48

About 75 percent of the caustic produced is concentrated. The remainder is used directly as alkaline cell liquoras, for example, in the conversion of propylene to propylene oxide by the chlorhydrin process. Similarly, there is some chlorine produced by methods that do not produce caustic, as shown in Table 12.18. Fused salt electrolysis of metal chlorides, most prominently sodium and magnesium, accounts for approximately 0.6 percent of U.S. chlorine production.27 Chlorine is a by-product of the reaction of potassium chloride and nitric acid to produce potassium nitrate. This accounts for less than 0.5 percent of chlorine production.27 Electrolysis of HCl produces chlorine and hydrogen. Several cell designs are available commercially for effecting this transformation. Chemical oxidation of HCl aided by catalysts such as nitrogen oxides, manganese dioxide, and copper oxychloride. The most recent commercially practiced approach of this type, the Kel-Chlor process,49 used nitric acid to oxidize HCl through intermediates NOCl and $HNSO_5$. The U.S. DuPont plant using this process has been shuttered. Recently, DuPont announced a new process discovery based on a copper chloride catalyst.27 Another developmental chlorine recovery process, the MT-Chlor process, is currently in commercial demonstration in Japan by Mitsui Chemicals. Anhydrous HCl is oxidized by a chromium-based catalyst in a gas phase fluidized bed.27

TABLE 12.18 U.S. Chlorine Capacity by ProcessJanuary 199927

Process	Percent
Diaphragm cells	62
Membrane cells	18
Mercury cells	16
Other	4

A small, but important, variation of NaCl electrolysis substitutes KCl as the feed. Both mercury cells and diaphragm cells are used, producing chlorine and KOH (caustic potash). The KOH is concentrated for sale as a 45percent solution or in solid form, containing 88 to 92 percent KOH. A big use for KOH is in the manufacture of liquid soaps and detergents. The analogous sodium soaps and detergents are generally solids. Approximately 1.5 percent of chlorine production results from electrolysis of KCl.27

Table 12.19 gives 2000 U.S. list prices for chlorine and various forms of caustic soda.28,50 Much chlor-alkali product is sold below list prices via long-term contracts and large volume spot purchases.

TABLE 12.19 2000 List Prices of Chlor-Alkali Products28,50

Product	$/ton, 100% Basis
Liquid chlorine	$250
50% caustic	$300
50% caustic (rayon grade)	$325
73% caustic	$350400
Caustic beads	$620

The entire chlor-alkali world must constantly perform a balancing act. The electrolysis of salt brine yields almost equal amounts of chlorine and caustic. It is rare that the growth or use-rates for the two products are equal. Furthermore, the chlor-alkali balance problem extends to related products such as soda ash, lime, salt cake, and hydrochloric acid, even to the extent of influencing (and being influenced by) the current economics of chlorinated hydrocarbons.

For example, in the 1960s the demand for chlorine was so much greater than for caustic that concerted efforts had to be made to move caustic into new markets, such as the glass industry. Another effect of the same economic climate was a glut of HCl from the tremendous growth in vinyl chloride and a host of other chlorinated organic derivatives. Necessity being the mother of invention, this situation led to the development of the oxy-chlorination process for the production of vinyl chloride, resulting in no net production of HCl; to a route for producing methyl chloride using HCl and methanol; to the Kel-Chlor process for production of chlorine without caustic; and to the recycle of byproduct HCl to chlorine by electrolysis.5153 There was a host of developmental activity in both the United States and Europe to convert caustic to soda ash and to improve the old Deacon process, which is a nonelectrolytic, catalytic conversion of HCl to chlorine.54 In another case, a large consumer of sulfuric acid, the steel industry, was converted to the use of HCl for pickling steel. Huge amounts of caustic cell effluent were used to replace lime in a variety of applications.

By the mid-1970s, the picture was changing. Environmental pressures on fluorocarbon derivatives (derived from chlorinated compounds and creating by-product HCl), and on some of the chlorinated solvents, impacted negatively on chlorine growth. During the late 1980s, the pulp and paper industry started to come under intense environmental scrutiny due to the presence of total organic halides (AOX) in effluents from bleaching operations in kraft mills. As a result the industry voluntarily reduced chlorine usage by 5070 percent in the early 1990s and used chlorine dioxide in its place. This downward trend continued as the U.S. EPA mandates effectively outlawed elemental chlorine pulp bleaching in April 2001. In the mid-1980s, the pulp and paper industry's peak consumption was 2 million tons of chlorine.28

In the latter half of the 1990s and into the new century, chlorine and caustic prices continued their dramatic and seemingly unpredictable fluctuations, as supply and demand for each of the co-products swung out of phase. From 1994 to 1999 chlorine fluctuated between $160 and $255/ton; 50% caustic varied between $270 and $330/ton on a 100 percent basis.28,29 Caustic prices fell as low as $65/ton in April 2000.29 The spot price for chlorine was $15/ton in December 2001.55 In 2000, there were 23 chlorine producers with 45 plants in the United States. The top five producers account for 77 percent of U.S. capacity, as indicated in Table 12.20.28

TABLE 12.20 Major U.S. Chlorine Producers200028

Company	Capacity (Thousands of Tons Company per Year)
Dow chemical	4,590
Oxychem	3,525
PPG	1,695
Olin	1,025
Formosa	990
18 Other companies	3,593
Total U.S. capacity	15,418

The most consistent outlet for chlorine will continue to be vinyl chloride, and future growth is tied to its fortunes. With its wide variety of uses, caustic soda demand generally reflects the health of the entire U.S. economy, though the most recent low in caustic pricing came during an economic expansion. The unpredictable financial returns attributable to the co-products will make it very difficult for producers to manage the large capital outlays required as aging diaphragm plants are replaced by membrane facilities.28,29

12.13 Hydrochloric Acid

The predominant source of HCl is as a by-product from the chlorination of hydrocarbons, including the incineration of chlorination organic wastes in thermal oxidizers. Hydrogen and chlorine can be burned together to produce HCl. Hydrochloric acid produced by the reaction of salt and sulfuric acid is not economically competitive in North America. Since 1990 in the United States, more than 93 percent of the HCl has been produced from the chlorination of hydrocarbons. HCl burners to convert chlorine plant tail gas to HCl accounted for 7 percent. Once 20 percent of the total production capacity, the salt and sulfuric acid process is no longer used in the United States.56 In 1999, the capacity of facilities that deal in merchant HCl was 2.47 million tons. U.S. demand in 1999 was 4.09 million tons. It is estimated that approximately 40 percent of HCl produced is traded on the merchant market, making total U.S. production approximately 610 million tons. During the 1990s, the HCl merchant market changed from being somewhat oversupplied to a rough balance as isocyanate producers committed their production to ethylene dichloride makers. Shifts in the type of chlorofluorocarbons required for refrigerants removed a source of co-product HCl as well.57 Accurate data for hydrochloric acid are difficult to obtain because of the large number of producing and consuming companies, the variety of grades, and the general reluctance of many producing companies to discuss HCl production and use. The production of HCl depends on demand for the primary product (vinyl chloride monomer, isocyanates, fluorocarbons, etc.). In the late 1990s, the average contract price of 20° Be aqueous HCl ranged as high as $88/ton on the Gulf Coast of the United States.57

Hydrogen chloride is marketed both as anhydrous HCl and as hydrochloric acid, usually a 32 percent solution. The absorption of HCl in water customarily is achieved in a falling film absorber, followed by a packed tails tower that vents inerts to the atmosphere. The weak acid from the secondary absorber is fed to the primary absorber. The degree of absorption of the HCl is dependent on the concentration of inerts in the gas stream. The heat of absorption of HCl in water is about 700 BTU/lb of HCl, which requires that the primary absorber be water-cooled.

Acid made by burning hydrogen in chlorine tail gas is quite pure and requires no further treatment. By-product acid requires purification in packed scrubbers in which the scrubbing medium is the hydrocarbon, which is circulated in the scrubbing tower before being sent to the chlorinator. A refrigerated cooling system frequently is employed to maintain a low temperature in the circulating scrubber liquor in order to maximize absorption efficiency.

12.14 Bromine and Brine Chemicals*

Although bromine once was recovered from seawater in the United States, and there still are seawater operations in France and Spain, the most important source of bromine is brines and bitterns (the mother liquor from a saltworks). Seawater contains only 65 ppm of bromine, so very large volumes of it must be processed to recover bromine. Major sources of commercial bromine are underground brines in Arkansas (which contain 30005000 ppm bromine), China, Russia, and the United Kingdom. Bitterns from mined potash in France and Germany, seawater bitterns in India, Italy, and Japan, and bitterns of potash production (which contain 12,000 ppm bromine) from Dead Sea brines in Israel are the other sources.[58] The United States now is the largest producer, with most U.S. production in the area around Magnolia, Arkansas. In 1997, U.S. production capacity was 635 million pounds per year of elemental bromine with another 50 million pounds per year announced and about to come on line.[59] Israel is the second largest producer. The Former Soviet Union is third in production.

*The material in this section was contributed by Philip F. Jackish, Ethyl Corp., Baton Rouge, LA, and John Pavlick, Dow Chemical U.S.A., Ludington, MI.

There are four major steps in bromine production: (1) oxidation of bromide to bromine with chlorine; (2) stripping of bromine from the aqueous solution; (3) separation of bromine from the vapor; and (4) purification of the bromine. The two major stripping vapors are steam and air. Steam is used when the concentration of bromine in brine is greater than 1000 ppm. The advantage of using steam is that bromine can be condensed directly from it. Air is used when seawater is the source of bromine because very large volumes of stripping gas are needed, and steam would be too expensive. When air is used, the bromine needs to be trapped in an alkaline or reducing solution to concentrate it.

In a newer steam-based process, vacuum is used, which, by matching the vapor pressure of the hot brine, causes bromine to be stripped out without the need for steam to heat the brine. Claimed benefits are increased plant capacity, reduced chlorine loss, less lime needed to treat the spent brine, lower plant maintenance costs, and decreased atmospheric emissions.[60,61]

Table 12.21 gives the U.S. market distribution for bromine, which in 1997 had a list price of 56 cents/lb. 1997 U.S. demand was 435 million pounds.[59] The largest use of bromine is in the production of flame retardants. Additive-type flame retardants, such as decabromobiphenyl oxide, are physically blended into polymers; reactive types, such as tetrabromobisphenol A, chemically react during polymer formation. Brominated polymers are used in flame-retardant applications. Other large-volume bromine chemicals are methyl bromide, used as an agricultural fumigant, and ethylene dibromide, used in lead antiknock fluids outside the United States. Ethylene dibromide was once the largestvolume bromine chemical, but the phase-out of lead in gasoline and its banning as an agricultural fumigant have reduced its markets. Other bromine compounds find use as pesticides and as intermediates in various chemical processes.

TABLE 12.21 1997 U.S. Market Distribution of Bromine[59]

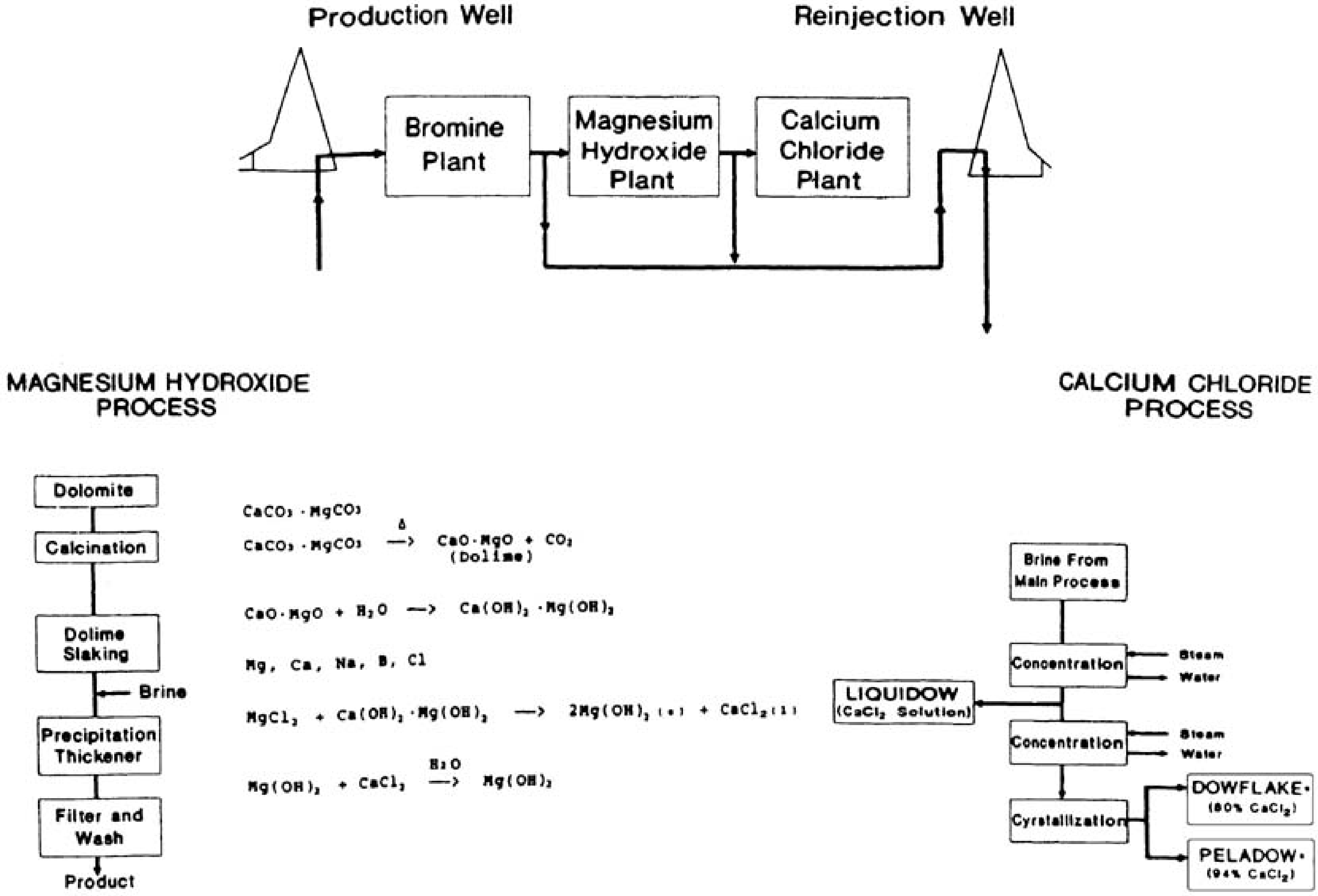
Market	Percent
Flame retardants	38
Drilling fluids	22
Sodium bromide solutions	10
Brominated pesticides	8
Water treatment chemicals	7
Other (i.e. photographic chemicals and rubber additives	15

A fairly new major use for bromine is in clear completion fluids used in crude oil production. These are concentrated solutions of $CaBr_2$ or $CaBr_2ZnBr_2$, which have a high density. Several hundred pharmaceuticals contain bromine atoms. Other moderate-sized markets for bromine compounds are in photography, water treatment chemicals, extinguisher fluids, and dyes.

Miscellaneous bromine uses are in catalysts, fluxes, precious metal recovery, hair care products, food additives, flotation agents in ore treatment, solvents, refrigerants, quartz-halide light bulbs, some lasers, some photovoltaic batteries, and some electrically conductive polymers.

In Michigan this debrominated brine (containing 1020% $CaCl_2$ and 310% $MgCl_2$) is processed further to produce calcium chloride and magnesium hydroxide, as shown in Fig. 12.16.

Fig. 12.16 Brine chemicals in Michigan. (*Courtesy Dow Chemical Company*)



Magnesium hydroxide is produced by precipitation of the magnesium ions with $Ca(OH)_2$ obtained from dolomitic limestone. The limestone is calcined in a rotary kiln to produce lime; and the lime is reacted with water to give $Ca(OH)_2$, which reacts with the Mg ions in the brine to give $Mg(OH)_2$. One-half of the $Mg(OH)_2$ comes from the dolomitic lime. The slurry of $Mg(OH)_2$ in $CaCl_2$ brine is thickened in settling tanks and filtered, and the $CaCl_2$ is washed out of the cake with water.

The major use of magnesium hydroxide is as the raw material for magnesium oxide production, which is used in refractory brick. The hydroxide is converted to the oxide by heat treatment in a rotary kiln or multiple hearth furnace. Other uses are in pulp and paper, Epsom salts, waste water treatment, flame retardants, and smoke supressants.

The effluent from the magnesium hydroxide process is a 25 percent calcium chloride brine. The water in the brine is boiled away with steam in triple-effect vacuum evaporators to make 3245 percent $CaCl_2$. Alkali chlorides precipitate in the evaporators and are removed by settlers and centrifuges. Additional water is removed in evaporators and dryers to attain anhydrous calcium chloride at concentrations from 7894 percent.

Calcium chloride is both hygroscopic and deliquescent. Thus, solid material will absorb moisture from the air or snow and ice until it dissolves, and the solution will continue to absorb moisture until an equilibrium is reached between the vapor pressure of the solution and that of the air. At 95 percent relative humidity, one pound of anhydrous calcium chloride will absorb 17.3 lb of water. The ability of calcium chloride to absorb moisture makes it useful in many construction, commercial, and industrial applications. Table 12.22 gives the market distribution of calcium chloride.[62] U.S. demand for calcium chloride in 2000 was 1.3 million tons. The list price for 78 percent flake early in 2002 was $245/ton.[62]

TABLE 12.22 2001 Market Distribution of Calcium Chloride[62]

Market	Percent

Pavement deicing	22
Road stabilization dust control	20
Industrial processing	20
Oil and gas well fluids	17
Concrete	12
Tire ballast	5
Miscellaneous	4

12.15 Bleaches

Since the fledgling days of industrial chemical practice, the bleaching properties of chlorine have been in demand. It is noteworthy that the beginnings of the industry, some two hundred years ago, centered upon materials and methods of bleach manufacture. Much of modern chemical engineering practice has derived from the struggles and developments of this heritage, particularly in the closing years of the 19th century, when the electrolytic production of chlorine and caustic began to take hold. Rampant spread of the great killer diseases such as cholera and typhoid was shown to be easily eliminated by the treatment of wastewater and drinking water with small amounts of bleach. Textile manufacturers and the paper industry had established the value of bleach in their operations early on.

Prior to the development of the liquid chlorine industry, the only means of transporting chlorine was as bleach. Common practice in the early years was to absorb the chlorine in hydrated lime, $Ca(OH)_2$, to form chloride of lime, $CaOCl_2$. Easily transported as a solid, this chemical contains the equivalent of 35 percent Cl_2. When dissolved in water, it forms equal moles of $CaCl_2$ and $Ca(OCl)_2$:

$$2CaOCl_2 \text{ (dissolved)} \rightarrow CaCl_2 + Ca(OCl)_2$$

It is not the same as calcium hypochlorite, which contains over 99 percent available chlorine. (The term "available chlorine" refers to the oxidizing power of a bleach equivalent to that of free chlorine, as determined by the standard acidic-KI-thiosulfate titration.)

The most practical route to achieving a high available-chlorine content in a lime-based bleach involves chlorinating a slurry of lime in caustic,19 followed by cooling to -10°F. Crystals are centrifuged off and added to a slurry of chlorinated lime containing $CaCl_2$ in an amount equivalent to the NaOCl content of the original crystals. Warming this solution produces crystals of $Ca(OCl)_2 \cdot 2H_2O$, containing around 3 percent lime as an impurity. This material, which is granulated, dried, and packaged for sale, has an available chlorine content over 70 percent.

Another approach uses chlorine monoxide and water to form a solution of HOCl, which is neutralized with a lime slurry, spray-dried, and granulated to yield a product containing 70 percent available chlorine.48

The lime-based products have the disadvantage of containing insoluble components, which has led to the necessity for settling and sludge disposal in some applications. These components are especially objectionable in laundry and dishwashing uses. A granular bleach based on LiOCl46 is quite useful in these applications because it is completely soluble. It is prepared as a 35 percent solution of LiOCl by mixing solutions of LiCl and NaOCl, from which NaCl precipitates. The solution is evaporated, and the solid (containing LiOCl and NaCl) is dried and formulated with salt cake to a white, free-flowing, granular product that retains its bleaching power (35% available chlorine) reasonably well during storage.

In the bleaching of fibers for the manufacture of paper and textiles, sodium chlorite ($NaClO_2$) and/or chlorine dioxide (ClO_2) often is used because of the ability to achieve excellent whiteness without as much fiber degradation as occurs in chlorine or hypochlorite bleaching. For example, in the production of kraft paper, known for its strength, the product is brown when conventional bleach is used because bleaching to whiteness would destroy its strength. By using ClO_2, a high-strength, white product is obtained.

An additional benefit of bleaching with ClO_2 is that the waste lignin degradation products generally present fewer disposal problems than do waste products of chlorine-based bleaches.63 Kraft pulp bleaching sequences that involved chlorine in the first stage (C), followed by caustic extraction (E), followed by two stages of hypochlorite bleaching (H), are being changed to vary the chlorine/chlorine dioxide ratio (D) in the first stage, followed by two stages of chlorine dioxide (i.e. CEHEH replaced by CDEDED). Regulations are being contemplated regarding maximum discharge levels of chlorinated organics and dioxins. Sodium hypochlorite has been linked to the formation of harmful chloroform emissions from bleach plants. Chlorine dioxide use in bleaching has increased, both as a partial substitute for elemental chlorine and as a replacement for sodium hypochlorite. It has been shown that chlorine dioxide substitution in the first-stage bleaching sequence above the 50 percent level can reduce total organic halide (AOX) discharge by up to 50 percent and eliminate the formation of dioxins.64

The generation of ClO_2 from chlorate involves a complex reduction of the chlorine atom of chlorate from the +5 oxidation state to the +4 oxidation state in an acid medium. Chlorine dioxide is a very unstable molecule. It cannot be economically transported, but is easily produced in situ. Treatment of a solution of $NaClO_3$ and NaCl with acid (H_2SO_4 or HCl) produces ClO_2, which is immediately absorbed for use in bleaching. The ClO_2 generator must be operated with care to avoid ClO_2 concentrations above 10 percent, which can lead to explosion from self-decomposition. The molar ratio of ClO_2 to Cl_2 formed in the generator is 2:1. If this level of Cl_2 is objectionable, the gas mixture from the generator, which is mostly air (for safety), is passed through a tower through which chilled water is circulated. This dissolves the ClO_2, but only 25 percent of the Cl_2; so the remaining Cl_2 is scrubbed with alkali. This treatment raises the molar ratio to $8ClO_2:1Cl_2$.

Over the years a number of commercial chlorine dioxide generators have been developed. These generators can be classified according to the type of acid media (i.e. sulfuric acid or hydrochloric acid) and the type of reducing agent (i.e. sulfur dioxide, methanol, sodium chloride, or hydrochloric acid). In most processes, some elemental chlorine is liberated as a by-product. The generator using sulfuric acid produces sodium sulfate and/or waste sulfuric acid, which must be disposed. A new process (R-8 generator65) has been commercialized that is capable of eliminating by-product chlorine (i.e. so that there is no hypochlorite) and produces reduced quantities of by-product salt cake.

Although chlorine dioxide has replaced sodium hypochlorite use in kraft pulp bleaching, hypochlorite bleach is widely used as an ordinary household bleach because of its ease of manufacture, cost, and handling convenience. Sodium hypochlorite is produced mostly by reaction of liquid chlorine with 20 percent NaOH. Soda bleaches vary in their NaOCl content up to about 15 percent; and the higher the NaOCl content is, the higher the required excess of caustic to maintain stability. For this reason, household bleach contains 5.5 percent NaOCl. In the trade it is common to speak of the available chlorine level of soda bleach as the "trade percent." This is related to the NaOCl concentration, expressed in grams per liter, as:

$$(\text{trade \% available } Cl_2)$$
$$= (\text{GPL NaOCl}) \times 0.0953$$
$$(\text{weight \% available } Cl_2)$$
$$= (\text{trade \%})/(\text{specific gravity})$$

Household bleach has a weight percent available Cl_2 of 5.2, and a trade percent of 5.6, which is the percent of available chlorine per volume of solution. It is to be noted that not only is sodium hypochlorite an oxidant, but it yields oxygen on decomposition:

$$2NaOCl \rightarrow O_2 + 2NaCl$$

12.16 Sodium Chlorate

The rapidly growing use of ClO_2 in the pulp and paper industry has led to the rapid growth of sodium chlorate, $NaClO_3$, production in recent years. Sodium chlorate is rapidly produced in solution form by the electrolysis of NaCl brine in a cell that is very similar to a diaphragm chlor-alkali cell, except that it has no diaphragm. The overall reaction is as follows:

$$NaCl + 3H_2O \rightarrow NaClO_3 + 3H_2$$

In this process the chlorine and caustic produced in the cell react immediately, and the cell liquor is kept hot in an additional vessel that provides sufficient residence time for all the hypochlorite to disproportionate into chloride and chlorate:

$$3NaOCl \rightarrow NaClO_3 + 2NaCl$$

For optimum operation, the cell temperature is maintained at 6575°C with a pH of 6.9. Current efficiencies of over 90 percent have been reported with a power consumption of 45005800 kWhr per metric ton of sodium chlorate produced.

Various cell designs are offered by licensers of sodium chlorate technology, the anode design being the most significant variable in the chlorate cell. Coated titanium metal anodes and steel cathodes are used, and small amounts of dichromate are added to the cell liquor to passivate the steel so that corrosion is minimized. The cell liquor contains about 50 percent $NaClO_3$, which may be used as is, or evaporated to 75 percent to drop out salt, filtered, and cooled to crystallize out $NaClO_3$.

In 1999, the total North American sodium chlorate capacity was 2.1 million tons per year with 56 percent of this capacity in Canada because of its lower energy costs. Ninety-eight percent of North American sodium chlorate is used for the manufacture of onsite CIO_2 for wood pulp bleaching, with the remainder being used for the manufacture of other chlorates and in perchlorates and chlorites, herbicides, and water treatment and mining. The 1999 price for bulk sodium chlorate crystal was approximately $450 ton. Sodium chlorate demand growth has significantly exceeded U.S. gross domestic product (GDP) expansion in the past decade as pulp and paper producers moved to comply with U.S. EPA cluster rules which call for elemental chlorine-free pulp bleaching. That switchover has nearly run its course with future sodium chlorate demand growth more closely following GDP growth of 23 percent.66

Akin to $NaClO_3$ is the similar manufacture of bromates, starting with NaBr, or KBr, cell feed. Bromates also are produced by bromination of the respective carbonates. The bromate business is extremely small by comparison to other oxidants, but bromates are important in the bread-making industry for maturing flour and in conditioning dough. They also are used in permanent wave lotions and in the manufacture of dyes. The bromates usually are sold as specialty chemicals. Both chlorates and bromates have substantial oxidizing power, and care must be taken in storage and handling of these solids. Contact with reducing agents and combustibles can be diastrous.

12.17 References

Kauffman, D. W., Sodium Chloride, Van Nostrand Reinhold, New York, 1960.
Leddy, J. J., J. Chem. Ed., 47, 386 1970.
Chemical Economics Handbook, Sodium Chloride, SRI International, Menlo Park, CA, 770.8000 U, 1998.
Chemical Economics Handbook, Sodium Chloride, SRI International, Menlo Park, CA, 770.8000 P, 1998.
Martuchou, M., and Ruiz, A., World Salt Symposium, 8th (2000), R. M. Geertman, editor, Elsevier Science, Amsterdam.
Chemical Economics Handbook, Sodium Carbonate, SRI International, 770.4001 D, 1998.
Eggemann, T., "Sodium Carbonate," KirkOthmer Encyclopedia of Chemical Technology, John Wiley & Sons, New York, 1997.
*Chemical Profiles*Soda Ash, Schnell Publishing Company, Web site: www.chemexpo, June 5, 2000.
*Chemical Profiles*Sodium Bicarbonate, Schnell Publishing Company, Web site www.chemexpo.com, June 12, 2000.
*Chemical Profiles*Sodium Sulfate, Schnell Publishing Company, Web site www.chemexpo.com, Aug. 6, 2001.
*Chemical Profiles*Sodium Hydrosulfide, Schnell Publishing Company, Web site www.chemexpo.com, Feb. 14, 2000.
Chemical Marketing Reporter, Nov. 14, 1994.
Miscellaneous Sulfur Chemicals-United States Chemicals Economics Handbook 780.4000M-X, SRI International, Menlo Park, CA, May, 1992.
Bush, D. R., "Sodium Sulfides," KirkOthmer Encyclopedia of Chemical Technology, John Wiley & Sons, New York, 1997.
Chemical Economics Handbook, Sodium Thiosulfate, 780.4001 M, SRI International, Menlo Park, CA, 1999.
*Chemical Profiles*Sodium Sulfate, Schnell Publishing Company, Web site www.chemexpo.com, January 25, 1999.
Weil, E. D., and Sandler, S. R., "Sulfur Compounds," KirkOthmer Encyclopedia of Chemical Technology, John Wiley & Sons, New York, 1997.
*Chemical Profiles*Sodium Hydrosulfite, Schnell Publishing Company, Web site www.chemexpo.com, Feb. 21, 2000.
Sconce, J., "Chlorine", ACS Monograph No. 154, p. 180 (1962).
Chemical Economics Handbook, Industrial Phosphates, SRI International, Menlo Park, CA, 760.2501 B, 1999.
Van Wazer, J. R., Phosphorus and Its Compounds, Interscience, New York, 1958.
*Chemical Profiles*Sodium Silicates, Schnell Publishing Company, Web site www.chemexpo.com, Jan. 18, 1999.
U.S. Patent 1,385,595.
U.S. Patent 1,132,640.
German Patent 249,222.
Chemical Marketing Reporter, Schnell Publishing Company, New York, (Oct. 1990).
Chemical Economics Handbook, Chlorine and Sodium Hydroxide, SRI International, Menlo Park, CA, 733 1000F, 2000.
*Chemical Profiles*Chlorine, Schnell Publishing Company, Web site www.chemexpo.com, Sept. 4, 2000.
*Chemical Profiles*Caustic Soda, Schnell Publishing Company, Web site www.chemexpo.com, Sept. 11, 2000.
Leddy, J. J., J. Chem. Ed., 57, 640 (1980).
Burney, H. S., and White, R. E., Dow Chemical Company, presented at AIChE Symposium: Tutorials in Electrochemical 1971.
Mantell, C. L., Electrochemical Engineering, 4th ed., McGraw-Hill, New York, 1960.
Kuhn, A., Industrial Electrochemical Processes, Elsevier Publishing Co., 1971.
U.S. Patent 3,718,627.
Grot, E. G., Chem.-Ing. Tech., 44, 167 (1972).
Seko, M., New York A.C.S. Meeting, Apr. 1976.
Beer, H., Chem. Tech., 9, 150 (1979).
U.S. Patent 3,632,498.
German Patent 71937.
Latimer, W. M., Oxidation Potentials, Prentice-Hall, Englewood Cliffs, NJ, 1952.
Leddy, J. J., et al., KirkOthmer Encyclopedia of Chemical Technology, John Wiley & Sons, New York, 1978.
Keating, J. T., and Gerner, H. M. B., in Modern Chlor Alkali Technology, Vol. 7, p. 137, The Royal Society of Chemistry, Cambridge, UK, 1998.
Iacopetti, L., in Modern Chlor Alkali Technology, Vol. 7, p. 85, The Royal Society of Chemistry, Cambridge, UK, 1998.
Kapoor, R., and Martin, J. J., "Thermodynamic Properties of Chlorine," Eng. Res. Inst., University of Michigan, 1957.
Chem. Met. Eng., 51, 119 (1944).
U.S. Patent 3,983,215.
Considine, D. M., Chem. Proc. Tech. Encyclopedia, p. 233, McGraw-Hill, New York, 1974.
White, G. C., Handbook of Chlorination, Van Nostrand Reinhold, New York, 1972.
Van Dyk, C. P., Chem. Eng. Progr., 69, 47 (1973).
Chemical Marketing Reporter, Jan. 1, 2001.
U.S. Patent 3,635,804.
Grosselfinger, F. C., Chem. Eng. (Sept. 14, 1964).
U.S. Patent 3,242,065.
U.S. Patent 3,210,158.
Chemical Marketing Reporter, Dec. 17, 2001.
Hisham, M. W. M., and Bommaraju, T. V., "Hydrogen Chloride," KirkOthmer Encyclopedia of Chemical Technology, John Wiley & Sons, New York, 1995.
*Chemical Profiles*Hydrochloric Acid, Schnell Publishing Company, Web site www.chemexpo.com, Nov. 22, 1999.
Lyday, P. A., "Bromine," in Minerals Yearbook 1987, Vol. 1, p. 172, Bureau of Mines, Washington, DC, 1987.
*Chemical Profiles*Bromine, Schnell Publishing Company, Web site www.chemexpo.com, Mar. 31, 1997.
U.S. Patent 4,719,096.
U.S. Patent 4,725,425.
*Chemical Profiles*Calcium Chloride, Schnell Publishing Company, Web site www.chemexpo.com, Mar. 4, 2002.
Reeve, D. W., and Earl, P. F., Pulp and Paper Canada, 90, 4 (1989).
Fleming, B. I., et al., Pulp and Paper Canada, 89, 12 (1988).
Krauza, K. E., Canadian Pulp and Paper Association, Spring Meeting, 1989.
*Chemical Profiles*Sodium Chlorate, Schnell Publishing Company, Web site www.chemexpo.com, Feb. 22, 1999.

13.1 Overview

Industrial gases may actually be used as gases, liquids, or cryogenic liquids. Industrial users generally accept them as those gases used primarily in their pure form in large quantities.

Most of the gases we consider to be industrial gases have been in use for many years. Processes for the cryogenic separation of the air gases were developed as early as 1895, with commercial production of oxygen beginning in 1902. Nitrous oxide was used as an anesthetic as early as 1799. Carbon dioxide had been identified as a specific substance by 1608. Methane has been used as an energy source since the 1700s.

The reactivity of gases can be summarized into three classes: oxidizers, inert gases, and flammable gases. The gases that fall into the category of inert are nitrogen, argon, helium, and carbon dioxide. The oxidizers are oxygen, chlorine, and nitrous oxide. Acetylene, liquefied natural gas (LNG), and hydrogen are the flammable gases. These elements and compounds by no means encompass all gases considered to be industrial gases, but they represent the major gases used and produced in industry (see Table 13.1).[1]

TABLE 13.1 Properties of Industrial Gases

Gas Property		Helium-4 (He)	Para-hydrogen (p-H_2)	Normal Hydrogen (n-H_2)	Air	Nitrogen (N_2)	Oxygen (O_2)	Argon (Ar)	Carbon Dioxide (CO_2)	Acetylene (C_2H_2)	Methane (CH_4)
Atomic/ molecular weight		4.00	2.016	2.016	28.975	28.01	31.9988	39.95	44.01	26.04	16.042
Normal boiling point	°F	-452.1	-423.2	-423.0	-317.8	-320.4	-297.3	-302.6	-109.33[c]	-103.4 (10 psig) -118[c]	-258.68
Triple point	°F psia	-455.75 0.730[a]	-434.8 1.021	-434.55 1.045		-346.0 1.81	-361.8 0.0216	-308.8 9.99	-69.9 60.4	-116 17.7	-296.5 1.69
Critical point	°F psia	-450.3 33.0	-400.31 187.5	-399.93 190.8	-221.1 547	-232.4 493	-181.4 731.4	-188.1 711.5	87.9 1070.6	96.8 907	-115.78 673.1
Gas density at NTP	Lb/cf	0.0103	0.00521	0.00521	0.07493	0.072	0.08279	0.103	0.1144	0.0678	0.04235
Gas density at STP	Lb/cf										
Vapor density at NBP	Lf/cf	1.0543[b]	0.084	0.083	0.2805	0.2879	0.2795	0.3606	0.1462		0.1134
Liquid density at NBP	Lb/cf	7.802	4.42	4.43	54.56	50.48	71.23	87.02	73.5[d]	24.0 (70°F)	26.57
Specific heat at NTP	BTU/ lb-°F	1.24	3.555	3.425	0.241	0.249	0.2197	0.125	0.203	0.383	0.5271
Specific heat ratio at NTP		1.66	1.38	1.42	1.40	1.41	1.40	1.67	1.304	1.26	1.307
Latent heat of vaporization at NBP	BTU/lb	8.72	191.6	191.7	88.2	85.6	91.7	69.8	245.5[c]	264	219.22
Ref: 8 *CGA Handbook*, 1999		p. 408	p. 416	p. 415	p. 234	p. 528	p. 555	p. 262	p. 296	p. 225	p. 488

NTP = 14.696 psia and 70°F; STP = 14.696 psia and 32°F; Lb/cf = pound per cubic foot.

[a] Lower lambda point; [b] NIST Technical Note 631, "Thermophysical Properties of Helium-4.", Nov. 1972; [c] Sublimation point; [d] Triple point.

CGA References from NIST Technical Notes 1025 (1980), 1048 (1982), 1079 (1985), 361, Monograph 168 (1981).

The cryogenic air separation process, in which air is liquefied and separated into its major constituents (oxygen, nitrogen, argon) by the use of cryogenic technology, is the major source of nitrogen, oxygen, and argon produced for industry, as well as the noble gases Krypton, Neon, and Xenon. Table 13.2 gives the cryogenic boiling point temperatures and concentrations for some of the gases present in the atmosphere.

TABLE 13.2 Cryogenic Gases: Boiling Point and Concentration in Air

	Normal Boiling Point (°F)	Dry Air Concentration (% v/v)
Argon	-302.6	0.93
Nitrogen	-320.5	78.1
Oxygen	-297.3	20.94
Helium	-452.1	0.0005
Hydrogen	-422.99	0.00005
Neon	-410.7	0.0018
Krypton	-244.0	0.00011
Xenon	-163.0	0.000008

The cryogenic air separation process starts with air entering the plant through air filters, where it is compressed and cooled. The air is passed through heat exchangers for further cooling and for removal of water vapor and carbon dioxide by freezing. Solid adsorbents, such as molecular sieves, silicas, and aluminas can also be used to remove the water and carbon dioxide by adsorption before separation of the air. When the air reaches a temperature of -338°F, it is sufficiently free of water and carbon dioxide and ready to be passed into the distillation column.[4]

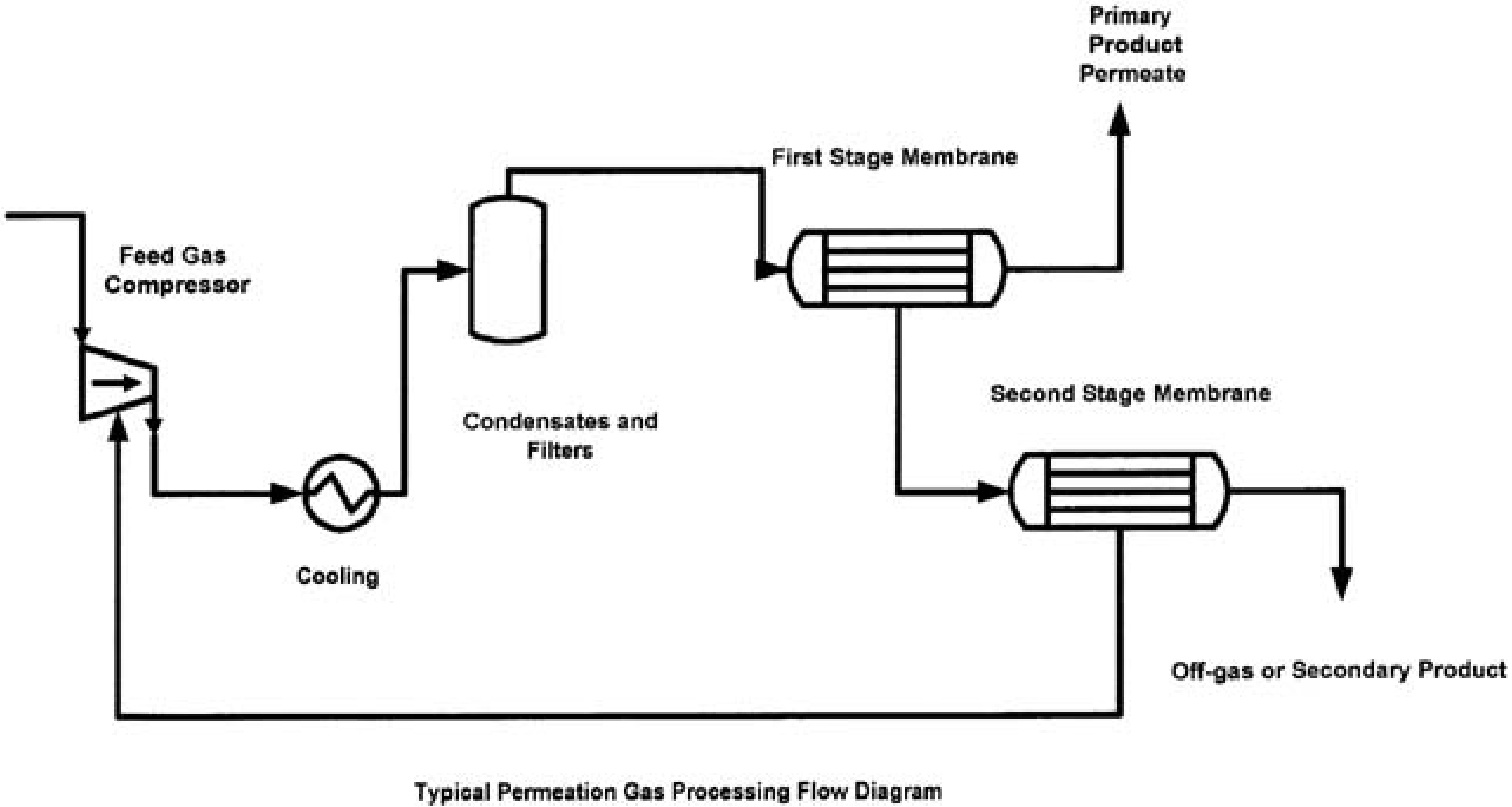
Separation occurs by distillation, which is the physical partition of compounds by the differences in their boiling points (see Fig. 13.5 below in the section "Argon"). The distillation method used is referred to as "double column," as it contains two separate distillation columns operated at two different pressures. The cooled air is fed to the bottom of the high-pressure column. Air is distilled in this column where the higher boiling temperature fluid oxygen stays at the bottom and the lower boiling temperature nitrogen boils to the top as a distillate product. The crude liquid oxygen from the high-pressure column is flashed into the low-pressure column as the two columns are thermally linked by a common reboiler/condenser. Vapor is withdrawn from the middle of the low-pressure column and is passed as feed to the crude argon column. The vapor in that column is condensed and taken off as liquid crude argon. Pure nitrogen vapor is extracted from the top of the low-pressure distillation column, and liquid oxygen is taken from the bottom of this column. Figure 13.1 shows an air separation plant with a molecular sieve front-end cleanup and liquid storage tanks.

Fig. 13.1 Air separation plant with a front-end cleanup, cold box tower and liquid storage tanks. (*Courtesy Air Liquide*)

Alternatives to the cryogenic separation process include the membrane separation process and adsorption processes. The adsorption processes include pressure swing adsorption (PSA)10 and vacuum swing adsorption (VSA). These methods are noncryogenic and produce a vapor product only. This reduces the cost of production considerably when the local use of gas-phase product is the primary objective.

The membrane separation process is used to produce a number of gases for industrial use. Many membrane technologies have been developed over the last 10 years, including various polymer and pure carbon membranes.9,11 A common polymer membrane may consist of a microporous substructure of cellulose acetate and a thin layer of dense cellulose acetate (active layer) on the upper surface, resembling a sponge covered by a plastic film. The membrane is cast into a supporting cloth for added mechanical strength. The active layer serves as the separating barrier and, because of its thinness, provides very high transport rates. The membrane selectively allows the smaller molecular-sized components to pass through the membrane substructure, thereby accomplishing the separation process. Carbon and zeolite-based membranes depend on the relative molecular size of the gases to effect the separation of the gases. A typical membrane separation flow system is shown in Fig. 13.2.

Fig. 13.2 Typical permeation gas processing flow diagram.



The VSA system is a noncryogenic process that makes use of pressure that is less than atmospheric (vacuum) and adsorbents to separate nitrogen, and carbon dioxide from air. The PSA system is a similar process for producing a particular gas by passing a pressurized gas stream through a set of adsorbent beds. The impurities are selectively adsorbed using a molecular sieve or other selective adsorbents to trap the undesired components. In the production of oxygen using the VSA process, the VSA system separates air at ambient temperature by passing it through a column of adsorbent. The adsorbent, which is an inorganic crystal material, selectively adsorbs nitrogen molecules on its surface, while the oxygen molecules pass on through the adsorbent column. The final oxygen product exits the adsorber column. The nitrogen adsorbed is released when the adsorber column is reactivated by use of differential pressure from a vacuum. Once the adsorber column is reactivated, the cycle is ready to be repeated.

In the PSA system, compressed air is passed through a column of chemical adsorbent, which is a carbon-based material that essentially traps the oxygen molecules and allows the nitrogen molecule to pass through to the use point. Figure 13.3 shows the process flow of a PSA system, which can produce 9899.5 percent pure nitrogen without the need for downstream purification. These adsorption systems usually are located at customer sites. See Fig. 13.4 for an example of an on-site PSA.

Fig. 13.3 Process diagram for nitrogen PSA system. (*Courtesy Air Products and Chemicals, Inc.*)

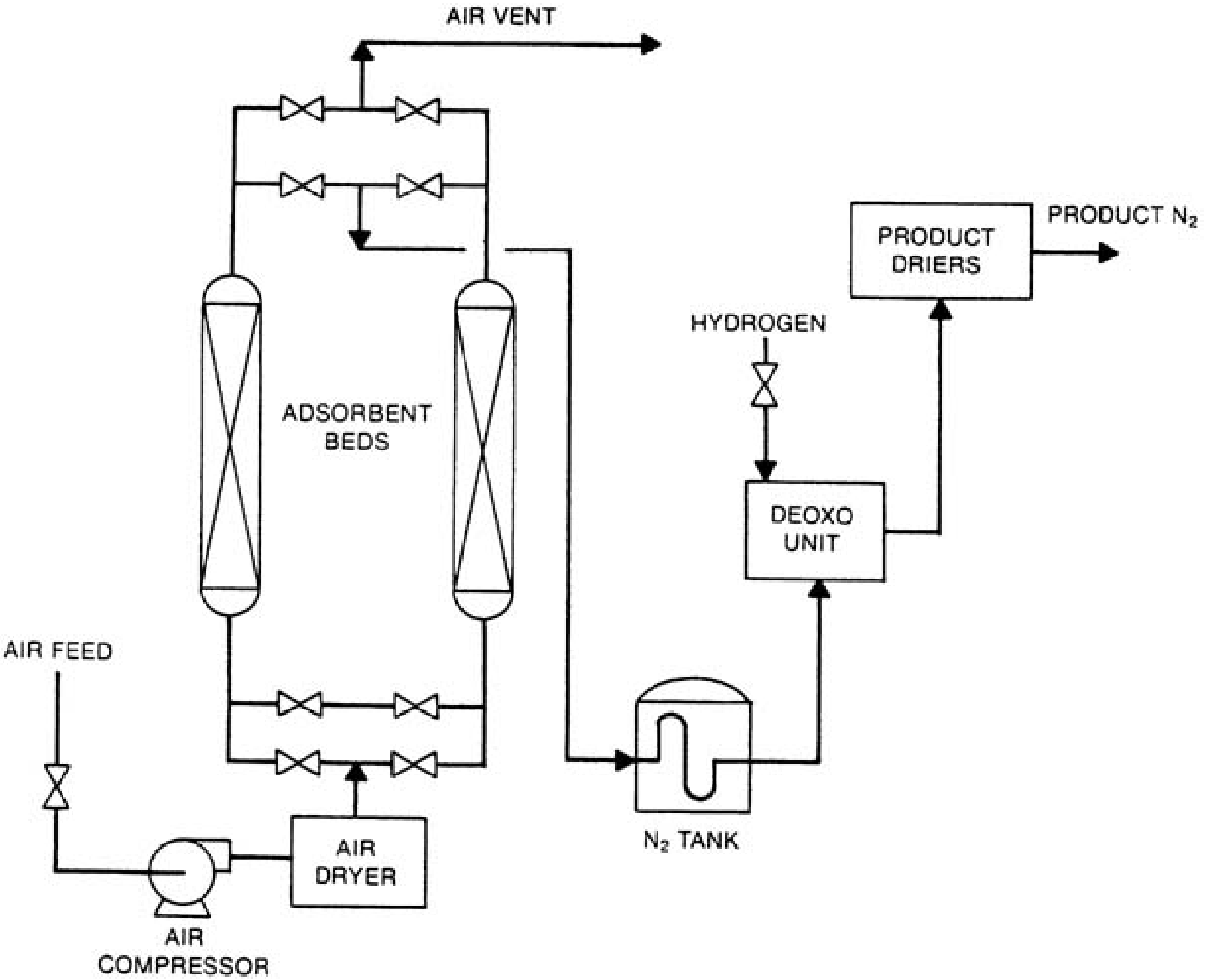
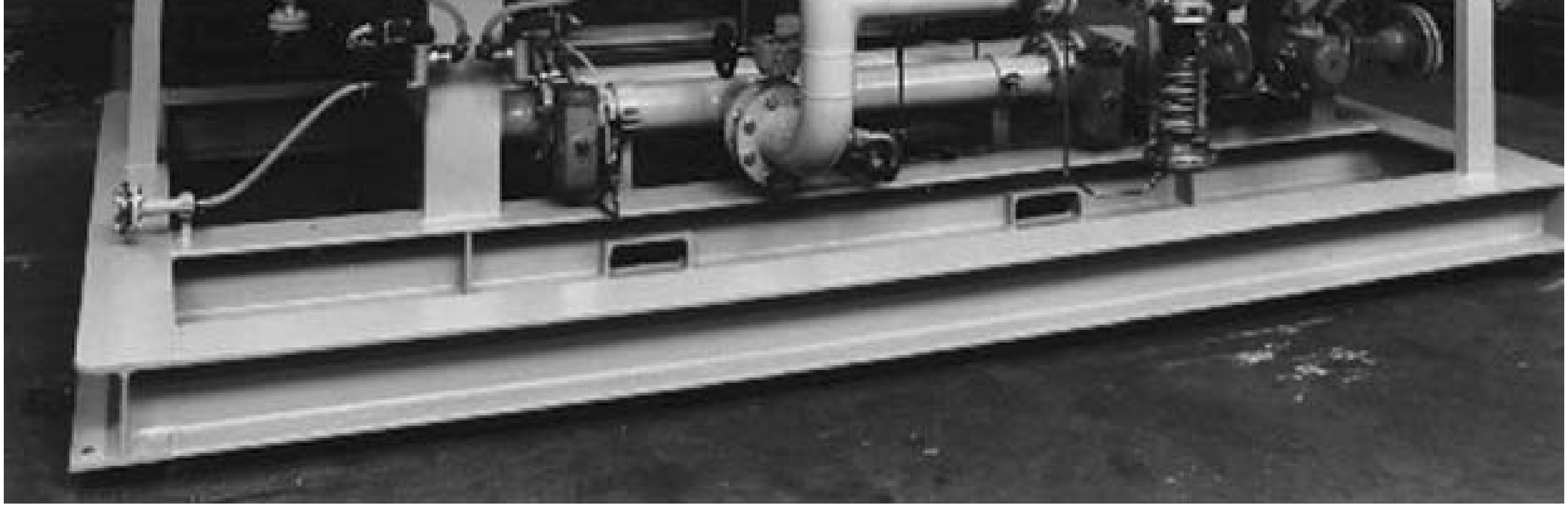


Fig. 13.4 An on-site nitrogen PSA supply system. The large vessels are absorbers and the smaller ones are dryers. (*Courtesy Cirmac International, BY, Netherlands*)



Small-volume users of the industrial gases are supplied by high-pressure gas cylinders filled with the desired product. These cylinders then are loaded onto flatbed trucks for delivery. For larger-volume gas transportation, long-length gas cylinders are permanently mounted on trailer frames. These high-pressure-cylinder vehicles commonly are referred to as tube trailers. The cylinders are constructed of alloy steels, and are designed, built, tested, and maintained in accordance with U.S. Department of Transportation (DOT) regulations. The typical small-volume cylinders are constructed to a DOT 3A or 3AA specification, and typically have rated service pressures from 2015 to 2640 psig and water volumes from 0.39 to 1.75 ft3. The larger cylinders often have similar pressure ratings but water volumes of approximately 75 ft3.

The containers used for transporting and storing the majority of cryogenic liquids are insulated cryogenic tankers, liquid cylinders, and stationary storage tanks. These containers are similar in design to a Thermos® bottle. There is an inner vessel, which contains the cryogenic product, and an outer vessel. The space between the two vessels is under a vacuum and usually contains a wrapped insulation material such as aluminized Mylar, or it may contain a powdered insulation such as expanded perlite. The liquid cylinders, usually 160250 L in capacity, are constructed to a DOT 4L specification and operate with pressure up to 235 psig. The cryogenic tank trucks may be constructed to a DOT MC-338 specification, and range in capacity from 3000 to 14,000 gal, usually depending on the product. The operating pressures may range from 25 to 150 psig, and the vehicles often are equipped with cryogenic pumps used to deliver the product to higher-pressure storage tanks. The stationary storage tanks are designed and built to ASME standards, typically are constructed for pressures of 250 psig or less, and may range in capacity from 500 to 70,000 gal.2

< previous page

next page >

13.2 Nitrogen

Nitrogen makes up 78 percent of the earth's atmosphere. It is a colorless, tasteless, nontoxic, and relatively inert gas. Nitrogen does not support combustion nor does it support respiration; but it reacts with the more active metals such as lithium and magnesium to form nitrides, and at high temperatures it will combine with hydrogen, oxygen, and other elements. As a cryogenic liquid, nitrogen is nonmagnetic, stable against mechanical shock, and free from toxic or irritant vapor.

Nitrogen usually is produced by cryogenic air separation, liquefaction of atmospheric air, and removal of the nitrogen by distillation, as described earlier. Noncryogenic adsorption technology, including PSA and VSA systems, economically provides nitrogen for requirements between 4,000 and 30,000 standard cubic feet per hour (SCFH).9

Nitrogen has many applications, both as a gas and as a liquid. Gaseous nitrogen is widely used for inerting and pressurizing systems. Applications include the agitation of color film solutions in photographic processing; blanketing and deaeration of oxygensensitive liquids, volatile liquid chemicals, and nonferrous metals; and in the production of semiconductor electronic components, primary metals heat treating, and brazing of copper tubing. It is used to pressurize aircraft tires and emergency bottles to operate landing gear; and often to purge and pressurize pipelines, reactor vessels, or storage tanks throughout the petrochemical industry.1

Gaseous and liquid nitrogen have applications in the foods industry in the areas of food processing and packaging. Liquid nitrogen is used for rapid freezing of expensive or highly perishable foods, such as shrimp and other meat products, as well as to refrigerate foods in long-distance hauling. A gaseous nitrogen atmosphere can help to retard the spoilage of some food products.

Liquid nitrogen is used in the cold-trapping of materials such as carbon dioxide and volatile organic carbons (VOCs) from gas streams, as a coolant for electronic equipment, for pulverizing plastics or rubber material, for deflashing of rubber tires, and for simulating the conditions of outer space. It is used to freeze liquids in pipelines during repairs to the pipeline. It can be used to generate a very high-pressure gaseous nitrogen (15,000 psig) by liquid pumping.

Liquid nitrogen is frequently used in medicine to remove skin blemishes in dermatology, and to preserve whole blood, livestock sperm, and other biological specimens.

Liquid nitrogen gas is transported in liquid cylinders, nonpressurized Dewars, cryogenic tankers, and trailers. Gaseous nitrogen is a transported in high-pressure gas cylinders and tube trailers. Both forms of nitrogen also are transported via pipelines when the sources are at economical locations and distances from the applications.

13.3 Oxygen

This gas, which occupies 21 percent of the earth's atmosphere, has a number of very important uses. Oxygen is a colorless, odorless, and tasteless gas that is essential to the support of life. All elements except the inert gases combine directly with oxygen to form oxides. Oxygen is nonflammable, but is an oxidizer that readily supports combustion.

A large percentage of the oxygen produced commercially is made by cryogenic air separation. It may also be produced by PSA and membrane systems, as described in the production of nitrogen. Oxygen also may be made by the most common decomposition reaction, electrolysis of water. Small quantities of pure oxygen may also be made by other chemical reactions. These reactions produce not only oxygen but by-products as well, and they are not usually as economical an alternative when compared with the other technologies.

Applications in enhanced combustion and life-support systems provide the majority of the demand for oxygen. It is used extensively in medical applications for therapeutic purposes, for resuscitation in asphyxia, and with other gases in anesthesia. Also it is used in high-altitude flying and deep-sea diving, and it is used for life-support and as a fuel oxidizer in the U.S. space program.1

For industrial applications, oxygen is in large demand for use with acetylene, hydrogen, and other fuel gases for purposes such as metal cutting, welding, hardening, and dehydrating. Oxygen enhances the capability of steel and iron furnaces. Oxygen is used in the production of synthesis gashydrogencarbon monoxide mixtures from coal, natural gas, or liquid fuels; the synthesis gases are used to make gasoline, methanol, and ammonia. Oxygen may be used for the cracking of methane or natural gas by partial oxidation to produce acetylene, and in the production of nitric acid, ethylene, and other basic chemical feedstocks.

One of the largest commercial applications for oxygen is the Basic Oxygen Furnace (BOF) for refining iron and making steel. Another common application is its use in oxy-fuel burners in the steel industry. It is used in electric arc furnaces, which melt scrap metals by passing an electric current through large graphite electrodes.

Oxygen enrichment is a process whereby oxygen is added to an air stream to improve a plant's efficiency. Combustion results in the generation of heat and gaseous by-products, usually carbon monoxide, carbon dioxide, and water vapor. Oxygen enrichment decreases the amount of nitrogen in the air stream, improving the combustion reactions while reducing the formation of undesired byproduct gases. Oxygen enrichment is used in pulp mills to increase lime production from existing equipment and eliminate the need to purchase lime from other industries. It is also used in pulp mills to reduce airborne sulfur emissions to comply with federal environment regulations.

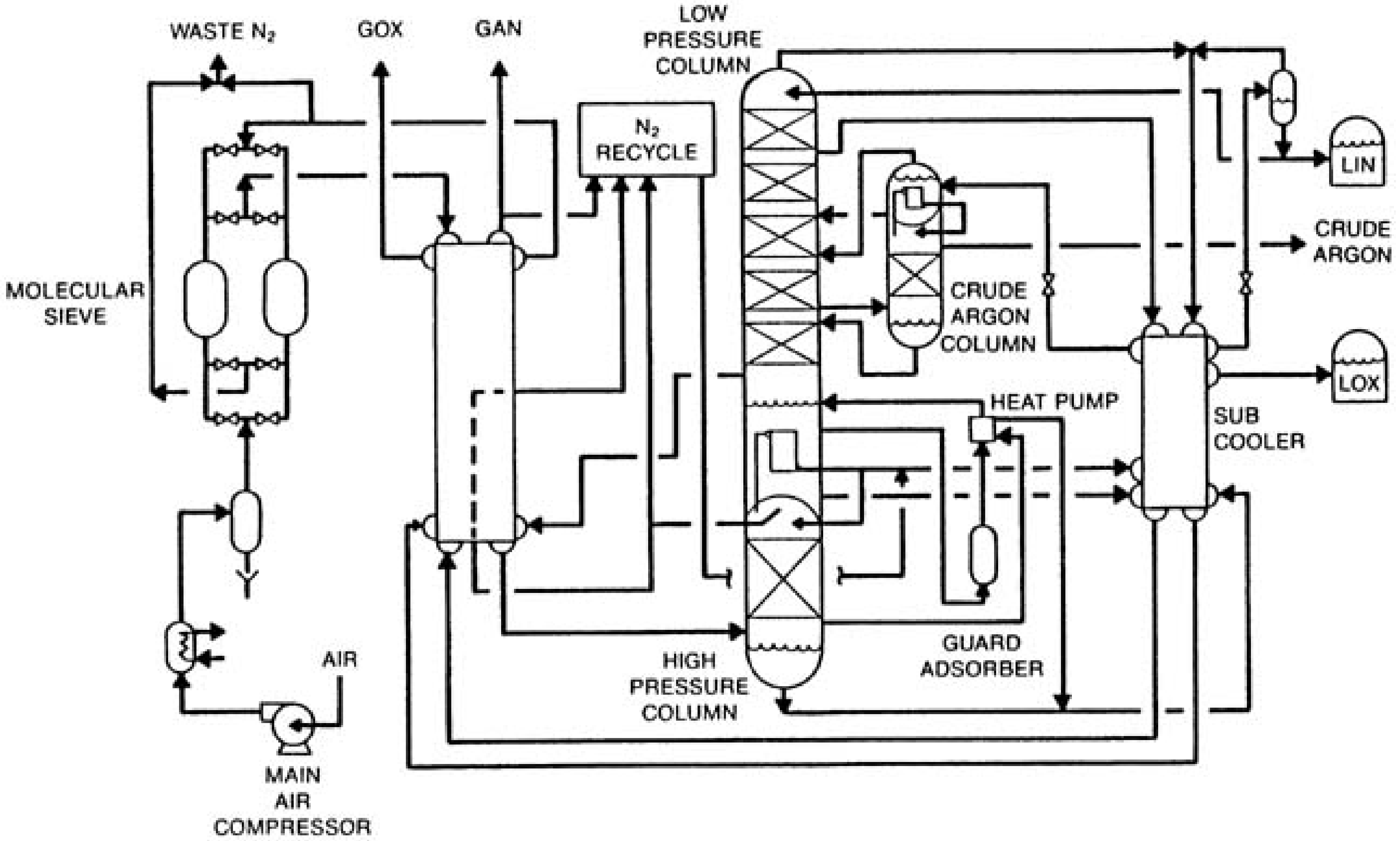
Gaseous oxygen is transported in highpressure gas cylinders, tube trailers, and portable cylinder modules. Liquid oxygen is transported as a cryogenic fluid in liquid cylinders and insulated cryogenic tankers. Customer on-site storage is usually accomplished in large insulated cryogenic tanks.

13.4 Argon

Argon, a relatively scarce gas, is colorless, odorless, tasteless, and nontoxic. It forms no known chemical compound; thus it is extremely inert. Representing only 0.93 percent of the earth's atmosphere, it is the most valuable of the commodity air gases. Argon is utilized in a variety of applications. In the manufacture of lighting devices, it is used to fill lamp bulbs and to produce colors in display tubes for lighted signs. Argon provides an inert atmosphere for the production of semiconductors and specialty metals, such as titanium and zirconium. It is often used to provide a protective shield for the growing of silicon and germanium crystals. Some metal welding and cutting applications require an inert gas shield, and argon gas often is used to fill this requirement.1

Argon is manufactured in oxygennitrogen plants by means of fractional distillation after the liquefaction of air as described above, in the cryogenic air separation process. In the distillation process where air is separated into oxygen and nitrogen, a stream of gas is withdrawn from the low-pressure column to enter the crude argon column. The argon content of the gas, initially 10 percent, is increased to 95 percent by subjecting the gas to several stages of distillation. This "crude" argon is further purified to yield the final pure argon product. Figure 13.5 shows a typical air separation process flow diagram.

Fig. 13.5 Flow diagram for an air separation plant. (*Courtesy Air Products and Chemicals, Inc.*)



Argon is most economically shipped as a liquid. Liquid argon commonly is transported in liquid cylinders, cryogenic trailers, and portable tanks. Argon can be shipped in high-pressure gas cylinders and high-pressure cylinder trailers.

13.5 Hydrogen

Hydrogen, the lightest element known, is extremely flammable and is found in 0.00005 percent concentration in the air. The hydrogen molecule exists in two distinct forms: ortho and para, named according to their types of nuclear spins. Ortho-hydrogen molecules have a parallel spin, and para-hydrogen molecules have an anti-parallel spin. There is no difference in the chemical or transport properties of these forms, but there is a difference in the thermodynamic properties. Para-hydrogen is the form preferred for rocket fuels. Hydrogen consists of about three parts ortho and one part para as a gas at room temperature, which is referred to as normal or equilibrium hydrogen. The concentration of the para form increases with decreasing temperature until, for the liquid, the para concentrations is nearly 100 percent. If hydrogen should be cooled and liquefied rapidly, the relative three-to-one concentration of ortho to para would not immediately change.2

Conversion to the para form takes place at a relatively slow rate and is accompanied by the release of heat. For each pound of rapidly cooled hydrogen that changes to the para form, enough heat is liberated to vaporize approximately 1.5 lb of liquid hydrogen. However, if a catalyst is used in the liquefaction cycle, para-hydrogen can be produced directly without loss from self-generated heat.

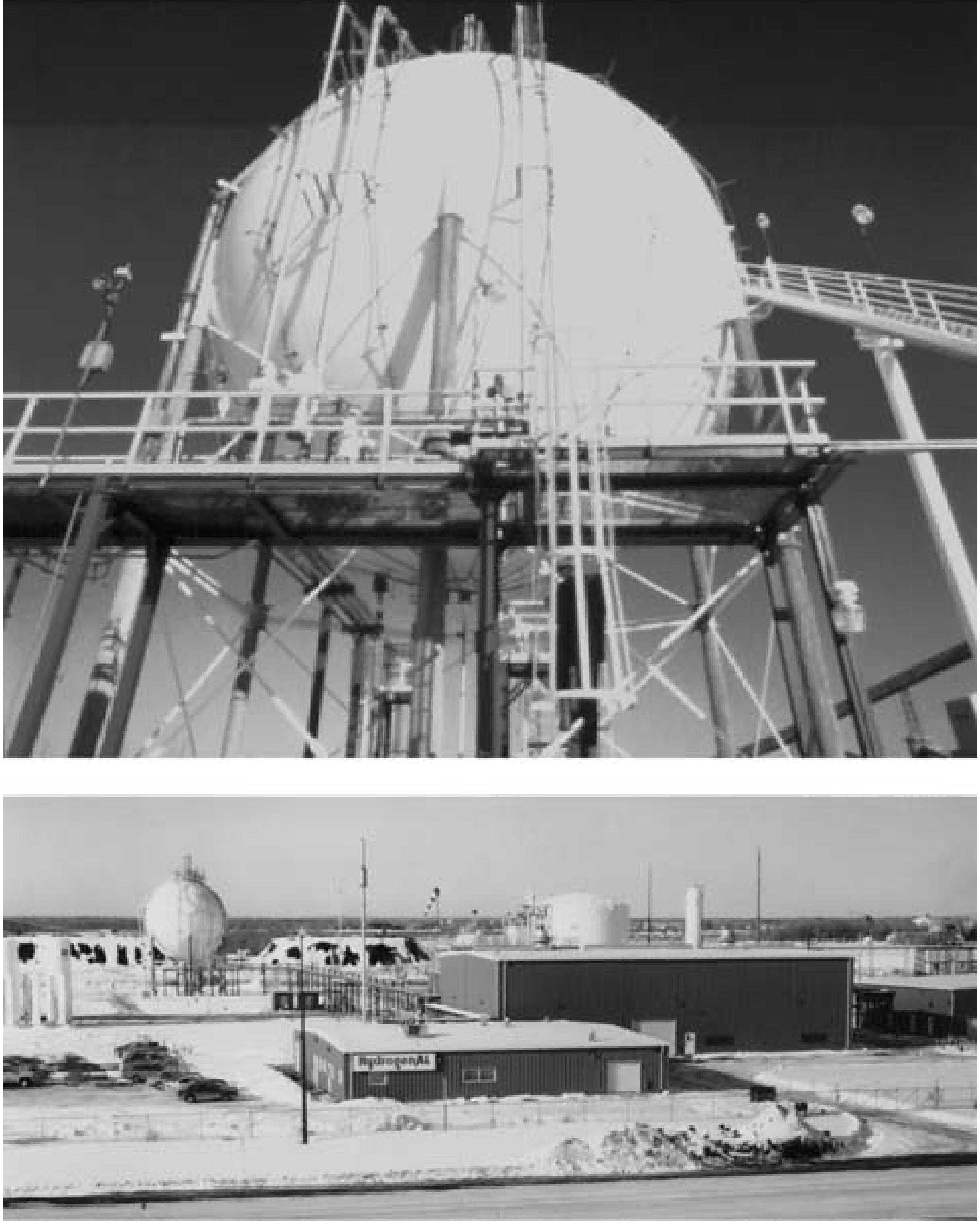
Hydrogen can be produced by steam methane reformation (the most widely used process) or by off-gas purification. Hydrogen can be produced at normal purity (9598%) or at ultra-high purity (99.99%). The first step to obtain pure hydrogen in the steam reformer process when using a natural gas feedstock is to perform a pretreatment. There are small amounts of sulfur in natural gas that must be removed. Sulfur removal is done by preheating the natural gas to 700° F, and then sending it through a catalytic reactor. The catalyst inside the reactor adsorbs the sulfur. The second step is steam methane reforming. In this procedure, natural gas and steam are sent through a set of reformer tubes that are packed with a reforming catalyst. A furnace inside the reformer heats the reaction. The methane and steam react, and the results are 45 percent hydrogen, 40 percent steam, 8 percent carbon monoxide, and 7 percent carbon dioxide. The third step to obtain normal purity hydrogen is gas shifting. Carbon monoxide, hydrogen, and steam are sent through a shift converter where the gas is heated between 400°F and 700°F. The carbon monoxide and steam react to yield hydrogen and carbon dioxide.

To produce pure hydrogen, the carbon dioxide must be removed. The gas passes through a carbon dioxide removal system, which contains a chemical solvent that selectively absorbs the carbon dioxide as the gas passes through the solvent.12 Heat then is added to the solvent to discharge the carbon dioxide. The regenerated solvent is returned to the system to continue the removal of carbon dioxide.

To obtain ultra-pure hydrogen, a pressure swing adsorption (PSA)10 system may be used. In this process, the hydrogen, steam, and carbon dioxide are introduced to the system. At high pressure, the gas is passed through an adsorbent bed, and the adsorbent picks up the impurities and allows hydrogen to pass through the bed. In the regeneration process, the pressure is lowered to purge out the impurities. Then the hydrogen is used to rinse the adsorbent and repressurize the vessel. The efficiency of the hydrogen purification system is optimized at an operating pressure that allows use of a catalyst tube in the reformer.

There are four other distinct processes by which hydrogen may be produced under the off-gas purification method: PSA, membrane separation, catalytic purification, and cryogenic separation. The cryogenic separation process uses cryogenic principles to condense and separate the impurities from hydrogen. Impure hydrogen is introduced to the cryogenic system or "cold box." The impurities are condensed out, and pure hydrogen is obtained. Figure 13.6 shows a typical facility where liquid hydrogen is stored and produced.

Fig. 13.6 A typical facility where liquid hydrogen is produced and stored. (*Courtesy Air Liquide Canada*)



Membrane separation makes use of two layers of membrane material; the first layer is nonporous, and the second layer is porous.11,13,14 A gas mixture under increased pressure is subjected to the surface, which allows a smaller molecule such as hydrogen to permeate faster than a larger molecule such as carbon dioxide.

The catalytic purification process requires the use of many catalysts to convert impurities into compounds. Hydrogen, chlorides, oxygen, and carbon oxides are passed through several catalytic reactors. In the first reactor, the chlorides are absorbed. In the second reactor, oxygen and hydrogen react to obtain water/

steam, which is easily removed. In the third reactor, the carbon oxides react with hydrogen to yield methane, which would be considered an acceptable impurity in this particular hydrogen product stream.

Some other sources of hydrogen are the dissociation of ammonia, steam reforming of light hydrocarbons, and by-product streams of chemical and petroleum processing. The "steamiron" process is another method used to make hydrogen. Passing steam over heated, spongy iron reduces the steam to hydrogen with additional formation of iron oxide. There are several varieties of this process. The watergas reaction is a reaction of steam with incandescent coke or coal; it is a source of hydrogen with carbon monoxide as an additional product. In the catalytic version of this reaction, excess steam breaks down to form more hydrogen while oxidizing the carbon monoxide to form carbon dioxide. Hydrogen frequently is obtained as a by-product of cracking operations using petroleum liquids or vapors as feedstock. Also hydrogen may be produced through the electrolysis of water.

There are many applications for hydrogen. It is used to make fertilizers, as the active ingredient in detergents, and in the manufacture of polyurethanes. It finds application in the production of semiconductor devices, in the process of refining crude oil, in food processing for edible oils and shortening, and as fuel for the space shuttle. The hydrogenation of edible oils in soybeans, fish, cottonseed, and corn produces solids used in shortening and other foods. Many alcohols also are produced by the hydrogenation of the corresponding acids and aldehydes. It is possible to use hydrogen in low concentrations as a physiologically inert gas for a breathing atmosphere where multiple gases are involved. Hydrogen is an important gas in several semiconductor-manufacturing stages. It is used as a protective atmosphere in silicon crystal growth, as a reactant and carrier gas during semiconductor chip manufacturing, and as a reducing atmosphere for sintering and the bonding of leads to the finished chip.1

In the metals industry, hydrogen is used for applications such as annealing metals. It is also used as a fuel in underwater oxy-hydrogen torches and with oxy-hydrogen welding and cutting systems. Hydrogen serves as a nonoxidizing shield alone or with other gases in furnace brazing and in welding aluminum, magnesium, and lead. The temperature in the oxy-hydrogen flames is about 4000°F, which is suitable for low-temperature welding and brazing. Oxy-hydrogen flames are used in the fabrication of quartz and glass, and hydrogen is used as a protective atmosphere in the float glass process whereby molten glass is floated on a pool of liquid tin. Atomic hydrogen welding is suitable for very thin stock and can be used with virtually all nonferrous alloys. In the process, an arc with a temperature of about 11,000°F is maintained between two nonconsumable metal electrodes. Molecular hydrogen fed into the arc is transformed into atomic hydrogen, which sends heat from the arc to the weld zone. At the surface of the welding area the atomic hydrogen recombines with molecular hydrogen, with the release of heat.

Liquid hydrogen is very important as a fuel for powering missiles and rockets. It is used in laboratory research on the properties of some materials at cryogenic temperatures, work often associated with studies of superconductivity. Liquid hydrogen can be used as a fuel, for either propulsion or heating. Hydrogen gas is shipped in high-pressure gas cylinders and high-pressure cylinder trailers. Liquid hydrogen is shipped in insulated portable containers, and in cryogenic tankers. The normal liquid hydrogen supply system utilizes a cryogenic stationary storage tank rated at 150 psig, just below the critical pressure of hydrogen. A cryogenic pumping station, which has a liquid hydrogen storage tank, a high-pressure cryogenic pump capable of supplying liquid hydrogen at 3000 psig, and high-pressure gas storage tubes, often is used to supply large quantities at pressure above 150 psig.

13.6 Helium

Helium was discovered during a solar eclipse as an orange line in the spectrum of the sun's atmosphere. Physically, helium is colorless; odorless, and tasteless; and it is the second lightest element, second only to hydrogen. Helium, is chemically inert and has only a 5 part-per-million (ppm) concentration in the atmosphere.

Small amounts of helium are extracted from the atmosphere by fractionation methods, but not commercially because of the small amount of helium in the atmosphere. A number of natural gas wells contain helium, which can be recovered by a liquefaction and stripping process. Natural gas containing at least 0.2 percent helium has been found in the American Southwest, where the natural gas fields are the major U.S. source of helium. Those helium-rich fields are within 250 miles of Amarillo, Texas; other helium-bearing fields have been found in Saskatchewan, Canada, and in areas near the Black Sea. Helium cannot be synthesized; so conservation and cleanup recycle systems for "spent" gas are important means of preserving the earth's helium resources.

One of helium's first uses was as a nonflammable replacement for hydrogen for inflation of lighter-than-air aircraft. Today liquid helium is vitally important in cryogenic research, as it is the only known substance to remain fluid at temperatures near absolute zero. It has a unique use as a refrigerant in cryogenics and is the only nuclear reactor coolant that does not become radioactive. Liquid helium is used extensively with superconducting magnets; in the medical field, it is used to cool the superconducting magnets for magnetic resonance imaging (MRI). It is used to purge and pressurize NASA vehicle liquid hydrogen tanks because it is the only gas that remains a vapor at liquid hydrogen temperatures.

Helium has a variety of other uses in welding and lighting. It is used as an inert gas shield in arc welding, for filling cold weather fluorescent lamps, and to trace leaks in refrigeration and other closed systems. Also helium is used for specialized purging and pressurizing applications. Helium can be used in place of or together with nitrogen as the inert portion of the atmosphere in sealed environments such as space vehicles, deep-sea submergence vehicles, and diving suits. The use of helium shield gas in a plasma arc furnace is a popular method for melting precious metal ores and scrap. Helium is very important for use in lasers and fiber-optic production, as well as in the production of germanium and silicon crystals for semiconductors.

Gaseous helium is stored and transported in high-pressure gas cylinders, high-pressure gas trailers, and portable cylinder modules. Liquid helium is transported in liquid cylinders, portable tanks, and cryogenic tankers. Because of the extremely low temperature of liquid helium and the cost of producing it, special design considerations must be given to storage containers in order to maintain the helium in the liquid phase. Some containers are designed to include a sacrificial liquid nitrogen shield that intercepts heat before it is transferred into the liquid helium.

13.7 Carbon Dioxide

The carbon dioxide market in the United States is served by gaseous, liquid, and solid carbon dioxide. The consumption of carbon dioxide gas is approximately 25 million tons per year with the enhanced oil recovery and urea production markets being the primary consumers. Liquid and solid consumption is about 7.5 million tons per year, serving primarily the food and beverage processing industries.15 Carbon dioxide is used extensively in a variety of areas. It is used for pressurizing, and as a source of the bubbles and the acidic taste in soft drinks. Carbon dioxide is used to fill a type of fire extinguisher that literally depends on the inertness of the compound. Carbon dioxide is also used in freezing specialty and quality foods. Supercritical carbon dioxide (a dense, high pressure, single-phase form) is finding many new applications in pharmaceutical processing, plastics recycling, dry cleaning, flavor and fragrance extractions, and other solvent-based processes.1

More specifically, each physical phase of carbon dioxide has uses in many areas. Gaseous carbon dioxide is used for pH control in water treatment and as a growth stimulant for plant life. It has been used successfully as a grain storage fumigant against pests and as a wide area mosquito lure to enhance chemical spraying effectiveness. Alone and in combination with other gases, carbon dioxide in food packaging extends the freshness and shelf-life of many products. It is second only to argon as a welding shield gas, and is used for flammable tank purging and inert blanketing of reactive liquids.

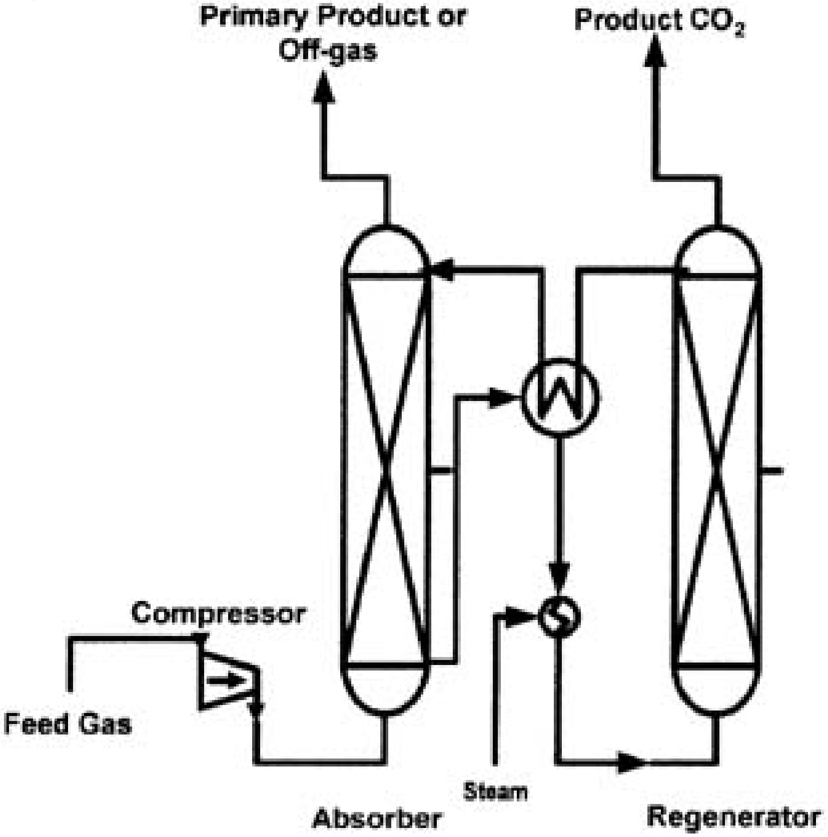
Liquid carbon dioxide has multiple applications as a rapid, controllable refrigerant. It is used in one case as an expendable refrigerant for low-temperature testing of aviation, missiles, and electronic components. Carbon dioxide is also used in controlling chemical reactions and for stimulation of oil and gas wells. It is used extensively in food chilling and freezing applications, both in processing and in transportation.

Solid carbon dioxide, commonly called "dry ice," is used extensively as a refrigerant for dairy products, meat products, and other frozen foods while in transit. It is also used as a cooling agent in many industrial processes such as grinding heat-sensitive materials, cold-treating metals, shrink-fitting machinery parts, in vacuum cold traps, in cryo-surgery, and specimen preservation and storage. It is used as a residue-free abrasive cleaner for many industrial equipment cleaning applications through the use of special air-driven particle blasting machines.

Carbon dioxide is usually nonreactive and nontoxic. At normal atmospheric pressure and temperature, it is colorless and odorless. Carbon dioxide will not burn, nor will it support combustion. Unrefined carbon dioxide gas is typically obtained from the combustion reaction of coal, coke, natural gas, and other carboniferous fuels. Other major sources are ethanol fermentation plants and the development of landfill gases as renewable sources. The gas obtained is liquefied and purified by several different processes to a purity of about 99.99 percent.

The major source of carbon dioxide is as a by-product of steammethane reforming. The resulting reformer syngas may utilize one of the following three processes in the production of carbon dioxide. Recovery of pure carbon dioxide from reformer syngas has most often been accomplished through absorption by liquid solvents. The solvents used are: monoethanolamine (MEA), diethanolamine (DEA), and methyl diethanolamine (MDEA). A hot potassium carbonate solution may also be used in some applications (see Fig. 13.7 for the process flow diagram). In this process, the syngas stream from the reformer is sent to the stripper column where the carbon dioxide is absorbed by the solvent. The residual gas stream components are discharged from the stripper column. The carbon dioxide in solvent is sent to the regeneration column where the carbon dioxide is regenerated and discharged for further purification by compression dehydration and refrigeration steps. The solvent is returned to the stripper column. The final product is essentially 99.99 percent pure.

Fig. 13.7 CO2 recovery by solvent adsorption process. (*Courtesy Process Systems Consulting, Inc.*)

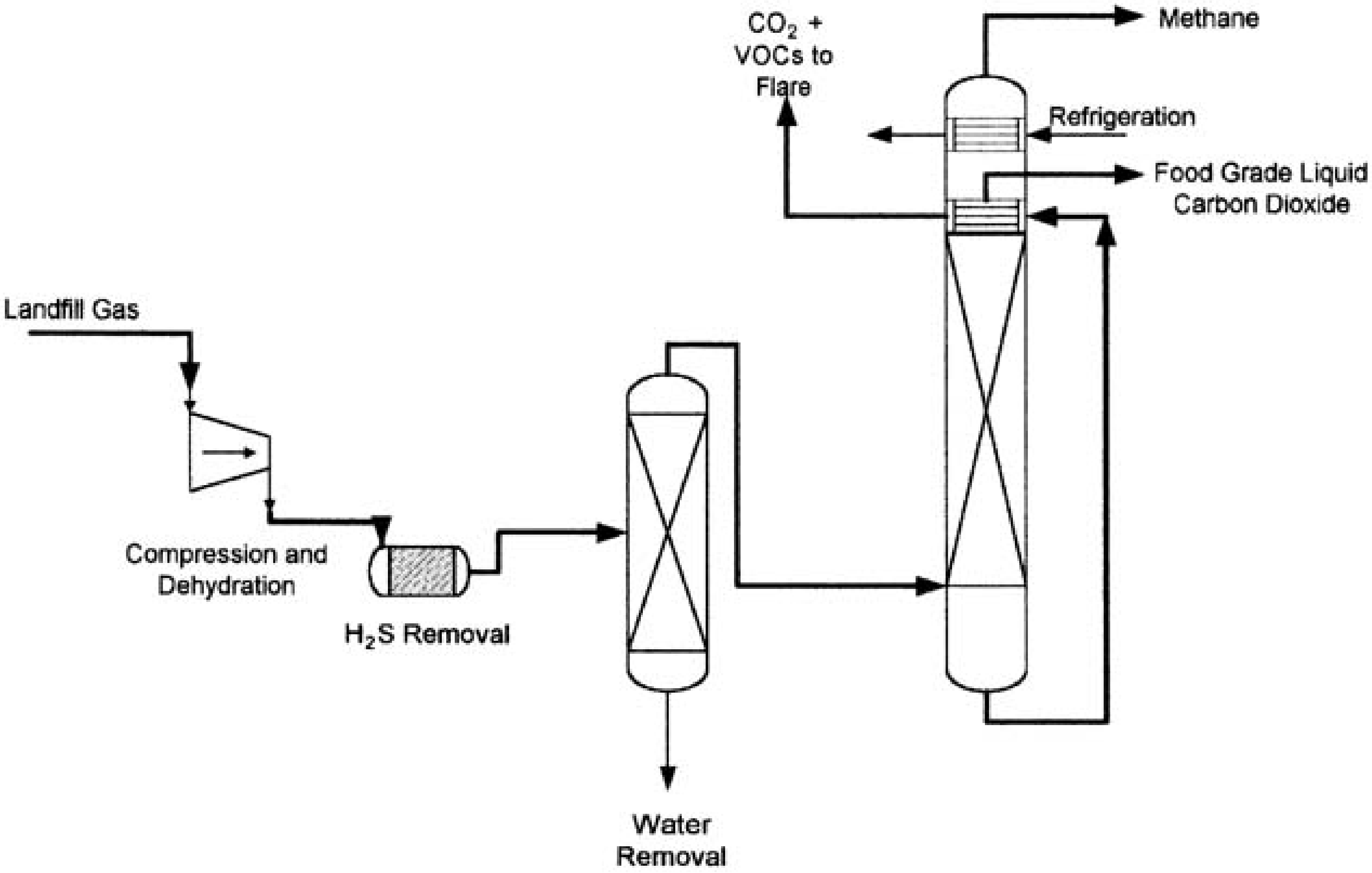


Carbon dioxide may be manufactured at 95 percent purity by the membrane separation process6 from a similar gas or from natural gas.12,13,16 In the recovery of high-purity carbon dioxide from the syngas or natural gas, and before any separations can occur, a pretreatment step removes any contaminants and any condensate liquids that may be present. The feed gas then enters the membrane system at 3001000 psig. The permeate gas is delivered at pressures between 50 and 150 psig and at a carbon dioxide purity of about 95 percent. The high-pressure residual gas containing 1030 percent carbon dioxide enters a conventional treatment process such as monoethanolamine or diethanolamine for a final cleanup. The carbon dioxide off-gas from the conventional treatment process is then elevated to the same pressure as that of the permeate gas from the membrane systems. The combined streams may be further compressed to higher pressures, such as 2000 psig for applications in carbon dioxide flooding in enhanced oil recovery.

The third process used in the production of carbon dioxide is pressure swing adsorption. The feed gas usually contains approximately 20 percent carbon dioxide, 70 percent hydrogen, and the remainder methane, carbon monoxide, nitrogen, and water. The feed gas is typically under a pressure of 125400 psig at temperatures of 80120°F. The carbon dioxide and water are strongly adsorbed in the adsorb beds and the residual gas stream is depressurized for further recovery. The adsorber vessel is then evacuated through vacuum blowers where the carbon dioxide, which has been adsorbed by the bed, is released at purities of essentially 99+ percent pure.

Carbon dioxide produced from ethanol fermentation plants or landfill gas may be recovered with similar processes. Unique to the fermentation plant is the ability to recover the carbon dioxide directly from the ethanol distillation tower, followed by a secondary water wash. Final purification and liquefaction stages then follow the normal process flow. Landfill gas recovery is unique in requiring essentially the removal of the methane and trace impurities.7,17 Several processes exist to provide two gas products (via membrane separation), a liquid methane and gaseous carbon dioxide stream (via the CryoFuel® system), or a gaseous methane and liquid carbon dioxide stream (via the Acrion® process). The Acrion® system is unique in that it uses the carbon dioxide as part of the purification process instead of relying on additional chemical treatments. Figure 13.8 shows the process flow diagram for the Acrion® system. Liquid carbon dioxide is normally stored in a foam-insulated tank that has a refrigeration unit to maintain the tank temperature and pressure.

Fig. 13.8 Acrion CO2 wash process flow diagram. (*Courtesy Acrion Technologies, Inc.*)


CO$_2$ + VOCs to Flare
Methane
Refrigeration
Food Grade Liquid Carbon Dioxide
Landfill Gas
Compression and Dehydration
H$_2$S Removal
Water Removal

13.8 Liquefied Natural Gas

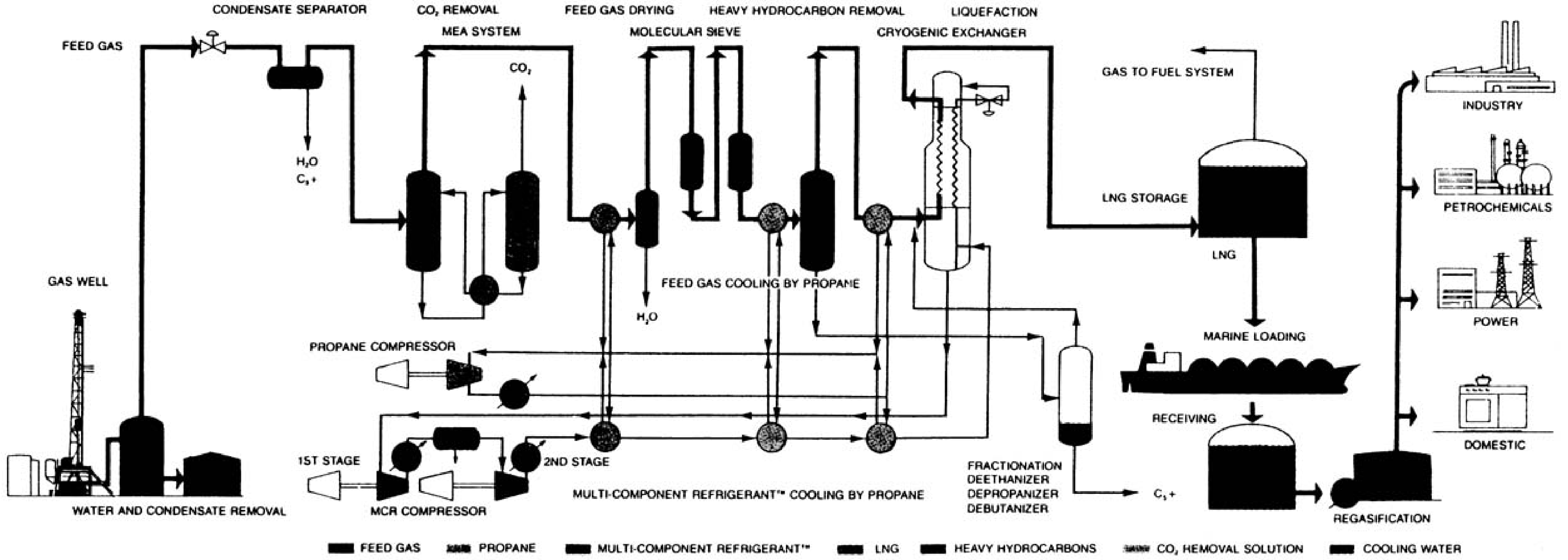
The primary advantage of liquefied natural gas (LNG) over gaseous natural gas is the cost and space savings due to the tremendous change in the product volume. Over 600 standard cubic feet of natural gas occupy only 1 ft^3 of space in its liquid form at -260°F and atmospheric pressure. This significant volume reduction reduces the storage and transport volume, and allows shipment of natural gas in areas where pipelines were never feasible. The primary vehicle for transporting LNG is the cryogenic tanker. It also is transported in specially designed ships.

LNG comes primarily from natural underground reservoirs. Significant quantities are now being recovered from both landfill sources and bioreactor sources around the world. LNG is composed predominantly of methane, which may contain minor quantities of ethane, propane, nitrogen, helium, and other components normally found in natural gas.

In the past the liquefaction of natural gas used a classic cascade cycle. The process required 120,000 hp for liquefaction of over 150 million standard cubic feet (mmscf) per day. Provisions are made for some of these cycles to use seawater for cooling. Later, baseload LNG plants utilized mixed refrigerant cycles, such as Air Products and Chemicals, Inc.'s propane precooled mixed refrigerant system. Baseload plant capacities range from about 70 mmscf/day to about 350 mmscf/day of LNG. Baseload plants move LNG from remote sites by ship to populated areas. For example, Indonesia supplies LNG to Japan for electric power generation.[7]

At baseload LNG production plants, natural gas enters the plant at near 85°F and 600 psig. The gas is treated to remove carbon dioxide and any sulfur compounds, and it is cooled with high-level propane refrigerant to remove as much water as possible. Dehydration to a dew point of about -100°F is done by carrying out regeneration with dried outlet gas by way of a high-pressure steam heater and a regeneration gas compressor. The dry, treated gas is passed through a stationary bed of activated carbon for removal of mercury, and is cooled with two successively lower levels of propane refrigeration. A scrub column flows to a fractionation train, which consists of a de-ethanizer, a depropanizer, and a debutanizer where relatively pure ethane and propane are produced. Excess ethane, propane, and butane from the fractionation process are re-injected into the main gas stream just prior to liquefaction. The overhead from the scrub column flows to the main heat exchanger for liquefaction to LNG. Figure 13.9 shows the process flow for an LNG plant.[18]

Fig. 13.9 Schematic of process flow for a baseload LNG plant. (*Courtesy Air Products and Chemicals, Inc.*)



For years LNG has been used for peak-shaving purposes, which is the storage of excess capacity of LNG to be re-vaporized for the coldest days of the year.[19] In recent years LNG has been used in heating and vehicle propulsion. Also LNG is used to produce carbon black, which has applications in the manufacture of rubber products and printing ink. The burning of high-purity LNG (methane) is done to make carbon black for particular use in electronic devices.

13.9 Acetylene

Acetylene (C2H2) is a colorless and flammable gas. At 100 percent purity, acetylene is odorless, but at commercial purity it has a garlic-like odor. Acetylene can be liquefied and solidified with ease, although in both cases it explodes with extreme violence when ignited.

Acetylene can be inhaled in rather high concentrations without chronic, harmful effects. In fact, it has been used as an anesthetic. However, it is a simple asphyxiant if present in concentrations high enough to deprive the lungs of oxygen and produce suffocation.1

A primary method for the manufacture of acetylene is to react calcium carbide, the principal raw material, with water to produce acetylene with calcium hydroxide as a by-product:

$$CaC_2 + H_2O \rightarrow C_2H_2 + Ca(OH)_2$$

Acetylene also is manufactured by thermal or arc cracking of hydrocarbons and by a process utilizing the partial combustion of methane with oxygen. The acetylene produced from calcium carbide is basically pure, with only water and air as impurities.

Eighty percent of the acetylene produced is used for chemical synthesis. Acetylene is an important raw material for a series of organic compounds such as acetaldehyde, acetic acid, acetic anhydride, acetone, and vinyl chloride. Those compounds are used to manufacture a large group of products, including plastics, synthetic rubber, dyestuffs, solvents, and pharmaceuticals, and in the manufacture of carbon black. The other 20 percent is used for oxy-acetylene welding, cutting, heat treating, lighting purposes, buoys, and beacons.1

In the primary method for producing acetylene, calcium carbide is fed into a generator containing water. The resulting wet acetylene gas at 120140°F is piped to a cooler condenser where it is cooled to 8090°F and the water vapor is condensed. This is the first step in the drying process. The gas then goes to a low-pressure dryer filled with calcium chloride. The calcium chloride absorbs the moisture from the acetylene gas, which then enters a compressor, where the pressure is increased to 200375 psig. An oil separator filters out excess impurities and oil. The gas next goes to a high-pressure dryer, where all remaining water is removed. The gas is then pumped into cylinders for shipment. The by-product, calcium hydroxide, is sent to a storage tank, where water is decanted off the top, and is recovered.

The only method of distributing acetylene, other than pipeline, is by means of portable steel tanks containing a porous solid filler saturated with acetone or other suitable solvent, in which the acetylene is dissolved under pressure. Acetylene alone is not handled at pressures higher than 30 psig because of its tendency to decompose explosively. Dissolved in acetone it may be stored at pressures of 150250 psig.

13.10 Nitrous Oxide

Nitrous oxide (N2O), commonly known as laughing gas, is a colorless and virtually odorless, tasteless, nontoxic gas. Nitrous oxide is an oxidizer that will support combustion, but it is a nonflammable gas. It is water-soluble; but when it is dissolved, its acidity does not change.

The primary commercial method for producing nitrous oxide is by thermal decomposition of ammonium nitrate, with nitrous oxide and water in the primary reaction:

$$NH_4NO_3 \rightarrow N_2O + 2H_2O$$

The impurities formed are mostly high oxides of nitrogen and are highly toxic. After the water is removed in a condenser, the gas is washed in a solution of potassium dichromate to remove nitric oxide, in caustic to remove nitric acid, and finally in water.

Nitrous oxide is an inhalant type of anesthetic or analgesic gas. It serves as a propellant for various aerosol products, particularly with foods such as whipped cream; and it is used as an oxidizing agent in blow-torches, atomic spectroscopy instruments, and in the manufacture of various compounds, both organic and inorganic. It also serves as fuel oxidant in rocket fuel and high-performance vehicles, and it is used as part of the working fluid in hypersonic wind tunnels that have been investigated recently. It has applications in cryosurgery similar to nitrogen.1

Nitrous oxide is contained and transported in its liquid phase in high-pressure gas cylinders or in liquid cylinders. It is transported as a liquefied compressed gas under high pressure in cylinders and at lower pressures and reduced temperatures in refrigerated cargo tanks and insulated portable tanks. Nitrous oxide is stored in a foam-insulated tank accompanied by a refrigeration unit, which is similar to the tank used for carbon dioxide.

13.11 References

Compressed Gas Association, Inc., Handbook of Compressed Gases, Van Nostrand Reinhold, New York, 1990.
Compressed Gas Association, Inc., Handbook of Compressed Gases, Van Nostrand Reinhold, New York, 1966.
Air Products Industrial Gases Fact Book and Safe Practices Guide, Air Products and Chemicals, Inc., Allentown, PA, 1985.
Vance, R., and Weinstock, H., Applications of Cryogenic Technology, Tinnon-Brown, Inc., Los Angeles, CA, 1969.
Technical Orientation Program, Air Products and Chemicals, Inc., Allentown, PA, 1986.
Schnell, W., Houston, C. D., and Hopper, W., "Membranes Can Efficiently Separate CO2 From Mixtures," Oil & Gas J., 81(33), 5256 (Aug. 15, 1983).
Speir, G. A., "Indonesia's Badak LNG Project Sets New Records," Pipeline & Gas J. (205), 2224, 26, 28 (June 1978).
Compressed Gas Association, Inc., Handbook of Compressed Gases, Kluwer Academic Publishers, Boston, 1999.
Castro, R. P., Cohen, Y., and Monbouquette, Y. G., in Proceedings of the 1994 National Conference on Environmental Engineering, p. 82, Boulder, CO, July 1994.
Hydrocarbon Processing, pp. 107108, Gulf Publishing Co., April 1998.
Depuis, G. E. et al., "Hydrogen Membrane Recovery Estimates," Hydrocarbon Processing, pp. 6164, Gulf Publishing Co., April 1993.
Hydrocarbon Processing, pp. 120124, Gulf Publishing Co., April 1998.
Zolandz, R. W., and Fleming, G. K., in Membrane Handbook, W. S. W. Ho, and K. K. Sirkan (Eds.), Chapter 3, Van Nostrand Reinhold, New York, 1992.
Humphrey, J. L., and Keller, G. E., II: Separation Process Technology, Chapter 5, McGraw-Hill, New York, 1997.
Marshall, T. L., and Garvey, M. D., "Special Report," CryoGas Internat., November 2000.
Singh, R., "Industrial Membrane Separation Processes," Chemtech, 3344, April 1998.
Thorneloe, Cosulick, Pacey, and Roqueta, Landfill Gas UtilizationSurvey of United States Projects. Presented at the Solid Waste Association of North America's Twentieth Annual International Landfill Gas Symposium, Monterey, CA March 2527, 1997. Published in Conference Proceedings. EPA-ORD 1997.
Hydrocarbon Processing, p. 112, Gulf Publishing Co., April 1998.
Hydrocarbon Processing, pp. 120124, Gulf Publishing Co., April 1998.

14.1 Sulfur

Sulfur is one of the few elements that is found in its elemental form in nature. Typical sulfur deposits occur in sedimentary limestone/gypsum formations, in limestone/anhydrite formations associated with salt domes, or in volcanic rock.1 A yellow solid at normal temperatures, sulfur becomes progressively lighter in color at lower temperatures and is almost white at the temperature of liquid air. It melts at 114119°C (depending on crystalline form) to a transparent light yellow liquid as the temperature is increased. The low viscosity of the liquid begins to rise sharply above 160°C, peaking at 93 Pa·s at 188°C, and then falling as the temperature continues to rise to its boiling point of 445°C. This and other anomalous properties of the liquid state are due to equilibria between the various molecular species of sulfur, which includes small chains and rings.

Sulfur also is found as sulfide minerals in combination with iron or base metals (e.g. pyrites) and as sulfates in combination with alkali metals and alkaline earths (e.g. gypsum). Hydrogen sulfide, with its "rotten egg" odor, is the primary sour component of sour gas. Crude oil and coal contain a variety of complex sulfur-containing organic species. These sulfur compounds are removed from the liquid fuels by treatment with hydrogen to convert the sulfur to hydrogen sulfide, which is taken off in the gas stream. The recovery of sulfur from sour fuels for environmental reasons is the largest source of sulfur today.

World elemental sulfur production in 2000 was 42 million tons.2 Over 99 percent of the sulfur that is marketed is sold as crude sulfur. The two primary grades are "bright," which is bright yellow and at least 99.8 percent pure (typically 99.9+ %; pure with a maximum of 0.02% carbonaceous material), and "dark," which at the time of production can contain in excess of 0.25 percent carbon, is typically sold as 99.5 percent (min) sulfur with a carbon content not to exceed 0.25 percent and ash of less than 0.25 percent. To achieve those specifications certain types of mined sulfur must be both filtered and blended with "bright" sulfur. Bright sulfur, which today represents more than 95 percent of world production of elemental sulfur, is almost exclusively obtained as a derivative of oil and gas processing. Certain deposits of mined sulfur, such as those still being produced in Poland (also the now-idle sedimentary deposits in West Texas), also produce bright sulfur. However, many sulfur mines are associated with oil fields found in conjunction with geologic formations known as salt domes. In situ cross-contamination normally leads to the production of dark sulfur with varying degrees of hydrocarbon components. Volcanic sulfur deposits yield bright sulfur, but it is often contaminated with toxic metal oxides. Sulfuric acid production accounts for 93 percent of the elemental sulfur used in the United States. Small quantities of several specialty sulfurs are produced for a variety other applications, including bleaching, fumigation, pharmaceuticals, pyrotechnics, rubber manufacture, and cutting oils. Applications of elemental sulfur as a fertilizer represents a growing use with significant potential and much commercialization left to be done.

Transportation and Storage

Although all sulfur is producedand most is consumedin a molten state, the preponderance of international commerce takes place in solid form. Vancouver, Canada, is the world's leading sulfur-exporting port, with volumes in excess of 5 million tons per year, all solid. Sulfur is solidified into a variety of "forms," which will be discussed in the following section. "Formed" solid sulfur is easily transported by truck or railcar but caution must be exercised in order to avoid fugitive dust and spillage. Sulfur dust is highly visible and, under certain circumstances, can lead to corrosion. In the case of marine transport, which may last several weeks, the threat of corrosion during transit becomes a significant concern. To that end it is imperative that the steel of the vessel's holds be coated to preclude direct sulfursteel contact. The need for such preparation to avoid a range of potentially severe consequences cannot be underestimated. Water accumulation in the hold must also be eliminated throughout the voyage to avoid the creation of a watersulfursteel interface.

In contrast, sulfur destined for internal use in the United States, Canada, and Europe is almost exclusively transported in liquid form. This has led to the establishment of an extensive sulfur infrastructure consisting of molten sulfur terminals, tanker vessels, barges, rail tank-cars, and tank-trucks. In the cases of marine transport, most vessels range from 23,000 to 9500 tons and barges from 1000 to 2500 tons. Sulfur is maintained in its molten state during transit with the use of heating coils. But, when transported by tank-car, sulfur is allowed to solidify in transit and must be re-liquefied prior to discharge. That process, which can take up to three days, is accomplished by passing steam through specially built coils built into each tank-car. Trucks operate within close enough ranges as to preclude the need for steaming. In Tampa, FL, many trailers carry sulfur within an inner core on the way out of the port and phosphoric acid, solid fertilizers, or phosphate rock for the return trip. Japan exercises yet another point of view by prohibiting solid sulfur, even in the case of export activities. That has led to a fleet of coastal vessels, some as small as 1000 tons, for exporting molten sulfur to users in Korea and China. Certain countries have regulations that require molten sulfur to be treated (de-gassed) at the production point in order to reduce the liquid's hydrogen sulfide level to under 10 ppm weight. The City of Vancouver requires formed sulfur transshipped at the port to test below 30 ppm weight.

Storage of molten sulfur requires insulated tanks equipped with heating coils through which steam or glycol is constantly circulated. Exposed tank surfaces, particularly the roof, must be protected against corrosion and carefully monitored. A means of injecting steam into the tank void should be provided for fire control. Finally, depending on environmental regulations, tank vents may require scrubbing systems. In the case of marine terminals, these must be sized to accommodate significant inflows of product. "Formed" solid sulfur, on the other hand, can be stored in a pile exposed to the open air. Runoff water containment and neutralization is required around such sites. In certain localities it must conform to maximum H_2S standards. Another means of storage, employed to minimize cost and maximize volume in any given area, is to "vat" or "block" the sulfur. In this case molten sulfur is poured onto a specially prepared area which is surrounded by a rectangular metal barrier (forms). Once the layer of sulfur solidifies, the forms are raised and the process is repeated. This leads to the establishment of a "block" of sulfur which resembles a box or inverted bathtub (vat). Vats containing several million tons of sulfur can be up to 20 m high and of almost any length. As of December 2001, 15 million tons of sulfur was present at vats located within the Canadian province of Alberta, still well below the record of 22 million metric tons in the late 1970s.

One final consideration when storing solid sulfur is the almost inevitable presence of sulfuric acid. Sulfur can become naturally contaminated with sulfuric acid through the presence of thiobacilli thiooxidans3 or continuous exposure to direct sunlight.4 Recent research has demonstrated the short-term effectiveness of certain bactericides in delaying bacterial colonization. Nevertheless, discrete pockets of weak (highly corrosive) sulfuric acid should always be presumed to exist within a sulfur storage pile. Hydrochloric acid, which may also be present when solid sulfur has been transported by vessel,5 must be neutralized to avoid potentially disastrous corrosion of downstream equipment.

Solidification and Melting

For many years, the standard industry practice was to ship sulfur as a crushed bulk solid. Currently accepted best available technologies for solidifying sulfur create pelletized, prilled, or granulated products. These types of "formed" sulfur were developed in the late 1970s to minimize the creation of, and pollution from, sulfur dust during transport operations. As the product names imply: *pelletized* sulfur is generated when a discrete amount of sulfur is deposited and cooled on a conveyor belt. The most common processes are offered by Sandvik6 and Berndorf7 but other licensors exist. *Prilled* sulfur is created when sulfur droplets are cooled by a surrounding fluid, which may be air or water. *Air-prills* are created when spraying sulfur from the top of a tower into a strong updraft of cool air. It is an extremely smooth product which resembles prilled urea or ammonium nitrate. No air-prilling installations have been built recently due the cost of the tower and requisite high-volume air scrubbing system. The last installations erected at Jubail, Saudi Arabia, were destroyed by fire, believed to have been ignited by static electricity. Wet-prilled sulfur, created when sulfur droplets enter a water bath, have an irregular texture caused by water's more rapid cooling action. Wet prill resembles small nuggets but, despite their name, do not encapsulate water within the prill's structure. Various licensors of such technology exist, two of these being DEVCO8 and Enersul.9

Granular sulfur, as the name implies, is created through the gradual buildup of successive layers of sulfur around a central core. As a result the ultimate product size can be controlled. Enersul10 possesses the most widespread granulation process, with installations primarily in Europe and the Middle East. Kaltenbach-Thuring11 provides another alternative.

"Slate," another type of formed sulfur, was the sulfur industry's first attempt at developing a "formed" product which would generate less dust than the traditional "crushed bulk" sulfur commonly used until the 1960s. Slate is produced by pouring a thin (usually up to 1 cm) layer of molten sulfur onto a moving belt while simultaneously cooling the belt's underside. The solidified sulfur breaks off into irregular shapes when falling off the end of the belt, leading it to resemble pieces of "peanut brittle." Owing to its greater dustiness, slate sulfur is a less desirable form of material which now faces commercial resistance.

"Crushed bulk" is the least acceptable type of solid sulfur. It is created when using earth-moving equipment to recover sulfur which has been allowed to solidify in sheet or vats. Obviously, the product is extremely dusty. Very few producers (Former Soviet Union, Iran) still ship crushed bulk. Most consumers refuse to accept shiploads even at significant price discounts due to dustiness, related contamination, and product loss. A U.S. patent was recently issued to d'Aquin for a revolutionary solid sulfur unloading system in marine application which eliminates dust and controls acidity.12

Despite developments in forming technology, moisture remains a necessary tool for dust suppression. To that end, a mixture of water and surfactant should be applied during bulk transfer operations. Disposing of that moisture during transport, storage, and at the time of melting can be an added cost.

Melting or re-melting solid sulfur that has been transported by vessel requires proper knowledge and equipment. Contaminantswhich can include sand, dirt, rocks, and rusted metalmixed with solid sulfur during transit and storage must be removed. Sulfuric acid is another issue: weak acid is almost always present in solid sulfur stored for some time, let alone stored, transported by vessel, and then stored again. As a result, customers may include a specification "below 100 ppm free acid." The currently accepted practice is to add lime prior to melting for acid neutralization. Liming, in turn, requires the introduction of a filtration stagelarge installations use steam jacketed pressure leaf filters, containing a diatomaceous earth medium, to remove the lime/ash residue. Finally, carbon contained within sulfur tends to solidify into extremely abrasive particles when the sulfur is reliquefied. This CarSul abrades burner spray nozzles.13 Users in the sulfuric acid industry therefore prefer receiving sulfur in its molten state rather than incurring the foregoing operating costs and losses of sulfur as filtrate residue. Owing to the volume of production, legislation, and a developed liquid infrastructure, most of the sulfur consumed in North America, Europe, Japan, and Korea is never solidified.

Development of the Sulfur Industry

Early humans doubtless found elemental sulfur in volcanic craters, encrusting the edges of hot sulfur springs, and embedded in limestone formations. They discovered that it would burn and used it for medicinal purposes, as a bleach, as a fumigant, as a colorant, and as incense. Its use for these purposes is mentioned in ancient writings. The Romans produced incendiary weapons from sulfur. In the thirteenth century, the Chinese invented gunpowder using sulfur, nitrate, and charcoal.

The earliest commercial sulfur came from limestone deposits, of which those in Sicily and the Italian mainland developed world markets in the eighteenth and nineteenth centuries. Traditional mining methods were used to produce sulfur ore, which was burned slowly in a pile (Calcarone) to yield crude sulfur.14 Steam smelting in autoclaves came into use about 1859. In 1890 Gill built a multi-chamber furnace to improve the process's production rate and efficiency. Italian monopoly of the sulfur markets continued until the early 1900s when the Frasch process brought previously unrecoverable sulfur deposits on the North American Gulf Coast into production. Oil exploration efforts in Texas and Louisiana in the late 1800s uncovered sulfur deposits in limestone at depths of 200300 m. Mining was complicated by intervening layers of quicksand and the presence of hydrogen sulfide gas. Numerous conventional mining attempts at Sulphur, LA, proved disastrous. Finally, in December 1894, Hermann Frasch demonstrated the hot water process for mining underground sulfur deposits. With its favorable economics, the Frasch process completely displaced the Italian sulfur industry. The ready availability of low-cost sulfur opened the way for commercial sulfuric acid production by burning sulfur. This process largely supplanted the long-standing iron pyrite combustion process for sulfuric acid production by eliminating its extensive gas-cleaning operations.

In 1883 Claus patented a process for producing sulfur from hydrogen sulfide through partial combustion over an iron oxide catalyst. A number of improvements on the Claus process were made over the next 50 years, but it enjoyed limited success as a method for producing sulfur. Its primary driving force was, and remains, the need for a means to remove the sour component of sour gas for processing reasons and for environmental compatibility.

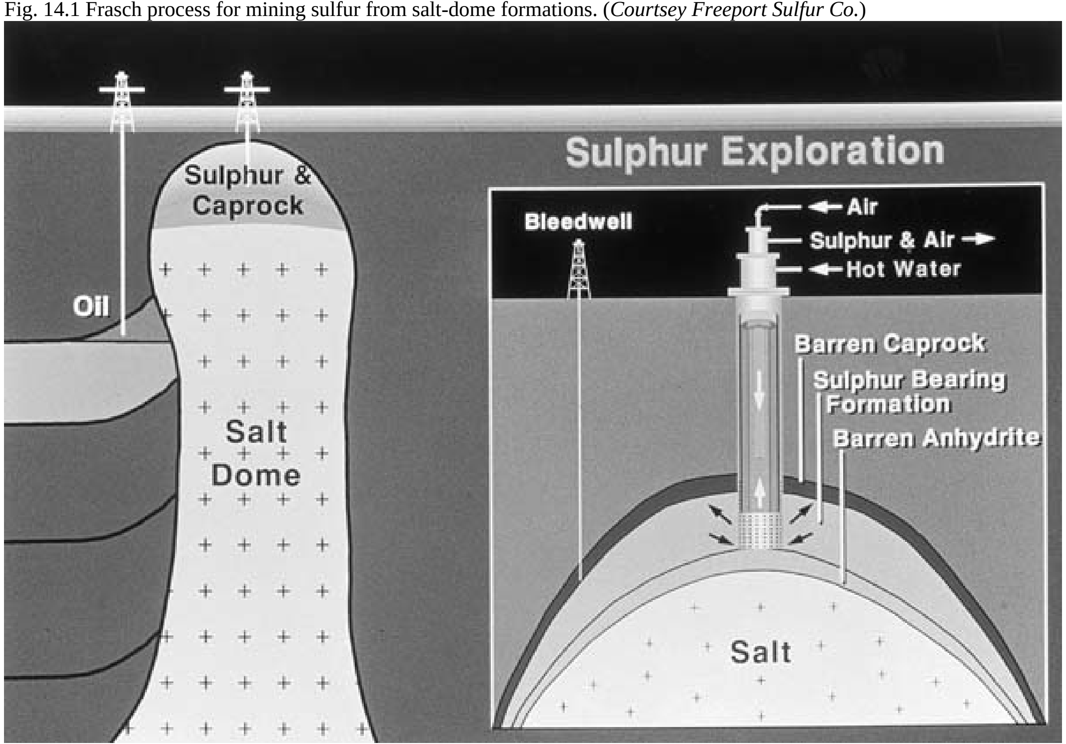
The number of Claus installations grew during the second half of the twentieth century as technical and environmental requirements led to lower sulfur content in hydrocarbon fuels. The trend towards Claus-produced "recovered" sulfur accelerated rapidly after the 1970s when environmental legislation finally took hold. In 1985, Frasch and recovered sulfur in the United States each accounted for about 5 million tons of production. By 1995 Frasch had declined to 3.1 million tons and recovered represented 7.3 million tons. Continuation of the trend changed the industry dramatically: by 2001, faced with oil companies' disposal of recovered sulfur at low prices, all U.S. sulfur mines had closed.15 Frasch sulfur technology, with its colorful and inventive history, lasted just over 100 years. Sulfur output now exceeds world demand by more than 3 million tons and the surplus is increasing. No respite is in sight from either production or demand, leading several companies to consider re-injecting either hydrogen sulfide or elemental sulfur into geologic strata, or other means, for permanent disposal. If such efforts are not successful the amount of sulfur stored in vats will eventually exceed the level which local communities are willing to tolerate, leading possibly to lower use of sour hydrocarbon reserves.16

Sulfur is also produced from sulfide ores (pyrites) by thermal decomposition in the absence of air, by roasting/smelting under reducing conditions, or by reaction of the ore with SO_2. Hydrometallurgical processes have produced sulfur from metal pyrites as a by-product. Except for China, pyrites roasting no longer accounts for significant quantities of sulfur values. And even China has embarked on a rapid conversion to elemental sulfur use in pyrites roasters in order to lower production costs and pollution. Canadian sulfur exporters, who initiated this trend with the development and introduction of innovative technology, have gained the most. Canadian exports of sulfur to China rose from 31,000 tons in 1995 to 1.8 million tons in 2001.17

Sulfur Production Processes

Despite the economically driven closure of the Main Pass 299 sulfur mine in 2000, just eight years after initial output, the Frasch process remains the most economical method for extracting sulfur from native deposits. Certain Constraints on the geological formations required for the Frasch process limit its use to deposits along the Gulf of Mexico, in Poland, in the Former Soviet Union, and in Iraq. Other sulfur deposits may yield to the Frasch process but they have marginal economics. Figure 14.1 shows the structure needed for Frasch mining from salt domes. The sulfur-bearing limestone must have sufficient porosity to allow the sulfur to migrate upon melting. Both the caprock and the underlying anhydrite formation should be impervious to prevent the loss of the hot water pumped into the mine. These salt-dome deposits are typically lens shaped and are 175 m thick with diameters of a few hundred meters up to several kilometers.

Fig. 14.1 Frasch process for mining sulfur from salt-dome formations. (*Courtsey Freeport Sulfur Co.*)

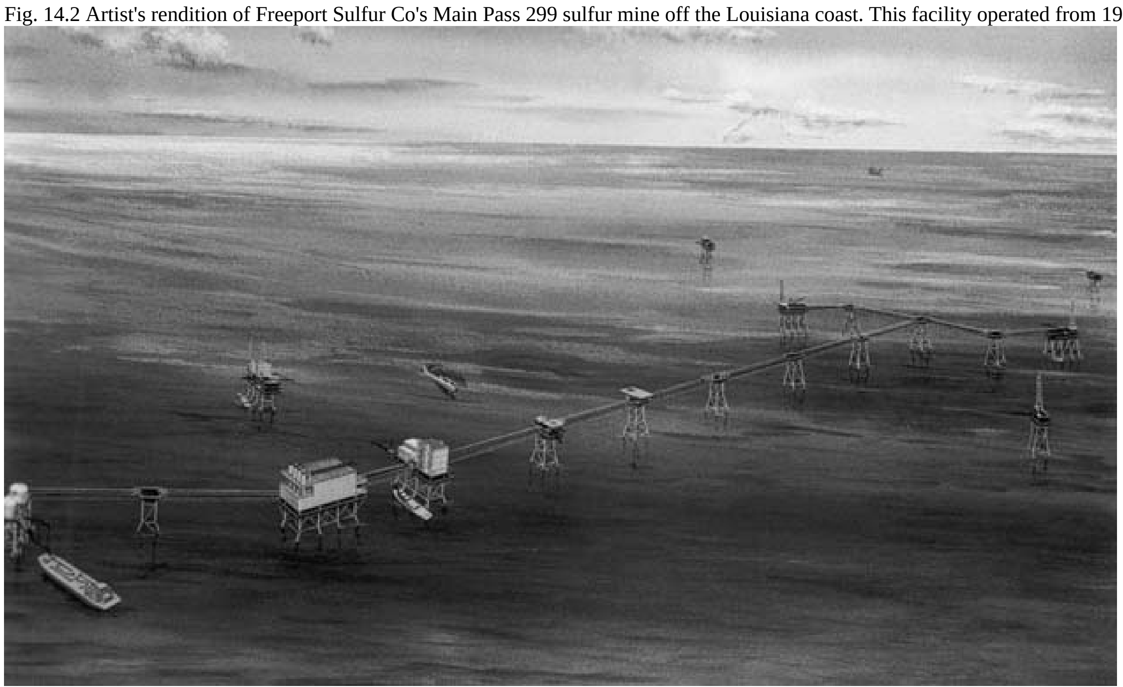


A sulfur well consists of a casing and three concentric pipes reaching into the sulfurbearing strata. The outer 8- to 10-in. pipe carries 165°C water pumped into the formation to melt the sulfur. An inner cement-lined 3- to 6-in. pipe is used to transport the melted sulfur to the surface. Compressed air is passed through the 1-in. tube in the center to air lift the sulfur. Without the air lift the molten sulfur would rise only part way in the middle pipe. The compressed air produces a lowdensity sulfur froth that rises to the surface.

The superheated water melts the sulfur in the vicinity of the well, forming a molten sulfur pool at the bottom of the well. As production continues, the formation fills with water. To continue production, bleed wells are drilled at the periphery of the formation to allow for discharge of the cooled mine water. In some mine fields, sufficient mine water is lost to the geological formation to provide for continued production. To limit mine water loss, mud or synthetic foam sometimes is pumped into the formation to seal major crevices.

Although most U.S. Frasch mines were located inland, Freeport Sulfur Company pioneered offshore sulfur mining in the 1960s.15 The company utilized offshore oil drilling techniques to access several shallow water deposits. Developing a means to utilize saltwater to provide heated mine water presented the company with unique challenges in corrosion and scaling control. Figure 14.2 shows an artist's impression of Freeport's now defunct 2.5 million tons per year Main Pass 299 Frasch mine. This facility, located in 200 ft of water 10 miles from the Mississipi Delta, commenced production in 1992 and closed in 2000.

Fig. 14.2 Artist's rendition of Freeport Sulfur Co's Main Pass 299 sulfur mine off the Louisiana coast. This facility operated from 1992 to 2000. (*Courtesy Freeport Sulfur Co.*)



Sulfur produced from salt-dome structures can be quite pure, but it often contains up to 1 percent of bituminous residues, which render it dark and can make it unacceptable for sulfuric acid production. Some purification is obtained by filtering the dark sulfur through diatomaceous earth. Nevertheless, the most effective means of meeting maximum commercial specifications for carbon is through blending dark sulfur with bright, recovered sulfur containing virtually no carbon. That practice was widely used by Louisiana Frasch producers. The carbonaceous material can be formed into larger, filterable particles (Carsul) by treating the sulfur with heat or sulfuric acid. Freeport's submerged combustion distillation process was used from 1966 until 1979 to purify Frasch sulfur with up to 2 percent carbonaceous material.

Recovered Sulfur

Hydrogen sulfide is recovered from natural gas or refinery gases by absorption in a solvent or by regenerative chemical absorption.18,19 In either case a concentrated hydrogen sulfide stream is produced that is treated further by the Claus process. A typical Claus plant has a feed stream of at least 45 percent H2S, but with modifications can handle streams containing as little as 5 percent H2S. For gas streams with low concentrations of hydrogen sulfide, direct conversion of the hydrogen sulfide to sulfur is accomplished in the solvent system, for example, the Stretford process or CrystaSulf process.

The Claus process is based on the reaction of H2S with sulfur dioxide according to the highly exothermic reaction:

$$2H_2S + SO_2 \longrightarrow 3S + 2H_2O$$

In practice, sulfur dioxide is produced in situ by partial oxidation of the hydrogen sulfide with air or oxygen in a furnace. In the split-flow arrangement, one-third of the H2S stream is burned and then recombined with the remaining two-thirds before entering the Claus reactor. In the straight-through version, the entire H2S stream is sent through the burner and the extent of H2S combustion is controlled by the air feed rate.

A flow diagram for a typical Claus process is shown in Fig. 14.3. The hydrogen sulfide is burned in a fuel-fired furnace (9501250°C) with air to produce sulfur and a gas stream containing H2S and SO2. Process controls maintain the H2S: SO2 ratio near 2, in accordance with the stoichiometry. Heat is removed from the gas stream in a waste heat boiler to control the process gas temperature. The process gas is passed through one or more catalyst beds to convert the H2S and SO2 to sulfur, which is removed in condensers between each bed. A high temperature shifts the equilibria toward the reactants, whereas a low temperature causes sulfur condensation on the catalyst bed, leading to decreased catalyst activity, requiring bed switching/ regeneration to achieve modern recovery standards. In practice the temperature is closely controlled for each Claus reactor, with higher temperatures at the first reactor where compounds such as COS and CS2 are converted to S. Much research and development into H2S conversion has occurred in the past decade. Three-stage Claus units are now capable of achieving recoveries of 98+ percent, a significant improvement over the 8590 percent range of the late 1980s.20

Fig. 14.3 Claus process flow diagram.

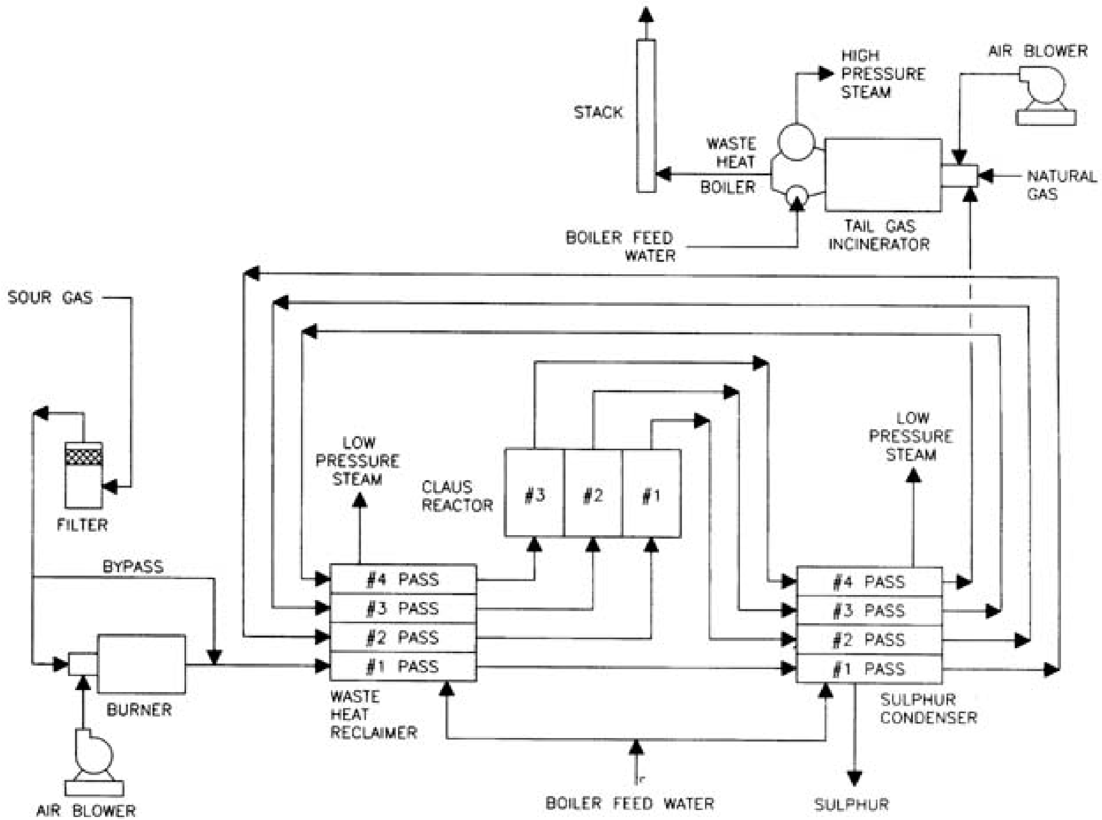


Figure 14.4 shows a typical sulfur recovery plant based on the Claus process. The tail gas from the Claus reactors may be further processed to remove any remaining sulfur compounds. Combined H2S removal efficiencies of 99.599.99 percent are achievable.20 This may be done by low-temperature Claus-type solid-bed processes (e.g. the Sulfreen process), wetClaus absorption/oxidation processes (e.g. the Clauspol 1500 process), or hydrogenation of the off-gas to form H2S for recycle (e.g. the SCOT process). Residual sulfur compounds in the tail gas are then incinerated to SO2. The residual SO2 in the oxidized tail gas may be scrubbed by any of several processes (e.g. the WellmanLord process) before being vented to the environment. It is feasible to bring the H2S content of the treated tail gas to as little as 15010 ppm volume, depending on the solvent used.20

Fig. 14.4 Republic Refining Co. 440 metric ton per day Claus sulfur recovery unit (left) and SCOT tailgas cleanup unit (right). (*Courtsey Ortloff Engineers, Ltd.*)

Production and Consumption of Sulfur

World production of sulfur in all forms reached 57.7 million tons in 1999,21 compared with 58.1 million tons in 1990.22 This similarity in volumes conceals the dramatic structural change in the industry's source of supply. Table 14.1 summarizes the dramatic shifts in the sources of sulfur during this period. It shows, for example, the virtual demise of Frasch and native sources of supply, and a 50 percent reduction in sulfur values derived from sulfuric acid pyrites roasting offset by huge increases in recovered sulfur plus added sulfuric acid and SO2 recoveries from metallurgical smelting.

TABLE 14.1 Sources of SupplyAll Forms of Sulfur23

	1999 (tons)	%	1990	%	Difference (tons)	%
Base processes						
Frasch	3.0	5	11.0	19	(8.0)	(73)
Native	0.7		3.1		(2.4)	(76)
Pyrites	5.5		10.0		(4.5)	(45)
Subtotal	9.2					
By-product						
Recovered	33.7	58	24.0	41	9.7	40
Metallurgy	10.2		7.7		2.5	32
Unspecified	4.5		2.4		2.1	88
Subtotal	48.4	84	34.1	59	14.3	
Total	57.6		58.2		(0.6)	(1)

Note: The 2000 USGS Mineral Industry Survey of Sulfur contains detailed country information (http://minerals.usgs.gov/minerals).

From 1990 to 1999, voluntarily produced sulfur (Frasch, native mining, and on-purpose pyrites roasting) declined almost 15 million tons. At the end of the decade it accounted for only 15 percent of the world's sulfur output. On the other hand, by-product sulfur recovered from hydrocarbon and metal smelting climbed from 59 to 84 percent of the total supply by 1999. Both trends continue to diverge.

The shift devastated commercial enterprises in the voluntary extraction industry. Sulfur prices embarked on a decline which culminated with spot sulfur prices falling to the $10s per ton in Tampa, FL, and Vancouver, Canada, by mid-2001. In contrast, prices during 1990 were $140/ton and $108/ton, respectively. The primary difference was that, in 2001, the newly developed disparity between voluntary and involuntary volumes of output precluded a curtailment in sulfur output in the face of a dramatic, yet temporary, decline in demand (see Figs 14.514.7).

Fig. 14.5 Trends in elemental sulfur production in the United States.24

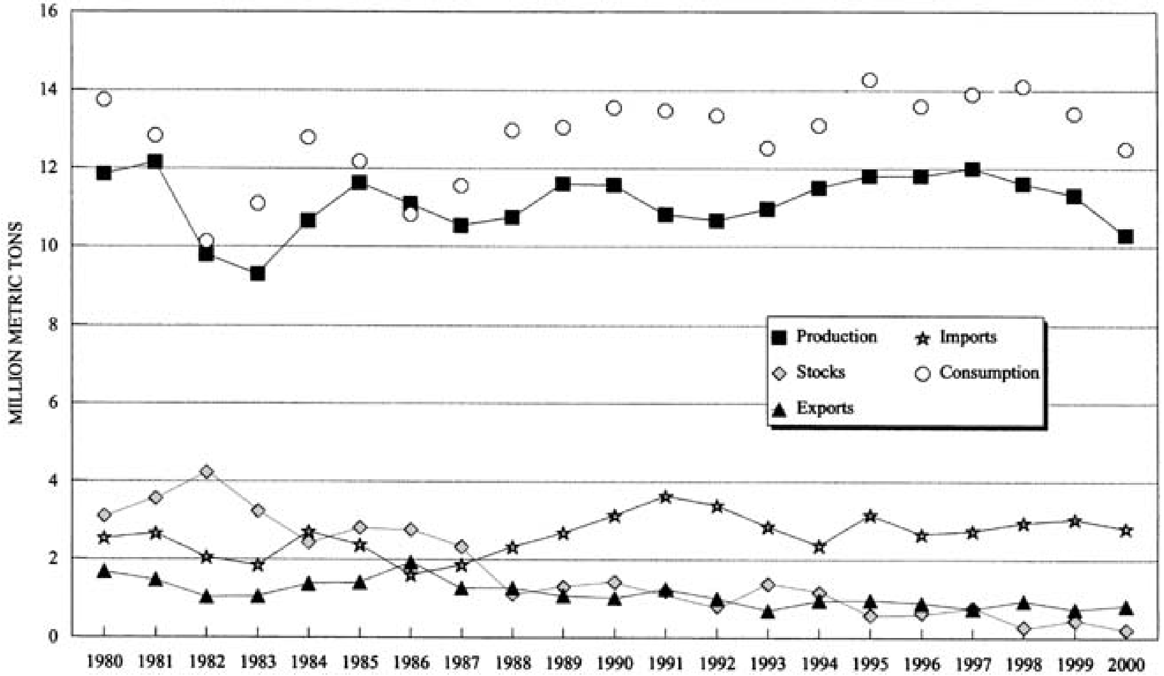


Fig. 14.6 Trends in the production of all forms of sulfur in the United States.24

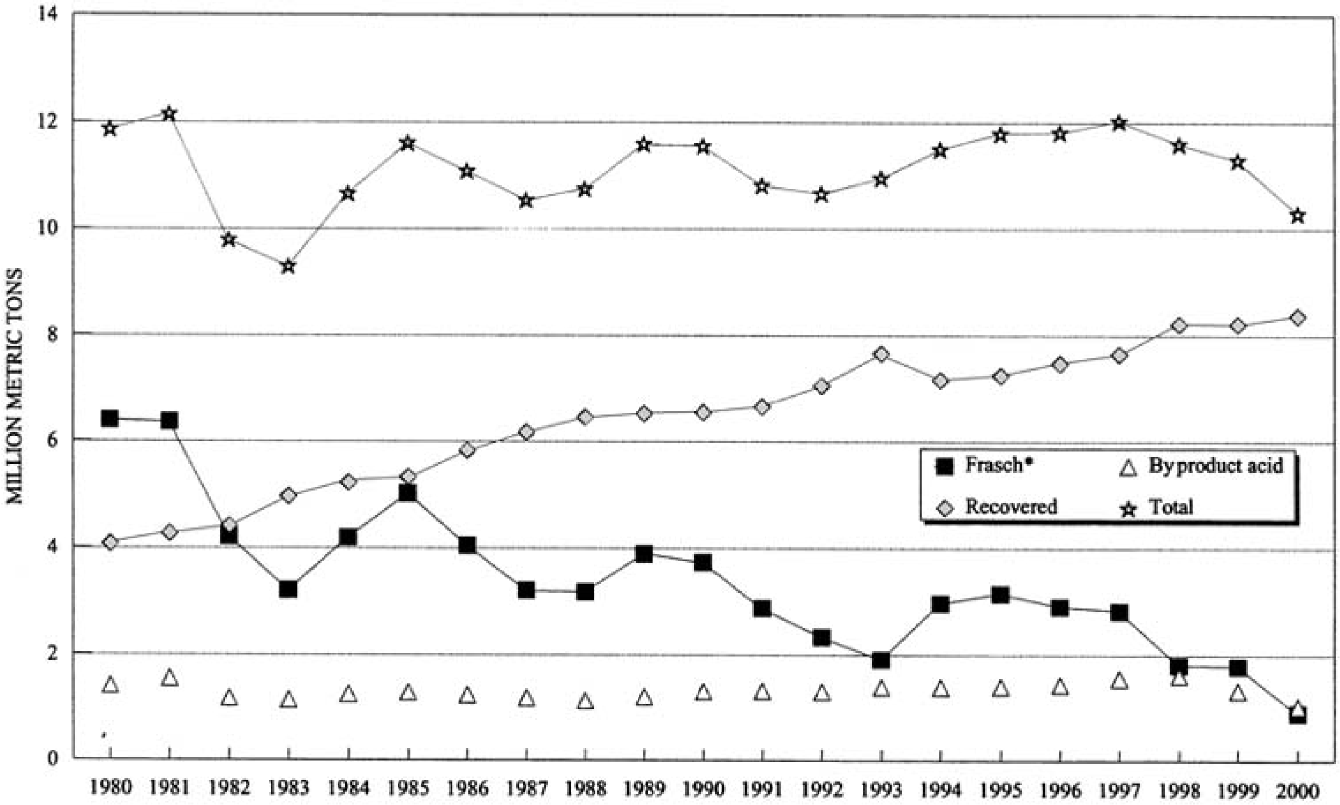
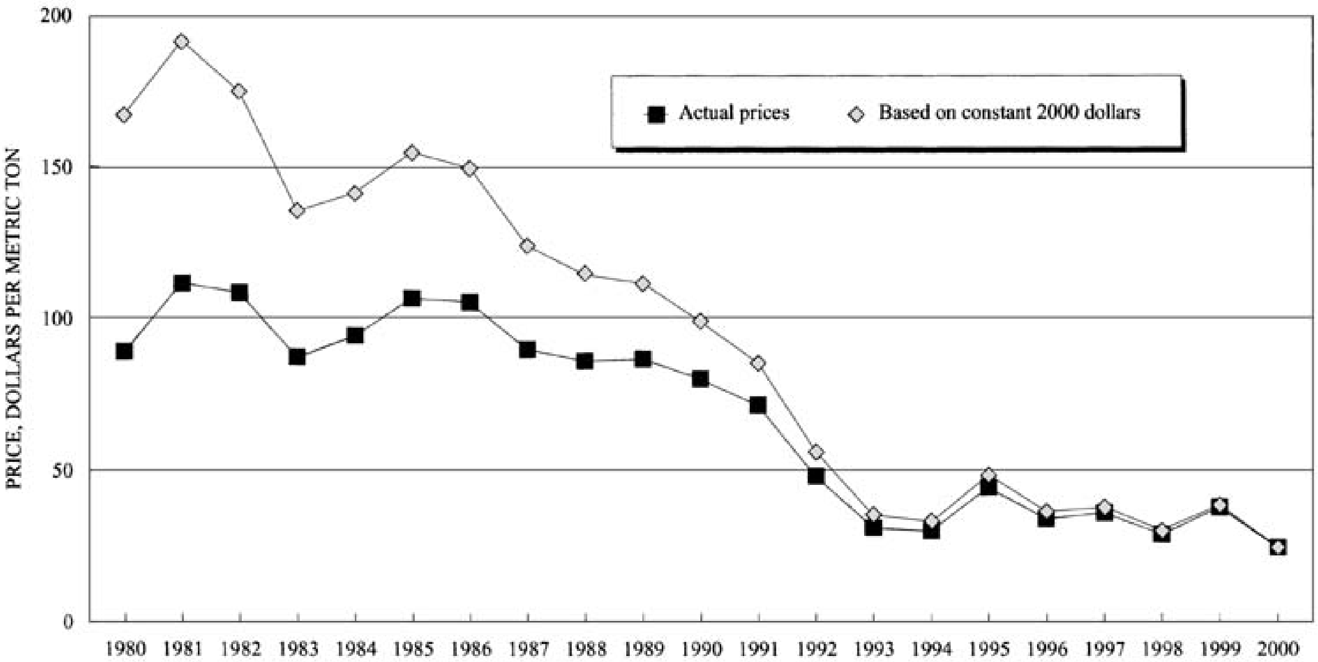


Fig. 14.7 Trends in sulfur pricing.24



Despite the dramatic cessation of North American Frasch mining during the 1990s, the global elemental sulfur output remains almost equally divided between the Americas (North America) and the rest of the world (RoW) (Table 14.2). Substantial shifts occurred within each hemisphere as the Frasch industries in Mexico, the United States, Poland, Eastern Russia, and Iraq collapsed and recovered sulfur volumes increased (Canada, the United States, Central Asia, the Middle East, and Japan, primarily). In 1999, the United States and Canada, at 10 and almost 9 million tons, respectively, are the first and second largest producers of elemental sulfur in the world. Canadian output rose almost 3 million metric tons during the decade, U.S. output declined only marginally despite the closing of all Frasch mines, and Mexico's output collapsed concurrent with mine shutdowns. European volumes fell: increased refinery and gas extraction was unable to offset the effect of Polish Frasch curtailments and reduced volumes in France's Lacq gas field. Increases in Central Asia and the Middle East are both primarily linked to gas processing. The perhaps tempered growth in Middle East output must be viewed in the context of rebuilding many oil and gas processing centers lost in the IraqIran and IraqKuwait conflicts. A significant number of projects are planned for this region. In Asia, Japan, and Korea became increasingly important crude processing centers, with attendant increases in sulfur output.

TABLE 14.2 World Production of Elemental Sulfur24 (Million Metric Tons)

	1999	1990	Difference

	Tons	%	Tons	%	Tons	%
Canada	8.8	*21*	5.9	*16*	2.9	*49*
United States	10.0	*24*	10.3	*27*	(0.3)	nm
Mexico	0.9	*2*	2.1	*5*	(1.2)	*(57)*
Subtotal	19.7	*47*	18.3	*48*	1.4	*7*
Europe	5.8	*14*	8.5	*22*	(2.7)	*(32)*
Central Asia	5.9	*14*	4.5	*12*	1.4	*31*
Middle East	5.4	*13*	4.3	*11*	1.1	*26*
Asia	3.5	*9*	1.8	*5*	1.7	*94*
Subtotal	20.6	*50*	19.1	*50*	1.5	*8*
Other	1.0	*3*	0.6	*2*	0.4	*67*
Total	41.3		38.0		3.3	*9*

The United States is the largest producer and consumer of sulfur in all forms. U.S. production and consumption data are provided in Table 14.3. Fertilizers and agro-chemicals represent the largest use for sulfur, 69 percent. Phosphatic fertilizers consume 7.1 million metric tons, 57 percent, in the form of sulfuric acid. Agro-chemicals account for 1.3 million metric tons, 10 percent, all in the form of elemental sulfur. Sulfur, surprisingly, represents the largest fungicide and pesticide product applied in the United States refining use, for alkylation, is the second largest segment of consumption. It is also one of the most difficult segments to track accurately. This arises from refineries not reporting a portion of their production and using that material directly for internal use. Accordingly, a portion of actual sulfur output does not get reported within the scope of "production of sulfur," causing total uses to exceed production by 1.1 million metric tons.

TABLE 14.3 U.S. Production and Consumption of Sulfur in All Forms in 200024 (Million Metric Tons)

Production		*Uses*		
Frasch	0.9	Fertilizers and agro-chemicals		
Recovered	8.4	Phosphatic	7.1	H2SO4
Other (nonelemental S)	1.0	Nitrogenous	0.2	H2SO4
Total	10.3	Other agro-chemicals	1.3	S
			8.6	
Supply				
Elemental S Produced	9.3			
emsp; Inventory Withdrawal	0.3	Petroleum refining	1.5	S
Other	1.0		0.5	H2SO4
Imports Elemental S	2.3		2.0	
Acid	0.5			
Exports Elemental S	(0.8)	Copper ores (leaching)	0.7	H2SO4
Acid	(0.1)			
Total	12.5	Inorganic chemicals	0.2	S
			0.4	H2SO4
			0.6	
Consumption				
Elemental S	11.1			
Acid/other	1.4			
All forms	12.5	Other	0.3	S/H2SO4
		Unidentified	1.4	S
		Total	13.6	

Source: Ober, J. *Sulfur-2000*, Mineral Industry Surveys, U.S. Dept. of the Interior.

U.S. production of elemental sulfur has declined dramatically in recent years due to the cessation of Frasch mining operations. Output went from 10.0 million tons in 1999, to 9.3 million tons in 2000 and 8.3 million tons in 2001. As a result Canada has become the world's largest producer of both elemental and "sulfur in all forms." From this point forward, U.S. output should rise although re-injection of acid gasses might lead to occasional declines. Such a project is underway at an ExxonMobil gas processing installation in Wyoming.

14.2 Sulfuric Acid

Sulfuric acid is the largest-volume chemical manufactured in the world and its consumption is often cited as an indicator of the general state of a nation's economy.28 About 41 million tons of sulfuric acid were produced in the United States in 1999, of which approximately 70 percent was used in fertilizer production. Its use extends to nearly every major chemical sector. This versatile acid is truly the "work horse" of the chemical industry.

Pure sulfuric acid is an oily, water-white, slightly viscous liquid with a melting point of 10.4°C and a boiling point of 279.6°C. It is infinitely miscible with water, forming sulfuric acid solutions characterized by their weight percent H_2SO_4. Oleum may be formed by dissolving SO_3 in sulfuric acid to attain fuming sulfuric acid, with concentrations nominally greater than 100 percent H_2SO_4. Historically, sulfuric acid concentrations were determined by measuring the solution density using hydrometers calibrated in degrees Baume (Be). This practice is waning, although some specifications and tables of properties still include this measurement.

Uses of Sulfuric Acid

The primary industrial uses of sulfuric acid are in phosphate fertilizer manufacture, petroleum refining, copper ore leaching, synthetic rubber and plastics, and pulp and paper mills.45 It is used as a solvent, a dehydrating agent, a reagent in chemical reactions or processes, an acid, a catalyst, and an absorbent, and in many other applications. In spite of its wide usage, sulfuric acid rarely is contained in the final product. Sulfuric acid ends up as gypsum in phosphate fertilizer manufacture, for example. In many other processes the sulfuric acid is converted to a waste product that requires disposal or reuse. Because disposal of waste sulfuric acid is becoming increasingly unacceptable environmentally, the recycle of sulfur values from waste sulfuric acid has become more widespread.

Nearly all sulfuric acid is manufactured by the contact process in which sulfur trioxide is absorbed into 9398 percent H_2SO_4. The acid may be sold at various strengths, usually depending on the requirements of the consumer. It generally is marketed on a 100 percent basis, but normally is shipped as 93 percent H_2SO_4 (66°Be), as 98 percent acid, or as 2022 percent fuming oleum. Table 14.3 shows common acid strengths and end uses. Concentrated acid may be stored in mild steel tanks, but dilute acid must be contained in lead-lined or plastic tanks. Bulk shipments of concentrated acid are made in steel tanks on ships, tank barges, or railcars. Reagent grade acid is commonly sold in 5-L glass bottles.

Development of the Sulfuric Acid Industry

Sulfuric acid is formed in nature by the oxidation and chemical decomposition of naturally occurring sulfur and sulfur-containing compounds. It is made by the action of bacteria (*Thiobacillus ferrooxidans*) on coal brasses or iron disulfide discarded on refuse dumps at coal and copper mines; it is produced in the atmosphere by the oxidation of sulfur dioxide emitted from the combustion of coal, oil, and other substances; and it also is formed by chemical decomposition resulting from geological changes.

Although there were vague references to "spirits" expelled from alum by Arabian alchemists in the tenth century and by Roman alchemists in the thirteenth century, the first distinct mention of sulfuric acid has been credited to Basil Valentine in the late 1400s.44 He burned sulfur with saltpeter in glass retorts or bell jars with a little water, and he also calcined copperas (ferrous sulfate heptahydrate) with silica, with both processes yielding sulfuric acid, although he took them to be different substances. These processes for making sulfuric acid continued until 1746 when Dr Roebuck constructed a lead chamber in England for sulfuric acid manufacture. This marked the beginning of the "chamber process" for sulfuric acid, which was to continue in use for the next two centuries.43

The first lead chamber was 1.8 m square, and 8:1 mixtures of sulfur and saltpeter in iron carts were rolled into it and burned with intermittent admission of air.47 As in the glass retorts, the sulfur trioxide that was formed combined with water to produce sulfuric acid, which condensed on the walls and collected in pans. Steam was introduced into the chambers in 1774, and continuous addition of air was begun in 1793. It then was recognized that the sulfurous acid from the burning of sulfur was oxidized by air and needed saltpeter only as a catalyst. In 1827 Gay-Lussac invented a tower for recovering the nitrogen oxides escaping from the chamber. The nitrogen oxides were condensed in sulfuric acid but could not be economically recovered from the acid until the invention of the Glover denitrating tower in 1859. The introduction of these two towers completed the chamber process except for various refinements to reduce costs. The process could produce acid with up to 77 percent H_2SO_4 but generally yielded strengths in a 6568 percent range. Higher-strength acid was produced by boiling chamber acid to remove water. A platinum still for producing concentrated sulfuric acid from chamber acid was first built for the Harrison Works in Philadelphia in 1814.

The developing markets for sulfuric acid in the late eighteenth century increased the demand for Sicilian sulfur. By 1832, sulfur prices had risen to $80/ton, and stocks rose; then in 1833 the market broke, with the price at $15/ton. Sicilian government attempts to stabilize sulfur at $70/ton failed, in part because of an 1833 discovery by a French chemist that sulfur dioxide could be obtained by roasting pyrites in a furnace. Process improvements in pyrite roasting by 1870 made pyrites competitive with sulfur as a raw material. By 1880, with the single exception of the United States, the sulfuric acid industries had gone to a pyrites basis. By 1909 virtually all U.S. sulfuric acid came from pyrites and as by-product acid from zinc and copper smelters. In the late 1890s, development of the Frasch process for sulfur mining lowered the price of sulfur. This change, coupled with new developments in sulfur burners, led sulfuric acid manufacturers back to sulfur as a raw material source. During World War I gypsum was burned with coal to produce SO_2 for sulfuric acid production. This process continued in Europe until the 1950s, with the last plant (in Great Britain) shut down in 1975. Today sulfuric acid production from pyrite roasting remains important in Spain, the Former Soviet Union, China, Japan, South Africa, Turkey, and some eastern European nations.42 The Bosveld No. 2 plant in South Africa operated on gypsum until the early 1980s when it was shut down.

The contact process for sulfuric acid dates from 1831, when a Briton patented a method for converting sulfur dioxide to sulfur trioxide by passing the gas through a heated tube filled with finely divided platinum. The sulfur trioxide was adsorbed in chamber acid to produce concentrated and fuming sulfuric acids. Commercialization of the new process was delayed by lack of markets for the concentrated acid and poor understanding of the process parameters. About 1870, demand for fuming sulfuric acid spurred German development of the contact process. In 1901, BASF reported that the governing principles for the successful manufacture of contact acid were well understood. These included cleaning of the sulfur dioxide gas stream, use of excess oxygen, and temperature control of the catalyst bed. However, the reliance of the contact process on feed acid produced by the chamber process limited its development until about 1930. It became known that the acid in the final SO_3 absorption tower had to be kept in the range 98.599.0 percent H_2SO_4 to be effective. Accurate dilution techniques to allow recycle of the absorber acid were developed in the late 1920s. The contact process then could continuously produce diluted acid, and no longer required chamber acid. The development of poison-resistant vanadium catalysts for SO_2 conversion allowed the use of smelter gases in the contact process. With its lower capital and operating costs, the contact process then supplanted the chamber process for sulfuric acid production. By 1940 the United States had equal numbers of chamber and contact processes. Today no commercial chamber plants are operating in the United States.

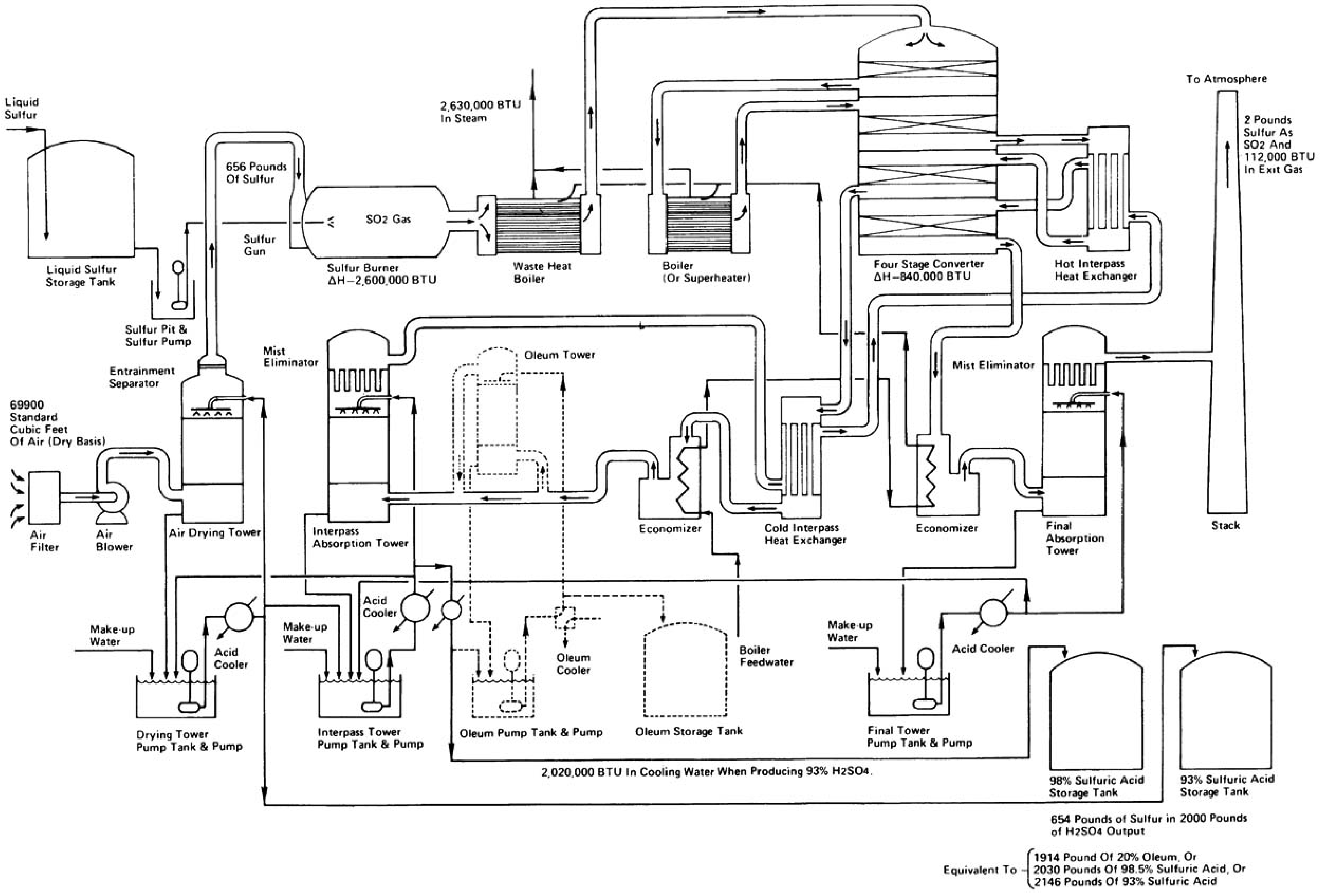
Since the 1940s most developments in the contact process have focused on energy recovery and pollution abatement. The production of SO_2, its conversion to SO_3, and the dilution of H_2SO_4 are exothermic processes that are exploited to reduce energy costs at sulfuric acid plants. Sulfur emissions have been sharply reduced by using two SO_3 absorbers, although one absorber is the economic choice. Tail gas scrubbing processes have been developed to further reduce sulfur emissions from sulfuric acid plants. No single absorber plants have been built in the United States since the 1970s, although there were still some in operation as recently as 2000. As of the year 2000, U.S. EPA regulations limited new sulfur-burning sulfuric acid plant emissions to 2 kg SO_2 and 0.075 kg acid mist per metric ton of H_2SO_4 produced, and metallurgical plants to 650 ppm volume SO_2. During the 1990s some local requirements exceeded the EPA regulations and several plants were constructed with SO_2 limited to 100 ppm SO_2 (about 0.5 kg SO_2 per metric ton H_2SO_4 produced). One double absorption plant followed by an ammonia based DynaWave O scrubber, built under Monsanto Enviro-Chem license in Ulsan, Korea, started operation in 1999 with SO_2 at 30 ppm.

Manufacture of Sulfuric Acid by the Contact Process

The basic steps in the contact process are: (1) production of sulfur dioxide; (2) cooling and, for smelters, cleaning of the process gas; (3) conversion of the sulfur dioxide to sulfur trioxide; (4) cooling of the sulfur trioxide gas; and (5) absorption of the sulfur trioxide in sulfuric acid.28 Figure 14.8 is a photograph of a contact process plant. A simplified diagram of a double absorption contact sulfuric acid process is shown in Fig. 14.9. Because sulfur dioxide is produced by several processes, it is convenient to separate the discussion of sulfur dioxide production from its conversion to sulfuric acid.

Fig. 14.8 Modern double absorption sulfuric acid plant with view of sulfur furnace in foreground. (*Courtesy Monsanto Enviro-Chem.*)

Fig. 14.9 Schematic flow diagram of a double absorption contact process for sulfuric acid manufacture.



Sulfur Dioxide Production

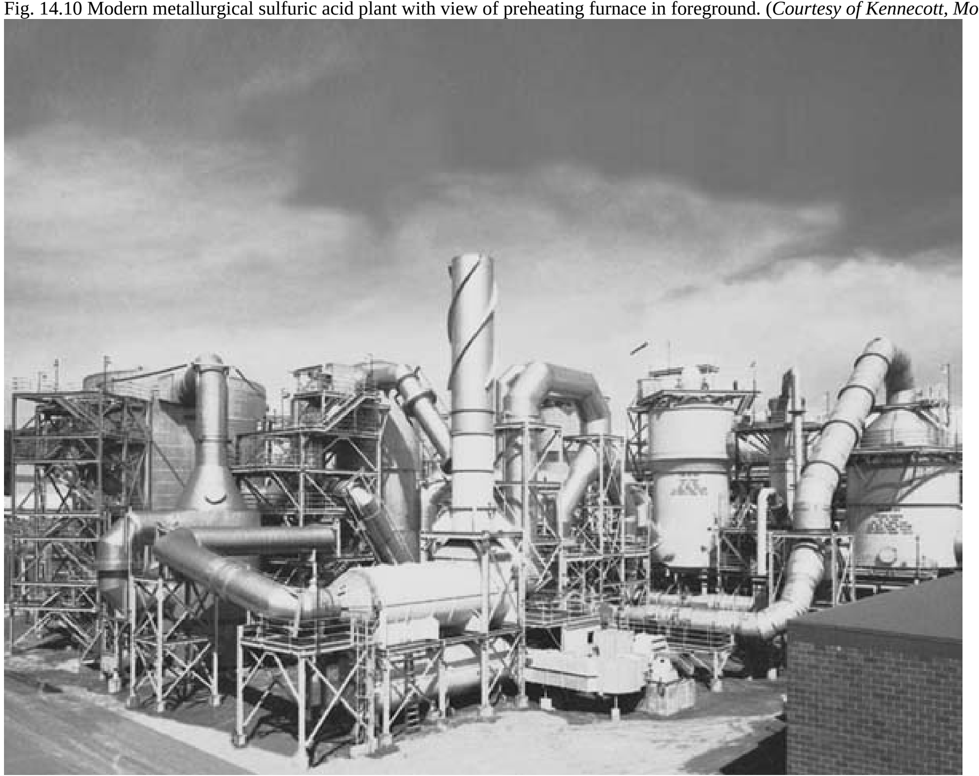
Sulfur is converted to sulfur dioxide by burning molten sulfur with dried air in a sulfur burner to yield a 10001200°C gas stream containing 1012 percent SO_2. The burner is mounted at one end of a sulfur furnace, and the gas passed through a waste heat boiler at the other end. The gas temperature is reduced to 420440°C on leaving the boiler, which produces 4060 bar steam.

In the simple pressure-nozzle burner, the liquid sulfur is atomized by pumping it at 815 bar through the nozzle. In a twocomponent burner, the sulfur is atomized primarily by the combustion air stream. It operates at lower pressure and has a wider sulfur throughput range, 5170 tons/day, than the simple burner. Lurgi's rotary burner can burn up to 400 tons/day of sulfur. Larger sulfuric acid plants, such as the Monsanto designed Anaconda plant in Western Australia, use multiple sulfur guns in a single furnace to burn 1400 metric tons per day of sulfur. A process gas with 18 percent SO_2 can be achieved in Lurgi's two-stage sulfur combustion process. However, SO_2 gas produced from burning sulfur in air at 18 percent cannot be used directly in a contact sulfuric acid plant without the addition of dilution air as there is not enough oxygen in the gas stream to react with the SO_2 to achieve a normal conversion to SO_3. In the two-stage process excess sulfur is burned in the first stage, consuming all of the oxygen. The low residual oxygen level limits the formation of nitrogen oxides, which otherwise would be very high at the combustion temperature of 1750°C. The combustion gas then is cooled to 620650°C, and additional dried air is injected to burn the residual sulfur in an afterburner. A second waste heat boiler cools the process gas to 420440°C before sending it to the converter.

Pyrites and other iron sulfides are roasted to produce an iron oxide cinder and an off-gas containing 714 percent SO_2 which is contaminated with varying amounts of arsenic, lead, zinc, and other metal oxides. The off-gas must be cleaned before it is sent to the converter. Various types of pyrite roasting equipment have been used in the past, including shaft furnaces, multiple hearth roasters, rotary kilns, and dust roasters. Fluid-bed roasters now are widely used for their superior process technology, throughput rates, and economics. The roasting process must be controlled between 850 and 940°C. At lower temperatures the reaction is incomplete, whereas at higher temperatures the iron oxides and sulfides form a eutectic melt that inhibits the reaction rate. Fluid-bed roasters surpass other types in temperature control and temperature uniformity throughout the bed. Fluid-bed technology for SO_2 production was introduced in the early 1950s by Dorr Oliver in the United States and by BASF in the Federal Republic of Germany.

Copper, lead, zinc, and other sulfide ores may be processed by roasting or smelting. Roasting or sintering of sulfide ore is essentially identical with pyrite roasting. Sulfur melting generally occurs at higher temperatures. Older reverberatory furnace smelters produce off-gas with only 12 percent SO_2, too low for its economical recovery as acid. By using an oxygen-enriched air feed, the off-gas can be raised to 68 percent SO_2. Bath smelters (Mitsubishi, Noranda), where the ore concentrates are heated and reacted in the slag/matte melt, produce an off-gas with 1020 percent SO_2.[25] Flash smelters (Inco, Outokumpu), which involve suspension and reaction of the concentrates in an oxidizing gas stream, operate at 12001300°C and produce a waste gas with 1015 percent SO_2. Oxygen enrichment of the feed air can raise the SO_2 level in the off-gas to 3080 percent. The Kivcet process smelts with pure oxygen and produces off-gas with 8085 percent SO_2.[26] Normally strong gases are diluted with air to 14 percent SO_2 in order to limit outlet temperatures in the converter first pass to less than 640°C and to provide sufficient oxygen to convert the SO_2 to SO_3. In 1996 the original Olympic Dam sulfuric acid plant in South Australia was modified to operate with 18 percent equivalent SO_2 gas strength. This plant used a Monsanto preconverter and a cesium-promoted catalyst.[27] Figure 14.10 is a photo of a metallurgical gas sulfuric acid plant.

Fig. 14.10 Modern metallurgical sulfuric acid plant with view of preheating furnace in foreground. (*Courtesy of Kennecott, Monsanto Enviro-Chem, and Manly Prim Photography*)



Off-gas from roasting and smelting operations may contain dust, SO_3, halogens, NO_x, arsenic and other toxic metal fumes, and mercury.[28,29] These components must be removed from the gas stream before it is sent to the converter. Although SO_3 is produced in the converter, its presence in the cooled gases upstream of the converter will cause excessive corrosion by forming sulfuric acid mists. The cleaning plant steps are: (1) hot-gas dedusting; (2) wet scrubbing; (3) gas cooling; (4) mist removal; and, if necessary, (5) mercury removal. The hot gases generally are passed first through a waste heat boiler to reduce the temperature to 250400°C. Cyclones followed by hot-gas electrostatic precipitators (5090 kV) remove nearly all (99+%) of the dust. The gas then is contacted with weak (530%) H_2SO_4 in an open spray tower which removes metal vapors and additional solids, cools the gas to 5080°C, and converts SO_3 to acid mist for later removal. If higher levels of particulate removal are required, a venturi, Swemco O, or DynaWave O scrubbers may be used. The saturated gas is next cooled in a packed tower or shell and tube heat exchanger to condense excess water. Shell and tube heat exchangers of graphite or alloy construction are generally used only in smaller size sulfuric plants or where the gas temperature has a tight approach to available cooling water. Silica packing is used in the towers, or sodium silicate is added to the weak acid circuits to remove fluorides as fluosilicic acid. In cases where the fluoride levels are very high, additional liquid-gas contacting stages are provided to reduce the gas phase concentration of fluorides. Acid mist is removed in wet electrostatic precipitators fitted with lead tubes and star wires or, more recently, with FRP or PVC tubes specially treated to maintain a conductive liquid film surface (Lurgi)[30] and composite wires fabricated with barbs to promote corona discharge.[29] If present, mercury is removed in an additional tower by scrubbing with mercury chloride solution (Boliden),[31] hydrochloric acid solution, or 7085 percent sulfuric acid.

Waste sulfuric acid sludges from petroleum refineries are disposed of by conversion to SO_2 for production of fresh sulfuric acid. The heavy organic components of the sludges can be decomposed thermally at 8001300°C (Lurgi, Monsanto Enviro-Chem) or reductively at 200600°C with coal in a rotary kiln (Chemico). Thermal decomposition is accomplished in a fuel-fired vertical or horizontal furnace. The acid sludge also can be injected into fluid-bed pyrite roasters as a means of disposing of the acid and reducing fossil fuel consumption. Dilute acid sludges must be concentrated to 6075 percent H_2SO_4 for economical conversion. This is generally done by using waste heat from the decomposition process. The sulfur dioxide gas stream from these processes requires cleaning, as described above for roasting and smelting plants.

Calcium sulfate may be decomposed to cement clinker and sulfur dioxide gas in a coke-fired rotary kiln at 9001400°C (Miiller Kiihne).[28] However, the unfavorable economics of this process relegate it to countries that do not have other sources of sulfur. Phosphogypsum (gypsum produced by the acidulation of phosphate rock) may be decomposed in this way as a means of recycling the sulfur values in the large waste phosphogypsum piles at fertilizer plants (OSWKrupp and FIPR/Davy McKee). This process is hampered by the more extensive gas cleaning requirements for decomposing phosphogypsum as compared with natural gypsum. During the early 1980s there was some interest in recycling phosphogypsum. By 2000, because of the high cost of these processes, there was little commercial interest. Environmental forces also are behind the recycling of ferrous sulfate from metals industry pickling liquors. This "green salt" is decomposed to sulfur dioxide and iron dioxide in pyrite roasters. Elemental sulfur, coal, or fuel oil may be used as supplementary fuels.

Refinery waste gases may be burned to eliminate hydrogen sulfide and other sulfur-containing contaminants. Streams containing small amounts of H_2S or constituents unsuitable for Claus plants may require combustion to SO_2 as the means of disposing of the toxic gas. The resulting effluent gas usually is low in SO_2 and contains water vapor and carbon dioxide. Flue gases from fossil fuel power plants also fall into this category. Recovery of the sulfur values from these dilute gases usually is driven by environmental considerations rather than economics. In the United States, power plant flue gas often is scrubbed with lime to convert the SO_2 to gypsum sludge for landfill disposal. Alternatively, the Bergbau Forschung process recovers sulfur dioxide by dry adsorption on activated coke at 130°C. The SO_2 is released by heating the coke to 600650°C. Sulfur dioxide can be absorbed in a sodium sulfite solution (WellmanLord) to produce sodium bisulfite. Pure moist sulfur dioxide can be recovered by heating the sodium bisulfite.

Single vs. Double Contact Process

The single absorption contact process for sulfuric acid is characterized by four main process steps: gas drying, catalytic conversion of SO_2 to SO_3, absorption of SO_3, and acid cooling. The maximum SO_2 conversion for a single absorption plant is about 97.598 percent. By adding a second SO_3 absorber with one or two catalyst beds between absorbers, the SO_2 conversion can be increased to 99.599.8 percent or even as high as 99.9 percent with a cesium-promoted catalyst, resulting in lower SO_2 emissions. The so-called double absorption process is now the industry standard.

If water vapor is present in the gas stream or the gas temperature or metal surface temperatures drop below the dew point, liquid acid is formed by condensation of H_2SO_4 vapor rather than by absorption of SO_3. Therefore, the SO_2 laden process gas sent to the converter must be dry to protect the downstream process equipment against corrosion. The drying generally is done in a packed tower with recirculating concentrated (9398%) sulfuric acid kept at 5060°C by indirect cooling. The tower acid stream is heated by condensation of the water and by dilution of the acid. The tower temperature is used to control the moisture level of the gas sent to the converter. Acid from the drying tower is cross flowed to the absorber or is sent to storage tanks for shipment. When sulfur is burned, the combustion air to the sulfur burner is dried because the combustion of sulfur does not produce water. Off-gases from pyrite roasters and metallurgical smelters are dried as part of the gas cleaning process. For roaster gases with low SO_2 concentrations, a predryer may be added upstream of the main dryer.

Oxidation of SO_2

Oxidation of SO_2 to SO_3 is accomplished in multi-stage, fixed-bed catalytic converters equipped with interstage boilers or heat exchangers to remove the heat of reaction. Typically, four stages are compartmented within a single vertical converter, which may be brick-lined, steel, or cast iron. Newer converters are stainless steel, and some have five stages for higher conversion. Isothermal tubular converters are no longer suited to modern high-capacity plants. The extruded cylindrical catalyst pellets are usually 49 percent V_2O_5 with alkali metal sulfate promoters on a silica carrier (diatomaceous earth, silica gel, or zeolites). The reaction temperature for vanadium catalyst is generally 410440°C. In the late 1980s a cesium-promoted catalyst became commercially available from Topsoe and Monsanto. These low-strike catalysts operate at 360400°C. Higher temperatures (•600°C) reduce the SO_2 conversion and lead to structural damage of the catalyst. High-pressure drops across catalyst beds from catalyst dust formed during processing require periodic catalyst removal and screening to remove dust. Ring-shaped catalysts developed by Topsoe and others have lower dust pressure drops and are now in wide use. Other catalyst shapes used are ribbed rings and cylinders. The usual catalyst loading per one ton/day sulfuric acid capacity is 150200 L in a double absorption plant and 200260 L in a single absorption plant. Bayer developed and operates fluid-bed converters that utilize special 0.31 mm abrasion-resistant catalysts.

Absorption of SO_3

Sulfur trioxide from the converter is absorbed in 98 percent H_2SO_4 recirculated countercurrently through a packed tower maintained at 6080°C by indirect cooling. The optimum concentration of the absorber acid is near the $H_2OH_2SO_4$ azeotrope, 98.3 percent H_2SO_4, where the SO_3, H_2SO_4, and H_2O vapor pressures are at their lowest values. Absorption efficiencies in excess of 99.9 percent generally are obtained. On leaving the converter, the process gas is cooled first with feed gas in a gasgas heat exchanger and then with a water feed preheater to 180220°C before it enters the absorber. An impingement separator, or Teflon or glass fiber mist eliminator, is placed in the top of the absorber to remove acid mists. If oleum is produced at the plant, it is made in a separate oleum tower upstream of the absorber. A portion of the SO_3 stream to the absorber is diverted to the oleum tower where it is absorbed in a recirculating stream of oleum.

In double absorption plants an additional absorber is placed between the second and third (or between the third and fourth) converter beds. By removing SO_3 from the gas stream at this intermediate point, higher SO_2 conversions are attained in the downstream converter beds, and the overall SO_2 conversion is increased. The cooled gas from the intermediate absorber is reheated by hot converter gas in gasgas heat exchangers before returning to the converter. An oleum tower may be placed before the intermediate absorber.

Acid Cooling

Absorption of SO_3 in concentrated sulfuric acid and the formation of H_2SO_4 from SO_3 and H_2O produce heat in the absorber, as does acid dilution from the addition of makeup water. Process control requires that the acid be cooled before it is recirculated to the dryer or absorber towers or sent to storage. Earlier acid coolers of parallel banks of stacked, irrigated, cast iron sections have been largely replaced by stainless steel shell and tube or plate exchangers, with or without anodic protection. Hastelloy, Sandvik SX, ZeCor, and Saramet alloys and Teflon linings are also used in acid piping and coolers.[3235]

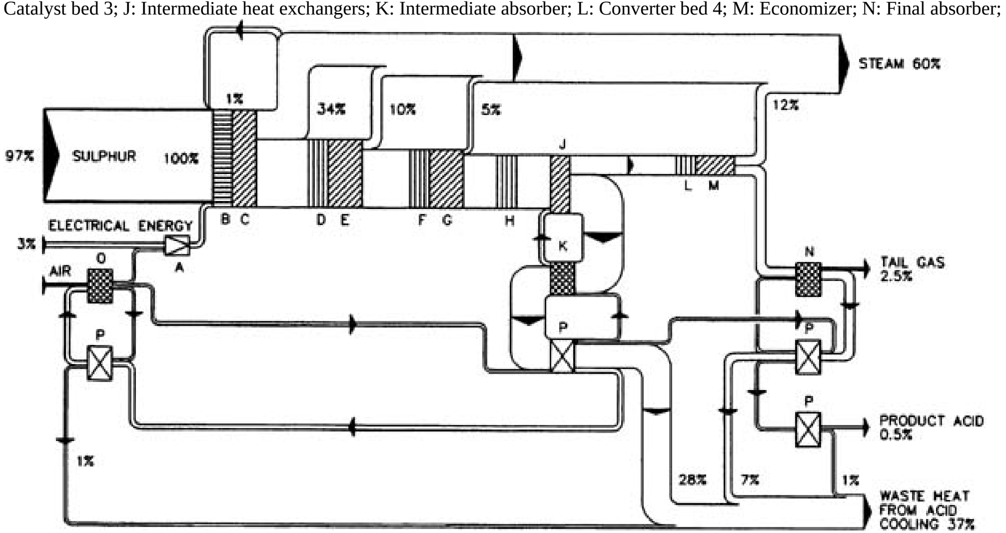
Tail gas emissions are controlled by improving the SO_2 conversion efficiency and by scrubbing the tail gas. In a double absorption process plant, a five-bed converter has 0.3 percent unconverted SO_2, as compared with 0.5 percent for a four-bed converter. A Lurgi Peracidox scrubber may be used to remove up to 90 percent of the residual SO_2 in the tail gas from a double absorption plant. Hydrogen peroxide or electrolytically produced peroxymonosulfuric acid is used to convert the SO_2 to H_2SO_4 in the Lurgi scrubber.

Other Modifications to the Sulfuric Process[36]

Tail gas from single absorption plants may be absorbed on activated carbon (Sulfacid) or scrubbed with ammonia (Monsanto AMMSOX) or sodium sulfite (Wellman Lord). Metallurgical acid plants differ from sulfur-burning plants in that the cleaned SO_2 process gas must be heated before it is sent to the converter. Many of these plants have weak SO_2 streams that require large gasgas heat exchangers for temperature control. Four plants in the Former Soviet Union processing 24 percent SO_2 use an unsteady-state oxidation process in which the cold (4070°C) SO_2 gas is reacted on hot catalyst beds without intervening heat exchangers.[37,46] As the temperature front moves through the bed to the exit side, the flow is reversed. Cycle times are 30120 min, and single-bed conversions of 8090 percent are reported as compared with 5560 percent for conventional processes at higher exit temperatures.

The thermal capacity of a 1000 ton/day sulfuric acid plant is about 63 MW. This heat liberation must be controlled in a manner that maintains optimum gas temperatures in the converter system and optimum acid temperatures in the dryer and absorber circuits. Tail gas emissions also are affected by the energy balance. Figure 14.11 shows an energy flow diagram for a contemporary sulfur-burning sulfuric acid plant. About 97 percent of the total energy input derives from burning sulfur, and 3 percent comes from the electricity consumed to drive the gases through the plant. Most plants can recover 5560 percent of the energy as high-pressure steam (4060 bar, 400480°C), but about 40 percent is lost as waste heat dissipated to the environment in the form of hot water from acid coolers.

Fig. 14.11 Sankey energy flow diagram for a 1000 ton/day sulfur-burning double absorption sulfuric acid plant (feed gas 10% SO_2). A: Blower; B: Sulphur furnace; C: Waste heat boiler; D: Catalyst bed 1; E: Steam superheater; F: Catalyst bed 2; G: Boiler; H: Catalyst bed 3; J: Intermediate heat exchangers; K: Intermediate absorber; L: Converter bed 4; M: Economizer; N: Final absorber; O: Air dryer; P: Acid coolers. (*Courtsey Lurgi GmbH, Frankfurt, Germany*)



During the late 1970s acid plants were optimized to generate more steam. Steam can be produced at pressures up to 80 bar from high-temperature sulfur burners such as the Lurgi two-stage combustion system. The high-pressure steam is let down to low pressure steam through a turbogenerator that co-generates electricity. In double absorption plants, economizers were installed upstream or downstream of the heat exchanger servicing the intermediate absorber. Boiler feed water is preheated to 9095°C in this economizer to increase steam production. The energy production from acid plants was increased to 70 percent by installing low-gas-temperature economizers, lowpressure-drop catalysts, and suction drying towers, by increasing the SO_2 feed gas concentration, and by preheating the boiler feed water with hot acid. Further energy recovery requires higher operating temperatures for the absorbers and acid coolers. Venturi concurrent absorbers operating at acid temperatures of 130140°C are installed in several plants. To recover acid heat directly as steam, it was necessary to increase the acid temperatures from 110120°C to about 200°C.

As of 1999 Monsanto Enviro-Chem had built 19 Heat Recovery System (HRS) units since demonstrating the first HRS at Namhae Chemical in South Korea in 1987.[38,39] The process is based on 310 stainless steel, which resists corrosion in 98.5 percent H_2SO_4 at temperatures up to 220°C. The intermediate absorber at Namhae takes 194°C gas from the converter third stage economizer and absorbs the SO_3 in 199°C, 99 percent acid. Recirculated acid from the absorber is cooled from 220°C in a 10-bar HRS boiler. The added energy recovery for this process is reported to increase the total recovery to 90 percent. Monsanto's proposed Monarch process combines HRS technology with the wet catalytic converter process (Lurgi) to increase heat recovery and shift it to high-pressure steam production for electric power generation.[36]

Other Sources of Sulfuric Acid

Spent sulfuric acid usually is diluted in the process in which it is used: titanium dioxide pigment processing, plastics manufacture, and so on. The dilute acid may be used in processes requiring dilute acid or may be concentrated for reuse by a number of vacuum evaporation processes (SimonsonMantius, Chemetics)[40] or by thin-film evaporation (DuPont, Bofors). In the submerged combustion distillation process, water is evaporated from the dilute acid by forcing hot flue gases from a fuel-fired burner below the acid surface (Chemico). The concentration of 75 percent acid to 9598 percent H_2SO_4 by the PaulingPlinke process is done by feeding the 75 percent acid to a stripping column fitted with a stirred cast iron pot mounted in a furnace. The acid concentration in the pot must be kept above 80 percent to minimize corrosion.

Chemetics has developed a process for treating spent alkylation sulfuric acid with nitric acid to produce a sulfuric acid that can be used to acidulate phosphate rock, the major use for sulfuric acid. The organic contaminants are converted to carbon particles that are removed with the gypsum on filtration of the phosphoric acid. Special alloys are used in the fabrication of the acid reactor. Topsoe developed and, by the year 2000 had built, more than 30 Wet Sulfuric Acid (WSA) process units. This process is especially suited for low-strength, less than 4 percent, SO_2 gas streams which would not be auto thermal or meet water balance conditions in the conventional dry sulfuric acid contact process. In the WSA process wet SO_2 gases pass through converter beds where the SO_2 is oxidized to SO_3. The SO_3 reacts with water vapor to form H_2SO_4 in the gas phase. The acid is condensed in proprietary WSA condensers. Sulfuric acid is produced at concentrations around 96 percent.[41]

Production and Consumption of Sulfuric Acid

The world production of sulfuric acid (1999 and 1989) is given in Table 14.4. There was virtually no change in the global use of sulfuric acid during the decade. But major regional shifts did take place. Asia, driven by the Chinese push towards self-sufficiency in phosphate fertilizers and increased manufacturing activity, nearly doubled its use during the decade. In Africa, Morocco added fertilizer capacity, and H_2SO_4 use was up 41 percent, although elemental sulfur use was unchanged during the start and end of the decade. Significant changes also affected the raw materials used to manufacture sulfuric acid.

TABLE 14.4 World Production of New Sulfuric Acid for 1999 and 1989 (Million Metric Tons 100% H_2SO_4)[42]

	1999		1989		Difference	
	Elemental Sulfur	Total	Elemental Sulfur	Total	Elemental Sulfur	Total
Western Europe	8.5	17.5	12.3	24.9	-3.8	-7.4
France	1.4	2.2	3.5	4.1	-2.1	-1.9
Germany	1.2	3.0	1.6	3.7	-0.4	-0.7
Spain	0.3	2.8	0	3.7	0.3	-0.9
Central Europe and FSU	10.3	15.6	24.5	38.5	-14.2	-22.9
Poland	1.2	2	2.5	3.1	-1.3	-1.1
FSU	8.3	12	18.7	28.3	-10.4	-16.3
North America	37.6	46.2	37.8	44.6	-0.2	1.6
Canada	1.6	4.8	1.2	3.5	0.4	1.3
U.S.	33.2	37.1	33.3	37	-0.1	0.1
Mexico	2.8	4.3	3.4	4.1	-0.6	0.2
Latin America	5.6	10.2	4.3	5.7	1.3	4.5
Brazil	4.0	4.8	3.0	3.7	1.0	1.1
Africa	17.1	18.1	10.0	12.8	7.1	5.3
Morocco	8.5	8.5	4.1	4.1	4.4	4.4
Tunisia	4.8	4.8	3.2	3.2	1.6	1.6
South Africa	2.5	3.1	1.1	3.1	1.4	0
Asia	16.5	42.0	12.7	30.7	3.8	11.9
China	4.3	21.6	1.0	11.4	3.3	10.2
Japan	2.2	6.9	1.2	6.9	1.0	0
India	5.2	6.1	3.6	3.9	1.6	2.2
Oceania	1.1	2.5	1.2	2.0	-0.1	0.5
World Total	102.8	158.7	102.8	159.2	0	-0.5

Source: Con-Sul, Inc. and Freeport Sulfur Marketing Department estimates.

A large shift occurred from pyrite roasting to acid recovered for environmental reasons. Almost all nations reduced their use of pyrites as a result of environmental considerations. China's increase was only temporary: by the late 1990s that nation was joining the rest of the world in shifting from pyrites to elemental sulfur. The production of sulfuric acid as a result of pollution abatement regulations represents the industry's only growing segment, with a 30 percent gain over the decade. That trend should continue. The closing of three copper smelters in the Southwestern United States will, however, reduce SOF production by around 1 million tons (Table 14.5).

TABLE 14.5 Sulfuric Acid Production from Pyrites and Other Forms (Million Tons 100% H_2SO_4)

	1999		1989		Difference	
	Pyrites	SOF	Pyrites	SOF	Pyrites	SOF
World	16.0	40.0	25.4	31.1	-9.4	8.9
Western Europe	2.1	6.9	6.9	5.7	-4.8	1.2
Eastern Europe FSU	1.0	4.3	5.9	8.1	-4.9	-3.8
North America		8.6		6.3		2.3
Asia	12.3	13.4	10.3	7.6	2.0	5.8

Source: U.S. Geological Survey, Con-Sul, Inc., Freeport Sulfur Marketing Department.

Table 14.6 lists the sulfuric-acid-consuming industries in the United States and shows the trends in their acid consumption rates through the 1980s and 1990s. Agriculture further increased its dominant use of sulfuric acid, accounting for 77 percent in 1999 vs. 70 percent in 1989. Even so, lower phosphatic fertilizer production reduced sulfuric acid use in agriculture by almost 10 percent. The U.S. economic shift from manufacturing to services and higher valued products is highlighted by the 37 percent decline in sulfuric acid consumed by nonagriculture industries.

TABLE 14.6 Sulfuric-Acid-Consuming Industries in the United States (Millions of Metric Tons, 100% H_2SO_4)

Consuming Industries	1999	1989	1984	1979
Agriculture	24.5	26.9	26.8	24.1
Phosphatic	23.7	26.4	26.4	23.2
Nitrogenous	0.6	0.3	0.3	0.5
Pesticides	0.1	0.1	0.1	0.2
Other agricultural	0.1	0.1	0.1	0.2
Other industries	7.2	11.5	11.1	14.1
Petroleum refining	1.6	2.1	2.1	2.4
Synthetic rubber and plastic		1.2	1	0.6
Rayon and cellulose		0.2	0.2	0.3
Other chemicals	1.4	2.7	3.2	4.2
Copper ore processing	2.2	1.8	1.0	2.1
Uranium and other ore processing	0.3	0.4	0.4	
Iron and steel		0.2	0.3	0.9
Other metals	0.2	0.2	0.2	0.1
Pulp and paper	0.4	1.0	0.8	0.8
Pigments and paint	0.5	0.4	0.3	0.6
Other	0.9	1.6	1.8	1.6
Total	31.7	38.4	37.8	38.1

Source: U.S. Geological Survey and U.S. Bureau of Mines: Sulfur Annual Reports.

Nearly all the sulfuric acid consumed in agriculture was reacted with phosphate rock (principally $Ca_9(PO_4)_6CaF_2$) to produce phosphoric acid. Granular phosphate fertilizers are produced by ammoniating phosphoric acid to yield mono- and diammonium phosphates; ammonium phosphate is also produced as a fertilizer. Petroleum refining is the largest nonfertilizer use for U.S. sulfuric acid. The acid competes with hydrogen fluoride as a catalyst in petroleum alkylation reactions for gasoline production. Sulfuric acid acts as a catalyst in synthetic rubber and plastics manufacture. Copper ore leaching is carried out by distributing the acid over leach piles of the ore and collecting the copper-rich leachate for processing. Sulfuric acid from nearby smelters is normally used in copper ore leaching.

14.3 References

British Sulphur Corp: Ltd, Sulphur, No. 193, 3336 (Nov./Dec. 1987).
International Fertilizer Industry Association, Paris, France, 2001.
(a) Sulfur Magazine, No. 259, p. 41, Nov.Dec. 1998, Gerrie du Plessis. (b) Hyne, J. B., "The Forming, Handling and Transportation of Elemental Sulphur," Presented at Sulphur 1990 Conference, The British Sulphur Corporation, London.
(a) Ibid, 45 (b). (b) Kemp, E., Hyne, J. B., and Rennie, W. J., "Reaction of Elemental Sulfur with Water Under U.V. Radiation," Internat. J. Sulfur Chem., Part A1(1), 6970 (971).
(a) Personal communication, Pennzoil Sulphur Co. Operations at Antwerp terminal. (b) Dowling, N. J., Clark, P. D., and Hyne, J. B., "Understanding and Mitigating Corrosion During Handling and Transportation of Elemental Sulfur," Sulphur 1996 Conference, The British Sulphur Corporation, London.
Sandvik Process Systems, Inc., 21 Campus Rd., Totowa, NJ, USA.
Berndorf Belt Systems USA, 920 Estes Avenue, Schaumburg, IL, USA.
Devco International, Inc., 8801 So. Yale Ave., Tulsa, OK, USA.
Enersul, Inc., 7210 Blackfoot Tr. SE, Calgary, Alberta, Canada.
Ibid. 51.
KaltenbachThuring SA, 9 Rue de l'Industrie, 6000 Beauvais, France.
d'Aquin, G. E., Transporting Sulfur Pellets, U.S. Patent 6368029, Apr. 9, 2002.
Hyne, J. B., "Some Impurities in Elemental Sulphur-Origins and Elimination," Sulphur 1991 Conference, The British Sulphur Corp, London.
Thieler, E., Sulphur, Theodor Steinkopff, Leipzig, 1936.
In 2001, Iraq and Poland were the only nations with operating Frasch mines.
d'Aquin, G., "North American Sulphur Perspective," paper read at Sulphur 1998 Conference, British Sulphur Corporation, London.
d'Aquin, G., North American Quarterly Sulphur Review, Con-Sul, Inc., Tulsa, OK (Apr. 2002).
British Sulphur Corp. Ltd, Sulphur, No. 193, 2630 (Nov./Dec. 1987).
British Sulphur Corp. Ltd, Sulphur, No. 192, 3039 (Sept./Oct. 1987).
Rameshi, M., "State of the Art in Gas Treating," Sulphur 2000 Conference, The British Sulphur Corporation, London.
Ober, J., Sulfur2000 Annual Report, Mineral Industry Surveys, U.S. Geological Survey, U.S. Dept of the Interior, Reston, VA.
Ober, J., Sulfur1991 Annual Report, Mineral Industry Surveys, U.S. Bureau of Mines, U.S. Dept of the Interior.
Ibid., 1991 and 2000.
Ibid., 2000.
Sohn, H. Y., George, D. B., and Zunkel, A. D., Advances in Sulfide Smelting, The Metallurgical Society of AIME, Warrendale, PA, 1983.
British Sulphur Corp. Ltd, Sulphur, No. 190, 3340 (May/June 1987).
Bhambri, N., Fell, R. C., Fries, R. M., and Ritschel, P. M., "Metallurgical Sulfuric Acid Plants for the New Millennium," Sulphur 98 Conference, British Sulphur Corporation, London.
Sander, U. H. F., Fischer, H., Rothe, U., and Kola, R. Sulphur, Sulphur Dioxide, and Sulphuric Acid, Verlag Chemie, Weinheim, 1984.
Friedman, L., "The Sulfuric Acid Plant Recent AdvancesReviews and Analysis," paper read at the AIChE Central Florida Meeting, Clearwater, FL, June 2, 1988.
British Sulphur Corp. Ltd, Sulphur, No. 203, 2432 (July/Aug. 1990).
British Sulphur Corp. Ltd, Sulphur, No. 207, 3741 (Mar./Apr. 1990).
Davies, M., Hodgson, D. S., and Rodda, J. "Applications of SARAMET in H2SO4 Plants," paper read at Sulphur 88, Vienna, Austria, Nov. 69, 1988.
Douren, L., "Sandvik SXThe Future Material for Concentrated Sulfuric Acid," paper read at Sulphur 87, Houston, TX, Apr. 58, 1987.
Holdschick, H., and Jurascheck, M. A., "Advances in the Development of Cast Materials for High Concentrated Sulphuric Acid at Temperatures up to 280° C," paper read at Sulphur 88, Vienna, Austria, Nov. 69, 1988.
Puricelli, S., Fell, R. C., and Randolph, D. R., "The State of Sulfuric Acid Technology," AlChE meeting, Clearwater, FL, May, 2000.
McAlister, D. R., Grendel, R. W., Schneider, D. R., Shafer, J. R., and Tucker, J. S., "A Sulfuric Acid Plant for the 1990s," paper read at Sulphur 1990, Cancun, Mexico, Apr. 14, 1990.
Blicharz, M., "Different Methods of Sulphur Removal from Gas from Non-ferrous Metal Works," paper read at Sulphur 88, Vienna, Austria, Nov. 69, 1988.
Shafer, J. R., "Namhae Heat Recovery System Update," paper read at Fertilizer Industry Round Table, Baltimore, MD, Nov. 15, 1988.
Smith, R. M., Sheputis, J., Kirn, U. B., and Chin, Y. B., "Sulfuric Acid Heat Recovery System (HRS) Operations at Namhae Chemical Corporation, Korea," paper read at Sulphur 88, Vienna, Austria, Nov. 69, 1988.
Lebel, G., "Pilot Plant Development of the Chemetics TiO2 Waste Acid Recovery Process," paper read at Sulphur 1990, Cancun, Mexico, Apr. 14, 1990.
British Sulphur Corp. Ltd, Sulphur, No. 253, 6769 (Nov./Dec. 1997).
The British Sulphur Corp. Ltd, World Sulphur and Sulphuric Acid Atlas, 5th ed., Purley Press Ltd, London, 1989.
Duecker, W. W., and West, J. R. (Eds.), The Manufacture of Sulfuric Acid, Van Nostrand Reinhold, New York, 1959.
Fairlie, A. M., Sulfuric Acid Manufacture, Reinhold, New York, 1936.
Heydorn, B., Aguiar, D., and Ferguson, A., "Sulfuric Acid," in Chemical Economics Handbook, SRI International, Menlo Park, CA, 1987.
Matros, Y. S., and Bunimovich, G. A., "Reverse Process of SO2 Oxidation in Sulfuric Acid Production," paper read at Sulphur 1990, Cancun, Mexico, Apr. 14, 1990.
Wyld, W., Sulphuric Acid and Sulphur Dioxide Raw Materials, D. Van Nostrand, New York, 1923.

15.1 The Nature of Petroleum

Petroleum is a diverse mixture of hydrocarbonschemical combinations of primarily hydrogen and carbon. Complete combustion of hydrocarbons yields the end products of carbon dioxide (CO_2) and water (H_2O). However, incomplete combustion results in a composite mixture of other products such as CO_2, H_2O, carbon monoxide (CO), and various oxygenated hydrocarbons. Since burning petroleum consumes air, nitrogen compounds are also formed. In addition, other elements are associated with hydrocarbon compounds such as sulfur, nickel, and vanadium.

Petroleum is found at great depths underground or below seabeds. It can exist as a gas, liquid, solid, or a combination of these three states, which is very common. Drilling efforts are used to reach and extract gaseous and liquid deposits. These products are brought to the surface via piping. Once found in a reservoir, gas usually flows under its own pressure. Conversely, discovered liquid hydrocarbons may flow on their own due to pressure from the reservoir or may be forced to the surface by submerged pumps. Also injection of fluids and gases provides a driving force to push liquid hydrocarbon through rock strata. Solid or semi-solid petroleum is brought to the surface through several methods: by digging with conventional mining techniques, by gasifying or liquefying with high-temperature steam, or by burning a portion of the material in the ground so that the remainder can flow to the surface.

Hydrocarbon Forms

As mentioned earlier, petroleum is any product that is primarily composed of hydrogen and carbon bonded compounds. These compounds can be further categorized by their characteristics.

Natural gas is the gaseous form of petroleum. It is mostly the single-carbon moleculemethane (CH_4). When natural gas is associated with liquid petroleum underground, the methane will come to the surface in admixture with some heavier hydrocarbons. The gas is considered a wet gas; the heavier hydrocarbons are isolated and purified in natural-gas processing plants. Gas processing yields ethane (an important petrochemical feedstock), propane [liquefied petroleum gas (LPG)], butane (refinery blending stock), and hydrocarbon liquids (natural gas condensate). When the underground natural gas is associated with solid hydrocarbons such as tar or coal, the methane will have few other hydrocarbons and is considered a dry gas.

Petroleum, Petroleum refining *Crude oil* is the common name for liquid petroleum. In some literature, one will see reference to "petroleum and natural gas," suggesting petroleum and crude oil are used as synonymous terms. Some crude oils have such a great density that they are referred to as heavy oils and tars.

Tar sands are small particles of sandstone surrounded by an organic material called bitumen. The bitumen is a highly viscous hydrocarbon that clings tenaciously to the sandstone; thus, it is easy to think of the mixture as a solid form of petroleum. Yet, it is a mixture of high-density liquid on a supporting solid.

Oil sands are true petroleum solids. Curiously, oil sands do not contain petroleum crude oil; it is an organic material called kerogen. The kerogen can be heated to yield a liquid called shale oil, which can be refined into conventional petroleum products.

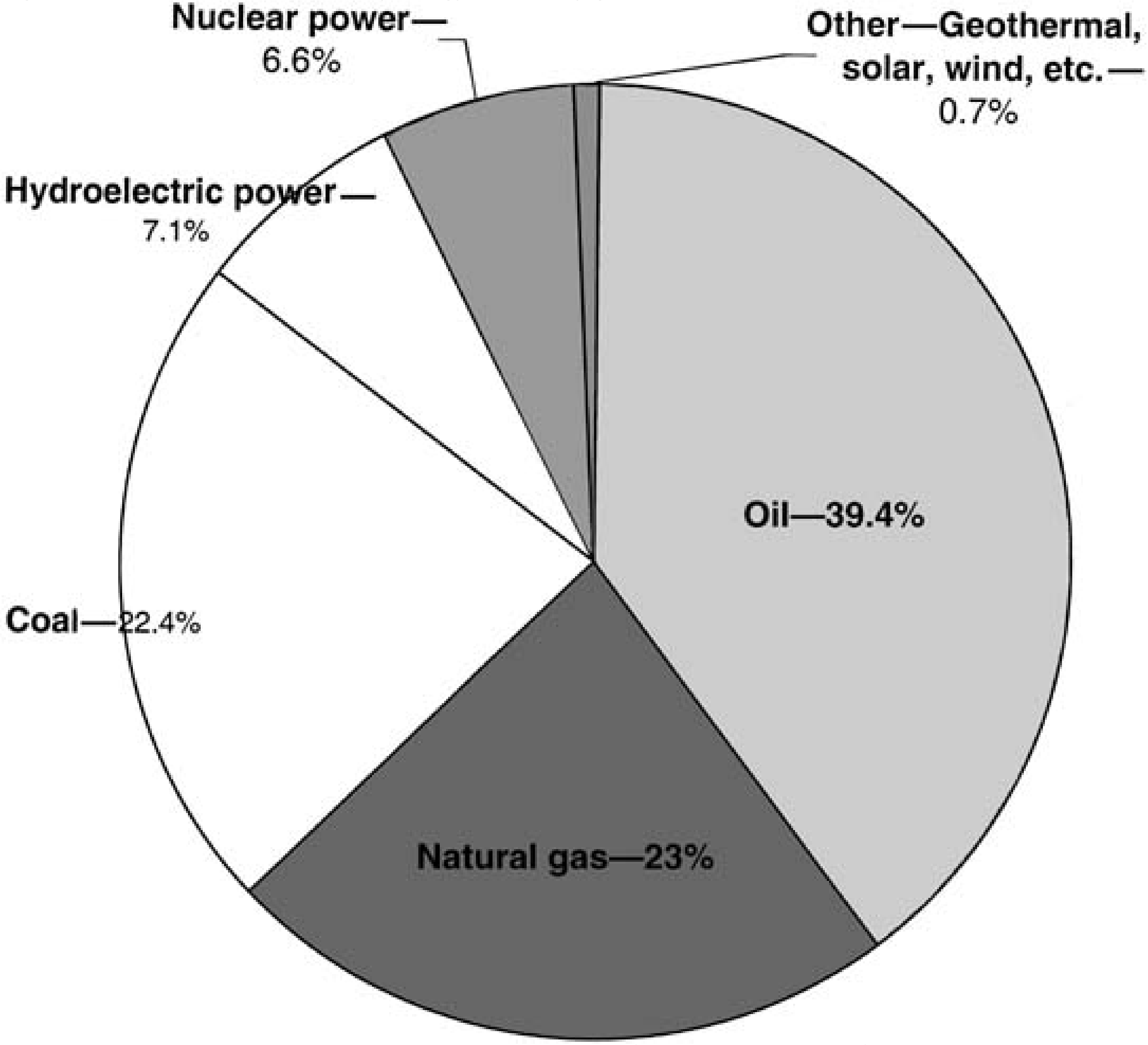
Largest Energy Supplier

In Chapter 17 the point is made that coal offers an abundant primary energy source. Yet, present and proposed environmental legislation deter future coal usage. Due to stringent stack-emission restrictions for power generation, utilities are seeking "cleaner fuel" options to replace coal. Utility companies view natural gas as the "cleaner fuel" option. Petroleum feedstocks contain sulfur, which is strictly regulated on emission permits. Natural gas usage is growing within the power/utility industry. Yet, petroleum remains the major fuel source used in transportation, manufacturing, and home heating.

Primary energy sources are defined as those sourced to natural raw materials. Electricity is not included because it is a secondary energy source; it is generated by consuming one or more of the other natural energy sources. To put petroleum consumption into perspective, the primary energy sources considered here are: petroleum crude oil, natural gas, coal, hydropower (water to generate electricity), and nuclear energy. The quantities reported here will exclude energy from wood, peat, animal waste, and other sourcesdespite their importance to some localities.

The common practice is to relate energy units to a common productin this case, to petroleum liquid. For example, world consumption of crude oil and liquids (condensates) from natural gas in 1999 reached 149.72 Quadrillion (1015 BTUs)Quad. If the amount of energy from other sources were converted to equivalent barrels of oil, the total world energy consumption in 1999 would be 380 Quads.1 The relative distribution of these sources is shown in Fig. 15.1. More energy comes from oil than from any other single source.

Fig. 15.1 World production of primary energy selected by groups.



Another view to consumption is that the world consumption of crude oil and liquids from natural gas in 2001 reached 76 million barrels per day (MMbpd). North America is the largest energy consumer at 24.1 MMbpd. The AsiaPacific regional demand is steadily increasing. In 2001, AsiaPacific comprised 27.2 percent of the world's oil demand, up from 20 percent in 1990.2 Crude oil's share of primary energy comsumption was 39.4 percent in 2001. Thus, petroleum oil and natural gas remain the steadfast energy sources globally, as shown in Fig. 15.1.2

From Well to Refinery

Crude oil production for various countries is shown in Fig. 15.2.2 The Middle Eastern countries produce more oil than they consume; the extra production is gated for export. Conversely, the United States and Western Europe consume much more crude oil than they produce (Fig. 15.3).2 This condition demonstrates the great importance of worldwide petroleum movements. The difference between production and consumption for any one period is balanced by oil being added to or removed from extensive areas around the world.

Fig. 15.2 World crude oil production by producing region. (*Cambridge Energy Research Associates. Reprinted with permission from Global Oil Trends 2002, by CERA, Copyright 2001, all rights reserved.*)

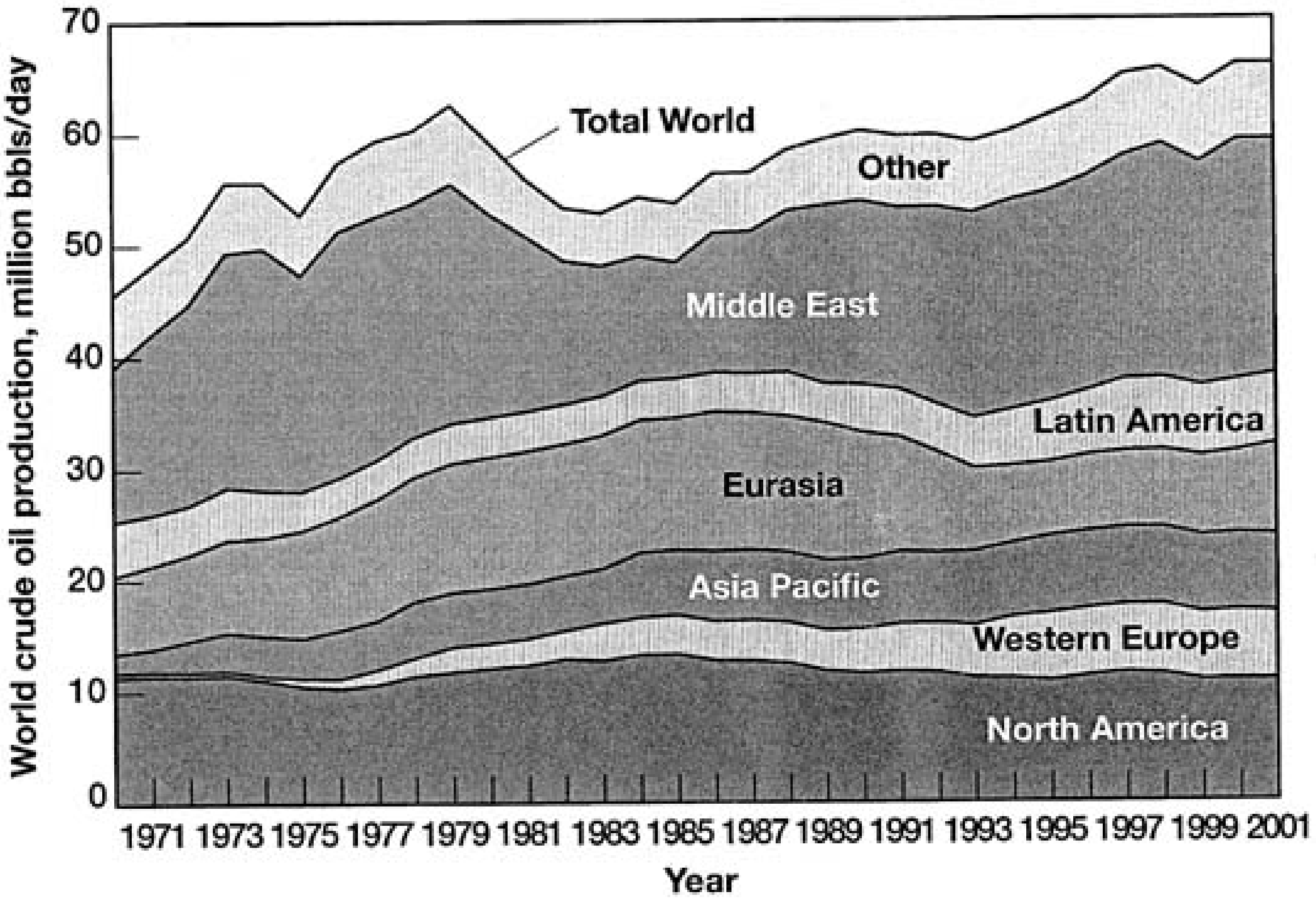
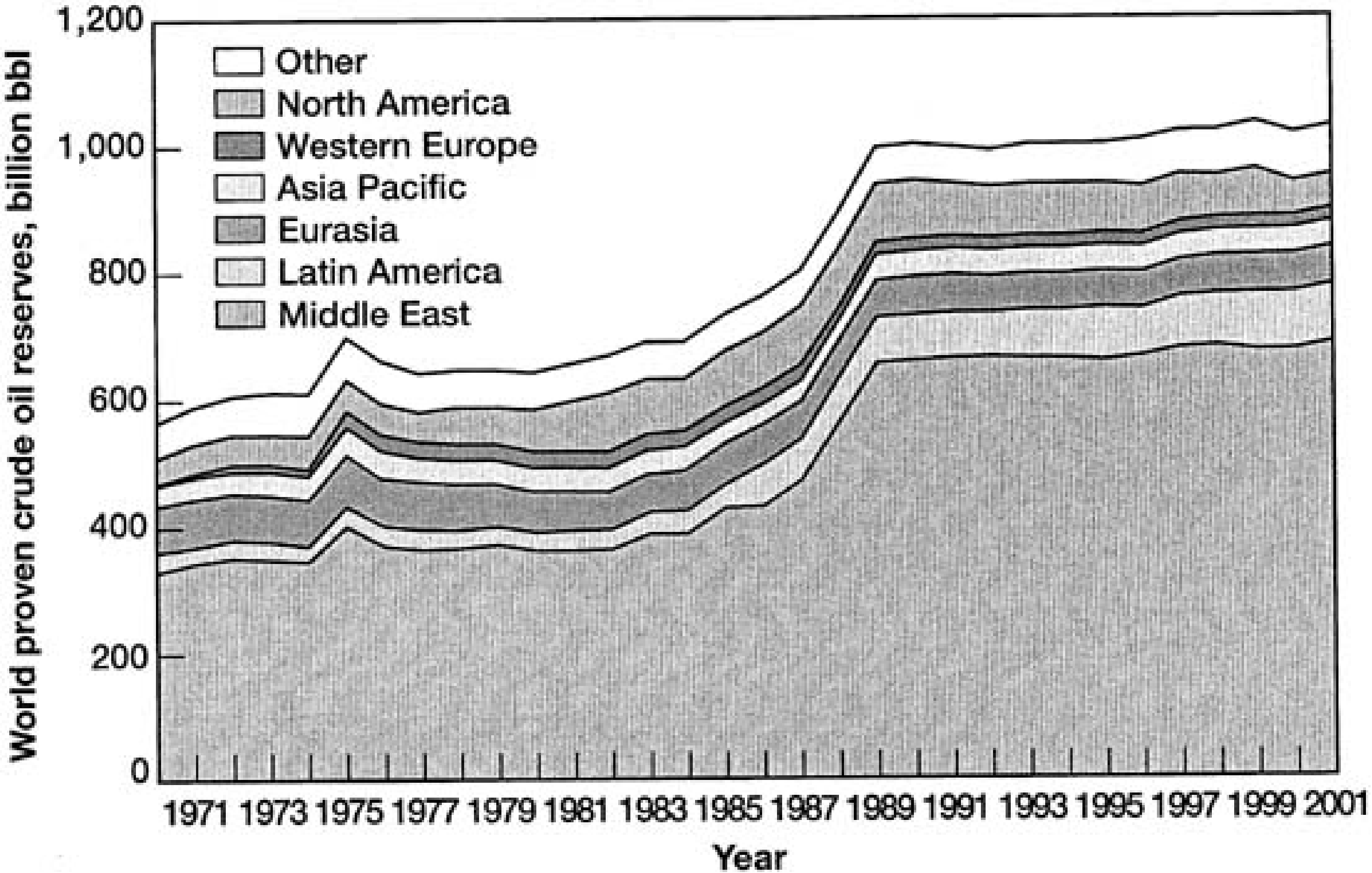


Fig. 15.3 World proven crude oil reserves. (*Cambridge Energy Research Associates. Reprinted with permission from Global Oil Trends 2002, by CERA, Copyright 2001, all rights reserved.*)

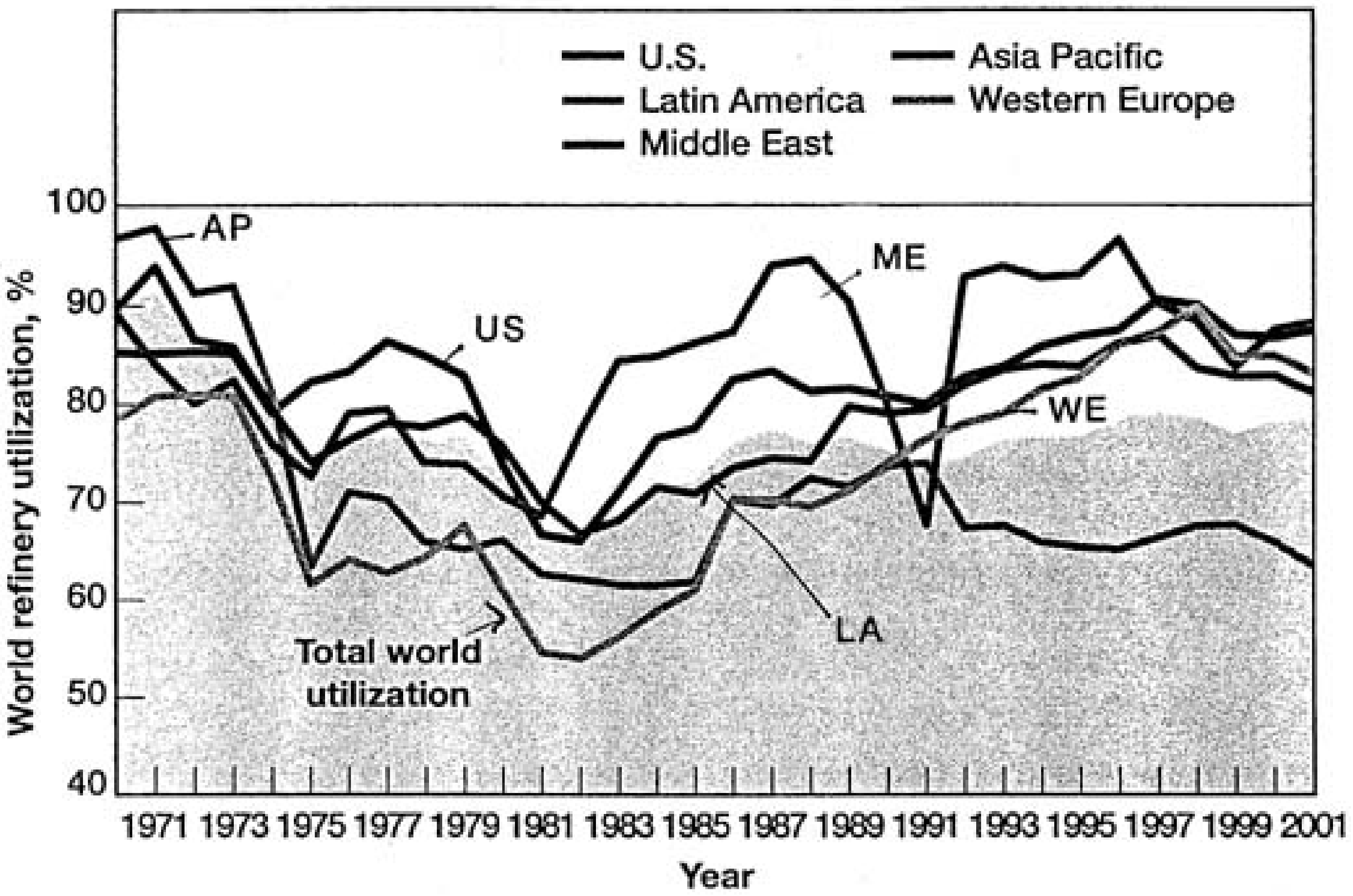


The growth of world refining capacity attempts to keep pace with rising demand for petroleum-based products.2 Curiously, refining capacity has surged ahead is some regions, notably the Middle East and the AsiaPacific region, over the last 30 years. However, in developed markets, refinery throughput was almost flat during the 1980s and 1990s. New and larger state-of-the-art refining facilities were more energy efficient and had lower operating costs per barrel of refined products. Consequently, smaller refiners could not compete against new facilities and subsequently had to shut down operations. Another factor in refining growth is the time to construct processing units. In highly industrialized countries such as the United States, Japan, and Western European countries, mounting environmental regulations and stiff emission and performance laws have all but stifled the construction of new grassroots refineries. Construction of new refining capacity must overcome a long list of federal, state, and local governmental requirements. In the United States, the last grassroots refinery was constructed in 1974, before the onslaught of the Clean Air Act of 1970. New governmental regulations are focused on strict reductions on emissions (air, water, and solids) to improve air quality for high-density populated areas and high-density industrial regions. Consequently, construction of new facilities in developed markets is negligible.

In industrialized nations, new refining capacity will be realized by the expansion of existing facilities that are permitted by the local regulatory agencies. Even expanding existing facilities is exempted from environmental constraints. With new construction and capacity expansion, operating companies must cut emissions below present permitted levels. Thus, operating companies must install more intrusive emission reduction/control technologies and equipment to eliminate release from new and existing plant equipment.

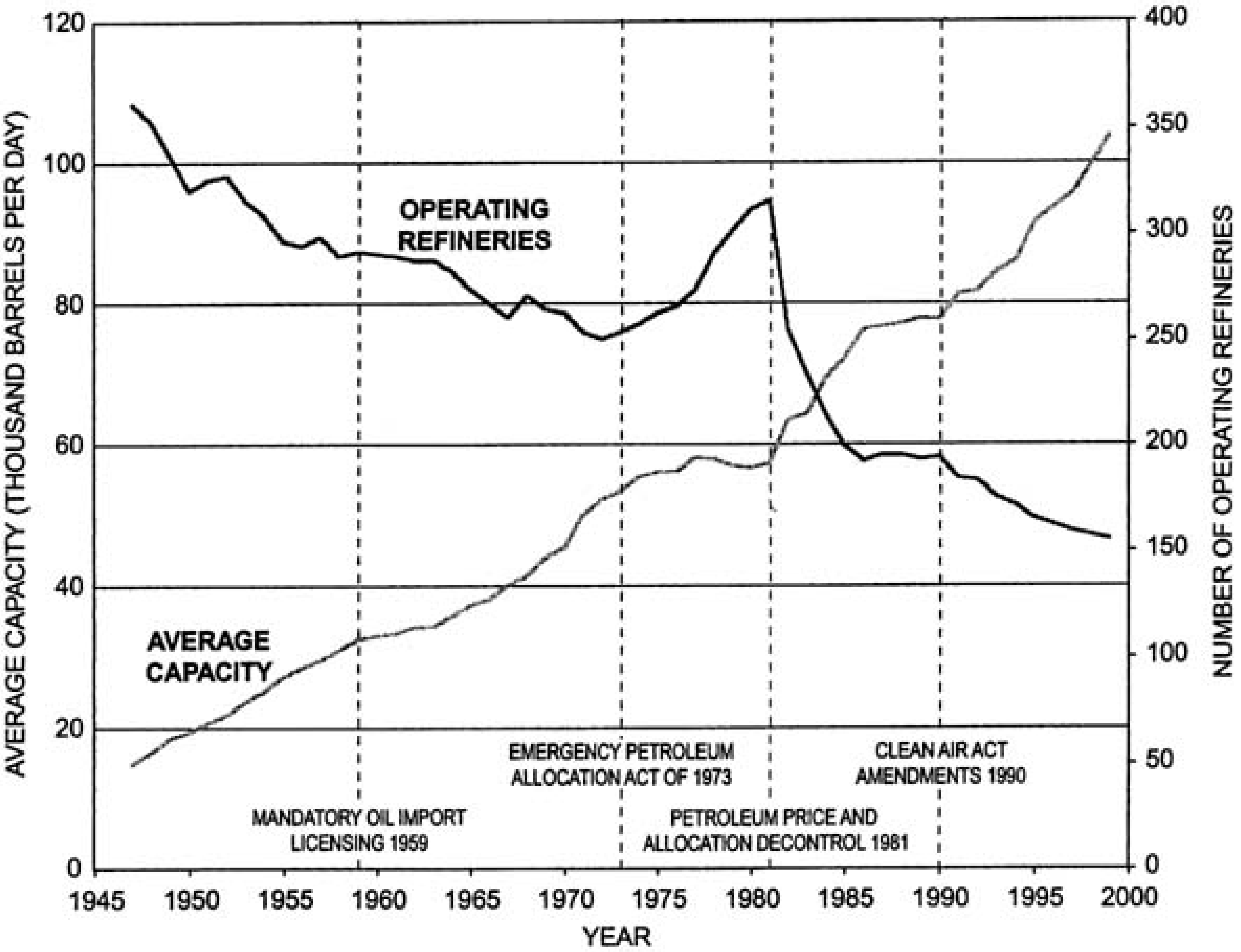
Refiners have become particularly adept in using technology to find incremental capacity from existing processing equipment. Thus, the refining industry can process more crude oil with present equipment. In spite of this, the number of refineries is decreasing; yet, capacity increases incrementally. Advancements and innovations in processing methods and catalyst systems have enabled construction of large, highly integrated and complex refining complexes. New materials of construction have aided in building larger facilities; thus, smaller, older refineries cannot compete with newer, larger facilities. The number of operating facilities continues to decrease; yet capacity rises. Less efficient facilities shut down their operations since they are handicapped in producing refined products (Fig. 15.4).

Fig. 15.4 World refinery utilization as percent. (*Cambridge Energy Research Associates. Reprinted with permission from Global Oil Trends 2002, by CERA, Copyright 2001, all rights reserved.*)



In 1984, the number of refineries operating in the United States peaked at 318 facilities with a refining capacity of 18.62 MMbpd.3 Technological advancements in processing methods and catalyst systems have enabled refiners to increase the capacity of existing units incrementally. Newer processing units are larger than the earlier versions. Refiners are applying economies-of-scale to disperse the product costs of refineries. Since 1981, the number of U.S. operating refineries has decreased to 155 in 2000 with a total operating capacity of 16.52 MMbpd, as shown in Fig. 15.5. Notably, smaller and less competitive refineries were shut down. Equally important, the utilization of operating U.S. refineries rose over this same period, from a low of 68.6 percent to 92.6 percent.3

Fig. 15.5 Average capacity and number of U.S. operating refineries. (*Source: National Petroleum Council*, U.S. Petroleum Refining: Assuring the Adequacy and Affordability of Cleaner Fuels, *June 20, 2000, Copyright by National Petroleum Council 2000; all rights reserved*.)



Technology helps refiners to push the boundaries of manufacturing, especially in the average size refinery. In 1975, the averaged U.S. refinery had an operating capacity of 60,000 barrels per day (bpd). Innovations in catalyst technology and equipment design enabled the construction of larger vessels and reactors, and the introduction of ancillary equipment to support processing operations. Thus, in 2000 the average refining capacity for a U.S. refinery exceeded 100,000 bpdnearly double the capacity from 1975. In the United States, refiners have avoided constructing grassroots facilities to meet rising demand for products.

As the wave of environmental regulations continues to be levied against the refining industry, more consolidation is anticipated. More companies will leave segments of fuel manufacturing due to diminishing returns from environmentally driven projects.

Distribution of Crude Oil and Refined Products

Crude oil and its refined products are viewed as commodity products; thus, they are easily traded and transported to market. Many methods can be used to deliver crude to oil refineries. For example, U.S. refinery can receive feedstock crude oil via pipelines, tank trucks, barges, and ocean-going vesselssupertankers. Most refineries are located near port facilities to receive feedstocks from barges and supertankers.

15.2 Product

Refining crude oil involves breaking carbon-to-carbon (CC) bonds of the complex hydrocarbon compounds and adding hydrogen. Such efforts are done thermally and catalytically. The distinction between refined products and petrochemicals often is subtle. In general, when the product fraction is from crude oil that includes a fairly large group of hydrocarbons, the fraction is classified as a refined product. Examples of refined products are: gasoline, diesel fuel, heating oil, lubricants, waxes, asphalts, and petroleum coke.

By contrast, when the product from crude oil is limited to only one or two specific hydrocarbons of fairly high purity, the fraction is called a petrochemical. Examples of petrochemicals are: ethylene, propylene, benzene, toluene, and styrene. Refined products are defined by the fraction's boiling point and may be composed of various hydrocarbons. Many compounds compose refined-product fractions. In contrast, petrochemicals are single-compound fractions, which are required for feedstocks for other petrochemicals and polymers. More processing and separation (distillation, extraction, etc.) operations are used to extract petrochemical products from processing streams. Thus, more identifiable petrochemical products are processed as compared to the categories of refined products. Many specific hydrocarbon compounds can be derived from crude oil. However, these hydrocarbons lose their individual identity when they are grouped together as a refined product.

Refined Products

Most refined products at the consumer level are blends of several refinery streams. Product specifications determine which streams are suitable for a specific blend. Part of the difficulty in learning about refining lies in the industry's use of stream names that are different from the names of consumer products. Refining is considered a "dirty" processing effort. Product separation of refined product streams is not as clean as efforts to process petrochemicals. Refiners have the flexibility to blend final fuel products. Thus, there is no exact recipe used by all refiners to produce consumer products. Multiple crude oils are processed and then blended to meet consumer fuel product specifications. Table 15.1 lists the refining streams that are blended to produce consumer products. The consumer products are familiar. However, within a refinery these products are blended from portions of crude oil fractions from the listed reforming process units. To complicate the situation further, not all refineries are configured identically. Many different processing operations can be used to refine and separate product streams to blend the products listed in Table 15.1.

TABLE 15.1 Several Names for the Same Material

Crude Oil Cuts	Refinery Blends	Consumer Products
Gases	Still gases	Fuel gas
	Propane/Butane	Liquefied petroleum gas (LPG)
Light/heavy naphtha	Motor Fuel	Gasoline
	Aviation turbine, Jet-B	Jet fuel (naphtha type)
Kerosine	Aviation turbine, Jet-A	Jet fuel (kerosine type)
	No. 1 Fuel oil	Kerosine (range oil)
Light gas oil	Diesel	Auto and tractor diesel
	No. 2 fuel oil	Home heating oil
Heavy gas oil	No. 4 fuel oil	Commercial heating oil
	No. 5 fuel oil	Industrial heating oil
	Bright stock	Lubricants
Residuals	No. 6 fuel oil	Bunker C oil
	Heavy residual	Asphalt
	Coke	Coke

For example, gasoline at the consumer level may be called benzol or petrol, depending upon the country where it is sold. In the early stages of crude oil processing, most gasoline components are called naphthas. Kerosine is another example. It may be called coal oil to denote that it replaces stove oil (or range oil) once derived from coal. Historically, kerosine gained significant importance as a replacement for whale oil for lamps. In the early 1900s, refining efforts were directed to supplying kerosinelamp oil. However, in the 1920s, a new energy formelectricitybegan to displace kerosine usage. Thus, early refiners sought other products to compensate for this market loss.4 Today, kerosine fractions are directed to jet fuel and high-quality No. 1 heating oil.

Product Specifications

Product applications and customer acceptance set detailed specifications for various products properties. In the United States, the American Society for Testing Materials (ASTM) and the American Petroleum Institute (API) are recognized for establishing specifications on both products and methods for testing. Other countries have similar referee organizations. For example, in the United Kingdom it is the Institute of Petroleum (IP). In Germany, it is the Deutsches Institute suer Normung (DIN). In Japan, it is the Ministry of International Trade and Industry (MITI).

Boiling range is the major distinction among refined products, and many other product properties are related to the products in these boiling ranges. A summary of ASTM specifications for fuel boiling ranges is given in Table 15.2.5

TABLE 15.2 Major Petroleum Products and their Specified Boiling Range5

		Specified Temperature. for Vol. % Distilled at 1 atm (°F)		
Product Designation	ASTM Designation	10%	50%	90%
Liquefied petroleum gas (LPG)	D 1835			
Commercial propane		a		b
Commercial butane		a		c
Aviation gasoline (Avgas)	D 910	167 max	221 max	275 maxd
Automotive gasoline	D 439			
Volatility class A		158 max	170250	374 maxe
Volatility class B		149 max	170245	374 maxe
Volatility class C		140 max	170240	365 maxe
Volatility class D		131 max	170235	365 maxe
Volatility class E		122 max	170230	365 maxe
Aviation turbine fuel	D 1655			
Jet A or A-1		400 max		f
Jet B		g	374 max	473 max
Diesel fuel oil	D 975			
Grade 1-D				550 max
Grade 2-D				540640
Grade 4-D			not specified	
Gas turbine fuel oil	D 2880			
No. 0-GT		h	not specified	
No. 1-GT				550 max
No. 2-GT				540640
No. 3-GT			not specified	
No. 4-GT			not specified	
Fuel oil	D 396			
Grade No. 1		419 max		550 max
Grade No. 2		h		540640
Grade No. 4			not specified	
Grade No. 5			not specified	

Grade No. 6	not specified

aVapor pressure specified instead of front end distillation; b95% point, -37°F max; c95% point, 36°F max; dFinal point, 338° F max; eFinal point, all classes, 437°F max; fFinal point, 572°F max; g20% point, 293°F max; hFlash point specified instead of front end distillation.

Boiling range is also used to identify individual refinery streamsin a later section we use the example of crude oil distillation. The temperature that separates one fraction from an adjacent fraction will differ from refinery to refinery. Factors influencing the choice of cut-point temperature include: crude oil feedstocks, type and size of downstream processes, and market demand for products.

Other specifications can involve either physical or chemical properties. Generally, these specifications are stated as minimum or maximum quantities. Once a product qualifies to be in a certain group, it may receive a premium price by virtue of exceeding minimum specifications or by being below maximum specifications. The only advantage of being better than the specifications is an increase in the volume of sales in a competitive market.

The evolution of product specifications will, at times, appear sadly behind recent developments in more sophisticated analytical techniques. Certainly, the ultimate specification should be based on how well the product performs. Yet, the industry has grown comfortable with certain comparisons, and these standards are retained for easier comparison with earlier products. Thus, it is not uncommon to find petroleum products sold under an array of tests and specificationssome seemingly measuring similar properties.

It is behind the scenes that sophisticated analytical techniques prove their worth. These techniques can identify the specific hydrocarbons responsible for one property or another. Suitable refining processes are devised to accomplish a desired chemical reaction that will increase production of specific hydrocarbon products.

When discussing refining schemes, major specifications will be identified for each product category. It should be kept in mind that a wide variety of specifications must be met for each product.

Product Yields

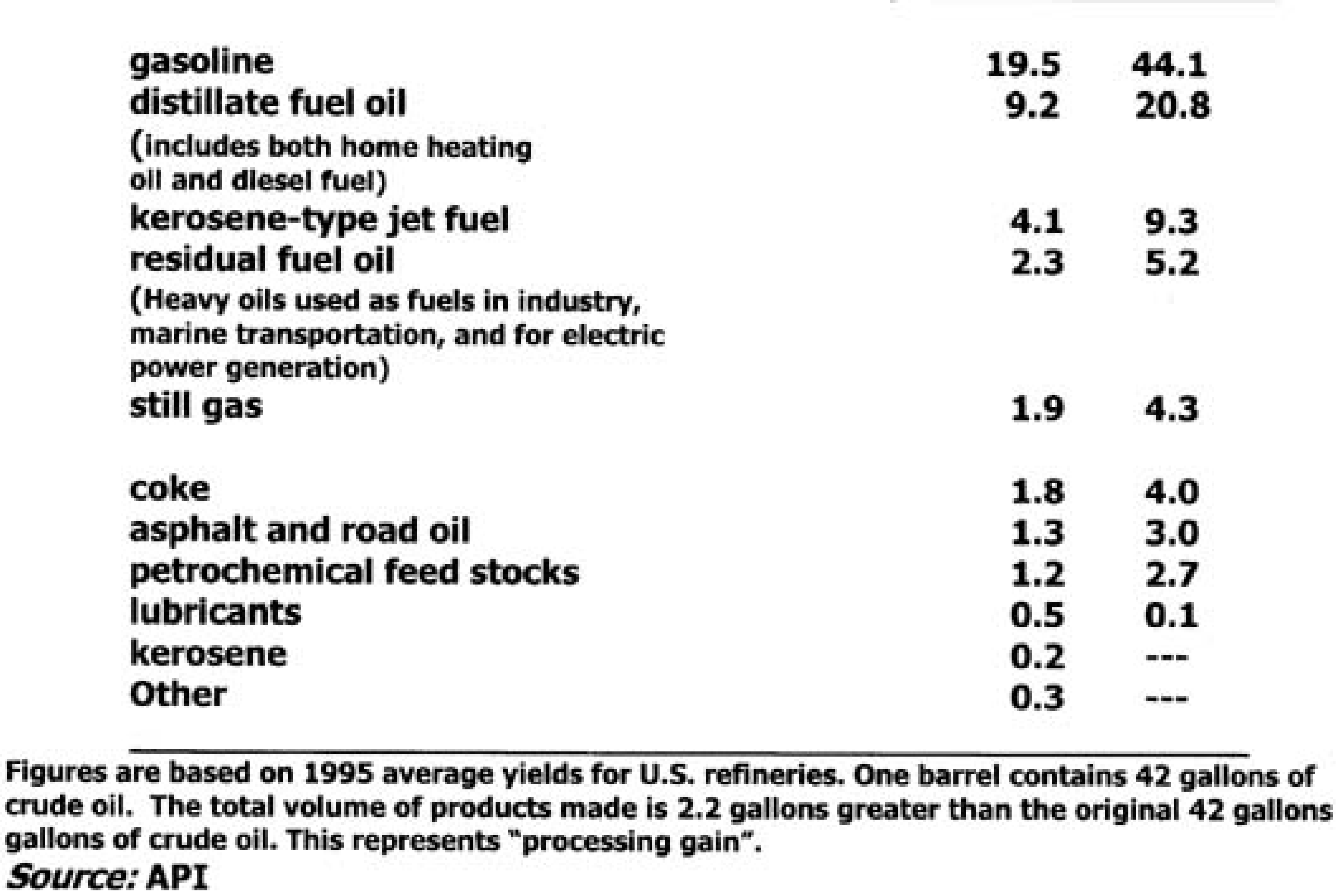
As changes occur in product demand and specifications, refiners continuously adjust the configuration of internal processing streams. The challenge remains that increasing the volume of one fraction of crude oil processing will lower volumes of other product fractions. Thus, adjustments of one processing stream, especially major processing units, impact downstream processing streams and end-product volumes.

Refined product demand is seasonal. Demand for heating oil is higher during winter than during mid-summer. Equally important, gasoline demand fluctuates from summer highs, known as the driving season, and then declines in fall and winter. Refiners begin ramping up gasoline production over heating oil in early spring to meet anticipated demand and have sufficient gasoline supply in the system for distribution. Refiners try to avoid storing products.

Notably, fuel specifications for industrialized countries mandate blending winter- and summer-grade gasolines. Regions that do not meet air-quality specificationsknown as nonattainment areasare strictly regulated on the sale and distribution of the proper gasoline types. Refiners constantly estimate how much of a particular gasoline type to blend without overprocessing. Linear program (LP) models are extensively used to evaluate how best to use a crude stock to process designated products with the available processing capabilities. LPs are models that refiners can use to predict product yields with available resources of the refinery. They are gaining increased importance to assist refiners in optimizing resources to avoid waste and maximize yields. These models are extensively used to estimate how changes in operating conditions, feeds, and new processing units/equipment will affect facility operations.

A barrel of crude oil has limited value, if any, to consumers. Its true value is the number of value-added products that can be extracted from the crude oil using various chemical reactions and separation methods. Thus, the refining operation is the first step in the transformation of crude petroleum oil into consumer products. So what are the possible products from a barrel of crude oil? Figure 15.6 lists the average breakdown of a barrel of oil by a U.S. refinery. As shown in Fig. 15.6, over 75 percent of the products yield from a refined barrel of oil is fuel based. In this example, U.S. refineries are focused on gasoline production, while European refineries focus on diesel product. Yet, refineries can also produce value-added petrochemicals for adjacent facilities.

Fig. 15.6 Product breakdown from one barrel of crude. (*Source: API.*)



Product	Gallons per Barrel	%
gasoline	19.5	44.1
distillate fuel oil (includes both home heating oil and diesel fuel)	9.2	20.8
kerosene-type jet fuel	4.1	9.3
residual fuel oil (Heavy oils used as fuels in industry, marine transportation, and for electric power generation)	2.3	5.2
still gas	1.9	4.3
coke	1.8	4.0
asphalt and road oil	1.3	3.0
petrochemical feed stocks	1.2	2.7
lubricants	0.5	0.1
kerosene	0.2	---
Other	0.3	---

Figures are based on 1995 average yields for U.S. refineries. One barrel contains 42 gallons of crude oil. The total volume of products made is 2.2 gallons greater than the original 42 gallons gallons of crude oil. This represents "processing gain".

***Source:* API**

Petrochemicals

The portion of crude oil going to petrochemicals may appear small compared with the volume of fuels yielded by refining operations; however, the variety of petrochemicals is large. Table 15.3 lists the many products derived for petrochemical applications. Many of these products are described in Chapter 22. A few are included here since they compete with the manufacturing of fuels. Despite their variety, all commercially manufactured petrochemicals account for the consumption of only a small part of the total crude oil processed.

TABLE 15.3 Petrochemical Applications

1. Absorbents
2. Activators
3. Adhesives
4. Adsorbents
5. Analgesics
6. Anesthetics
7. Antifreezes
8. Antiknocks
9. Beltings
10. Biocides
11. Bleaches
12. Catalysts
13. Chelating agents
14. Cleaners
15. Coatings
16. Containers
17. Corrosion inhib.
18. Cosmetics
19. Cushions
20. De-emulsifiers
21. Desiccants
22. Detergents
23. Drugs
24. Drying oils
25. Dyes
26. Elastomers
27. Emulsifiers
28. Explosives
29. Fertilizers
30. Fibers
31. Films
32. Finish removers
33. Fire-proofers
34. Flavors
35. Food supplements
36. Fumigants
37. Fungicides
38. Gaskets
39. Hair conditioners
40. Heat transfer fluids
41. Herbicides
42. Hoses
43. Humectants
44. Inks
45. Insecticides
46. Insulations
47. Lacquers
48. Laxatives
49. Odorants
50. Oxidation inhib.
51. Packagings
52. Paints
53. Paper sizings
54. Perfumes
55. Pesticides
56. Pharmaceuticals
57. Photographic chemicals
58. Pipe
59. Plasticizers
60. Preservatives
61. Refrigerants
62. Resins
63. Rigid foams
64. Rust inhib.
65. Safety glass
66. Scavengers
67. Stabilizers
68. Soldering flux
69. Solvents
70. Surfactants
71. Sweeteners
72. Synthetic rubber
73. Textile sizings
74. Tire cord

15.3 Refining Schemes

A refinery is a complex processing methodology involving a massive network of vessels, reactors, distillation columns, rotating/compression equipment, heat exchangers, and piping. The total scheme can be subdivided into a number of unit processes. In what follows, only the major flow streams will be shown, and each unit will be depicted by a simple block in the flow diagram.

Refined products establish the order in which each refining unit will be introduced. Only one or two key product(s) specifications are used to explain the purpose of each unit. Nevertheless, the reader is reminded that the choices between several types of units and size of these units are complicated economic decisions. The trade-offs among product types, quantity, and quality will be mentioned only to the extent that they influence the choice of one type of processing technology over another.

Feedstock Identification

Each refinery has its own range of preferred crude oil feedstocks from which a desired product portfolio can be obtained. Crude oil typically is identified by its source country, underground reservoir, or some distinguishing physical or chemical property. The three most frequently specified properties are density, chemical characterization, and sulfur content.

API *gravity* is a contrived measure of density:

$$\text{API} = \frac{141.5}{sp\ gr} - 131.5$$

where *sp gr* is the specific gravity, or the ratio of the weight of a given volume of oil to the weight of the same volume of water at a standard temperature, usually 60°F. An oil with a density the same as that of water, or with a specific gravity of 1.0, would then be a 10°API oil. Oils with a higher than 10°API are lighter than water. Since the lighter crude oil fractions are usually more valuable, a crude oil with a higher °API gravity will bring a premium market price. Heavier crude oils are receiving renewed interest as supplies of lighter crude oil dwindle and increase in price. Heavy crudes are those with an 20°API or less. Generally, heavier crudes fetch a lower price on the market. However, heavier crudes will require more processing to convert the high-boiling-point fractions into desired lighter products. Thus, refiners balance the cost of more expensive light, sweet feedstocks against capital investment to refine cheaper, heavy, sour crude oils.

A *characterization factor* was introduced by Watson and Nelson to use an index of the chemical character of crude oil or its fractions.6 The Watson characterization factor is defined as

$$\text{Watson } K = (T_{\text{B}})^{1/3}/(sp\ gr)$$

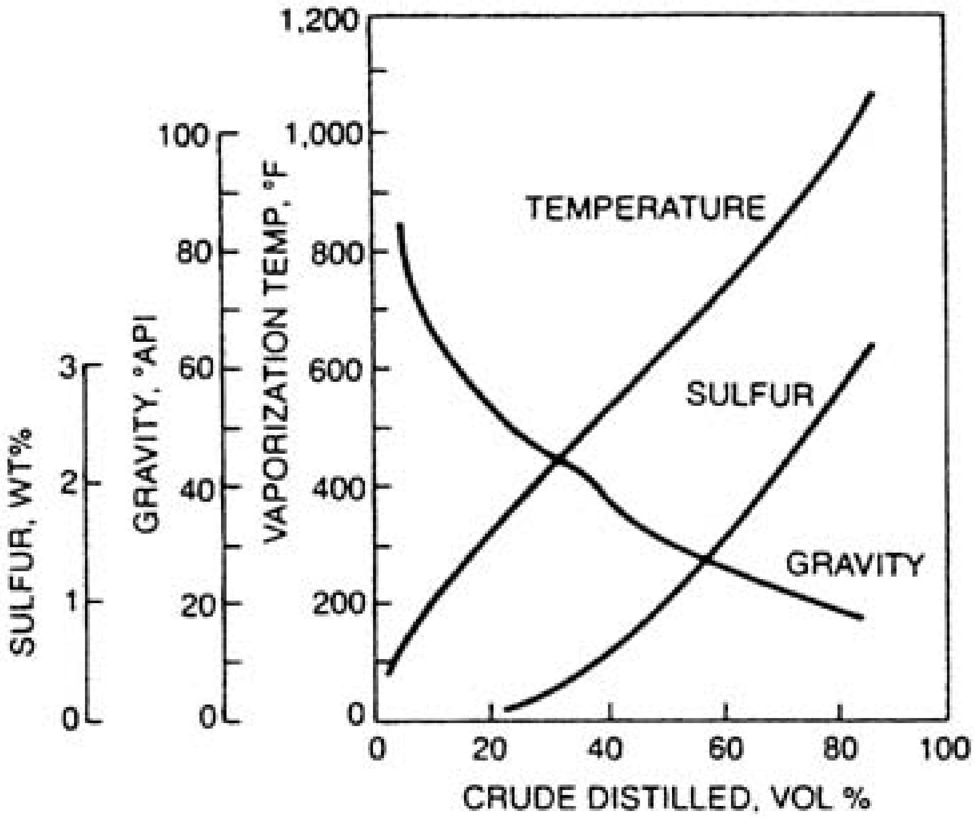
where T_B is the absolute boiling point in degrees Rankine (°R), and *sp gr* is the specific gravity compared with water at 60°F. For a wide boiling point range material like crude oil, the boiling point is taken as an average of five temperatures at which 10, 30, 50, 70, and 90 percent are vaporized.

A highly paraffinic crude oil might have a characterization factor as high as 13, whereas a highly naphthenic crude oil could be as low as 10.5. Highly paraffinic crude oils may also contain heavy waxes, which make the oil viscous and difficult to flow. Thus, another test for paraffin content is used to measure how cold a crude oil can be before it fails to flow under specific test conditions. The higher the pour-point temperature, the greater the paraffin content for a given boiling range.

Sweet and *sour* are terms that refer to the sulfur content of the crude oil. In the early days, those terms designated the smell of the oil. A crude oil with a high sulfur content usually contains hydrogen sulfidethe gas associated with rotten eggs. Such crudes with high sulfur levels were called sour. Without this disagreeable odor, the crude was judged as sweet. Today, the distinction between sour and sweet is based on analytical assessment of sulfur content. A sour crude oil is one with more than 0.5 weight percent (wt.%) sulfur, whereas a sweet crude has less than 0.5 wt. percent sulfur. It is estimated that 81 percent of the world's crude oil reserves are sour.7

ASTM distillation is a test prescribed by the American Society for Testing and Materials to measure the volume percent distilled at various temperatures.5 The results often are reported the other way around: the temperatures at which given volume percents vaporize.8 These data indicate the quantity of conventional boiling range products occurring naturally in the crude oil. Analytical tests on each fraction indicate the kind of processing that may be needed to make specification products. A plot of boiling point, sulfur content, and API gravity for fractions of Light Arabian crude oil is shown in Fig. 15.7. This crude oil is among those most traded in international crude oil markets.

Fig. 15.7 Analysis of Light Arabian crude.



From Fig. 15.7, this material in the midvolume range of the Light Arabian crude oil has a boiling point of approximately 600°F, a liquid density of nearly 30° API, and an approximate sulfur content of 1.0 wt. percent. These data are an average of eight samples. More precise values would be obtained on a specific crude oil if the data were to be used in design work.

Since a refinery stream spans a wide boiling range, the crude oil analysis data would be accumulated throughout that range to provide fraction properties. The intent here is to demonstrate the relationship between volume distilled, boiling point, liquid density, and sulfur content.

Crude Oil Pretreatment

Crude oil comes from the ground admixed with a variety of substances: gases, water, and dirt (minerals). The technical literature devoted to petroleum refining often omits crude oil cleanup steps. It is often assumed that the oil has been previously pretreated before entering the refining process. However, cleanup is important if the crude is to be transported effectively and processed without causing fouling and corrosion. Cleanup occurs in two ways: field separation and crude desalting.

Field separation is the first attempt to remove gases, water, and dirt that accompany crude oil extracted from the ground. As the term implies, field separation is done onsite at the production operation. The field separator is often no more than a large vessel that gives a quieting zone to permit gravity separation of the three phases: gases, crude oil, and water (with entrained dirt).

The crude oil is lighter than water, but heavier than the gases. Therefore, the crude oil appears within the field separator as a middle layer. The water is withdrawn from the bottom and is disposed of at the well site. Gases are withdrawn from the top and piped to a natural-gas processing plant or re-injected back into the reservoir to maintain well pressure. Crude oil from the middle layer is pumped to the refinery or to storage to await transportation by other methods.

Crude desalting is a water-washing operation done at the refinery to further clean up the crude oil before processing. The crude oil pretreated by field separators will still contain water and entrained dirt. Water-washing removes much of the water-soluble minerals and entrained solids.

If these crude-oil contaminants were not removed, they could cause operating problems during the refining process. The solids (dirt and silt) can plug equipment and deposit on heat-transfer surfaces, thereby reducing processing heat-transfer efficiency. Some solids, being minerals, can dissociate at high process temperatures and corrode major equipment. Other solids and minerals can deactivate catalysts used in refining processes.

Crude Oil Fractions

The importance of boiling range for petroleum products has already been discussed in Table 15.2. The simplest form of refining would isolate crude oil into fractions having boiling ranges that would coincide with the temperature ranges for consumer products. Some treatment steps might be added to remove or alter undesirable compounds, and a very small quantity of various chemical additives would be included to enhance final properties.

Crude oil distillation separates the desalted crude oil into fractions of different boiling ranges. Instead of trying to match final product boiling ranges, the fractions are defined by the number and type of downstream processes.

The desalting and distillation units are shown in Figs 15.815.10 along with the crude fractions from the crude distillation column. The relationships between some finished products and downstream processing steps will be expanded upon later in the chapter.

15.4 Gasoline

Gasoline is blended from several refining processes, as shown in Fig. 15.8. Depending on the individual refinery configuration, gasoline blending streams are separated and refined. Figure 15.9 depicts a light-fraction processing scheme.9 A straight-run gasoline stream is separated from the top portion of the atmospheric crude distillation column, which has a boiling range of 90200°F and is very paraffinic.

Fig. 15.8 Separating desalted crude oil into fractions.

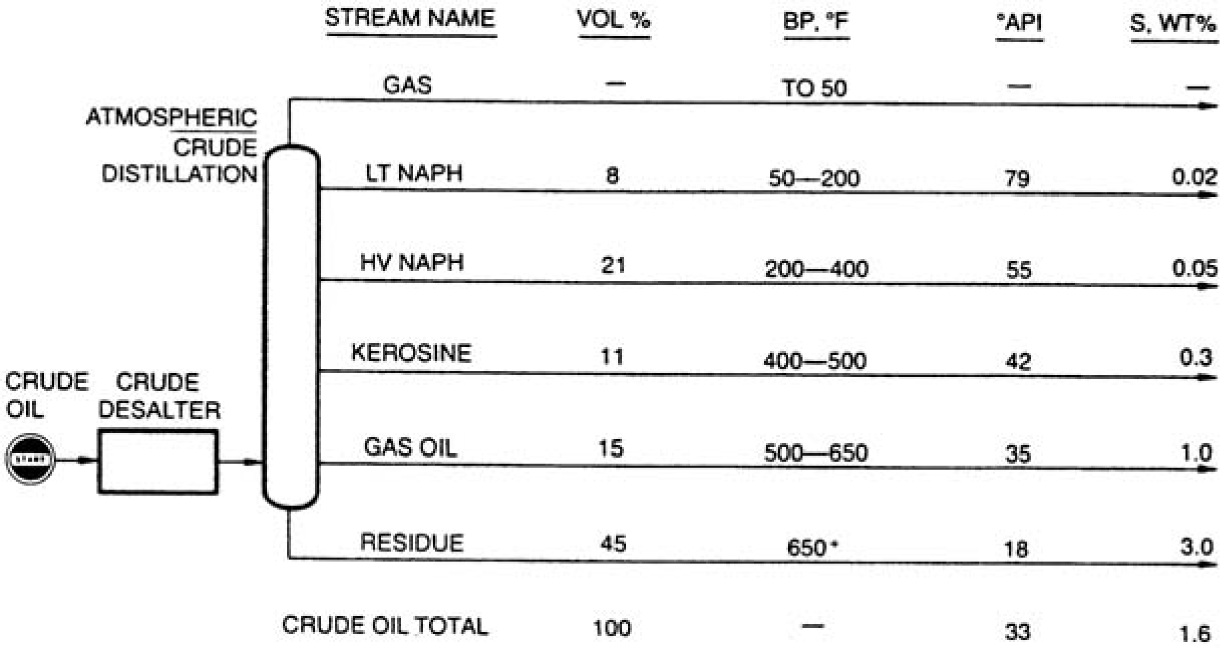
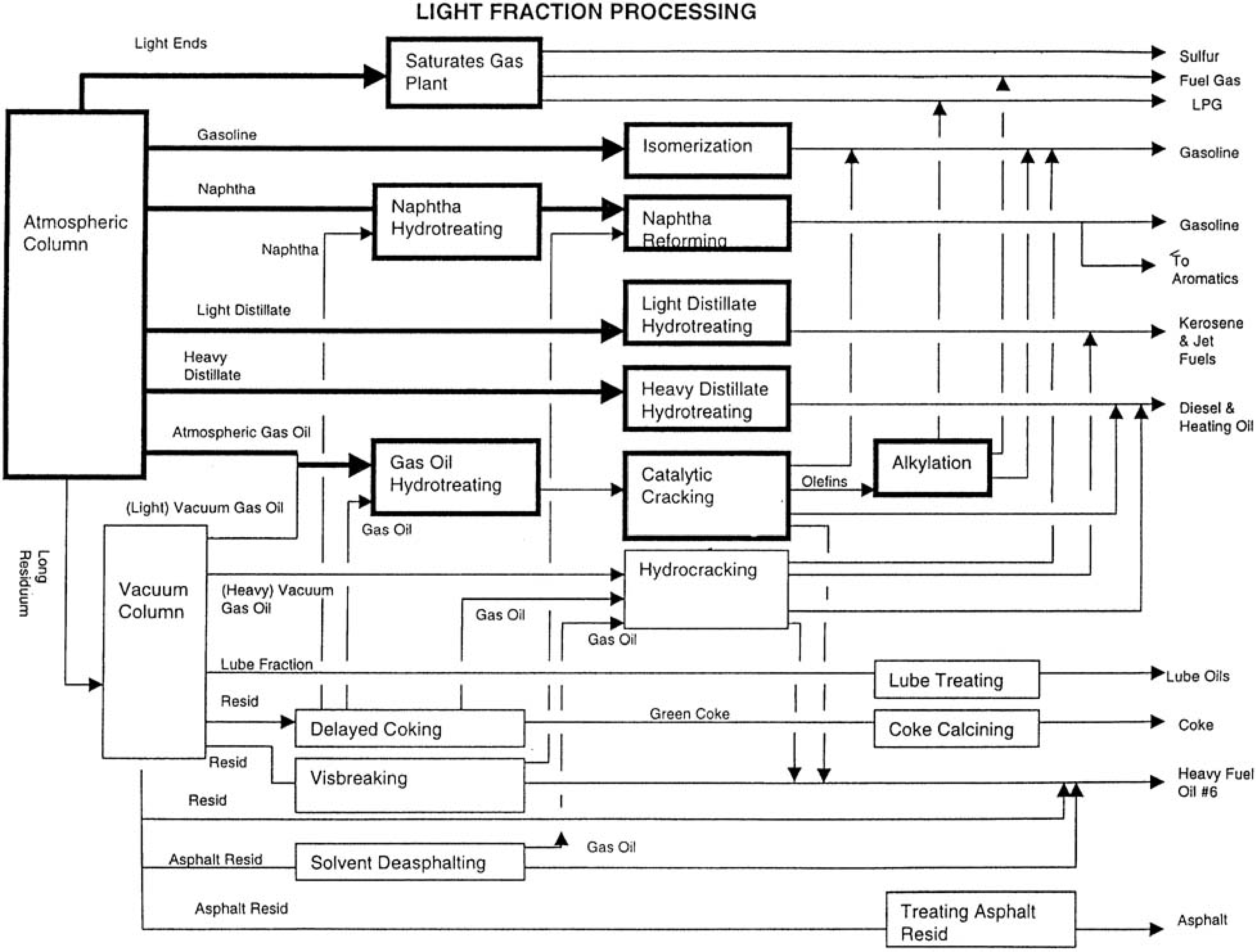


Fig. 15.9 Process flow diagram for a modern integrated refinery. *(Reprinted with permission of Ed Ekholm, Copyright 1999 by Ed Ekholm; all rights reserved.)*



The next cut gasoline stream from the crude distillation column is the naphtha cut. This stream has a boiling range of 200365°F and contains a significant portion of naphthenes, aromatics, and paraffins. Thus, the naphtha cut is hydrotreated and reformed to upgrade this stream into a gasoline blending stock.9

In present-day refineries, the fluid catalytic cracking (FCC) unit has become the major gasoline producing unit. The FCC's major purpose is to upgrade heavy fractions, that is, gas oil from the atmospheric and vacuum distillation columns and delayed coker, into light products. Atmospheric gas oil has a boiling range of between 650 and 725°F.9

The crude oil feedstock heavily influences the product slate for the refinery and the downstream processing required to meet the refinery's product goals. Fuels are blended to meet product specificationsof volatility, sulfur content, and octane number. Most important, refiners constantly seek to optimize their blending programs to meet product goals without giving up product.

The automobile's engine-drive train sets the specifications for gasoline. Notably, as automobile manufacturers design more sophisticated engines, in response refiners must adjust their operation to refine and blend fuels that are compatible with newer engines. Worldwide, new environment legislation has set product specifications for fuels. Table 15.4 lists the quality standards for automotive gasoline and diesel.10 These mandates are geared to lower tail-pipe emission from vehicles. Sulfur content and volatility will be strictly limited in future fuel requirements.

TABLE 15.4 Anticipated Fuel Specifications under Auto Oil Program and U.S. Requirements10

Period	Up to 1999	2000	2005	2008 (Expected Values)	2000 (U.S. only)	2006
Gasoline						
Sulfur content	<500 ppm	<150 ppm	<50 ppmb		<170 ppm	30 ppm
Benzene content	<5% v/v	<1% v/v	<1% v/va		<1% v/v	
Aromatics content		<42% v/v	<35% v/v		<25 v/v	
Diesel fuel						
Sulfur content, ppm	<500	<350	<50	<30	<500	<10
Cetane number, min	49	51	53a	5458	40	
Polynuclear aromatics		<11% w/w	<6% w/wa	<41% w/w		
Density, kg/l max	<0.86	<0.845	<0.845a	<0.8300.825		
Distillation T95, °C	<370	<360	<360a	<350340		

% v/v: percent by volume; % w/w: percent by weight; aExpected values; bGermany is promoting 10 ppm sulfur for gasoline by 2003.
Source: Linde Technische Gase GmbH.

Volatility

A gasoline's boiling point is important during its aspiration into the combustion chamber of a gasoline-powered engine. Vapor pressure is a function of the fuel's boiling point. Boiling range and vapor pressures are combined as the concept, *volatility*.11

The lighter components in gasoline are used as a compromise between two extremes: enough light components so that adequate vaporization of the fuelair mixture provides an easy engine start in cold weather, but too much of light components can cause the fuel to vaporize within the fuel pump and cause vapor lock.

Environmental studies suggest that light gasoline components are detectable in the atmospheres of large metropolitan areas. New environmental laws limit the volatility of gasoline, so refiners must use other processing streams to meet volatility requirements. However, the fuel must provide performance to consumers; for example, to minimize chamber deposits and spark-plug fouling in the engine.

Sulfur Content

Sulfur compounds are corrosive and foul-smelling. When burned in an engine these compounds form sulfur dioxide and other oxides referred to as SO*x* in engine exhaust. These compounds recombine and form sulfur trioxide and sulfuric acid mist, which is released as engine exhaust. Efforts to improve air quality are targeted at reducing vehicle engine exhaust of toxins and SO*x* compounds. Thus, many new environmental regulations (Table 15.4) are focused on reducing the sulfur content of fuels. All crude oils contain some sulfur concentration. How much desulfurization that is needed is dependent on the feedstock and product slate.

Caustic wash or other enhanced solvent-washing methods are a sufficient pretreatment to remove sulfur compounds from light naphtha. The sulfur compounds in light naphtha are mercaptans and organic sulfides that are readily removed by these washing processes. Heavy naphtha is harder to desulfurize. This stream has a higher sulfur content, and, equally important, the sulfur is embedded in complex hydrocarbon compounds and rings. Washing efforts are more effective on mercaptans, which are not usually present in heavy naphtha. More aggressive methods are needed to break the compound structures and release the sulfur. Hydrotreating is one effective method to reduce sulfur content in hydrocarbon streams.12

Hydrotreating is a catalytic process that converts sulfur-containing hydrocarbons, that is, sulfides, disulfides, and ring compounds such as thiopenes, into low-sulfur liquids and hydrogen sulfide. This technology is widely used throughout the refinery, as shown in Figs 15.9 and 15.10.9 This process is operated under a hydrogen-rich blanket at elevated temperatures and pressure. The process is a hydrogenconsuming step, since the hydrogen replaces the sulfur bond on the hydrocarbon molecule.9

Nitrogen and oxygen compounds are also dissociated by hydrotreating. Consequently, hydrotreating provides additional benefits of hydrodenitrification, which improves downstream operations. For nitrogen and oxygen compounds, the products from hydrotreating are ammonia and water, respectively. Thus, these contaminants will be separated in the off-gas and are easily removed by conventional gas-treating processes.

Octane Number

Another condition to keep gasoline engines running smoothly is that the fuelair mixture starts burning at a precise time in the combustion cycle. An electrical spark starts the ignition. The remainder of the fuelair mix should be consumed by a flame front moving out from the initial spark.

Under some conditions, a portion of the fuelair mix will ignite spontaneously instead of waiting for the flame front from the spark. The extra pressure pulses that occur from spontaneous combustion are usually audible above the background sounds of the engine running and give rise to a condition know as "engine knock." The engine pings and rumbles when under "knock conditions." This condition is undesirable; it is a waste of available power.

The *octane number* is a measure of a fuel's ability to avoid knocking. The octane number of gasoline is determined in a special single-cylinder engine where various combustion conditions can be controlled.5 The test engine is adjusted to trace the knock from the fuel being rated. Various mixtures of iso-octane and normal heptane (*n*-heptane) are used to find the ratio of the two reference fuels that will give the same intensity of knock as that from an unknown fuels. Defining iso-octane as 100 octane number and *n*-heptane as 0 octane number, the volumetric percentage of iso-octane in heptane that matches knock from the unknown fuel is reported as the octane number of the fuel. For example, 90 volume percent (vol.%) of iso-octane and 10 vol. percent *n*-heptane establishes a 90 octane number reference fuel.

Two types of octane number ratings are specified, although other methods are often used for engine and fuel development. Both methods use the same reference fuels for essentially the same test engine. Engine operating conditions are the difference. In the *research method*, the spark advance is fixed, the air inlet temperature is 125°F, and the engine speed is 600 rpm. The other method is called the *motor method*; it uses variable spark timing, a higher mixture temperature (300°F), and a faster engine speed (900 rpm).

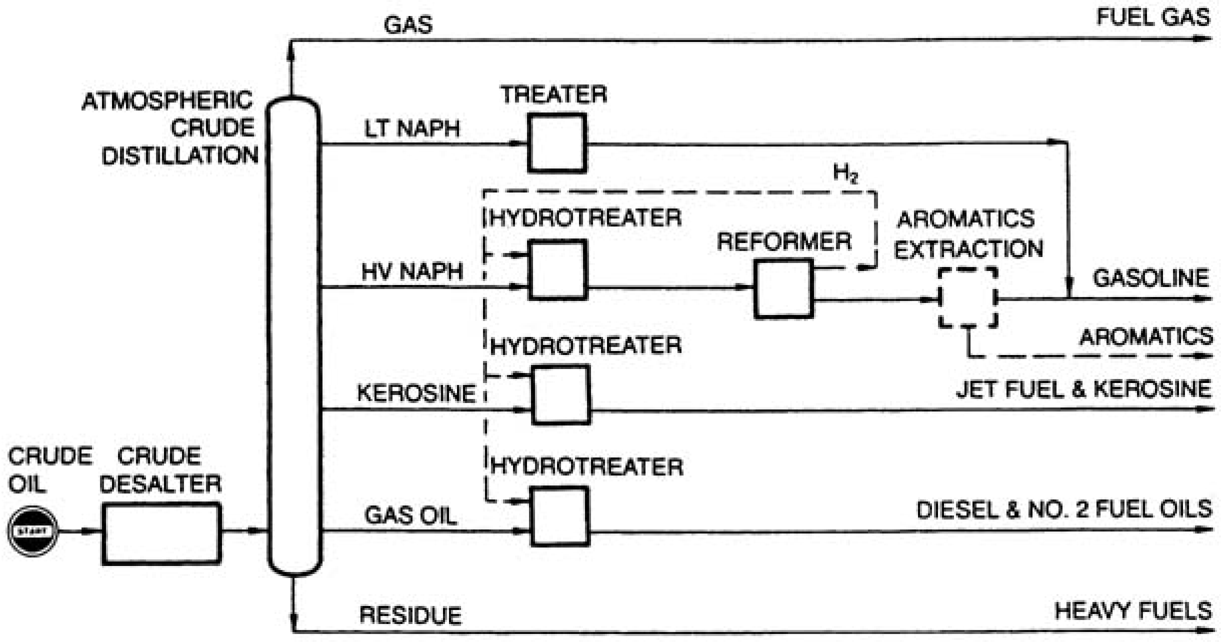
The more severe conditions of the motor method have a greater influence on commercial blends than they do on the reference fuels. Thus, a motor octane number (MON) of a commercial blend often has a lower research octane number (RON). Consequently, blended fuels use an arithmetic average of both ratingsMON and RONand can be abbreviated as (R + M)/2.

Catalytic reforming is a principal process used to upgrade the octane number of naphtha for gasoline blending.13 Reforming uses catalysts to reshape the molecular structure of hydrocarbons to raise the octane number of the process stream. Naphthenes are converted to aromatics; paraffins are isomerized to isomeric forms.9 Reforming efforts are most effective when used on heavier molecules; a greater increase in octane number can be attained by reforming heavy naphtha cuts.

Reforming catalysts typically contain platinum or a mixture of platinum and other metal promoters on a silicaalumina support. Only a concentration of platinum is used, averaging about 0.4 wt. percent. The reforming process is a highly endothermic process. Desulfurized feeds are preheated to 900°F, and the reactions are done at various pressures (50300 psig), which are dependent on the licensed process used.12 At elevated temperatures and pressures, the catalyst is susceptible to coking, which decreases catalyst efficiency. Thus, refiners must regenerate the catalyst to maintain process efficiency. Reforming catalyst can be regenerated in situ by burning off the coke from the catalyst. Newer developments now use continuous regeneration of the reforming catalyst in which three reforming reactors are stacked one on top of the other. Gravity flow moves the catalyst from the top to the bottom and sends it to a regeneration step in which a dry burn removes the coke. The regenerated catalyst is then returned to process. Also, reforming feeds are pretreated to remove poisons that can kill precious-metal catalysts.

Hydrotreating is an effective method to pretreat reforming feedstocks (Fig. 15.10). Combining hydrotreating with reforming is most effective. Due to cyclization and dehydrogenation of hydrocarbon molecules in the reformer, hydrogen is a by-product from this operation.14 Notably, by-product hydrogen from the reform can be directed to the hydrotreating operations. Thus, the reformer can provide the refinery with the hydrogen supply for hydrotreating. A rule of thumb is that the catalytic reformer produces 8001,200 scf/bbl for naphtha. The excess hydrogen is available for hydrotreating other fractions in separate hydrotreaters.

Fig. 15.10 Low-investment route to modern products.



15.5 Distillates

Jet fuel, kerosine (range oil), No. 1 fuel oil, No. 2 fuel oil, and diesel fuel are all popular distillate products from the 365650°F fractions of crude oil.9 Distillates are further classified as light distillates with a true boiling point range of 365525°F, and heavy distillate cuts have a true boiling point range from 525 to 650°F. Light distillates are blended into kerosine and jet fuels. Heavy distillate cuts are used to blend diesel fuels and home heating oils.9

Some heating oil (generally No. 2 heating oil) and diesel fuel are very similar and are sometimes substitutes for each other. Home heating oil is intended to be burned in a furnace for space heating. Diesel fuel is intended for compressionignition engines.

Distillates are lower cuts from the atmospheric crude distillation column (Fig. 15.10); thus, these refinery streams may have high sulfur concentrations due to the feedstock that is processed. Newer product specifications limit sulfur concentrations in consumer products, especially diesel. Consequently, distillate streams must be upgraded. Hydrotreating improves the product properties of distillate products; notably it reduces sulfur content. More important, hydrotreating hydrogenates unsaturated hydrocarbons so that they will not contribute to smoke and particulate emissionswhether the fuel is burned in a furnace or used in an engine.

Residuals

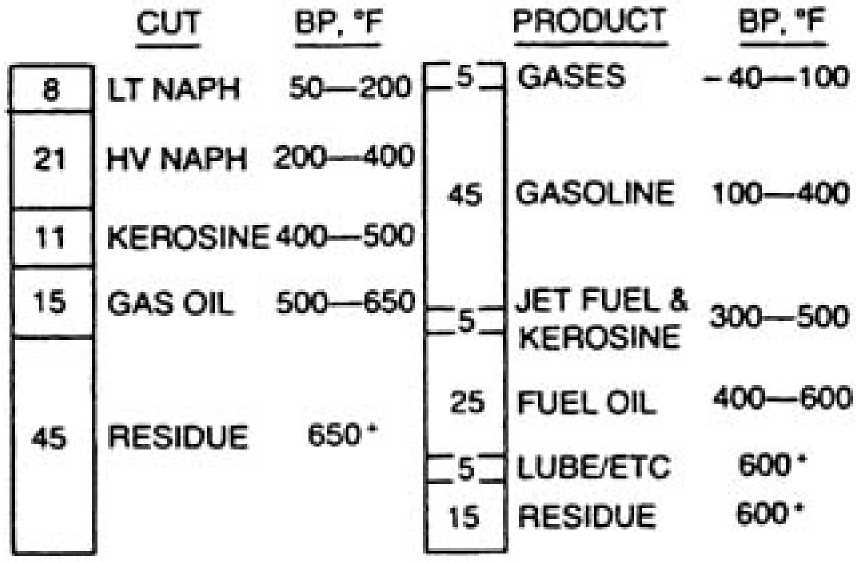
Crude oil is seldom distilled at temperatures above 650°F. At higher temperatures, coke will form and plug the lower section of the crude oil distillation tower. This bottom fraction from the atmospheric crude column has a true boiling point range of 650725°F and is often referred to as *atmospheric gas oil* or *residuals*. This fraction is traditionally not vaporized. Atmospheric gas oil must be upgraded extensively; it can be severely hydrotreated to break apart the complex ring compounds and saturate them into lighter products. This stream can be sent to a catalytic cracker to further upgrade this heavy fraction into gasoline, diesel, and home heating fuel.9 The heaviest cut of the atmospheric crude distillation is often referred to as the *long residuum*. This fraction is further processed via a vacuum distillation column.

15.6 Producing More Light Products

The refining scheme shown in Fig. 15.9 is a simplified view of an integrated refinery. More processing steps can be added and are dependent upon the product slate of the refinery and the cracking slate anticipated for its design. If the refinery is a gasoline refinery, the cracking process will be directed toward producing light productsgasoline at the expense of diesel and heating oil. Conversely, if diesel is the desired end-product, gasoline product is sacrificed to produce more distillate streams. The U.S. refineries are predominately gasoline-oriented. Conversely, Western European refineries are diesel-fuel oriented. In Western Europe, the demand for gasoline is projected to decline over the next 10 years, while demand for middle distillates increases.15 Therefore, Western European refiners must adapt operations not only to produce more diesel but also to minimize gasoline production.

For industrial areas, where the principal demand is for transportation fuels or high-quality heating oils, the simplified refining scheme is shown in Fig. 15.9. In the case of a U.S. refinery, the emphasis is directed into converting more of the crude oil into lighter products, namely gasoline. If this refinery processed a Light Arabian crude, nearly 8085 vol. percent of the products would be lighter than the boiling temperature of 650°F compared with the 55 vol. percent existing in the feed crude oil (Fig. 15.11). More than half of all U.S. products are gasoline and lighter distillates (boiling temperatures below 400°F) compared with 29 vol. percent in the crude oil. This comparison is unique for these circumstances. Refining technologies and catalysts have advanced from the earlier days. Notably, refiners can process a variety of crudes into valued end-products. Crudes that require more intense processing, that is, high-severity hydrocracking and hydrotreating, fetch a lower price on the market and are often referred to as *opportunity crudes*. Refiners processing lower-cost, sour, heavy crude must make more capital investment in processing capability. However, these refiners pay less for their feedstocks and risk attaining payback on capital investments through volume processing of lower-cost feedstocks. The design of an integrated refinery constantly balances the future price of oil against the return on investment for refining capacity. Thus, the decision on what types of crude to process will impact the design and operation of a refinery.

Fig. 15.11 Light Arabian crude oil compared with U.S. product deliveries.



Cracking

As mentioned earlier, the refining process involves adding hydrogen to carbon molecules. Notably, the desired hydrocarbons are much lighter products; thus, the refining process strives to break the large, complex hydrocarbon molecules into smaller molecules and add hydrogen to the open bonds. Cracking processes typically break hydrocarbon molecules into two or more smaller molecules. Thermal cracking uses high-temperature (above 650°F) and a long residence time to sever hydrocarbon bonds. Higher pressure facilitates the cracking process; however, the capital investment for a high-pressure reactor is greater than the expense for low- or medium-pressure reactors.

In thermal cracking, the formation and deposition of coke on piping and equipment walls is an unwanted side reaction. Another option to crack crude oil into desired products at lower pressures and temperatures is *catalytic cracking*. Catalytic cracking splits the molecules quicker and at lower temperatures. Catalysts are used to promote the desired reaction rate for the process.

Catalytic cracking involves large reactors with large fluidized catalyst beds. As in the case of thermal cracking, coke is also formed during the catalytic process, which can deposit on the catalyst and hinder its activity. Thus, this process uses reactorseparators and regenerators to remove the catalyst and regenerate it and return the catalyst back to process. The fluidized bed mixes the feed with the catalyst to optimize contact time. The catalyst is separated from the hydrocarbon products. A portion is regenerated; the remainder is returned to the catalytic cracker reactor.

Catalytic cracking is very effective in upgrading heavy refining streams, such as gas oils, into motor gasoline stocks thereby increasing the octane number for product streams. This process produces less gas and coke as compared with thermal cracking operations. Catalytic cracking also yields more liquid products, which can be tailored toward gasoline or diesel fuel and home heating oil products. Different operating conditions and catalysts will define the product mix from a catalytic cracker.

Several factors determine the best feeds for catalytic crackers. Heavy feeds are preferred; thus, the lower boiling point is about 650°F. The feed should not be so heavy that it contains an undue amount of metal-bearing compounds or carbon-forming material. Deposition of metals and coke can quickly deactivate the catalyst.

Visbreaking is a mild, once-through thermal cracking process. It is used to crack resid products into fuel-oil specifications. Although some light products such as naphtha and gasoline are produced, this is not the purpose of the visbreaker.

Coking is another matter. It is a severe form of thermal cracking in which coke formation is tolerated to attain additional lighter liquids from the heavier, dirtier fractions of crude oil. In this process the metals that would foul catalysts are laid down with the coke. The coke settles out in large coke drums that are removed from service frequently (about once a day) to have the coke removed by hydraulic methods. Several coke drums are used to make the process continuous; thus, one drum is online while the other is being emptied and readied for the next cycle.12

Hydrocracking converts a wide variety of heavy refining product streams into light productsfuels and distillates. A robust catalyst system is used to desulfurize, denitrify, and hydrocrack the feed.9,12,14 The process combines hydrotreating and catalytic cracking goals. However, hydrocracking is a more capital-intensive and operating-intensive step. The operating pressure is higher (up to 3000 psi); consequently, thick-wall vessels are used as reactors (up to 9 in. thick). Products from a hydrocracker are very clean (desulfurized, denitrified, and demetalized) and will contain isomerized hydrocarbons in greater quantity than from conventional catalytic cracking. This process consumes a large quantity of hydrogen, which adds considerably to its operating costs.

Vacuum Distillation

As mentioned earlier, most consumer products are light productsthose with boiling points less than 400°F. However, in the refining of a crude oil, a significant portion of the products has a true boiling point above 650°F. Atmospheric distillation is least effective in converting heavier products into lighter components. A second distillation column under vacuum is needed to further separate the heavier parts of crude oil into lighter fractions. Some fractions from the vacuum units have better quality than atmospheric distillation cuts since the metal-bearing compounds and carbon-forming materials are concentrated in the vacuum residue.

Reconstituting Gases

Cracking processes to convert heavy liquids into lighter products also create gases. Another option to make more liquid products is to combine the gaseous hydrocarbons. A gas separation unit may be added to a refinery to isolate individual types of gases. When catalytic cracking is part of the refining scheme, a large quantity of olefins (ethylene, propylene, and butylene) is co-produced. Two routes are available to reconstitute these gaseous olefins into gasoline blending stocks; as outlined below.

Polymerization uses a catalytic process to combine two or more olefins to make polymer gasoline. The double bond in only one olefin is changed to a single bond during each link between two olefins. Thus, the product will still have some double bonds. This process was developed in the 1940s to produce high-octane aviation fuel. However, the olefinic nature of polymer gasoline does have a drawback.4 The gum-forming tendencies of the polymer gasoline are problematic especially during long storage in warm climates. The olefins continue to link up and form larger moleculesgum or sludgewhich are undesirable. Some refiners still use catalytic polymerization of light ends. It is a lower-cost process, both in terms of operating and investment costs. However, due to new environmental specifications, hydrogenation of polymer gasoline may be necessary to meet emission standards for Bromine Number.9 Hydrogenation can reduce the octane number of the polymer gasoline.

Alkylation catalytically combines light olefinspropylene and butyleneswith isobutane to produce a branched-chain paraffinic fuel.14 Alkylate is a great blending component for the gasoline-blending pool. It has a high octane number (usually above 94), low vapor pressure, and is almost sulfur free.16 Present-day alkylation processes are carried out in the presence of sulfuric or hydrofluoric acids. New health and safety issues are promoting research on solid-acid technologies. Some successes have been demonstrated in the laboratory and in pilot studies; however, no commercial units have been built.

The *ether* process combines an alcohol with an iso-olefin. In the United States, a weight percentage of oxygenate (2 wt.%) content is mandated for reformulated gasoline (RFG). The most common oxygenate currently used is methyl tertiary butyl ether (MTBE). Methanol and the iso-olefin form of isobutylene are reacted to form MTBE. Other alcohols, such as ethanol, may be reacted with iso-butylene to form ethyl tertiary butyl ether (ETBE). Methanol can be reacted with iso-amylene, another iso-olefin, to form tertiary amyl-methyl ether (TAME). Of all the mentioned ethers, MTBE is the most widely used ether as a gasoline-blending component.16

15.7 A Modern Refinery

A refining scheme incorporating the processes discussed above is shown in Fig. 15.9. The variations in this flow diagram are numerous. Types of crudes processed, product slate, and competitive quality goals of products are just a few factors that influence the processing needs for a refining complex. Many other processes play an important role in the final scheme. A partial list of these processes would be: dewaxing lubricating oils, deoiling waxes, deasphalting heavy fractions, manufacturing specific compounds for gasoline blending (alcohols, ethers, etc.), and isolating specific fractions for petrochemical applications.

Petrochemicals

Refining crude provides many products, depending on the types of products sought. Lighter products from the refining of crude have higher value as petrochemical feeds than as gasoline-blending components. In particular, olefins and aromatic compounds possess higher values as petrochemicals than as gasoline components.

Ethylene is a key building block for the petrochemical industry. It is usually made by thermally cracking gasesethane, propane, butane, or a mixture of theseas they exist in refinery off-gases. When gas feedstocks are scarce or expensive, naphthas and even whole crude oils have been used in specially designed ethylene crackers. The heavier feeds provide significant quantities of higher molecular weight olefins and aromatics.

Aromatics are typically concentrated in product streams from the catalytic reformer. When aromatics are sought for petrochemical applications, they typically are extracted from the reformer product stream by solvent extraction or distillation extraction. A common solvent used is sulfolane; new processes now use *n*-formylmorpholin as the extractive solvent.12

The mixture of aromatics is typically referred to as BTX and is an abbreviation for benzene, toluene, and xylene. The first two componentsbenzene and tolueneusually are separated by distillation, and the isomers of the third componentxyleneare separated by partial crystallization.17 Benzene is the starting chemical for materials such as styrene, phenol, and many fibers and plastics. Toluene is used to make a number of chemicals, but most is blended into gasoline. Xylene usage is dependent on its isomer. *Para-xylene* (*p*-xylene) is a precursor compound for polyester. *Ortho-xylene* (*o*-xylene) is the building block for phthalic anhydride. Both compounds are widely used to manufacture consumer products.

15.8 Process Details

Thus far, the refining units have been described as they relate to other units and to the final product specifications. At this point, typical flow diagrams of some major processes will be presented to highlight individual features. In many cases the specific design shown is an arbitrary choice from the work of several qualified designers.

Crude Desalting

Salts such as sodium, calcium, and magnesium chloride are generally contained in water suspended in the oil phase of hydrocarbon feedstocks.9 Other impurities are also present in crude oils as mechanical suspensions of silt (dirt), iron oxides, sand, and crystalline salt.14 These contaminants must be removed before processing the crude oil feeds; thus, the best method is mixing the crude oil with water and creating an emulsion.12

A typical flow diagram is shown in Fig. 15.13 The desalter operation is incorporated into the preheat train of the crude distillation unit to conserve energy. Depending on the characteristics of the hydrocarbon feedstock, the crude oil is preheated to 150300°F. The crude oil is preheated with crude unit products and pumparound reflux to the desired temperature. The operating pressure is 40 psig or more. The elevated temperature reduces oil viscosity to improve mixing, and the elevated pressure suppresses vaporization. The washwater, 36 vol. percent, can be added upstream and/or downstream of the heat exchanger (s). The combined streams pass through a mixing devicea throttling valve or emulsion orificeand create a stable water-in-oil emulsion. The properties of the emulsion are controlled by adjusting the pressure drop across the mixing device. Trace quantities of caustic, acid, polymers, other chemicals are sometimes added to promote treatment.12

The emulsion enters the desalter vessel where a high-voltage electrostatic field is applied. The electrostatic field causes the dispersed water droplets to coalesce, agglomerate, and settle to the lower portion of the vessel. The various contaminants from the crude oil concentrate in the water phase. The salts, minerals, and other water-soluble impurities are discharged from the settler to the effluent system. Clean, desalted hydrocarbon product flows from the top of the settler and is ready for the next processing step.

Additional stages can be used in series to gain additional reductions in the salt content of the crude oil. Two stages are typical, but some installations use three stages. About 90 percent of the emulsified water can be recovered in one step, whereas 99 percent recovery is possible with a two-step process.9 The additional investment for multiple stages is offset by reduced corrosion, plugging, and catalyst poisoning of downstream equipment with the cleaner crude feed.

Crude Distillation

Single or multiple distillation columns are used to separate the crude oil into fractions determined by their boiling range. Common identification of these fractions was discussed using Fig. 15.12, but should only be considered as a guide. Many refining schemes can be used to alter the type of separation made at this point.

Fig. 15.12 High conversion refinery.

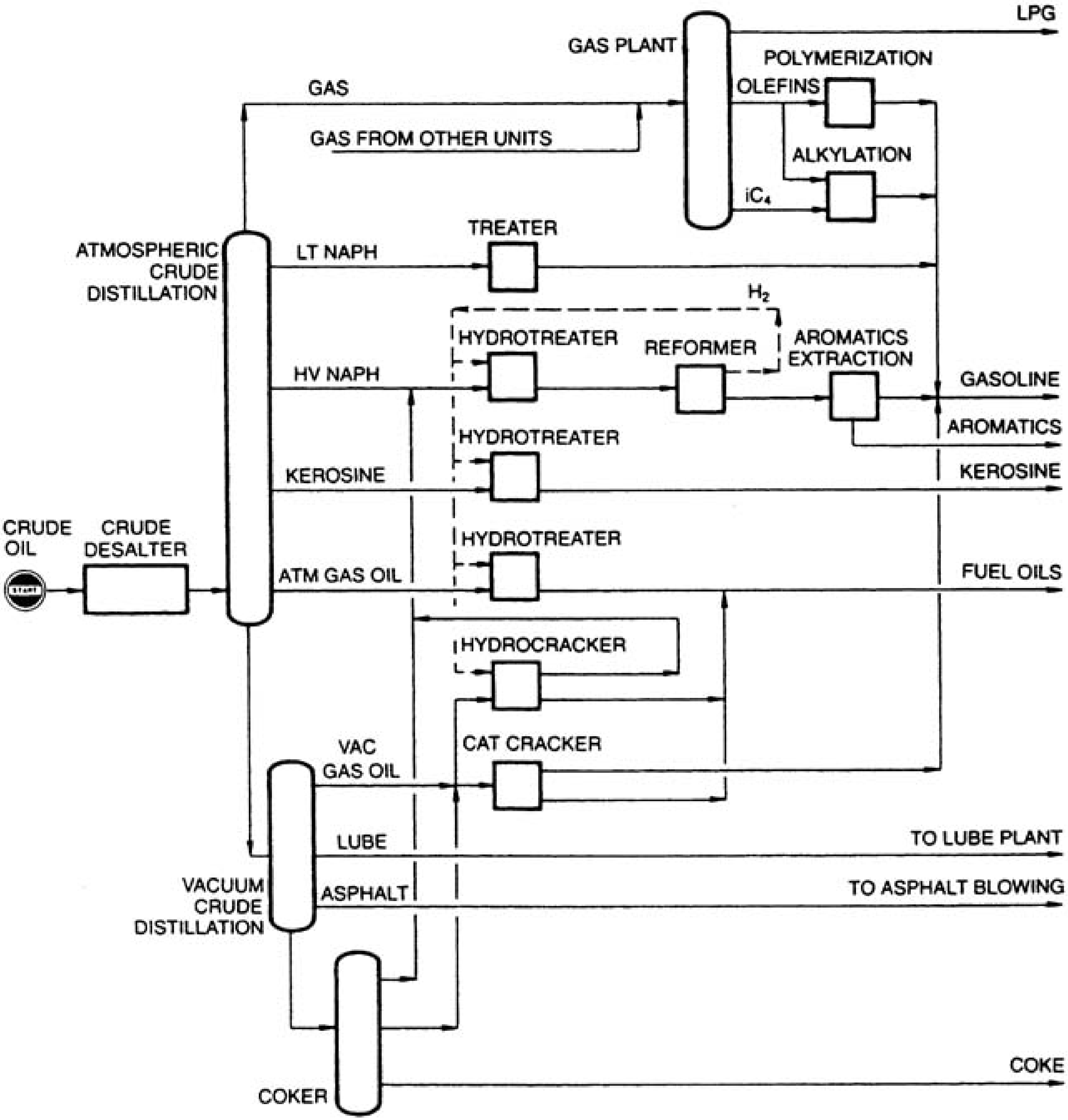
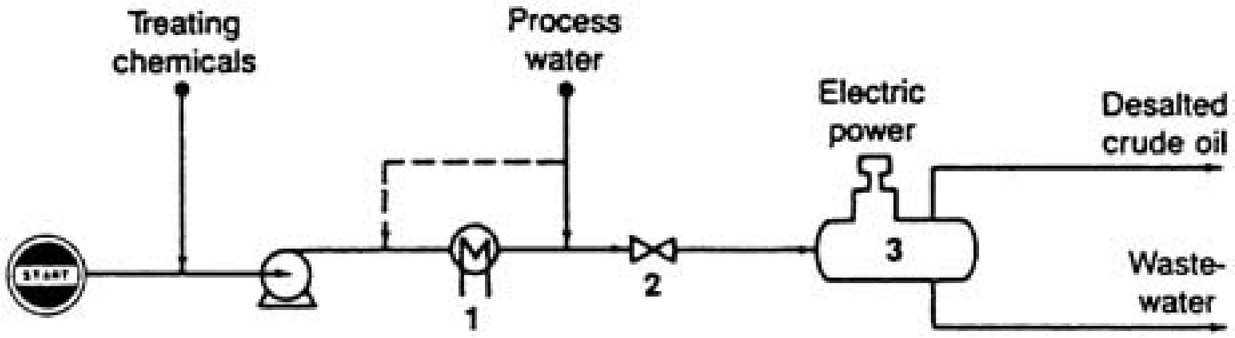
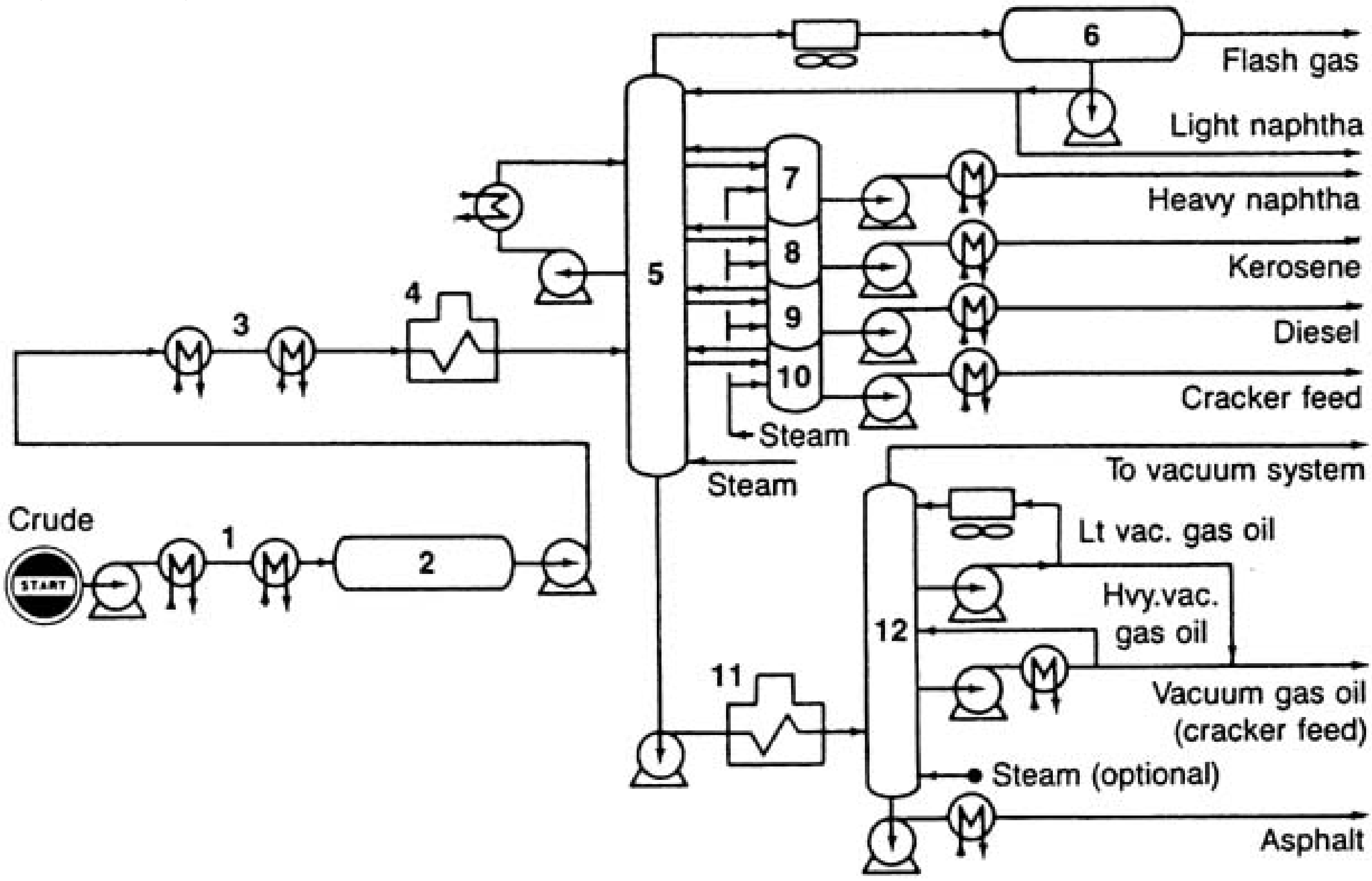


Fig. 15.13 Electric desaltingHowe Baker. Includes: heater (1), mixing valve (2), and electrostatic water settler (3). (Hydrocarbon Processing, 81*(11), 105 (Nov. 2002); Copyright 2002 by Gulf Publishing Co., all rights reserved.*)



A typical flow diagram of a two-stage crude oil distillation system is shown in Fig. 15.14. The crude oil is preheated with hot products from the system and desalted before entering the fired heater. The typical feed to the crude-fired heater has an inlet temperature of 550°F, while the outlet temperature may reach 657725°F. Heater effluent enters the crude distillation (CD) column, where light naphtha is drawn off the overhead tower. Heavy naphtha, kerosine, diesel, and cracking streams are sidestream drawoffs from the distillation column. External reflux for the tower is provided by several pumparound streams.12

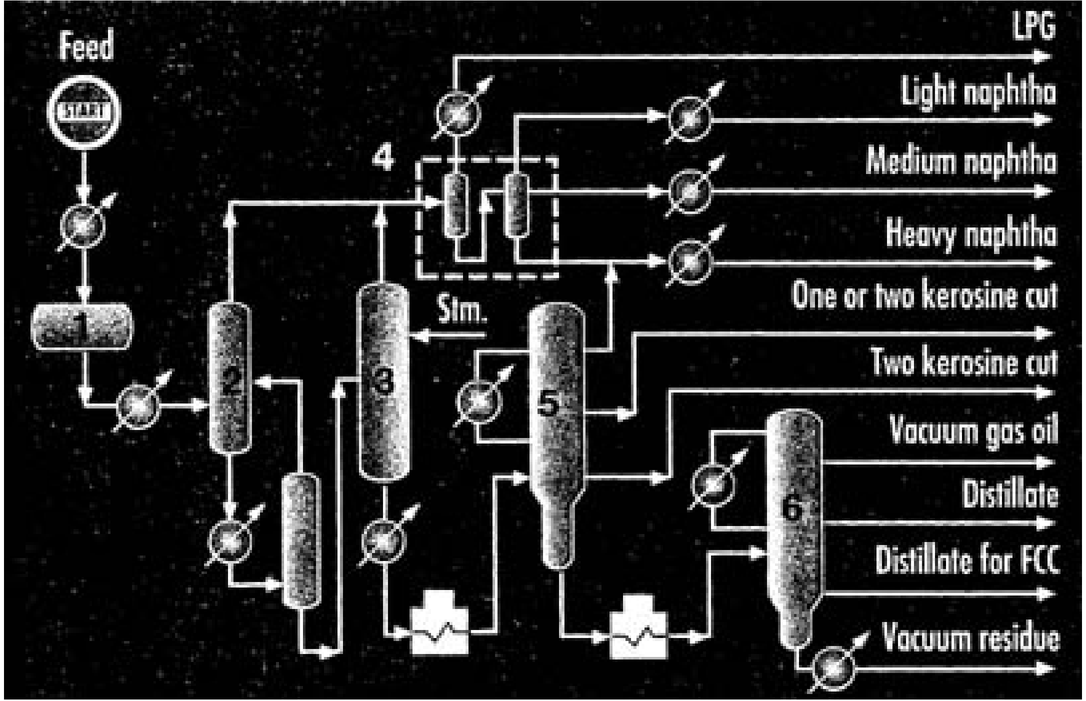
Fig. 15.14 Crude distillationFoster Wheeler. Includes: exchanger heaters (1,3), desalter process (2), fired heater (4), main fractionator (5), overhead gas accumulator (6), sidestream strippers (7,8,9,10), second fired heater (11), and vacuum fractionator (12). (Hydrocarbon Processing, 81(*11), 99 (Nov. 2002); Copyright 2002 by Gulf Publishing Co.; all rights reserved.*)



The bottoms of the CD, also known as atmospheric residue, is charged to a second fired heater where the typical outlet temperature is about 750775°F. From the second heater, the atmospheric residue is sent to a vacuum tower. Steam ejectors are used to create the vacuum so that the absolute pressure can be as low as 3040 mm Hg (about 7.0 psia). The vacuum permits hydrocarbons to be vaporized at temperatures below their normal boiling point. Thus, the fractions with normal boiling points above 650°F can be separated by vacuum distillation without causing thermal cracking. In this example (Fig. 15.14), the distillate is condensed into two sections and withdrawn as two sidestreams. The two sidestreams are combined to form cracking feedstocksvacuum gas oil (VGO) and asphalt base stock.

Atmospheric distillation is an energyintensive process. With pressure to reduce operating costs, new design efforts are investigating energy conservation on the CD column. A new atmospheric distillation process by TECHNIP FINAELF uses a progressive distillation strategy to minimize total energy consumption, as shown in Fig. 15.15. In this processing scheme, two pre-flash towers separate the light productsLPG, naphtha (light, medium and heavy), and kerosenefrom the crude feed to the main atmospheric distillation column. The light products are fractionated as required in a gas plant and rectification towers.

Fig. 15.15 Crude distillationTOTAL FINAELF. Includes: desalter and preheater (1), preflash towers (wet and dry) (3,2), gas plant and rectification tower (4), main fractionation tower (5), and vacuum fractionation tower (6). (Hydrocarbon Processing, 81(*11), 100 (Nov. 2002); Copyright 2002 by Gulf Publishing Co.; all rights reserved.*)



The topped crude is typically reduced by two-thirds of the total naphtha cut. The bottoms from the second pre-flash tower are sent to the charge heater and directed to the main distillation column and produce four product streamsheavy naphtha, several kerosene cuts, and bottoms residue. The residue is further processed in a vacuum column and produces VGO and several distillate streams.

Incidentally, the total refining capacity of a facility is reported in terms of its crude-oil handling capacity. Thus, the size of the first distillation column, whether a pre-flash or an atmospheric distillation column, sets the reported size of the entire refinery. Ratings in barrels per stream day (bpsd) will be greater than barrels per calendar day (bpcd). Processing units must be shut down on occasion for maintenance, repairs, and equipment replacement. The ratio of operating days to total days (or bpcd divided by bpsd) is called the "onstream" factor or "operating factor." The ratio can be expressed either as a percent or a decimal. For example, if a refinery unit undergoes one shutdown period for 1 month during a 3-year duration, its operating factor is (36 - 1)/36, or 0.972, or 97.2 percent.

Outside the United States, refining capacity is cited in metric tons per year. Precise conversion from one unit of measure to the other depends upon the specific gravity of the crude oil, but an approximate relation is 1 barrel per day equals 50 tons per year (tpy).

Hydrotreating

Hydrotreating is one of the more mature refining processes still practiced today. Refiners began using catalytic hydrotreating in the 1950s to remove undesirable materials from refining product streams, as shown in Fig. 15.9. This process effectively removes contaminants such as sulfur, nitrogen, olefins, metals, and aromatics.9,12,14 The chemistry of hydrotreating can be further divided into three categories: hydrodesulfurization, hydrodenitrification, and hydrodearomatization (saturation of olefins and saturation of aromatics).12,14

The utility of most hydrotreating efforts is desulfurization. Sulfur-containing hydrocarbons are present in crude oil with many varying forms and boiling points. New product specifications limit the amount of sulfur present in finished fuels; thus, greater effort will be needed to remove more sulfur-containing compounds throughout the fuel-blending pool. Notably, more complex, high-boiling-point sulfur compounds must be extracted from the blending pool to meet lower fuel specifications, as shown in Table 15.5.18

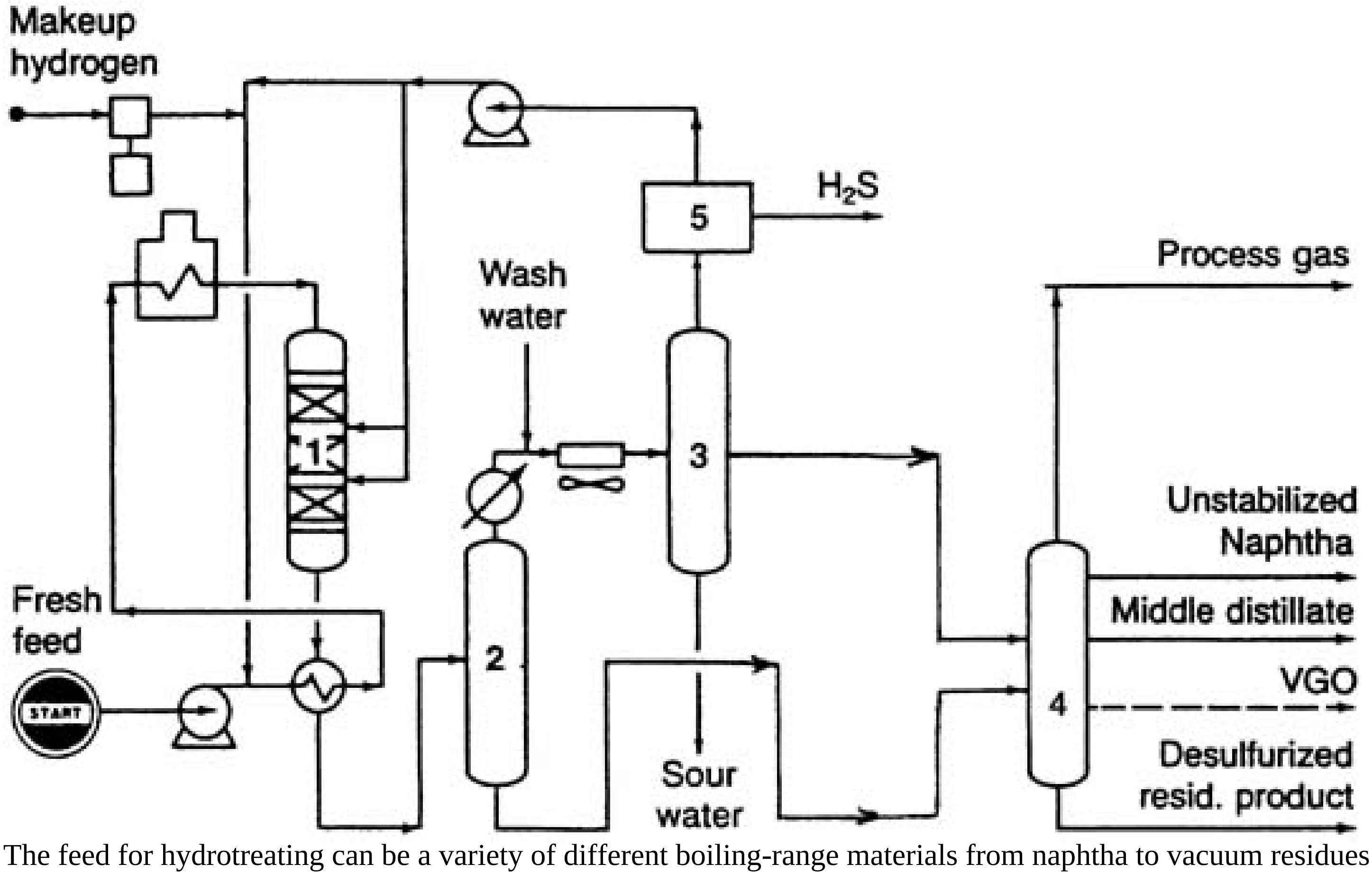
TABLE 15.5 Major Gasoline Sulfur Compounds Normal Boiling Point (NBP) and Hydrocarbon Boiling Range18

Component	NBP	Boiling Range (°F)
Ethyl mercaptan	95	7090
Dimethyl sulfide	99	7580
Iso-propyl mercaptan	126	110130
Tert-butyl mercaptan	147	120150
Methyl ethyl sulfide	151	130140
n-Propyl mercaptan	154	115130
Thiophene	183	140200
Iso-Butyl mercaptan	191	180200
n-Butyl mercaptan	204	185200
Dimethyl disulfide	230	190200
2-Methyl thiophene	234	200250
3-Methyl thiophene	239	210270
Tetrahydrothiophene	250	220260
1-Pentyl mercaptan	259	245255
C2 Thiophene	278	250310
C1 Tetrahydrothiophene	306	260320
Hexyl mercaptan	307	290340
C3 Thiophene	317	300340
C2 Tetrahydrothiophene	318	300340
C3 Tetrahydrothiophene	329	320340
C4 Tetrahydrothiophene	340	320360
C4 Thiophene	361	340380
C5 Thiophene	411	390420
Benzothiophenes and others	427+	400+

Reprinted with permission of Gulf Publishing Co., 2002. Copyright, all rights reserved.

Depending on the severity of the operation, hydrotreating is done at elevated temperatures and pressures. Higher temperatures and pressures are needed to open the complex ring compounds and remove the sulfur molecules. A typical flow diagram is shown in Fig. 15.16.12 This process converts atmospheric and vacuum residue into lighter products. The oil feed is preheated with product streams and a charge heater, and mixed with a hydrogen-rich gas. This mixture is charged to the main reactor and passed over a fixed-bed catalyst system where exothermic hydrogenation reactions occur. Proper internals are needed in the reactor to evenly distribute the feed throughout the catalyst bed and prevent channeling. Product separation is done by a hot high-pressure separator, cold high-pressure separator, and fractionator. In the first high-pressure separator, unreacted hydrogen is taken as the overhead; it is scrubbed to remove hydrogen sulfide (H_2S). The cleaned hydrogen is recycled back to process. In the second high-pressure separator, the remaining gases and light products are removed from the liquid product. If the feed is a wide-boiling range material from which several blending stocks are to be made, the hot, high-pressure separator is followed by a fractionation column. The fractionator separates the treated feed into several liquid product streamsnaphtha, middle distillate, VGO, and a very clean hydrotreated resid product.

Fig. 15.16 HydrotreatingChevron Research and Technology Co., Includes: reactor (1), hot highpressure separator (2), hot low-pressure separator (3), and product fractionator (4). (Hydrocarbon Processing, 79(*11*), *124 (Nov. 2000); Copyright 2000 by Gulf Publishing Co.; all rights reserved.*)



The feed for hydrotreating can be a variety of different boiling-range materials from naphtha to vacuum residues. Generally, each fraction is treated separately to permit optimum operating conditionsthe higher boiling-point materials require more severe treatment conditions. For example, naphtha hydrotreating can be done at 200500 psia and at 500650°F with a hydrogen consumption of 1050 scf/bbl of feed. Conversely, a residue-hydrotreating process can operate at 10002000 psia and at 650800°F, with a hydrogen consumption of 6001200 scf/bbl.[19]

Hydrotreating is a versatile cleanup step; however, it is a large hydrogen-consuming process. Most refineries are able to meet their hydrogen-processing demands with hydrogen recovered from the catalytic reforming process. However, as refiners intensify hydro-treating efforts to meet tighter specifications for products, hydrogen demand will increase. Consequently, one option to balance hydrogen consumption is to construct onsite hydrogen plants to meet present and future hydrogen needs. Purchasing hydrogen from over-the-fence suppliers is another option.[20]

Catalyst formation constitutes a significant difference among hydrotreating processes. Refiners must address reducing sulfur concentrations to lower levels15 ppm for diesel and 30 ppm for gasoline. Consequently, the activity and efficiency of the hydrotreating catalysts become even more vital. Presently, cobaltmolybdenum (CoMo) and nickelmolybdenum (NiMo) catalysts are the preferred hydrotreating systems. CoMo catalysts are very effective at breaking carbon sulfur (C-S) bonds, while NiMo catalysts are more effective at hydrogenation. Just using both systemsCoMo and NiMowill not guarantee optimum results.[14,21] To obtain to the very low ppm levels, desulfurization efforts become more specified at the compound that must be reacted to remove the sulfur from the product stream.

Catalytic Reforming

This process upgrades naphtha (light distillates) into aromatic-rich streams that can be used for octane enhancers for gasoline blending or as a petrochemical feedstock. Originally the process was developed in the 1950s to upgrade low-octane, straight-run gasoline to high-octane liquid, as shown in Table 15.6.[14] This process converts naphthenes into corresponding aromatics and isomerizes paraffinic structures to isomeric forms.[9] The naphtha charge is a varying mixture of C_6C_{11} paraffins, naphthenes, and aromatics. In a catalytic reformer, aromatic compounds pass through the system unchanged, while naphthalenes react selectively to form aromatics.[14]

TABLE 15.6 Aromatics have Higher Octane Numbers[6]

Hydrocarbon Homologs		Octane Number, Clear	
		Motor	Research
C7 hydrocarbons			
n-paraffin	C_7H_{16} (*n*-heptane)	0.0	0.0
naphthene	C_7H_{14} (cycloheptane)	40.2	38.8
	C_7H_{14} (methylcyclohexane)	71.1	74.8
aromatic	C_7H_8 (toluene)	103.5	120.1
C8 hydrocarbons			
n-paraffin	C_8H_{18} (*n*-octane)	-15[a]	-19[a]
naphthene	C_8H_{16} (cyclooctane)	58.2	71.0
	C_8H_{16} (ethylcyclohexane)	40.8	45.6
aromatic	C_8H_{10} (ethylbenzene)	97.9	10.4
	C_8H_{10} (*o*-xylene)	100.0	120[a]
	C_8H_{10} (*m*-xylene)	115.0	117.5
	C_8H_{10} (*p*-xylene)	109.6	116.4

[a] Blending value at 20 vol.% in 60 octane number reference fuel.

In the reformer, multiple reactions occur simultaneously. This process is endothermic and is subject to carbon laydown; thus, refiners must regenerate reforming catalysts. Several catalyst-regenerating approaches are possible. Semi-regenerative processes use moving-bed catalyst reactors. The catalyst bed reactors are placed side-by-side, and hydrogen is used to lift and convey the catalyst to the next bed, except for the last bed where it is regenerated, as shown in Fig. 15.17.[12] Other reforming designs use a continuous moving bed to continuously regenerate a portion of the catalyst. The reactors are stacked on top of each other, and gravity moves the catalyst through the bed. From the last reactor, the catalyst is lifted by nitrogen or hydrogen to a catalyst collection vessel. The catalyst is regenerated in a regeneration tower and returned to process as shown in Fig. 15.18.[12]

Fig. 15.17 Catalytic reformingAxens. Includes: multibed reactors (1,2,3,4). (Hydrocarbon Processing, 81(*11*), *95 (Nov. 2002); Copyright 2002 by Gulf Publishing Co.; all rights reserved*.)

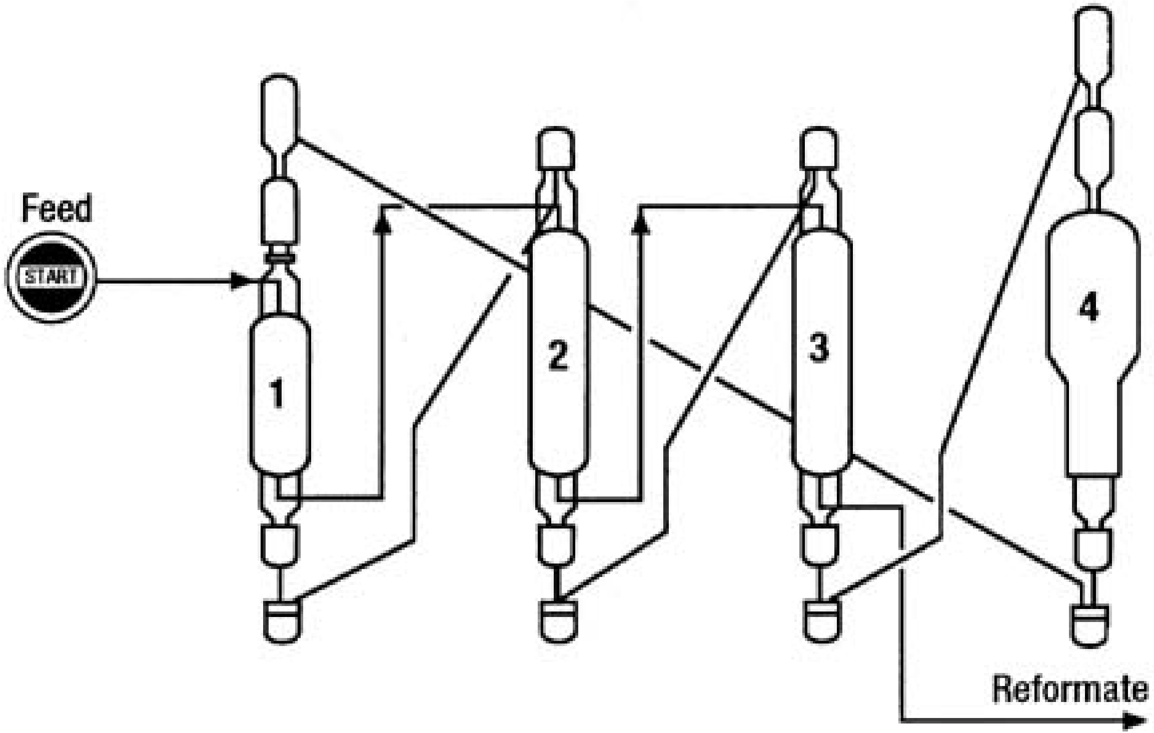
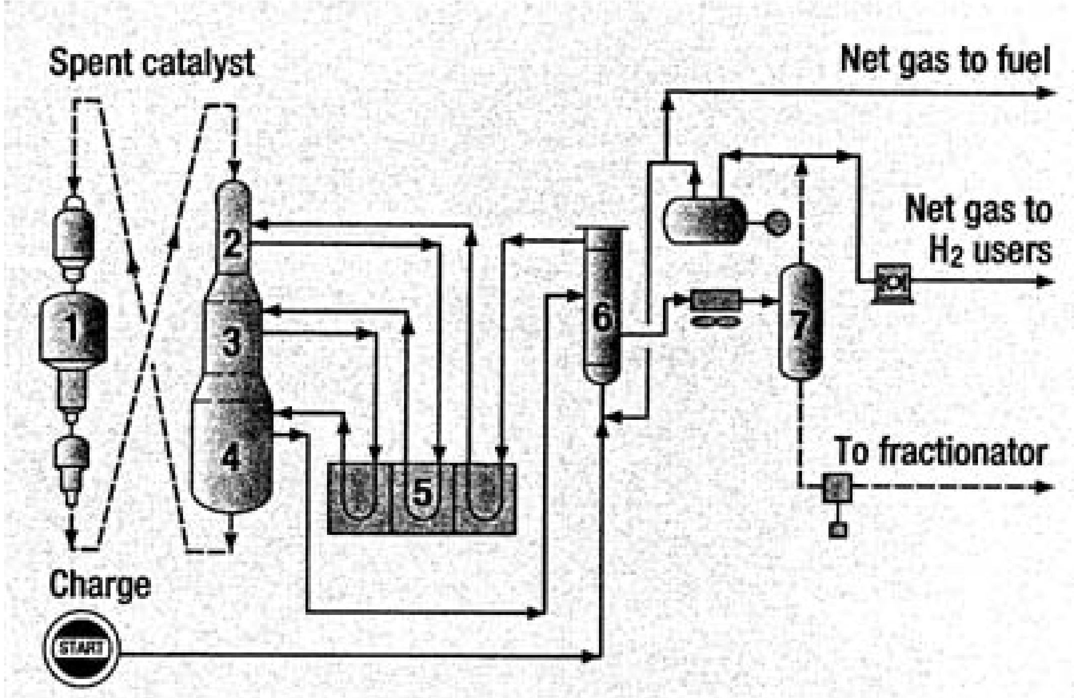


Fig. 15.18 Catalytic reformingUOP LLC. Includes: stacked reactors (1,2,4), and reregenerator (6) with product separation (7). (Hydrocarbon Processing, 81 (*11*), *96 (Nov. 2002); Copyright 2002 by Gulf Publishing Co.; all rights reserved*.)



In the catalytic reforming process, the feed is pumped to operating pressure and mixed with a hydrogen-rich gas before heating to reaction temperatures. The net hydrogen produced is a by-product of the dehydrogenation and cyclization reactions. Several reactions occur:

1. dehydrogenation of naphthene
2. isomerization of paraffins and naphthenes
3. dehydrocyclization of paraffins
4. hydrocracking and dealkylation of paraffins.14

Reforming catalysts promote these reforming reactions. Isomerization is a desired reaction, especially to raise the octane value of the products. However, hydrocracking is an undesired side-reaction that produces light gases. Higher operating pressures are used to suppress hydrocracking. Unfortunately, higher operating pressures suppress reforming reactions also. Generally, a compromise is made between the desired reforming and undesired hydrocracking. The effects of operating conditions on competing reactions are shown in Table 15.7.23

TABLE 15.7 Favored Operating Conditions for Desired Reaction Rates23

				To Get Desired Rate	
Feed	Reaction	Product	Desired Rate	Pressure	Temperature
Paraffins	Isomerization	Iso-paraffins	Inc.	Inc.	Inc.
	Dehydrocyclization	Naphthenes	Inc.	Dec.	Inc.
	Hydrocracking	Lower mol. wt.	Dec.	Dec.	Dec.
Naphthenes	Dehydrogenation	Aromatics	Inc.	Dec.	Inc.
	Isomerization	Iso-paraffins	Inc.	Inc.	Inc.
	Hydrocracking	Lower mol. wt.	Dec.	Dec.	Inc.
Aromatics	Hydrodealkylation	Lower mol. wt.	Dec.	Dec.	Dec.

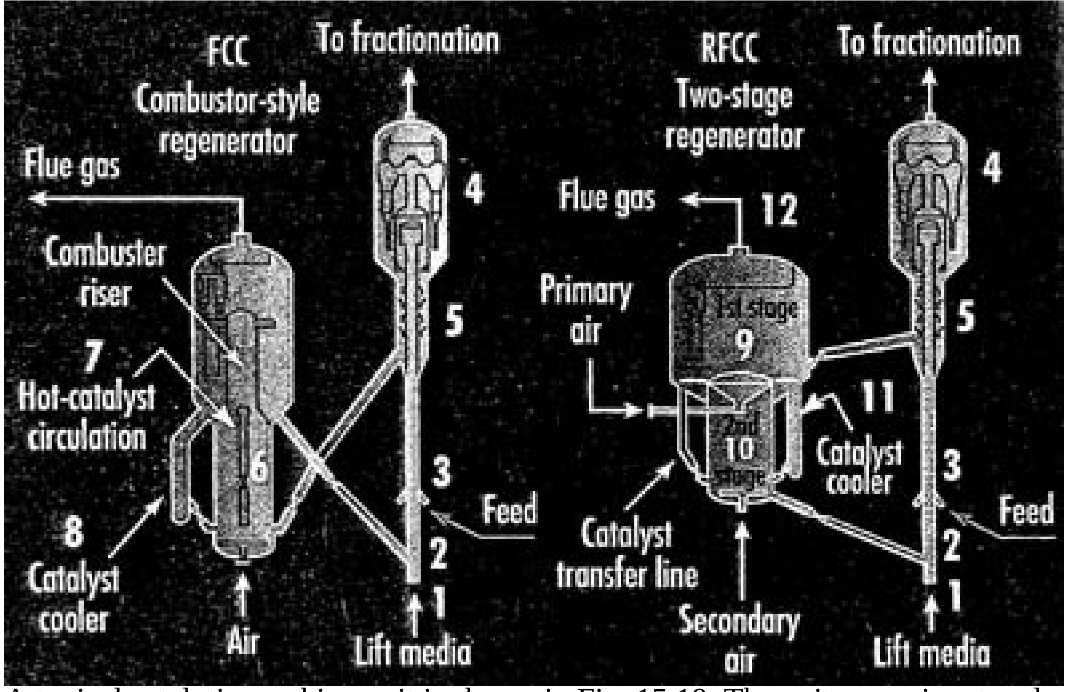
In the late 1960s, it was discovered that adding certain promoters such as rhenium, germanium, or tin to the platinum-containing catalyst would reduce cracking and coke formation. The resulting bi-metallic and tri-metallic catalysts facilitate a lower operating pressure without fostering hydrocracking conditions. Earlier reforming pressures ranged around 500 psig; with improved catalyst systems, such operations now use operating pressures of 170370 psig.12 Advances in continuous catalyst design permit using operating pressures as low as 50 psig.12

Operating temperatures are also critical. The listed reactions are endothermic. The best yields occur along isothermal reaction zones, but are difficult to achieve. Instead, the reaction beds are separated into a number of adiabatic zones operating at 5001000°F with heaters between stages to supply the necessary energy to promote heat of reaction and hold the overall train near or at a constant temperature. Three or four zones are commonly used to achieve high-octane products.

Catalytic Cracking

This process upgrades heavier products into lighter products, as shown in Fig. 15.19. Catalyst systems are used to catalytically crack the large, heavy hydrocarbons into smaller compounds. The products contain a higher hydrogen to carbon ratio than the feed. Consequently, excess carbon is produced, which can deposit on the catalyst and piping and equipment walls; both effects are undesirable.9

Fig. 15.19 Fluid catalytic crackingUOP LLC. Combustor style is used to crack process gas oils and moderately contaminated resids, while the two-stage unit is used for more contaminated oils. (Hydrocarbon Processing, 81(*11*,), *p. 112 (Nov. 2002); Copyright 2002 by Gulf Publishing Co.; all rights reserved.*)



A typical catalytic cracking unit is shown in Fig. 15.19. The unit comprises two large vesselsone to react feed with hot catalyst and the other to regenerate the catalyst by burning off the carbon deposits with air.12 The key feature of the catalytic cracking operation is the short contact time between the catalyst and feed to produce the desired products. Extended contact between the feed and catalyst promotes overcracking and the creation of gases. The activity of newer catalysts selectively promotes primary cracking reactions. New catalysts use a distributed matrix structure; zeolites are applied to the surface and pores of the catalyst, thus increasing the number of active sites.24 With more active sites, the activity of the catalysts increases and less contact time is necessary.

The short contact time is accomplished by using a transfer line between the regenerator and the reactor vessels. Most of the reaction occurs within the riser section.12,14,19 A termination device can be used to separate the catalyst from the products that are taken quickly as overhead. The main reactor vessels contain cyclone separators to remove the catalyst from the products and provide additional space for cracking the heavier fraction of the feed.

In the fluidized catalytic cracking (FCC) process, the feed is injected into the reactor through a feed-nozzle system and mixed with the catalyst. The atomized oil mixes with the catalyst and ascends the riser. The cracking process*riser cracking* or short-time contactinghas several advantages. This system can operate at high temperatures, thus promoting the conversion of feed into gasoline and olefins. It minimizes the destruction of any aromatics formed during cracking.

The net effect is gasoline production with two to three higher octane.

The catalyst is regenerated at high temperatures (13001400°F). Coke that is deposited on the catalyst is quickly burnt off with high-temperature air. Newer catalysts are rugged and can withstand the rigors of extreme heat and fluidizing.

The catalytic cracking unit is often referred to as the gasoline workhorse of a refining unit. As shown in Fig. 15.9, feeds to the catalytic cracking unit are gas oils from the atmospheric and vacuum distillation columns and delayed coker. These heavier fractions also carry metals such as nickel, vanadium, and iron. More important, sulfur compounds concentrate in the heavier product fractions. Table 15.8 lists a typical mass balance for sulfur.25 FCC blendstocks comprise 36 percent of the volume of the gasoline pool. However, this stream also contributes 98 percent of the sulfur concentration to blended procucts.25 As specifications on sulfur concentrations in diesel and gasoline tighten, more efforts are focused on how feeds and product streams from the FCC are pre- and post-treated for sulfur concentrations.

TABLE 15.8 Typical Gasoline Pool Composition of a Refinery25

Gasoline Blendstocks	Percent of Pool Volume	Percent of Pool Sulfur
Alkylate	12	
Coker naphtha	1	1
Hydrocracked naphtha	2	
FCC naphtha	36	98
Isomerate	5	
Light straight-run naphtha	3	1
Butanes	5	
MTBE	2	
Reformate	34	
Total	100	100

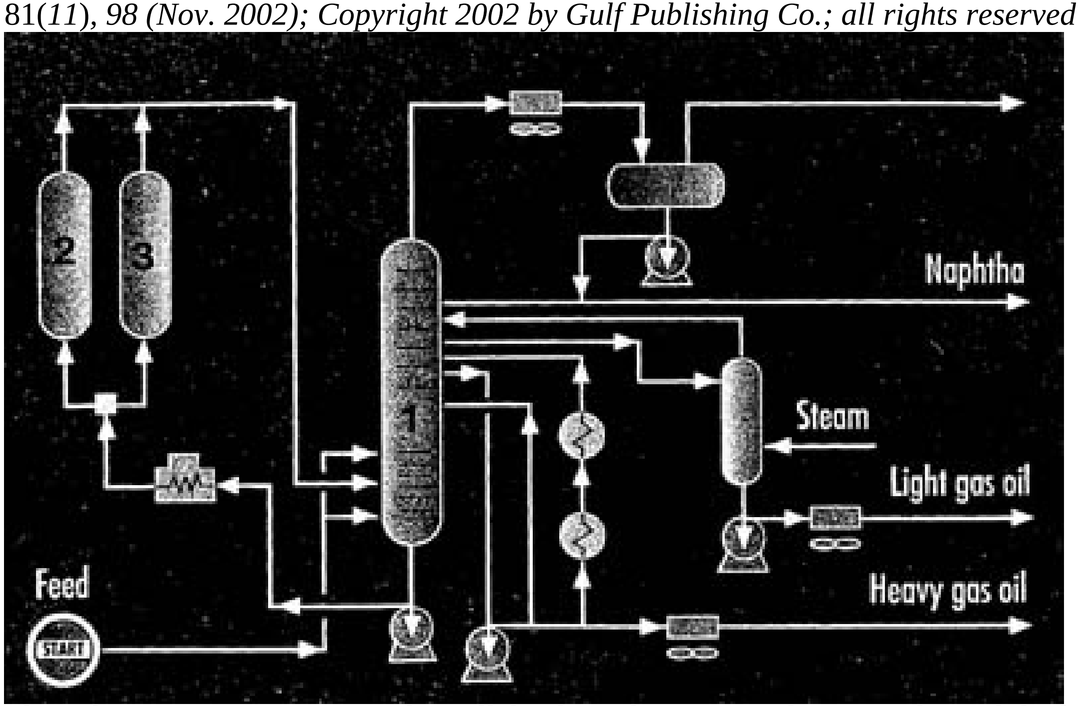
Reprinted with permission of Gulf Publishing Co., 1999. Copyright, all rights reserved.

Coking

Coking is an extreme form of thermal cracking. This process converts residue materialsproducts that are pumpablethat are not easily converted into lighter products through the catalytic cracking processes. Coking is a less expensive method to convert these residual fractions into lighter products. In the coking process, the coke is considered a by-product; its creation is tolerated in the interest of converting the bulk of the residuals into lighter products. The by-product coke can be sold as feedstock to power-generating utility companies. Electricity utility operations burn clean coke to generate high-pressure steam and power. Thus, refiners can sell coke to their over-the-fence power companies and, in return, purchase steam and electricity.26

A typical flow diagram for a delayed coker is shown in Fig. 15.20. Several processing configurations are possible. In this example, the feed is sent directly to the product fractionator to pick up heavier products to be recycled to the cracking operation. The term "delayed coker" indicates that the furnace adds the heat of cracking, and the cracking occurs during the long residence time in the coking drums. The feed and recycled products are heated by the coker heater to the desired operating temperature (900950°F) and then sent to the coking drum, where partial vaporization occurs in addition to mild cracking at pressures ranging between 15 and 90 psig. Overhead vapors from the coke drum are sent to the fractionator and separated into lighter products such as refinery fuel gas, LPG, (coker) naphtha, and light and heavy gas oils.12 The by-product coke accumulates in the coke drum.

Fig. 15.20 Delay cokingFoster Wheeler./UOP LLC. Includes: feed/product fractionator (1), coke drums (2,3), and vapor recovery. (Hydrocarbon Processing, 81(*11*), *98 (Nov. 2002); Copyright 2002 by Gulf Publishing Co.; all rights reserved.*)

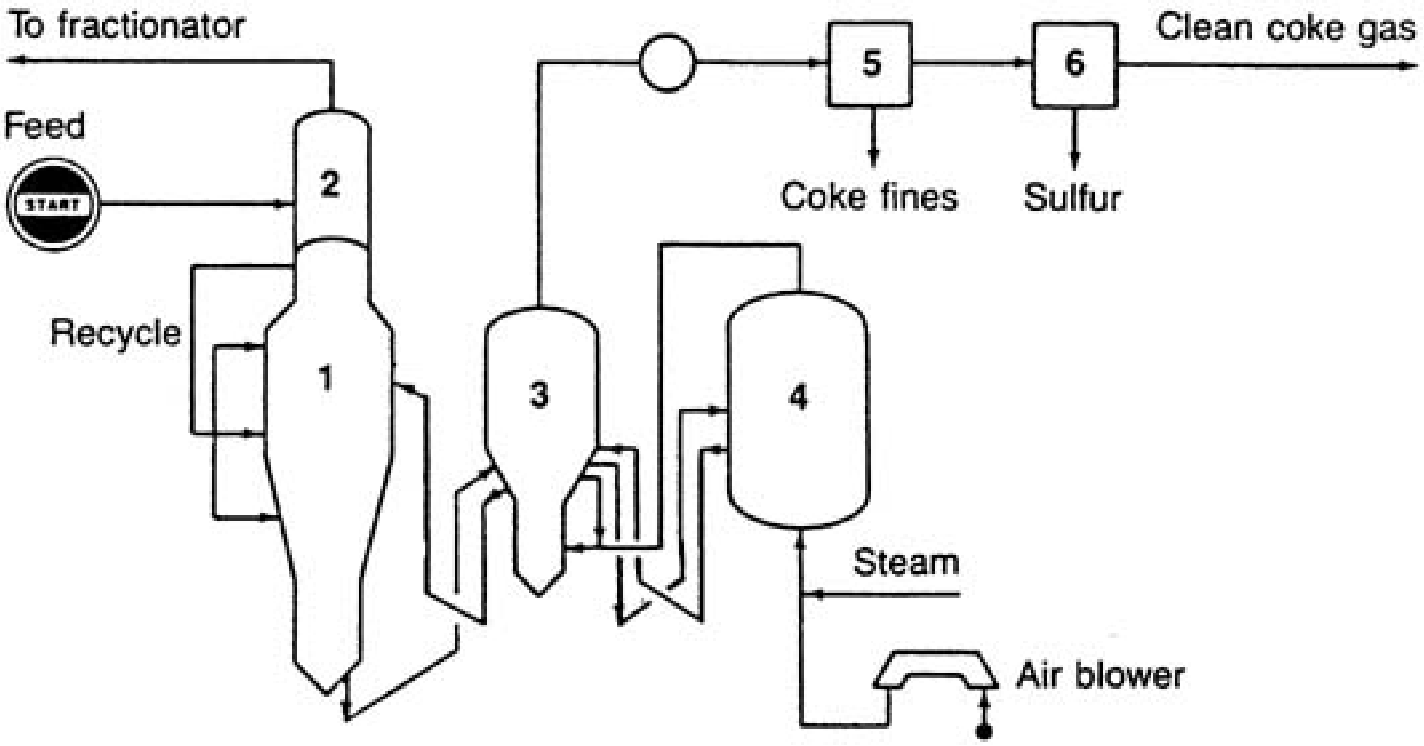


The coking process uses several parallel coking drums. One drum is online while the other is being emptied. At the end of the coking cycle, the coke is steamed to remove any residual oil-liquid.12,14 This mixture of steam and hydrocarbons is sent to the fractionator to recover hydrocarbons. The drum is cooled with water and then drained. After the coke drum is unloaded, high-pressure water jets are used to cut away the coke from the drum. Conveyors are used to move the coke to storage.

Fluid coking is a proprietary name given to a different type of coking process. In this process the coke is suspended as particles in fluids that flow from a reactor to a heater and back again. When the coke is gasified, the process is called *flexicoking*. Both fluid coking and flexicoking are proprietary process of ExxonMobil Research & Engineering Co.

A flow diagram for the flexicoking process is shown in Fig. 15.21. The first two vessels are typical of fluid coking, in which part of the coke is burned in the heater to provide hot coke nuclei to contact the feed in the reactor vessel. The cracked products are quenched in an overhead scrubber where entrained coke is returned to the reactor. Coke from the reactor circulates to the heater where it is devolatilized to yield light hydrocarbon gas and residual coke. A sidestream of coke is circulated to the gasifier, where, for most feedstocks, 95 percent or more of the gross coke is gasified at elevated temperature with steam and air. Sulfur entering the system is converted to hydrogen sulfide, exits the gasifier, and is recovered by a sulfur-removal step.

Fig. 15.21 Fluid coking (Flexicoking)Exxon Research and Engineering Co. Includes: reactor (1), scrubber (2), heater (3), gasifier (4), coke fines removal (5), and hydrogen sulfide removal (6). (Hydrocarbon Processing, 69(*11*), *106 (Nov. 1990); Copyright 1990 by Gulf Publishing Co.; all rights reserved.*)



Hydrocracking

Before the late 1960s, most hydrogen used in processing crude oil was for pretreating catalytic reformer feed naphtha and for desulfurizing middle-distillate products. Later, sulfur requirements for fuels were lowered and became an important consideration. The heavier fractions of crude oil are the fractions with the highest sulfur concentrations and are more difficult to treat. With a constant decline in demand for heavy fuel oils, refiners needed to convert heavier fractions into lighter products. Thus, hydrocracking became a possible solution to the problem.

Figure 15.22 is a typical flow diagram of a hydrocracking process. The process is similar to hydrotreating. The feed is pumped to operating pressure, mixed with hydrogen-rich gas, heated, passed through a catalytic reactor, and distributed among various fractions. Yet this process does significantly differ from hydrotreating. In the hydrocracking process, operating pressures are very high15003500 psia. Hydrogen consumption is also greater12001600 scf of hydrogen per barrel of feed, depending on the extent of cracking.19 If the refinery has a high hydrogen demand due to hydrocracking needs, construction of an onsite hydrogen plant may be necessary.

Hydrocracking catalysts perform a dual function. They provide both hydrogenation and dehydrogenation reactions and have a highly acidic support to foster cracking reactions. The hydrogenationdehydrogenation components of the catalyst are metals such as cobalt, nickel, tungsten, vanadium, molybdenum, platinum, palladium, or a combination of these metals. The acidic support can be silicaalumina, silicazirconia, silicamagnesia, aluminaboria, silicatitania, acid-treated clays, acidic-metals phosphates, or alumina, to name a few.27

Greater flexibility is attributed to most hydrocracking processes. Under mild conditions, the process can function as a hydrotreater. Under severe conditionshigh pressure and temperaturesthis process can produce a variety of motor fuels and middle distillates, depending on the feedstock and operating variables. Even greater flexibility is possible if the process is tailored to convert naphthas into liquefied petroleum gases or convert heavy residues into lighter products.

Hydrocracking is a swing process; it is a treater and a cracker. Thus, this process function can be incorporated into a number of different places within a refining scheme. As a cracker, it can convert feeds that are too heavy or too contaminant-laden to go to catalytic cracking. As a treater, it can handle high boiling point fractions such as heating oil and saturate this fraction to provide good burning quality.

With pending low-sulfur fuels specifications, hydrocracking efforts will be increased to break complex hydrocarbon compounds and expose embedded sulfur molecules. Hydrocracking significantly upgrades feeds to downstream processes and fuel products, especially diesel products.

Alkylation

Another method to convert light olefins into gasoline blending stocks is alkylation. In this process, light olefinspropylene, butylenes, and amylenes with isobutaneare reacted in the presence of strong acids to form branched chain hydrocarbons. These branched hydrocarbons, often referred to as alkylate, have a high-octane value; thus, it is an excellent contributor to the octane pool.9,12,14,16

A flow diagram of an alkylation unit using sulfuric acid is shown in Fig. 15.23. Alkylation traditionally combines isobutane with propylene and butylene using an acid catalyst, either hydrofluoric (HF) acid or sulfuric acid. The reaction is favored by high temperatures, but competing reactions among the olefins to give polymers prevent high-quality yields. Thus, alkylation is usually done at low temperatures to deter polymerization reactions. Temperatures for HF acid-catalyzed reactions are approximately 100°F, and for sulfuric acid they are approximately 50°F.14 Notably, some acid loss occurs with this process. Approximately 11.2 lb of HF acid/bbl of alkylate is consumed, while 2530 lb of sulfuric acid/bbl of alkylate is consumed. The alkylation feed should be dried and desulfurized to minimize acid loss. Since the sulfuric-acid-catalyzed reactions are carried out below normal atmospheric temperatures, refrigeration facilities are included.

Fig. 15.22 HydrocrackingChevron Lummus Global LLC. Includes: staged reactors (1,4), HP separators (2,5), recycle scrubber (3), LP separator (6), and fractionation tower (7). (Hydrocarbon Processing, 81(*11*), *116 (Nov. 2002); Copyright 2002 Gulf Publishing Co.; all rights reserved*.)

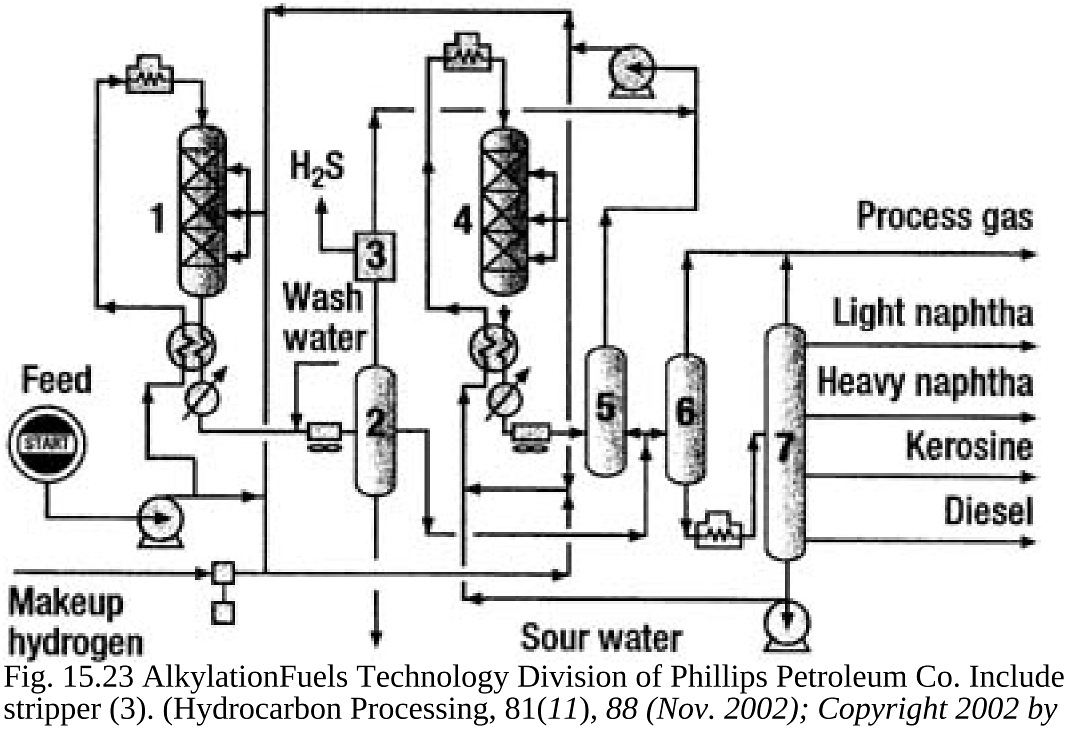
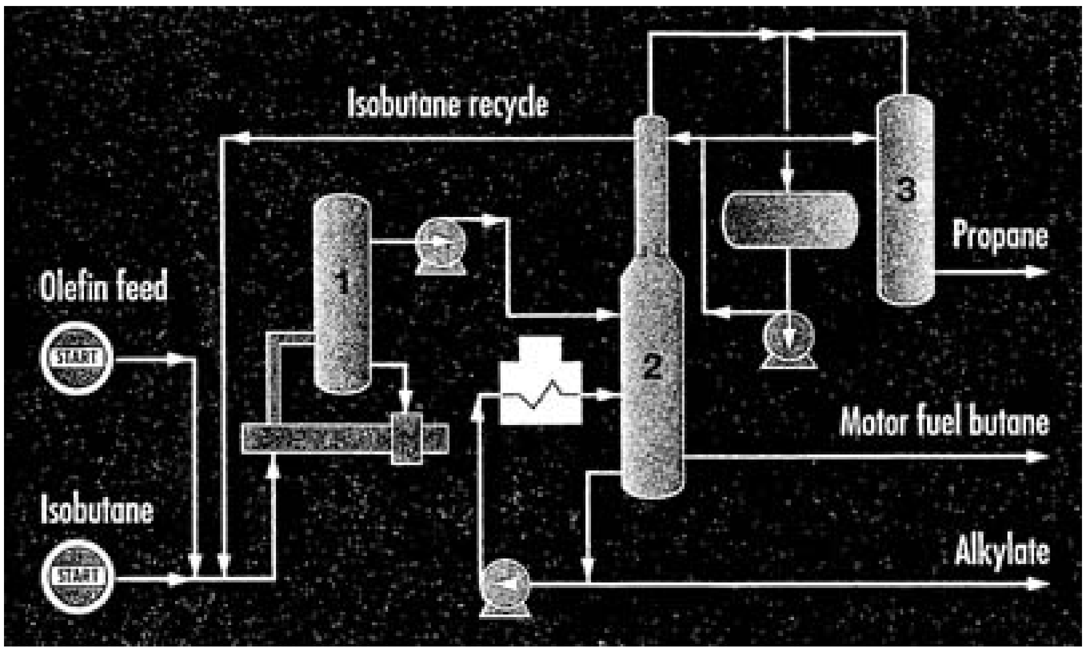


Fig. 15.23 AlkylationFuels Technology Division of Phillips Petroleum Co. Includes: combination reactor/settler (1), main fractionator (2), and small propane stripper (3). (Hydrocarbon Processing, 81(*11*), *88 (Nov. 2002); Copyright 2002 by Gulf Publishing Co.; all rights reserved*.)

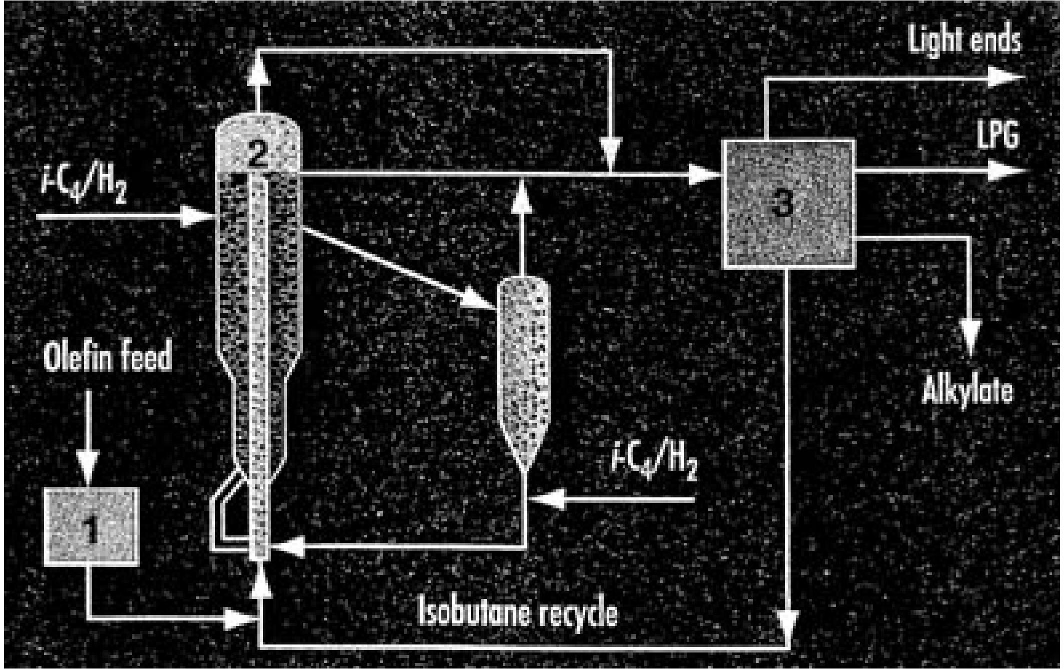


As shown in Fig. 15.23, dry liquid feed containing olefins and isobutane is charged to a combined reactorsettler. In this example, the reactor uses the principle of a differential gravity head to circulate through a cooler before contacting a highly dispersed hydrocarbon feed in the reactor pipe. The hydrocarbon phase, generated in the settler, is sent to a fractionator, which separates LPG-quality propane, isobutane recycle, *n*-butane, and alkylate products. A small amount of dissolved catalyst is also removed from the propane product by a small stripper tower.

Environmental and safety concerns on acid-based processes are promoting research and development efforts on solid-acid alkylation processes. Liquid catalysts pose possible risks to the environment, employees, and the general public from accidental atmospheric releases. Also, these acid catalysts must be regeneratedanother reliability and safety issue. Thus, research efforts are directed at investigating other methodologies to produce high-octane alkylation gasoline component streams.

UOP LLC has developed two alternate processes for liquid-acid alkylation. The direct alkylation method, Alkylene, uses a packed moving catalyst bed. The feed is pretreated to remove impurities such as diolefins, sulfur, oxygen, and nitrogen compounds. These components suppress catalyst activity and can also permanently deactivate the catalyst. As shown in Fig. 15.24, the olefinic feed and isobutane are combined and injected at the bottom of the riser, where the alkylation reaction occurs. At the exit of the riser the catalyst is separated from the hydrocarbons and flows by gravity to the reactivation zone.12,28 The hydrocarbon stream is sent to the fractionation (distillation) section, where alkylate product is separated from the light paraffins and LPG product. Isobutane is recycled back to the main reactor. For this process, the feed is partially dehydrogenated to remove diolefins.

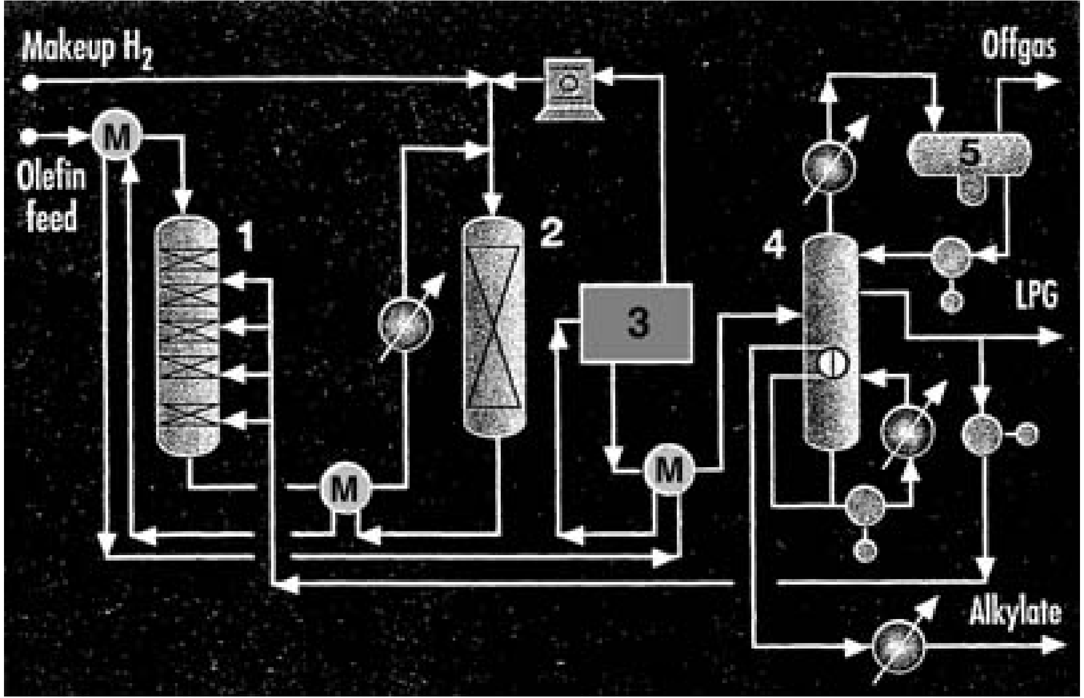
Fig. 15.24 AlkylationUOP LLC. Solid catalyst system removes impurities from the feed (1), clean feed, recycled isobutene, and catalyst and feed to main reactor (2), and final products are separated in fractionation section (3). (Hydrocarbon Processing, 81(*11*), *89 (Nov. 2002); Copyright 2002 by Gulf Publishing Co.; all rights reserved*.)



A second indirect alkylation process, InAlk, also is a solid catalyst process (Fig. 15.25).12,29 InAlk combines two commercially proven technologiespolymerization and olefin saturation. Isobutylene is reacted with light olefins (C3C5) in a polymerization reactor. The resulting mixture of iso-olefins is saturated in the hydrogenation reactor. Excess hydrogen is recycled and the product is stabilized to produce a paraffinic gasoline blending stream.

Yet, new solid-acid alkylation processes face technical challenges. Solid-acid catalysts remain more difficult to regenerate and have a shorter service life. Research to overcome these operating problems is continuing.

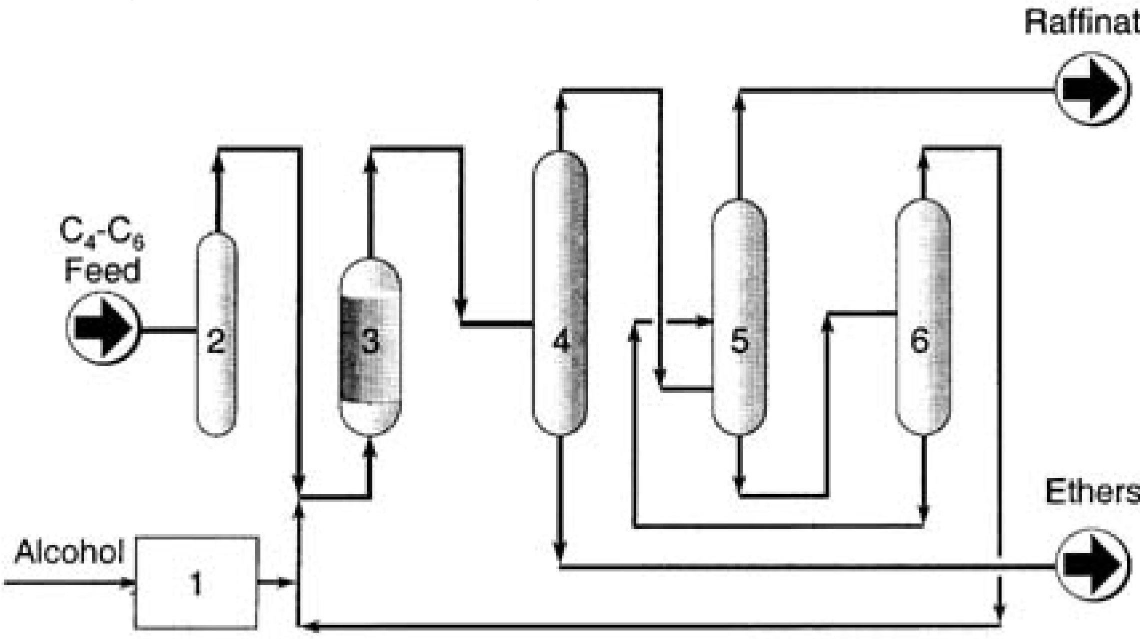
Fig. 15.25 Alkylation (indirect)UOP LLC. Process uses solid catalyst; major steps include: polymerization (1), hydrogenation reactor (2), hydrogen recovery (3), and product stabilization (4). (Hydrocarbon Processing, 81(*11*), *90 (Nov. 2002); Copyright 2002 by Gulf Publishing Co.; all rights reserved*.)



Ether Processes

Refiners have always incorporated ethers into the gasoline pool when needing to increase octane. Ethers provide a high-octane stream with low vapor pressure. Beginning in 1995, U.S. reformulated gasoline (RFG) was required to have 2 wt. percent oxygenate content. The choice of oxygenate was left to the refiners' discretion. Initially, the ethanol industry had hoped that ethanol would be selected as the primary oxygenate for RFG. However, refiners searched for other options. Methyl tertiary butyl ether (MTBE) became the oxygenate of choice for blending RFG. It is produced by reacting methanol with isobutylene, as shown in Fig. 15.26.38 Other ether compounds can be made in a similar process. Ethyl tertiary butyl ether (ETBE) is produced from ethanol and isobutylene; tertiary amyl methyl ether (TAME) is made from methanol and isoamylene. MTBE holds the lion's share of the oxygenate market.

Fig. 15.26 EthersAxens. Includes: alcohol purification (1), selective hydrogenation (2), main reactor (3), fixed-bed main reactor (4), reaction column actually operates as a conventional distillation column where ethers are the bottoms product (5), water wash, and (6) standard distillation column. (Hydrocarbon Processing, 81(*11*), *106 (Nov. 2002); Copyright 2002 Gulf Publishing Co.; all rights reserved*.)



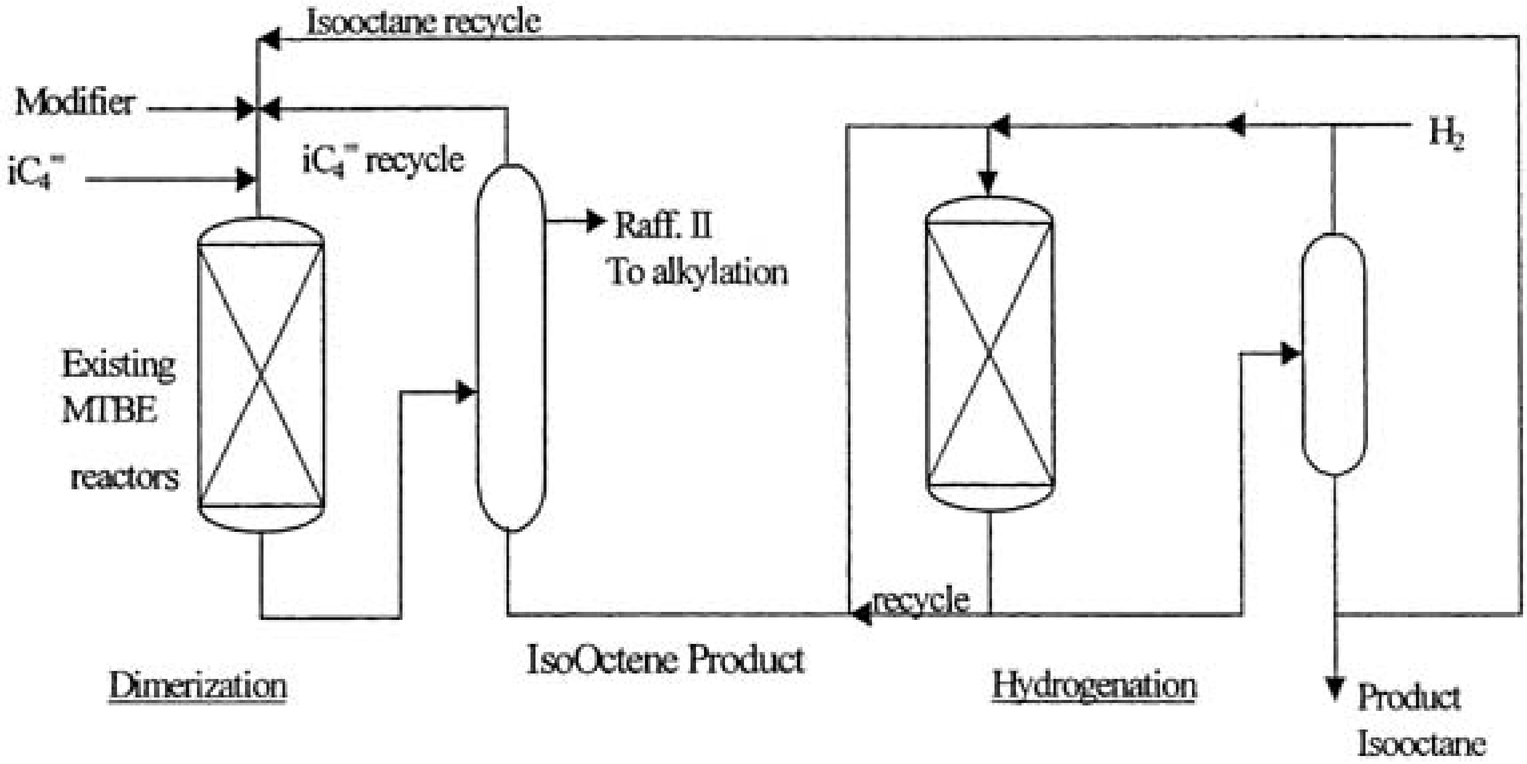
Presently, MTBE's usage is at the center of a highly emotional, political struggle. As refiners began using MTBE at the 2 wt. percent concentration, this ether was detected at trace levels in drinking water supplies in areas required to use RFG. Leaking underground storage tanks (USTs) were identified as the primary source for MTBE found in drinking water. The ether is found in water due its high affinity for water. Legislation was (and is) in place that required UST owners to install safeguards to prevent leakage; however, leakage did occur.

Because drinking water was affected by leakage from USTs, a political effort is underway to ban the use of MTBE in gasoline. At the time of this writing, much debate is ongoing. The State of California and several other states have banned using MTBE in RFG. However, the consequence of shortages in RFG supplies has raised some doubts and caused second thoughts on the MTBE ban.

Oxygenate usage is required also for gasolines supplied to nonattainment areas. The sister ethersETBE and TAMEare under the same disfavor as MTBE and are not logical replacements. Ethanol can be a substitute for MTBE; however, it also brings other problems. Ethanol usage raises the Reid vapor pressure of the fuel, which also is limited by fuel specifications. In the United States, ethanol receives an agricultural subsidy that elevates its price. In the U.S. market, ethanol capacity is located far from refining centers and fuel terminals; thus, transportation, distribution, and storage issues are challenges for ethanol.30 Of course, the ultimate fate of MTBE capacity constructed in the mid-1990s to meet RFG blending demand is yet to be determined.

Should MTBE be banned, what would be the logical replacement(s)? There are several options available. Several refiners opted to build MTBE capacity and avoid purchasing the ether on the open market. MTBE units were an option to use the facility's isobutylenes. Several licensed processes can be used to convert existing MTBE units. Kvaerner and Lyondell Chemical Co. offer technologies to convert an MTBE unit to produce iso-octane, as shown in Fig. 15.27.12 Snamprogetti SpA and CDTECH also have an iso-octene/iso-octane process. These processes can use various feedstocks such as "pure" iso-butane, steam-cracked C4 raffinate, 50/50 iso-butane/iso-butene feeds, and FCC butanebutane streams. The process selectively dimerizes C4 olefins to iso-octene and then hydrogenates the iso-octene (di-iso-butene) into iso-octane. The processes were developed to provide an alternative to MTBE. The dimerization reactor uses a catalyst similar to that for MTBE processes; thus, the MTBE reactor easily can be converted to iso-octane service. The product is low sulfur with 100 octane numbergreat for gasoline blending.31

Fig. 15.27 Iso-octaneLyondell Chemical and Kvaerner. (Hydrocarbon Processing, 78, *3 (Mar. 2001); Copyright 2001 by Gulf Publishing Co; all rights reserved*.)



15.9 Future Trends

The refining industry is constantly required to meet cleaner fuel specifications. The only certainty is that, globally, fuels are becoming cleaner. Yet, clean fuels comprise a very broad spectrum that is totally dependent on the market served. In Asia and some parts of Europe, lead removal still remains a key issue. These refiners strive to maintain octane while phasing out lead. In other markets, more complex issues remain to be solved.

In developed markets, cleaner-fuel issues revolve around several product specifications. The fuels market is at various stages of removing sulfur and aromatics from gasoline and diesel. For global organizations, formulating a clean-fuels agenda is a very market-dependent issue, as shown in Table 15.4.

Costs and Drivers

So how much investment will be made to process low-sulfur fuels? A recent Energy Information Administration (EIA) report estimates that U.S. refiners will invest $6.39.3 billion to meet full compliance with the ultra-low sulfur diesel (ULSD) rule through 2011.33 For Europe, refiners made investments of nearly $22.9 billion from 1997 to 2005 to meet gasoline and diesel specifications. From 2005 to 2015, an additional $9.714 billion may be spent on improving the quantity and quality of middle distillates in Europe.15 Thus, considerable investment will be necessary to bring cleaner fuels to market.

Really unknown, besides how much a clean fuels program will cost individual organizations, is how many refiners will make such an investment. Developed markets suffer from diminished demand growth for products. Mediocre returns from earlier environmental projects taint possible returns on future spending. With such a backdrop, hesitancy on expected spending is anticipated. Yet, the deadlines are getting nearer, and plans must be formulated.

Drivers for investment are directly linked to market demand. In the European Union, gasoline demand is declining, while demand for diesel is increasing with modest increases for treating gas oil. Heavy fuel oil demand is also declining. Under such market forces, the EU product market is shifting and demand for middle distillates is increasing from 310 million tons (310 MMt) in 1997 to 390 MMt in 2015.15 Production of LPG, naphtha, and other products is expected to increase.15

Under such conditions, the focus of retrofitting and revamping existing units will be optimizing middle distillate production. Consequently, to meet higher quality requirements for middle distillates, EU refiners are projected to invest in hydrocracking capacity. Such investments may include standalone units and moderate conversion of units upstream of the FCC.15 Refiners will also raise hydrotreating capability.

Technology Options

Due to the individuality of each refinery, multiple solutions are available. Selection will be directed toward final product slates for each facility. For gasoline-oriented facilities, several options are available. These are discussed below.

Sweetening

This application is most effective in treating straight-run (SR) gasoline streams. Amines are used to remove mercaptan species from the hydrocarbon stream. Caustic converts the mercaptans into disulfides, which are extracted by gravity separation or an extractive solvent.12

Hydrotreating

This treatment is used widely throughout the refinery. Hydrogen is reacted with the processing stream with a catalyst to remove sulfur compounds. Several licensed technologies are available at varying temperature and pressure ranges (Table 15.9).3436 Hydrotreating, depending on the severity of process conditions, effectively removes sulfur, nitrogen, metals, carbon residue, and asphaltene from the hydrocarbon stream. Mild hydrotreatinglow pressure and temperatureremoves sulfur and trapped metals and is done as a cleanup/upgrade step for downstream processes such as hydrocrackers, FCCs, resid catalytic crackers, and cokers. Processing streams typically treated include: naphthas, kerosenes, distillates, and gas oil (Fig. 15.9). This process consumes hydrogen.

TABLE 15.9 Available Licensed Post-treating Technologies3436

Approach	Process Name	Technology Provider
Conventional hydrotreating	Hydrotreating	Many
Selective hydrotreating	Scanfining	ExxonMobil
	Prime G	IFP
	Octgain 125	ExxonMobil
	Octgain 220	ExxonMobil
	ISAL	UOP LLC
Catalytic distillation	CDHydro/CD HDS	CD Tech
Adsorption	S Zorb SRT	Phillips Petroleum
Olefinic alkylation	Olefin alkylation of thiophenic sulfur	BP
Extractive mass	Exomer	ExxonMobil and Merrichem

For a conventional hydrotreating process, the process stream is heated and mixed with hydrogen. This stream is charged to a reactor filled with a high-activity catalyst. Single and multiple reactors can be used. Excess hydrogen is used. The reactor effluent is cooled and separated, and the hydrogen-enriched gas is recycled. Depending on sulfur levels of the feed, the hydrogen recycle may be amine scrubbed to remove hydrogen sulfide (H_2S). The liquid product is steam-stripped to remove lighter components and residual H_2S, and/or fractionated into multiple products.12

As sulfur levels for products continue to decrease, refiners strive to systematically remove sulfur from the blending pool. Notably, back integration to strategically and systematically remove sulfur compounds throughout the processing scheme is needed. Lower sulfur specs now mandate removing complex sulfur species from various blending streams.

Hydrotreating is an effective method to desulfurize products. However, as desulfurization requirements continue to increase, newer, high-activity catalyst systems will be required to meet desulfurization targets. Existing capacity can be retrofitted with better catalyst systems. Improved reactor internals can more efficiently distribute feed throughout the reactor, optimize contact of reactants with the catalyst, and increase desulfurization efficiency. Refiners may elect to raise operating severity by increasing operating temperatures. However, higher operating temperatures can reduce the service life of the catalyst. Raising operating pressures is not be viable and will depend on the pressure rating of the existing reactor. Yet, the new fuel specifications may warrant installing new reactor capacity.

As processing conditions increase, that is, pressure and temperature, the hydrotreating process emerges more as a hydrocracking/hydrodesulfurization/ hydrodearomatizing event. Deeper desulfurization of processing/ product streams is inevitable.

Sulfur Segregation

Another option is to concentrate sulfur compounds into various streams and selectively treat them. Refiners can undercut* product; however, such tactics will reduce yields. For diesel, undercutting will lower diesel yield and increase gas-oil productsan undesirable consequence.

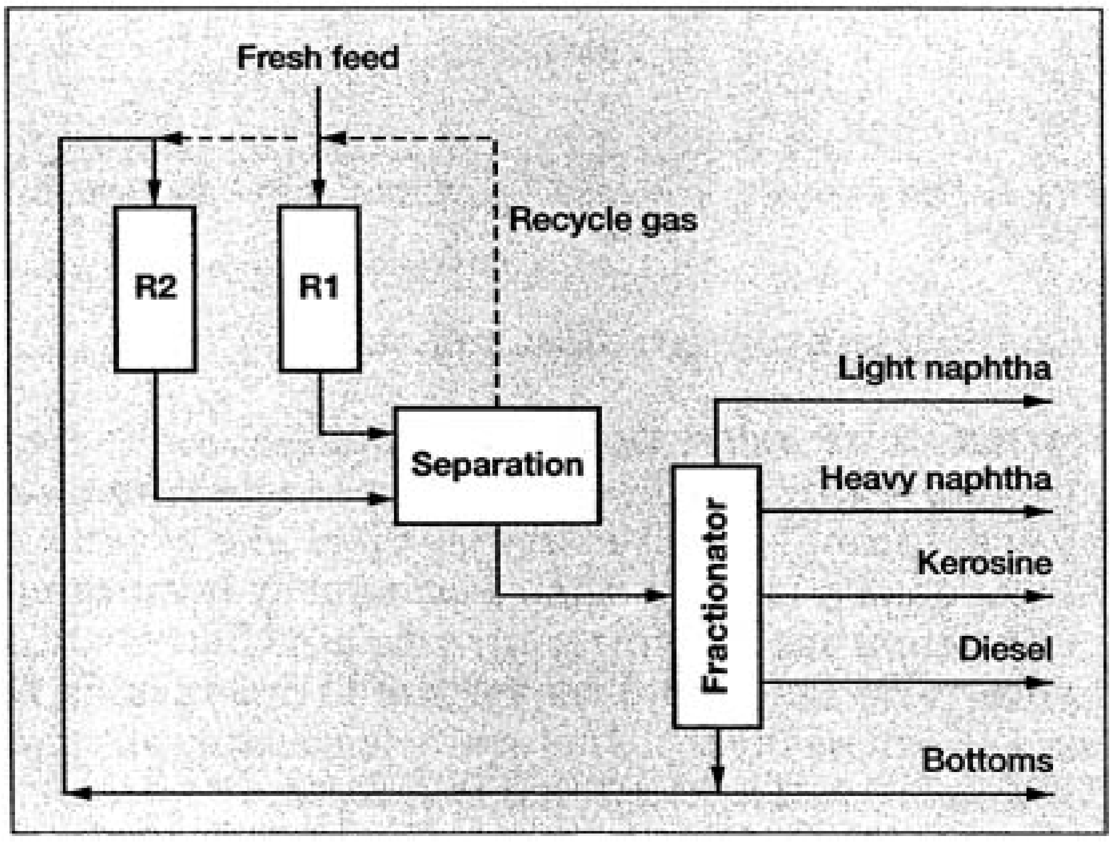
Undercutting is using distillation temperatures that are below the specification for the product. The refiner loses some light product to higher-temperature products as insurance to met light-product specifications. It is product "give-away" and not a good practice.

Desulfurization

As refiners strive to meet tighter restrictions on sulfur in product and blending streams, desulfurization technologies are attracting more interest. For gasoline, licensed post-treatment processes that are targeted at specific streams have high interest. Half the battle for clean fuels is identifying the highest sulfur contributors to the blending pool. For gasoline-oriented facilities, the FCC unit is the highest contributor. Nearly 9598 percent of the sulfur present in blended gasoline is linked to FCC product cuts. Notably, FCC technology is instrumental in upgrading heavier refining streams into desired products. The school of thought for FCC is split between pretreating FCC feed and post-treating. Pretreating offers some great benefits. In particular, hydrotreating feed streams to the FCC upgrades the feed by removing sulfur and nitrogen compounds and saturates some aromatics. It can increase LPG and gasoline production, while reducing regenerator-SO_x emissions, light-cycle oil (LCO), and clarified-slurry oil (CSO) yields, and minimizing coke formation.25,37 More important, it is reported that hydrotreating does not affect naphtha octane values.25 However, hydrogen consumption becomes a limiting issue. The available hydrogen balance will affect how much hydrotreating can be done economically.

Another pretreat option is to replace hydrotreating the FCC feed with partial conversion hydrocracking operations. There are several variations to the process. However, partial conversion hydrocracking operation splits the FCC feed. Difficult-to-convert materials such as LCO and coker gas oils can be upgraded with hydrotreating and hydrocracking before being sent to the FCC unit. UOP LLC's Unicracking process uses two reactors and separates the hydrotreating and hydrocracking processes into distinct zones, as shown in Fig. 15.28. The sulfur content determines the severity of the hydrotreating reactor. This process is estimated to yield more naphtha and distillates than other desulfurization methods.25 It also produces a higher quality diesel (cetane index of 50).

Fig. 15.28 Two-staged, partial-conversion hydrocracking process.25 (*Reprinted with permission of Gulf Publishing Co.; Copyright 1999 by Gulf Publishing Co.; all rights reserved.*)



The Axen's mild hydrocracking process combines an ebullated-bed, mild hydrocracking process with an inline fixed-bed to pretreat FCC feed. The T-Star process can treat vacuum gas oil (VGO) and deasphalted oils, and offers high selectivity toward diesel products. The efficient catalyst system enables more flexibility for processing various FCC feeds. The feed streams are cracked first and then hydrotreated.

Post-treating processes hold keen interest also. Table 15.9 lists several post-treating methods. Many of the newer processes recently have been commercially demonstrated and several commercial units have been completed and are now in operation. How effective these new processes will prove to be remains to be seen.

Tomorrow's fuels will contain less sulfur to be compatible with the more sophisticated engine designs of new automobiles. Fuels specifications will further reduce the concentrations of gasoline compounds that are listed and/or considered to be toxic. Refiners will continue to upgrade heavier components to lighter products and refineries will continue to use innovative catalysts and processing equipment to cost-effectively manufacture fuels and petrochemical feedstocks. We have barely touched on the sophisticated engineering needed to transform a barrel of crude oil into consumable products.

15.10 References

U.S. Energy Information Administration, International Energy Annual 2002, www.eia.doe.gov/oiaf/ieo.
Cambridge Energy Research Associates (CERA), Global Oil Trends 2002, 2001.
National Petroleum Council, "U.S. Petroleum Refining: Assuring the Adequacy and Affordability of Cleaner Fuels," June 20, 2000.
Hydrocarbon Processing, 34-D35-D (2002, February)
American Society for Testing and Materials, 1990 Annual Book of ASTM Standards, Vols. 05.01 and 05.02, 1990.
Watson, K. M., and Nelson, E. F., "Improved Methods for Approximating Critical and Thermal Properties of Petroleum Fractions," Ind. Eng. Chem., 25, 880 (1933).
Hoffman, H. L., "Sour Crudes Limits Refining Output," Hydrocarbon Processing, 107110 (1973, September).
Ferrero, E. P., and Nichols, D. T., "Analyses of 169 Crude Oils from 122 Foreign Oil Fields," U.S. Department of Interior, Bureau of Mines, Information Circular 8542, 1972.
Eckholm, E., Refining Fundamentals: Part 2 Light Fraction Processing, February 1999.
Hydrocarbon Processing, 1316 (2000, September).
Unzelman, G. H., and Forster, E. J., "How to Blend for Volatility," Petroleum Refiner, 109140 (1960, September).
"Refining Processes 2002," Hydrocarbon Processing, 85148 (2002, November).
Huges, T. R., et al., "To Save Energy When Reforming," Hydrocarbon Processing, 7580 (1976, May).
Meyers, R. A. (Ed.), Handbook of Petroleum Refining Processes, 2nd ed., McGraw-Hill, New York, 1997.
Birch, C. H., and Ulivier, R., "ULSG Diesel Refining Study," Purvin & Gertz, November 17, 2000.
Eastman, A., et al., "Consider Online Monitoring of HF Acid When Optimizing Alkylation Operations," Hydrocarbon Processing (2001, August).
"Petrochemical Processes 2001," Hydrocarbon Processing, 71146 (2001, March).
Golden, S. W., et al., "Use Better Fractionation to Manage Gasoline Sulfur Concentration," Hydrocarbon Processing, 6772 (2002, February).
Corneil, H. G., and Forster E. J., "Hydrogen For Future Refining," Hydrocarbon Processing, 8590, (1990, August).
Ratan, S., and Vales, C. F. "Improve your Hydrogen Potential," Hydrocarbon Processing, 5764 (2002, March).
Skiflett, W. K., and Krenzke, L. D. "Consider Improved Catalyst Technologies to Remove Sulfur," Hydrocarbon Processing, 4143 (2002, February).
American Petroleum Institute, Technical Data Book, 4th ed., Publ. No. 999, 1983.
Jenkins, J. H., and Stephens, J. W., "Kinetics of Cat Reforming," Hydrocarbon Processing, 163167 (1980, November).
"HP Innovations," Hydrocarbon Processing, p. 33 (2000, November).
Shorey, S. W., et al., "Use FCC Feed Pretreating Methods to Remove Sulfur," Hydrocarbon Processing, 4351 (1999, November).
Hydrocarbon Processing, 1516 (1999, December).
Sullivan, R. F., and Meyer, J. A., "Catalysts Effects on Yields and Product Properties in Hydrocracking," American Chemical Society, Philadelphia, April 611, 1975.
UOP, Product sheetAlkylene, 1999.
UOP, Product sheetIndirect Alkylation (InAlk), 1999.
"HP Insight," Hydrocarbon Processing, p. 13 (2002, April).
Tsai, M. J., et al., "Consider New Technologies to Replace MTBE," Hydrocarbon Processing, 8188 (2002, February).
Heck, R. M., et al., "Better use of Butenes for High-octane Gasoline," Hydrocarbon Processing, 185191 (1980, April).
EIA, "The transition to Ultra-low Sulfur Diesel: Effects on Process and Supply," May 2001.
Avidan, A., et al., "Improved Planning can Optimize Solutions to Produce Clean Fuels," Hydrocarbon Processing, 4753 (2001, February).
"2001 Clean Fuels Challenge," transcript-Question 2, NPRA Clean Fuels Conference, Houston, Aug. 2829, 2001.
Fredrick, C., "Sulfur Reductions: What Are the Options?," Hydrocarbon Processing, 4550 (2002, February).
Nocca, J. L., et al., "The Domino Interaction of Refinery Processes for Gasoline Quality Attainment," NPRA 2000 Annual Meeting, San Antonio, March 2628, 2000.
"Petrochemical Processes 1995," Hydrocarbon Processing, 109111 (1995, March).

16.1 Characteristics

Natural gas is a naturally occurring mixture of simple hydrocarbons and nonhydrocarbons that exists as a gas at ordinary pressures and temperatures. In the raw state, as produced from the earth, natural gas consists principally of methane (CH_4) and ethane (C_2H_4), with fractional amounts of propane (C_3H_8), butane (C_4H_{10}), and other hydrocarbons, pentane (C_5H_{12}) and heavier. Occasionally, small traces of light aromatic hydrocarbons such as benzene and toluene may also be present.

The ethane and heavier hydrocarbon components of natural gas are rather easily separated from the gas stream and liquefied under moderate pressure. Most of these components are separated and recovered because of their added value when they are sold as separate products. Unprocessed natural gas is described as "rich" ("wet") or "lean" ("dry"), depending on the amounts of liquefiable heavier components present in the unprocessed gas. (The dividing line is not specific.)

In addition to hydrocarbon components, raw natural gas from some reservoirs may contain varying amounts of nonhydrocarbon contaminants, or diluents, such as hydrogen sulfide (H_2S), carbon dioxide (CO_2), water (H_2O), nitrogen (N_2), and/or helium (He). Natural gas containing the first two of these compounds, hydrogen sulfide and carbon dioxide, is termed "sour" and the contaminants are referred to as "acid" gases. Natural gas that contains low enough concentrations of the acid gases to meet sales specifications is termed "sweet."

Although some natural gas is suitable for use as a fuel as produced, virtually all natural gas undergoes some processing to produce a merchantable gas the composition of which is principally methane and ethane. Water and acid gas components are removed to prevent freezing, corrosion, or other operating problems in transmission or utilization. Diluents that exist in significant quantities usually are removed to improve the combustion and/or heating properties of the gas. Much of the ethane and propane and essentially all of the butanes and heavier hydrocarbons are removed because of their greater value when sold as separate liquid products.

Table 16.1 gives typical raw gas compositions. "Casinghead" gas is typically collected from liquidgas separators in oil production facilities. Because the separators generally operate at low pressures, the gas from this source tends to be richer in liquefiable hydrocarbons content. "Gas well" gas and "condensate well" gas generally are produced at higher pressure and with little or no accompanying liquid, so their compositions show less liquefiable hydrocarbon content. These characteristics describe "traditional" natural gas or gas from commonly thought of sources. Today gas is available or becoming available from nontraditional sources. This "unconventional gas" tends to have a composition that is much more like the "gas well" gas than "casinghead" gas.

TABLE 16.1 Typical Raw Gas Compositions (Water-Free Basis)1

	Casinghead (Wet) Gas		Gas Well (Dry) Gas		Condensate Well Gas	
	(Mol. %)	(Gal/mcf)%	(Mol. %)	(Gal/mcf)%	(Mol. %)	(Gal/mcf)%
Carbon dioxide	0.63					
Nitrogen	3.73		1.25		0.53	
Hydrogen sulfide	0.57					
Methane	64.48		91.01		94.87	
Ethane	11.98		4.88		2.89	
Propane	8.75	2.399	1.69	0.463	0.92	0.252
iso-Butane	0.93	0.303	0.14	0.046	0.31	0.101
n-Butane	2.91	0.914	0.52	0.163	0.22	0.069
iso-Pentane	0.54		0.09		0.09	
n-Pentane	0.80		0.18		0.06	
		0.777		0.203		0.103
Hexanes	0.37		0.13		0.05	
Heptanes plus	0.31		0.11		0.06	
	100.00	4.393	100.00	0.875	100.00	0.525

Source: Gas Processors' Association.

Table 16.2 shows the physical properties of the principal constituents of natural gas as produced. Table 16.3 presents typical specifications for pipeline quality, or "merchantable," natural gas.

TABLE 16.2 Properties of Natural Gas and Natural Gas Liquids2

	Chemical Formula	Molecular Mass	Boiling Point @14.496 psi(°F)	Vapor Pressure @100°F (psia)	Gas Density Air = 1	Heating Value BTU/cf Gas
Methane	CH_4	16.043	-258.73	(5000)*	0.554	1010.0
Ethane	C_2H_6	30.070	-127.49	(800)*	1.038	1769.6
Propane	C_3H_8	44.097	-43.75	188.64	1.523	2516.1
i-Butane	C_4H_{10}	58.123	10.78	72.58	2.007	3251.9
n-Butane	C_4H_{10}	58.123	31.08	51.71	2.007	3262.3
i-Pentane	C_5H_{12}	72.150	82.12	20.45	2.491	4000.9
n-Pentane	C_5H_{12}	72.150	96.92	15.57	2.491	4008.9
Hexane	C_6H_{14}	86.177	155.72	4.96	2.976	4755.9
Heptane	C_7H_{16}	100.204	209.16	1.62	3.460	5502.5
Octane	C_8H_{18}	114.231	258.21	0.54	3.944	6248.9
Carbon dioxide	CO_2	44.01	-109.26		1.520	0
Hydrogen sulfide	H_2S	34.08	-76.50	394.59	1.177	637.1

*Extrapolated. Above critical temperature.

Source: Gas Processors' Association Standard 2145.

TABLE 16.3 Typical Pipeline Quality Natural Gas3

	Minimum	Maximum
Major and minor components, mol %		
emsp;Methane	75	
emsp;Ethane		10
emsp;Propane		5
emsp;Butanes		2
emsp;Pentanes and heavier		0.5
emsp;Nitrogen and other inerts		18
emsp;Carbon dioxide		3
emsp;Hydrogen		5
emsp;Total unsaturated hydrocarbons		0.5
emsp;Carbon monoxide		0.1
Trace components		
emsp;Hydrogen sulfide scf		0.25 gr/100
emsp;Mercaptan sulfur scf		0.50 gr/100
emsp;Total sulfur scf		1.0 gr/100
emsp;Water vapor		7.0 lb/mmcf
emsp;Oxygen		10 ppmv

Other characteristics		
emsp;Heating value, BTU/scf		
emsp;gross saturated	950	1150
emsp;gross dry	967	1170
Relative density	0.530	0.740

Liquids: Free of liquid water and hydrocarbons at delivery temperature and pressure.

Solids: Free of particulates in amounts deleterious to transmission and utilization equipment.

gr/100 = grains/100.

The common unit of measurement for natural gas is the "standard" cubic foot in the English system and the "standard" cubic meter in the metric system. Each of these "standards" is expressed at pressures and temperatures commonly used as standard to the system in the geographical area of concern. In the United States, where standards frequently vary from state to state, the cubic foot is frequently expressed in the English system at standard conditions of 14.73 pounds per square inch absolute (psia) and 60 degrees Fahrenheit (60°F), although there are a number of other "standard" conditions specified by various regulatory and government agencies at various governmental levels. In metric units, the cubic meter is commonly defined as being at one atmosphere of pressure (101.325 kPa) and 15 degrees Celsius (15°C).

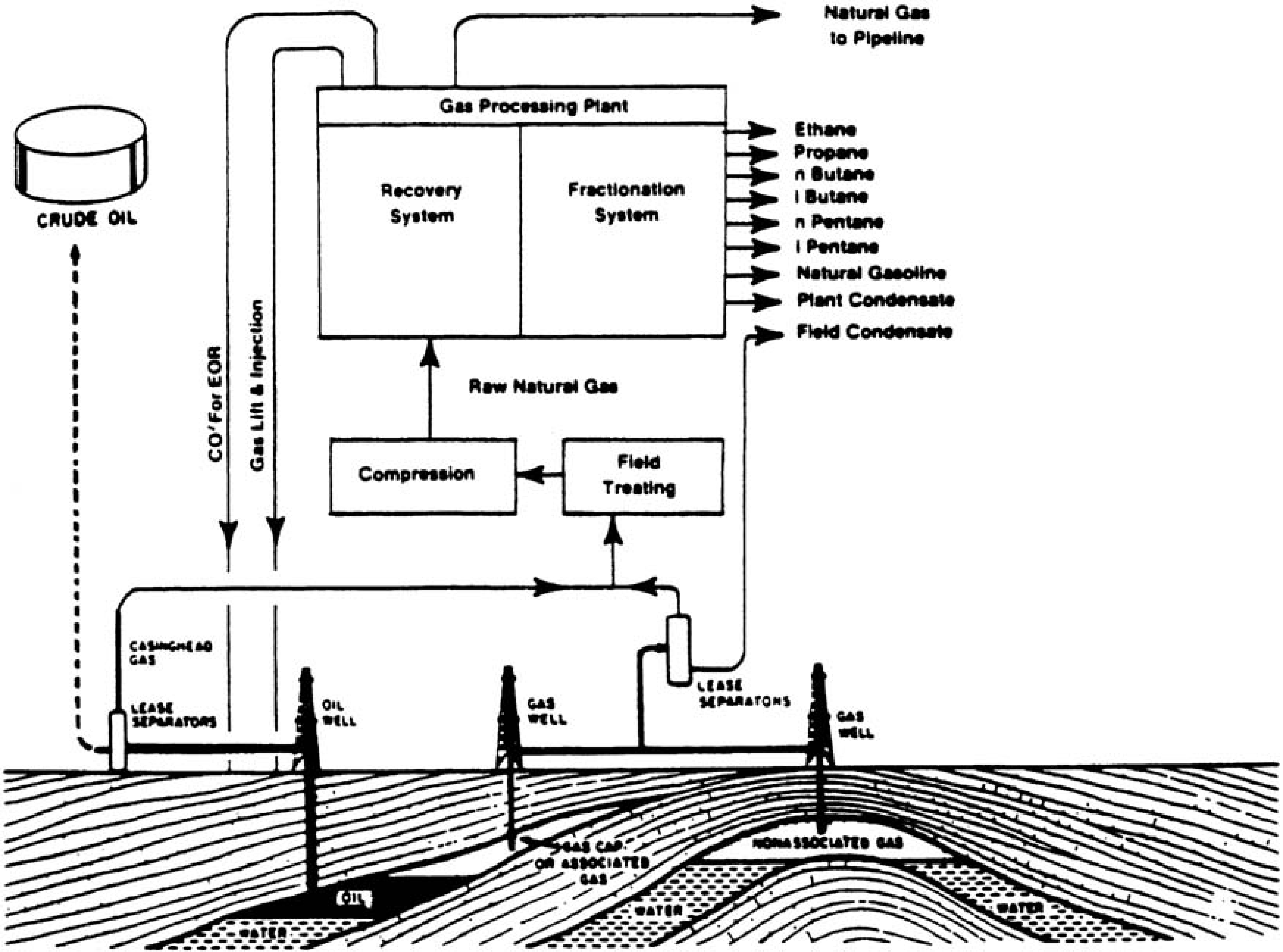
Another common unit of measurement for natural gas is by use of its heating value, expressed in British thermal units (BTU) per standard cubic foot in the English system, and in Joules (or calories) per unit volume in the metric system. Commercially used natural gas, after processing, yields the equivalent of about 9501050 BTU/ft3. Also used as a unit of measurement is the therm, equivalent to 100,000 BTU, or the nominal heat content of 100 standard cubic feet of lean, processed natural gas.

The most common unit of measurement in the English system is the mcf or thousand (103) cubic feet. Larger volumes, used to express production or pipeline volumes, are noted as bcf, or billion (109) cubic feet. Even larger volumes, such as reserve figures, usually are expressed as tcf, or trillion (1012) cubic feet.

16.2 Occurrence of Natural Gas

Natural gas occurs in a number of different forms, depending primarily on the source: associated gas is gas produced "in association" with crude oil; nonassociated gas is gas produced from a gas well with little or no accompanying liquid; gas produced in combination with liquid condensate which is much lower in molecular weight and lighter in color than crude oil; and "coal bed methane," which is gas produced from relatively shallow (30005000 ft) formations containing primarily coal but that also contain marketable quantities of high methane content natural gas. Typical compositions for the first three of these categories are shown in Table 16.1. Figure 16.1 is a simplified schematic layout of the flow of natural gas from the producing well through various stages of processing.

Fig. 16.1 Schematic flow of natural gas from well through processing.



Associated gas is found in crude oil reservoirs, either dissolved in the crude oil or in conjunction with crude oil deposits. Dissolved or associated gas provides the driving force or reservoir pressure necessary to produce oil from a reservoir. Dissolved gases are produced from oil wells along with the oil. The gas separates, or is separated from the crude oil at the well head. Such gas also may be called "casinghead gas" or "oil well gas." In the early days of the industry, virtually all the natural gas usefully utilized was associated gas from oil wells. Much of the very large gas reserves of several Persian Gulf countries and the Prudhoe Bay reserves of Alaska are associated gas.

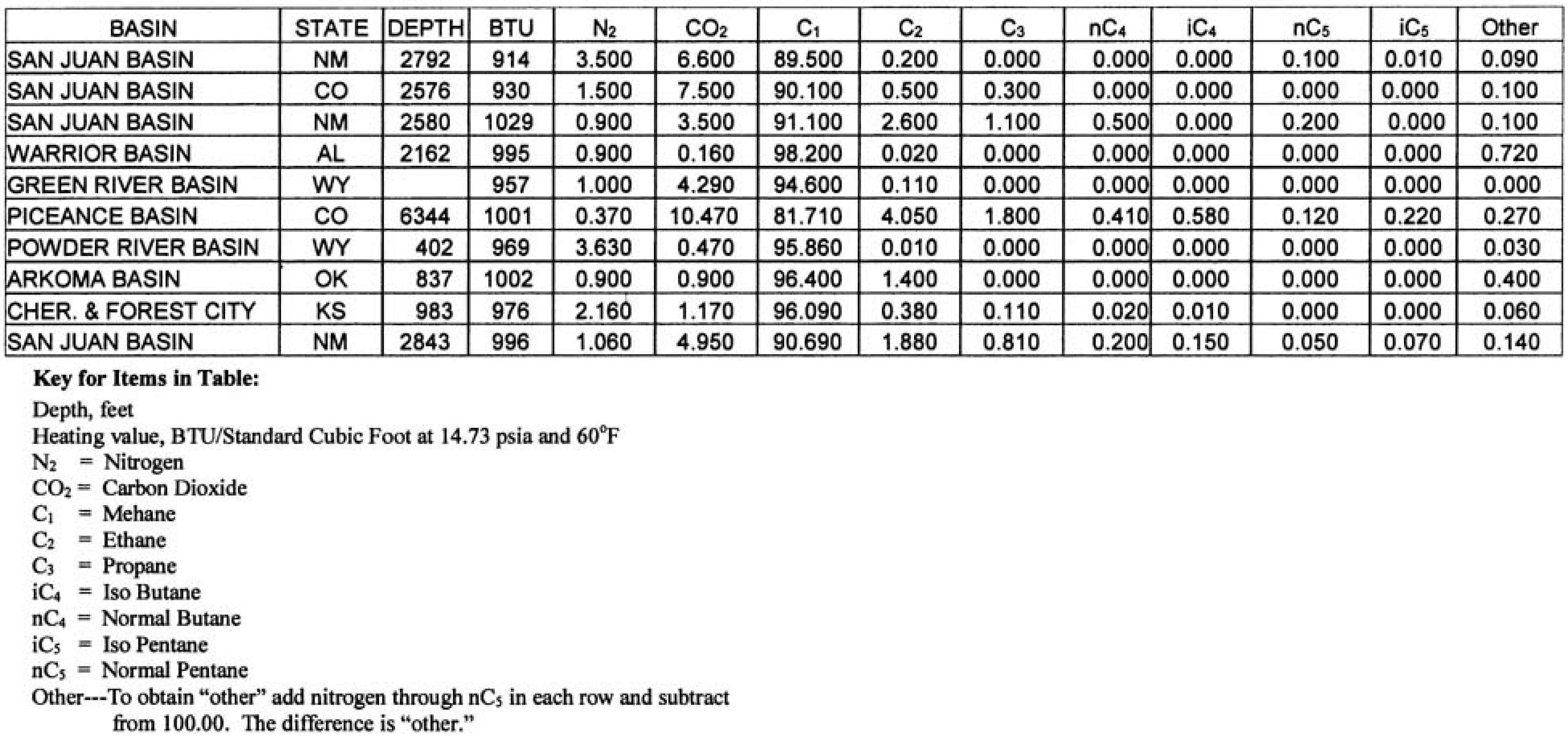
Nonassociated gas occurs in reservoirs separate from those of crude oil. It is commonly referred to as "gas well gas," and contains much less of the heavier, or condensable, hydrocarbons that are found in associated gas. Today, because of the wide demand for natural gas, most utilized gas is nonassociated gas.

The third form in which conventional natural gas reserves occur might be called a single-phase fluid, since it is neither a true gas nor a true liquid. It is not a gas or liquid because no surface boundary exists between gas and liquid. Reservoirs of this type are called "gas condensate" reservoirs, and usually are found in moderately deep formations, have very high pressures, and pose special problems in production and processing.

In recent years gas wells have been drilled at increasing depths. This greatly increases drilling costs and new wells may cost many millions of dollars to drill and bring into production. At least a partial remedy for this has been the rapid development of "coal bed methane" (CBM) as a major source of natural gas in the United States. The wells are relatively shallowaround 5000 ft maximum at presentand much less expensive to drill, complete, and bring to onstream production. In addition, CBM appears to be widely distributed across the lower 48 states of the United States.

Early on the subsurface coal mining industry learned that "swamp gas" or "firedamp" was prevalent in underground coal mines. The gas is largely methane formed from the decay and rotting of carbonaceous materials to ultimately form coal. When pressure is released (as by drilling into the coal seam) the methane-rich gas migrates to the low pressure and is produced. The composition and properties of several coal bed methane streams are shown in Fig. 16.2.15

Fig. 16.2 Composition of several coal bed methane streams.15



BASIN	STATE	DEPTH	BTU	N_2	CO_2	C_1	C_2	C_3	nC_4	iC_4	nC_5	iC_5	Other
SAN JUAN BASIN	NM	2792	914	3.500	6.600	89.500	0.200	0.000	0.000	0.000	0.100	0.010	0.090
SAN JUAN BASIN	CO	2576	930	1.500	7.500	90.100	0.500	0.300	0.000	0.000	0.000	0.000	0.100
SAN JUAN BASIN	NM	2580	1029	0.900	3.500	91.100	2.600	1.100	0.500	0.000	0.200	0.000	0.100
WARRIOR BASIN	AL	2162	995	0.900	0.160	98.200	0.020	0.000	0.000	0.000	0.000	0.000	0.720
GREEN RIVER BASIN	WY		957	1.000	4.290	94.600	0.110	0.000	0.000	0.000	0.000	0.000	0.000
PICEANCE BASIN	CO	6344	1001	0.370	10.470	81.710	4.050	1.800	0.410	0.580	0.120	0.220	0.270
POWDER RIVER BASIN	WY	402	969	3.630	0.470	95.860	0.010	0.000	0.000	0.000	0.000	0.000	0.030
ARKOMA BASIN	OK	837	1002	0.900	0.900	96.400	1.400	0.000	0.000	0.000	0.000	0.000	0.400
CHER. & FOREST CITY	KS	983	976	2.160	1.170	96.090	0.380	0.110	0.020	0.010	0.000	0.000	0.060
SAN JUAN BASIN	NM	2843	996	1.060	4.950	90.690	1.880	0.810	0.200	0.150	0.050	0.070	0.140

Key for Items in Table:

Depth, feet

Heating value, BTU/Standard Cubic Foot at 14.73 psia and 60°F

N_2 = Nitrogen
CO_2 = Carbon Dioxide
C_1 = Mehane
C_2 = Ethane
C_3 = Propane
iC_4 = Iso Butane
nC_4 = Normal Butane
iC_5 = Iso Pentane
nC_5 = Normal Pentane
Other---To obtain "other" add nitrogen through nC_5 in each row and subtract from 100.00. The difference is "other."

CBM has rapidly gained a major position in the lower 48 states natural gas production. Figure 16.316 shows, by year, the number of producing CBM wells in the lower 48 U.S. states. The total number producing for the year 2000 was almost 14,000, and the total gas production as shown in Fig. 16.416 was almost 1.4 tcf. This works out to an average per well production in the order of 40 mcf per day per well.

Fig. 16.3 Number of producing coal bed methane wells in lower 48 states (U.S.A.).16

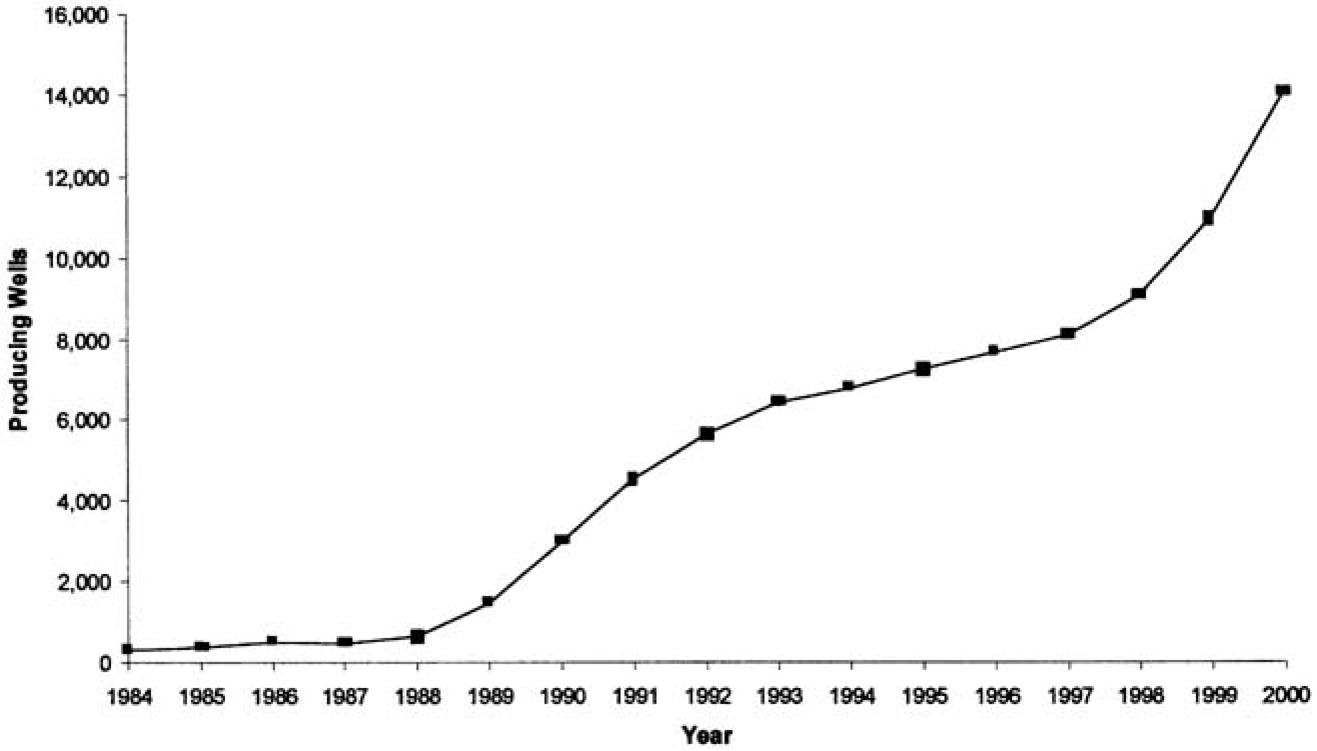
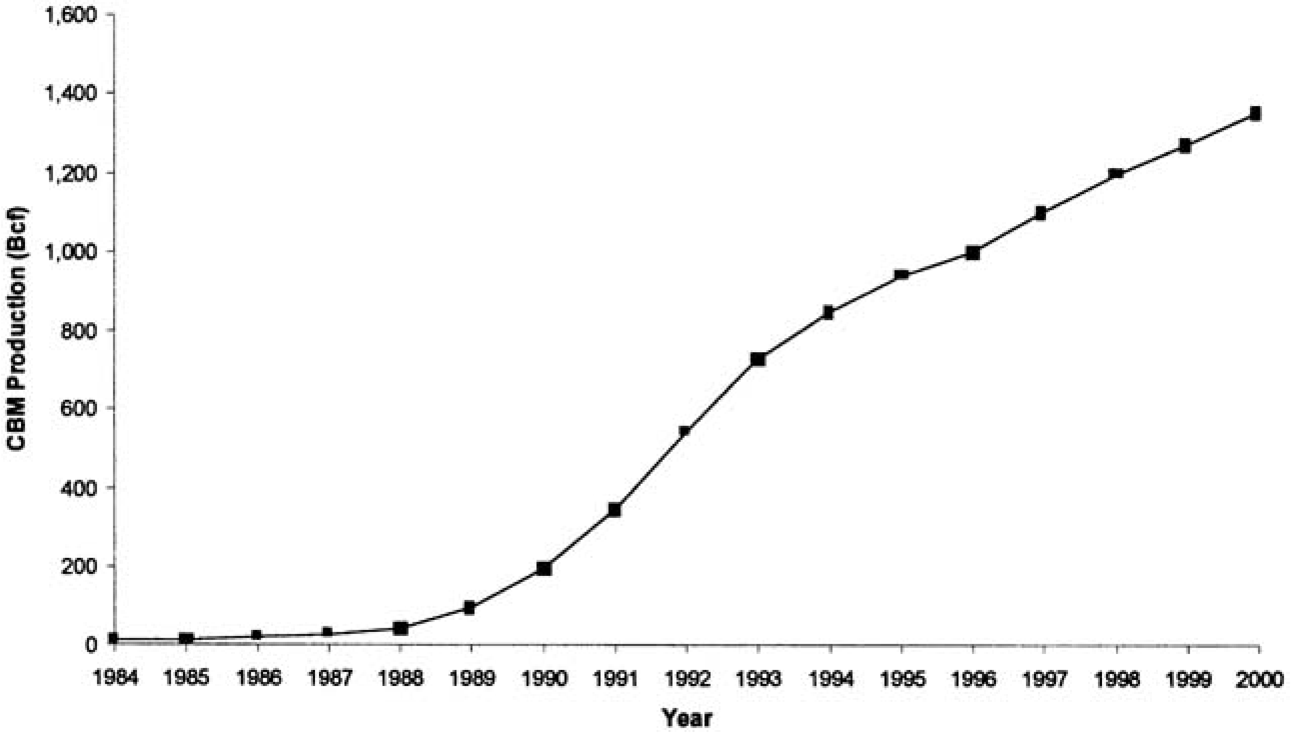


Fig. 16.4 Annual coal bed methane production for lower 48 states (U.S.A.).16



16.3 Evolution of the U.S. Natural Gas Industry

Although natural gas is the predominant product (98%) of the gas industry today, its utilization evolved out of the production and use of manufactured gas, and its major expansion came about through utilization of early "town gas" systems originally installed to distribute manufactured gas.

Manufactured gases are those obtained by destructive distillation of coal, by the thermal decomposition of oil, or by the reactions of steam passing through a bed of heated coal or coke. Prior to the development of large natural gas reserves, manufactured gas was used widely for illumination in the "gaslight" era of the United States.

Following the discovery of major oil and gas reserves in the early twentieth century, associated gas became plentiful in oil-producing areas, and natural gas began to displace manufactured gas in town systems. However, early natural gas utilization was confined largely to areas at or very near the source of supply because methods and materials for transporting the gas had not yet been developed.

Natural gas is relatively difficult to transport and store and is extremely bulky compared with other forms of energy. At atmospheric pressure, one cubic foot of natural gas contains about two-tenths the energy content of a cubic foot of solid or liquid fuel. This is the primary reason natural gas is transported through pipelines at very high pressures, typically 9001,000 psia. The development of high tensile strength, thin wall pipe and modern welding techniques in the mid-1930s permitted the construction of numerous long- distance, high-pressure gas transmission lines from the large and growing gas reserves of the southwestern United States to the major industrial and population centers of the country.

In rapid succession, a number of longdistance pipelines were laid during the period from 1930 to 1940, including three pioneer 1,000-mile lines from areas containing large natural gas reserves in the Southwest to major markets in the North and the Northeast. Spurred by abundant and inexpensive gas reserves, the U.S. natural gas industry expanded rapidly after 1945, and natural gas virtually replaced the more expensive manufactured gas in city distribution systems.

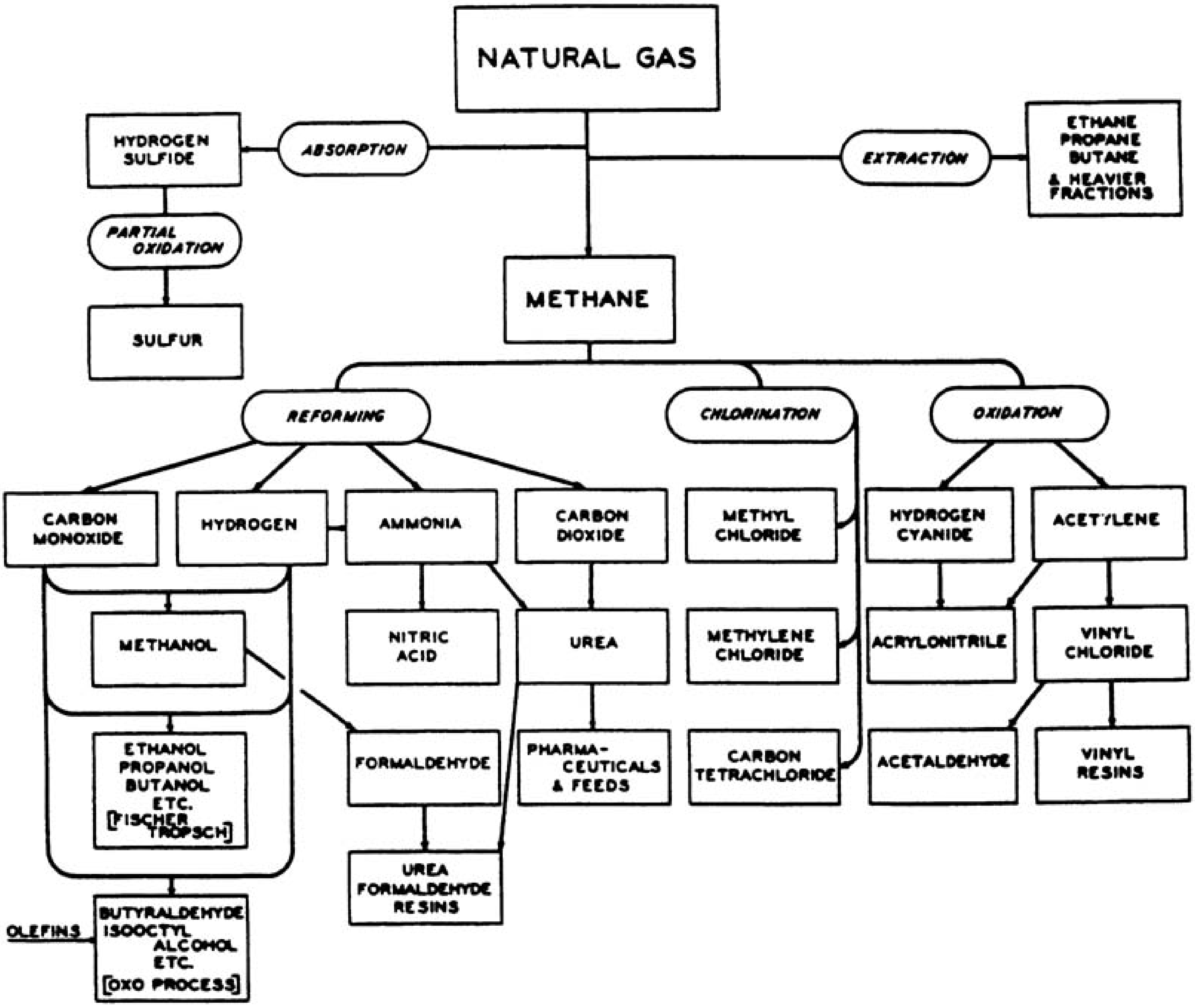
In 1930, marketed production of natural gas totaled 1.9 tcf; following early pipeline expansion, total consumption by 1940 had risen to 2.7 tcf. Then, in the decade between 1940 and 1950, gas consumption more than doubled to 6.3 tcf, and it doubled again to 12.8 tcf by 1960. Marketed production in the United States peaked at 22.7 tcf in 1973, followed by significant declines resulting from conservation efforts. By 1990 the marketed production of natural gas was about 18 tcf per year, and demand was increasing steadily because of the convenience of natural gas and its qualities as an environmentally desirable fuel.

16.4 U.S. Marketed Production[46]

Natural gas accounts for well over one-half of total U.S. petroleum energy production. Its major market segments are residential and commercial fuels, industrial fuel and chemical feedstock, and electric power generation.

Residential consumption is predominantly for space heating, cooking, and water heating. Commercial use includes space heating and other fuel needs for commercial establishments such as stores, hotels, and restaurants. Industrial use, accounting for about 35 percent of total gas consumption, includes fuels for manufacturing operations, industrial boiler fuel, and process heat. Oil refineries are major consumers of natural gas for refining operations. In addition, natural gas is the prime feedstock for the manufacture of ammonia fertilizers, methanol, and other basic chemical derivatives. Figure 16.5 is a simplified schematic that shows some of the principal chemical and commercial products derived from natural gas.

Fig. 16.5 Some petrochemical products derived from natural gas.[7]



Fuel for electric power generation accounts for about 20 percent of marketed natural gas production. This market segment is extremely price-sensitive and competes directly with fuel oil and coal. Given the current environmental emphasis on reduction of air pollutants, the use of cleaner-burning natural gas is expected to increase significantly in the foreseeable future.

In addition to what is termed "marketed production" for these uses, substantial quantities of natural gas are consumed annually as lease, processing plant and pipeline fuels.

At current production levels, a majority of U.S. petroleum energy production consists of the gaseous fuels natural gas and natural gas liquids. Moreover, since reaching peak production in 1985, U.S. crude oil production has consistently declined while gaseous fuel production has steadily increased. These trends suggest that the U.S. petroleum economy will become even more dependent on the gaseous fuels in the future.

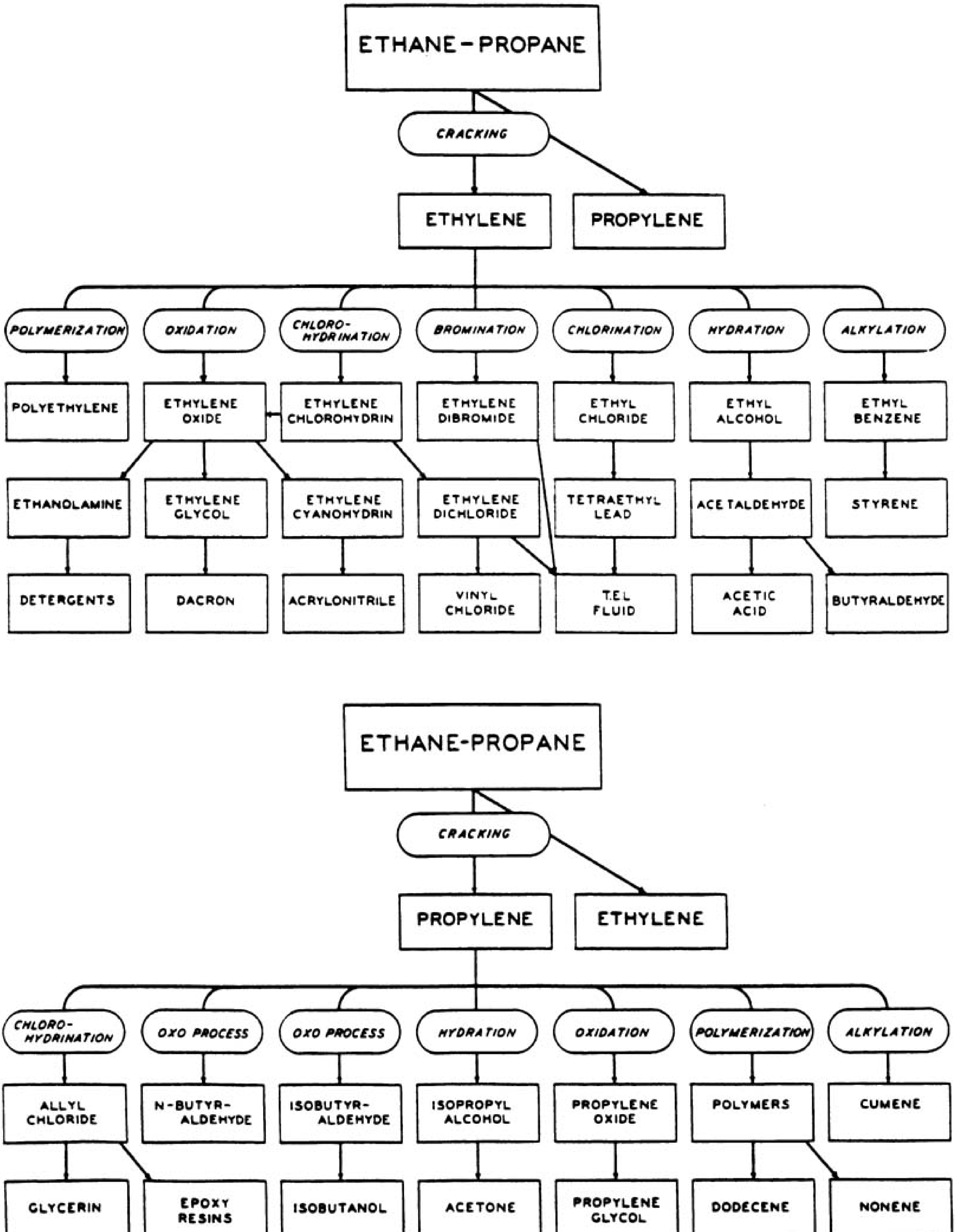
16.5 Natural Gas Liquids

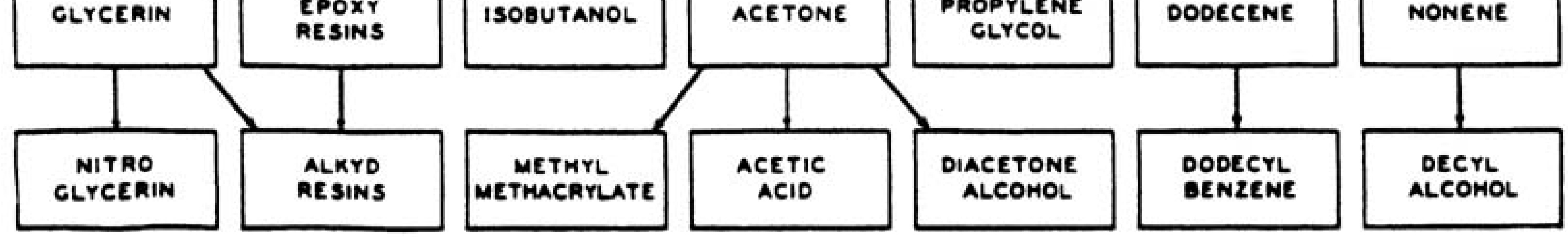
Included in the gaseous fuels segment of U.S. petroleum energy production are the natural gas liquids extracted from natural gas as produced and before sale. These natural gas liquids, commonly termed "NGLs," are ethane, propane, butanes and pentanes, and higher molecular weight hydrocarbons. The recovery and separation of natural gas liquids from raw natural gas is a relatively simple operation, which relies principally on the differences in boiling points and vapor pressures of the various components. (These differences are shown in Table 16.2.) The two most prevalent recovery processes are the absorption process and the cryogenic turboexpander process. Together, they account for approximately 90 percent of total U.S. natural gas liquids production. The bulk of U.S. gas liquids is produced as a raw mix stream in some several hundred gas-processing plants located in or near major gas-producing provinces. Typically, the raw mix streams are transported through an extensive pipeline network to large central fractionator complexes where the mixed liquid stream is fractionated into its components ethane, propane, butanes, and pentanes.

The total U.S. recovery of natural gas liquids is about 3035 barrels per million cubic feet of marketed natural gas. The total NGL supply is supplemented by about 200250 million barrels per year of refinery production, which equates to about 5 percent of the total crude oil charge to refineries. Some of the petrochemical products produced from natural gas and NGLs are shown in Fig. 16.5.

Major demand segments for NGLs indicate that around 3540 percent of gas liquids, principally ethane and propane, are consumed as cracking feedstocks for ethylene manufacture. Some of the chemical derivatives obtained from ethane and propane are shown in Fig. 16.6.

Fig. 16.6 Some products derived from ethane and propane.7





Traditionally, a major demand for natural gas liquids, mainly butanes and pentanes, has been in the manufacture of motor gasolines. However, the recent phasing out of leaded gasoline and the current, ongoing reduction in motor gasoline volatility have drastically reduced demand for butanes and pentanes as gasoline blending stocks. On the other hand, these same measures, enacted to reduce air pollutants from the automobile, have resulted in an increased demand for these products as prime feed stocks for the formulation of clean-burning reformulated gasolines.

A third major market is consumption of propane as both a residential and commercial fuel, principally in rural areas beyond the reach of natural gas distribution systems. Additional major uses indicated in the "other" category of Fig. 16.6 include consumption as internal combustion engine fuels and numerous agricultural uses, such as crop drying.

16.6 U.S. Natural Gas Reserves

At any given time proven U.S. natural gas reserves seem to remain at approximately 10 years of indigenous supply at the then current consumption rates. Proven reserves estimates are subject to constant change, either up or down, depending on reservoir operating experience, field extensions resulting from additional drilling, and revised estimates.

Not included in reserves data are undiscovered but probably recoverable reserves in unexplored or unproven provinces. Currently, these probable reserves are estimated by the U.S. Department of Energy at 307500 tcf, much of which almost certainly will become proven reserves when they become economically feasible.

Whatever the ultimate recoverable reserves may be, current estimates of proven reserves of natural gas comprise about 50 percent of U.S. proven reserves of total petroleum energy.

16.7 Structure of the U.S. Natural Gas Industry

The natural gas industry in the United States is made up of four major segments: production, processing, transmission, and distribution. The industry is physically interconnected by a pipeline network that extends through the lower 48 states and across international borders into Mexico and Canada.

The production function is a large part of the oil-producing industry, and most of the same firms are dominant in both oil and gas. Exploration and drilling technologies for both oil and gas are essentially identical. Production from both oil well gas and gas well gas is introduced into large field gathering systems for delivery to the processing facilities. A typical gathering system may include 2000 or more miles of gathering lines connected to 1000 or more producing wells.

The gas-processing function includes the gathering and delivery of gas streams from a field or fields into a central processing facility. There are currently several hundred gas-processing plants in operation in the United States. In these plants the raw natural gas is dehydrated to remove moisture, treated for removal of contaminants and, if necessary, compressed to pipeline pressure. During treating, most raw gas also is processed for recovery of liquid products, including ethane, propane, butanes, and natural gasoline. These liquid products are then sold separately to be used for petrochemical and gasoline feedstocks and other fuel uses.

The transmission function covers transport of the processed pipeline quality gas from the gas-processing plant to major markets. About two-thirds of U.S. marketed gas production reaches the ultimate consumer through a network of some 250,000 miles of pipelines. Other deliveries may be made directly to industrial consumers or to city distribution systems.

The distribution function receives natural gas from the transmission pipeline at the city gate of the local distribution system and delivers it to the ultimate consumer through a network totaling some 600,000 miles of distribution pipelines.

16.8 World Natural Gas

Natural gas, in recent years, has become an international commodity fuel. Current world reserves of natural gas are estimated at about 4000 tcf, the equivalent of about 40 percent of world petroleum energy reserves. About 40 percent of these reserves are in remote regions of the Former Soviet Union (FSU) and about 30 percent are located in the Persian Gulf countries of the Middle East.

The location of these reserves, far from the major populations of consuming countries, suggests that the world natural gas industry is still in a developmental stage and that further extension and expansion can be expected. Gas from the Groeningen field of the Netherlands, from Norway and the North Sea, and from the FSU is flowing by pipeline into industrialized western Europe. In addition, natural gas is piped from the FSU into eastern Europe.

Additional international movements of natural gas are in the form of liquefied natural gas (LNG), principally from Algeria into Europe and from Oceania into Japan.

It is certain that world use of natural gas will increase dramatically in the near future as industrialized countries replace coal-fired facilities with cleaner-burning natural gas. In addition, a number of countries remote from major markets are in the process of installing world-scale plants for utilization of natural gas and gas liquids for production of fertilizers, methanol, premium gasoline blending stocks, and other basic petrochemical derivatives that will result in higher-value products from natural gas and gas liquids for which there are no local markets.

Already, the gaseous fuels account for approximately 40 percent of world petroleum energy production and nearly one-half of world petroleum energy reserves. Moreover, the growth of world natural gas consumption is already increasing rapidly.

Despite a world surplus of crude oil capacity for the foreseeable future, it seems certain that the world has entered the "Age of Gaseous Energy," and that natural gas and gas liquids will continue to move into a position of petroleum energy dominance in the next decade.

16.9 Gas-to-Liquids Technology

In contrast to the United States, many oil- and gas-producing areas of the world do not have a well-developed gas distribution system and large numbers of consumers want a clean-burning, environmentally friendly industrial and residential fuel. Also, many of these areas have the potential capability for producing large quantities of natural gas for long periods of time. One way of solving this problem is to market LNG. Atmospheric pressure tankers are the popular choice, and they must operate at or near the atmospheric boiling point of methane, which is -161.52°C or -258.74°F. Huge quantities of compression horsepower are required to produce the LNG, and huge quantities of heat are required to gasify the LNG after it has reached port and before it is suitable for consumer use.

Despite these difficulties and obstacles, there are several successful LNG operations in various parts of the world. Japan, which has no petroleum or natural gas reserves, has been particularly active in trying to import LNG and other products.

Before World War II, German scientists developed processes for converting gas primarily formed by burning coal to hydrocarbons and other chemicals with molecular weights in the motor fuel range. Processes using these reactions and catalysts were used for some fuel and chemicals production during World War II. After the hostilities a good share of the FischerTropsch "know-how" and technology was liberated and put within the purview of the U.S. Bureau of Mines. In the early 1950s the South African government made a political decision to produce gasoline and chemicals from their plentiful internal supply of coal and not be dependent on imported petroleum. This led to the design and construction of multiple integrated chemicalmotor fuel plants to serve the internal requirements for motor fuel and simultaneously provide a variety of chemicals for both export and internal consumption. Derivatives of the original FischerTropsch and later Sasol processes are currently being used to try to make marketable natural gas discoveries in areas where there is no industrial or home-based demand for clean-burning natural gas fuel.

As crude oil exploration spread to cover more of the globe, discoveries of huge natural gas reserves spread to more inhospitable places and thoughts of plants to produce long-chain liquid hydrocarbons became more enticing. In the early 1980s planning for such plants began in earnest and in the space of a few years three separate plants utilizing three different liquidation technologies were well under way. Table 16.417 shows some of the key parameters for the three plants. All the investment numbers are too large for the projects ever to be profitable. Estimates now are that about $50,000 investment per daily capacity barrel is about the maximum the traffic can bear.

TABLE 16.4 Summarized Data for the First Three Major GTL Complexes Built

Plant Site	Date of Commissioning	Design Capacity (Mbpd)	F-T Unit Technology	Investment/Daily Capacity Barrel ($/bbl)
Montunui, New Zealand	1985	14.5	Mobil	102,000
Mossel Bay, South Africa	1993	27.5	Sasol	127,000
Bintulu, Malaysia	1993	12	Shell	125,000

Based on information in "Gas-to-Liquids; much smoke, little fire," Samsam Bakhtiari, A. M., *Hydrocarbon Processing*, December, 2001. Copyright 2001 by Gulf Publishing Company, all rights reserved.

Plans for several more liquidation plants had developed but the cost, construction, and operating problems of the first three slowed investor ardor and caused reassessment of the possibilities and goals. Expectations had been that by the mid 1990s a dozen or more plants would be in operation producing more than a million barrels per day of automotive fuel range hydrocarbons. Instead, the New Zealand plant was switched to producing methanol and the other two were continuing to encounter operating problems.

Three second generation gas-to-liquid (GTL) plants are known currently to be underway. Some details for each are shown in Table 16.5.18 Investment per daily capacity barrel is much lower than for the first three. But the total capacity of the three new plants is only 80,000 barrels per day. If GTL technology proves out, it will be at a much more sedate pace than originally envisioned.

TABLE 16.5 Summarized Data for the Three Second-Generation GTL Complexes Presently under Construction

Plant Site	Scheduled Startup	Design Capacity (Mbpd)	F-T Unit Technology	Investment Estimate/ Daily Capacity Barrel ($/bbl)
Australia, NW Shelf	2004	11.5	Syntroleum	52,000
Nigeria, Escravos	2005	34	Sasol	23,500
Qatar, Ras Laffan	2005	34	Sasol	23,500

Based on information in "Gas-to-Liquids; much smoke, little fire," Samsam Bakhtiari, A.M., *Hydrocarbon Processing*, December, 2001. Copyright 2001 by Gulf Publishing Company, all rights reserved.

16.10 Preparing Natural Gas for Transmission and Sale

To make a product suitable for sale, the various contaminants and undesirable constituents contained in the raw gas must be removed. The first step is to separate the gas from any liquids and/or solids. This initial separation usually is carried out in near proximity to the producing well. The separators used depend primarily upon the difference in density between the gas and the liquid for separating the two phases. These separators may be classified as vertical, horizontal, or spherical, depending on their geometry and physical orientation.8

The liquids produced are sold as oil or condensate, depending upon the type of reservoir from which the product is obtained. The natural gas stream after this first stage of separation contains only gaseous components, though some of these may be liquefied by later processing.

Typical compositions for gas streams at this point are shown in Table 16.1. (The compositions in Table 16.1 are on a water-free basis.) As typically produced, each of the gas streams would be saturated with water vapor. Developing the technology for handling gas streams that contain significant quantities of water vapor was one of the keys necessary to the development of the natural gas industry as it exists today.

In the presence of water, natural gas can, under the proper conditions of temperature and pressure, form hydrates. To the naked eye the hydrates are similar in appearance to snow; and they can result in plugging of pipelines and flow systems. Hydrates can form at temperatures well above the freezing point of water. In the early days of pipelining natural gas under pressure, shutdowns of the lines during the winter months frequently were caused by plugging with hydrates.

A hydrate is a type of chemical compound called a clathrate, defined as a solid molecular compound in which one component is trapped in the cavities of cagelike crystals of another component. In the natural gas hydrate, water molecules form the "cage," and hydrocarbon molecules are the trapped component (s).

Hydrate formation can be prevented or avoided in two ways: (1) by dehydration, in which water is removed from the gas stream so that under conditions of processing or transporting, no liquid water will condense; and (2) by inhibition, which involves injecting into the gas stream a component that will dissolve the water and thus interfere with the ability of the water to form hydrates with the gas. Both procedures are widely used in the gas industry.

Of the wide variety of materials that can effectively depress the hydrate-forming tendency of a natural gas, threemethyl alcohol, ethylene glycol, and diethylene glycolare used almost exclusively. Ethylene glycol ($C_2H_6O_2$) and diethylene glycol ($C_4H_{10}O_3$) are both polyhydroxy alcohols. They are high-boiling-point materials that mix in all proportions with water and both remain predominantly in the liquid phase. At temperatures approaching the freezing point of water, solutions containing more than 50 percent by weight ethylene glycol (or diethylene glycol) have viscosities approaching 10 cp. This limits their utilization in some low-temperature applications.

Methyl alcohol (CH_3OH) has a much lower boiling point and consequently a much higher vapor pressure than either ethylene or diethylene glycol. Hence procedures used for estimating the amount of methyl alcohol required must take into account the amount of methyl alcohol that vaporizes into the gas phase.9,13,14 Although the necessity for vaporization requires the injection of greater quantities of methanol, it does have potential advantages. If the gas phase were cooled in further processing or transmission, one would expect some methanol to condense along with additional water and/or hydrocarbons. This would provide a safety factor for hydrate prevention that does not exist with the glycols.

Dehydration of the gas can be accomplished through the use of either a liquid or a solid desiccant. The solid desiccants most often employed include alumina, silica gel, and molecular sieves. The liquid desiccant most frequently used is triethylene glycol. Any of the solid materials used as desiccants must be regenerated periodically, so there must be at least two adsorption towers if dehydration of the gas is to be continuous. A schematic of a typical two-tower adsorption plant is shown in Fig. 16.7. The adsorbent must be heated to high temperature (about 400°F, 200°C) to remove the water and regenerate the solid desiccant. This usually is accomplished by heating a slip stream of the gas and circulating the heated gas through the off-stream adsorbent bed.

Fig. 16.7 Two-tower adsorption dehydration unit.

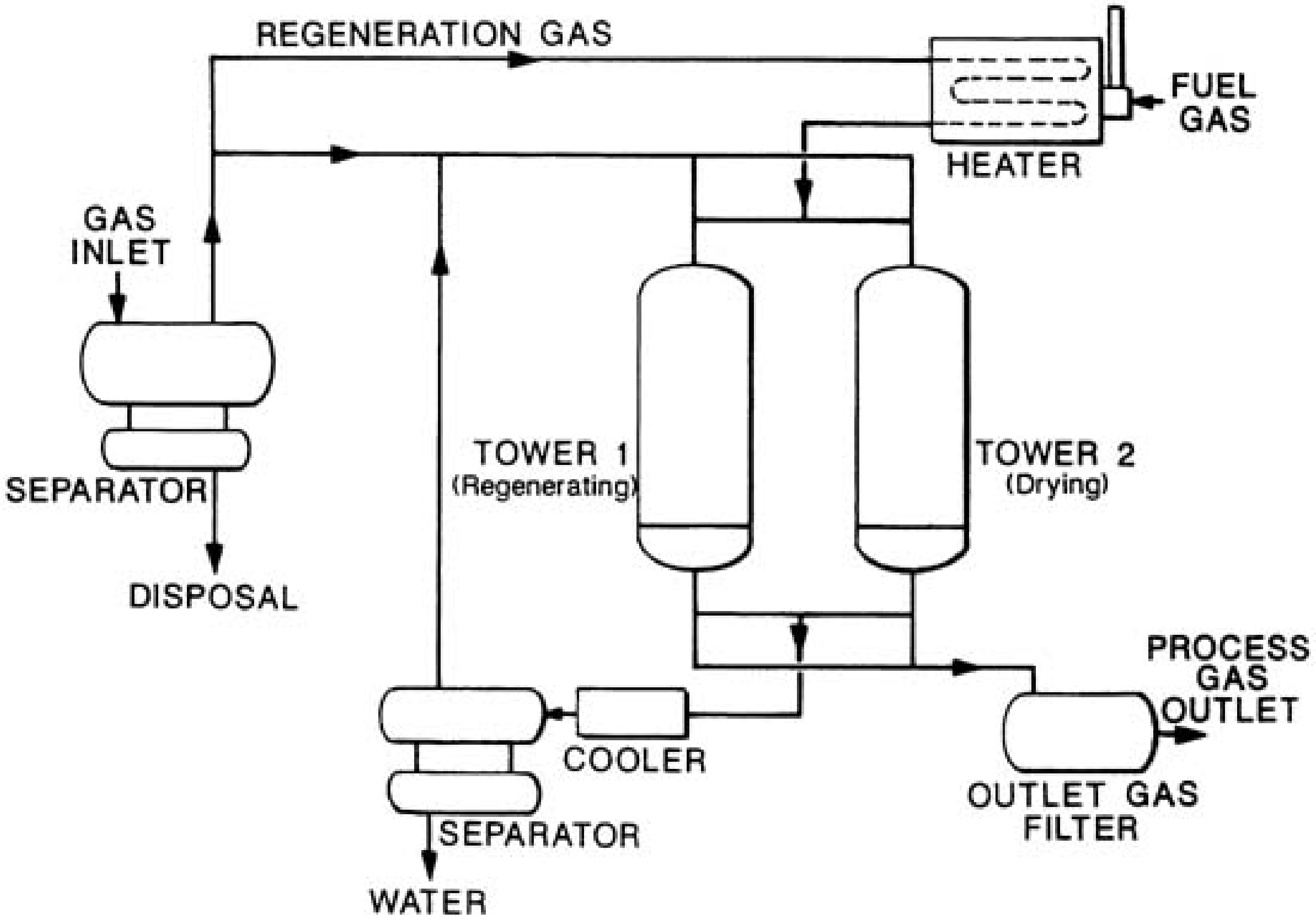
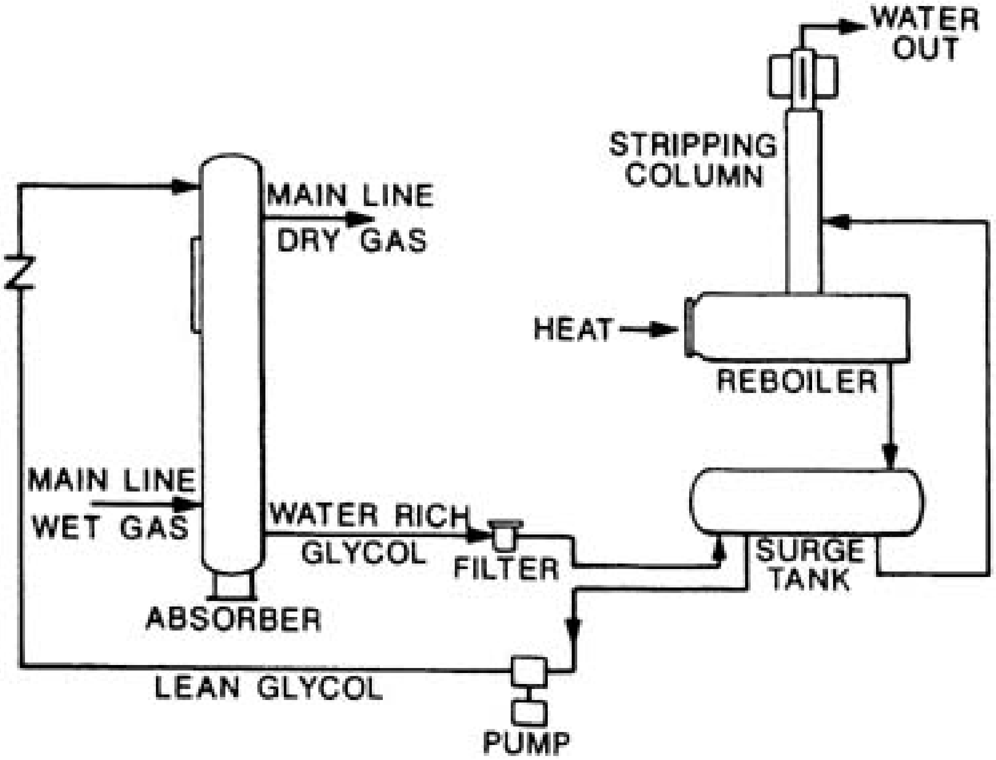


Figure 16.8 shows a schematic flow diagram for a typical triethylene glycol dehydration system. The lean glycol is pumped to the top of an absorber column and flows downward in countercurrent contact with the waterwet gas entering the bottom of the tower. The absorber may contain either plates or packing. Dehydrated gas leaves the top of the absorber and the glycol diluted by the absorbed water flows from the bottom into a regenerator. After the absorbed water has been fractionated from the triethylene glycol solution, the solution is recirculated to the absorber. The concentration of the glycol leaving the reboiler will be approximately 98.598.7 percent by weight triethylene glycol if the reboiler operates at 400°F (205°C) and one atmosphere pressure.

Fig. 16.8 Glycol dehydration unit.

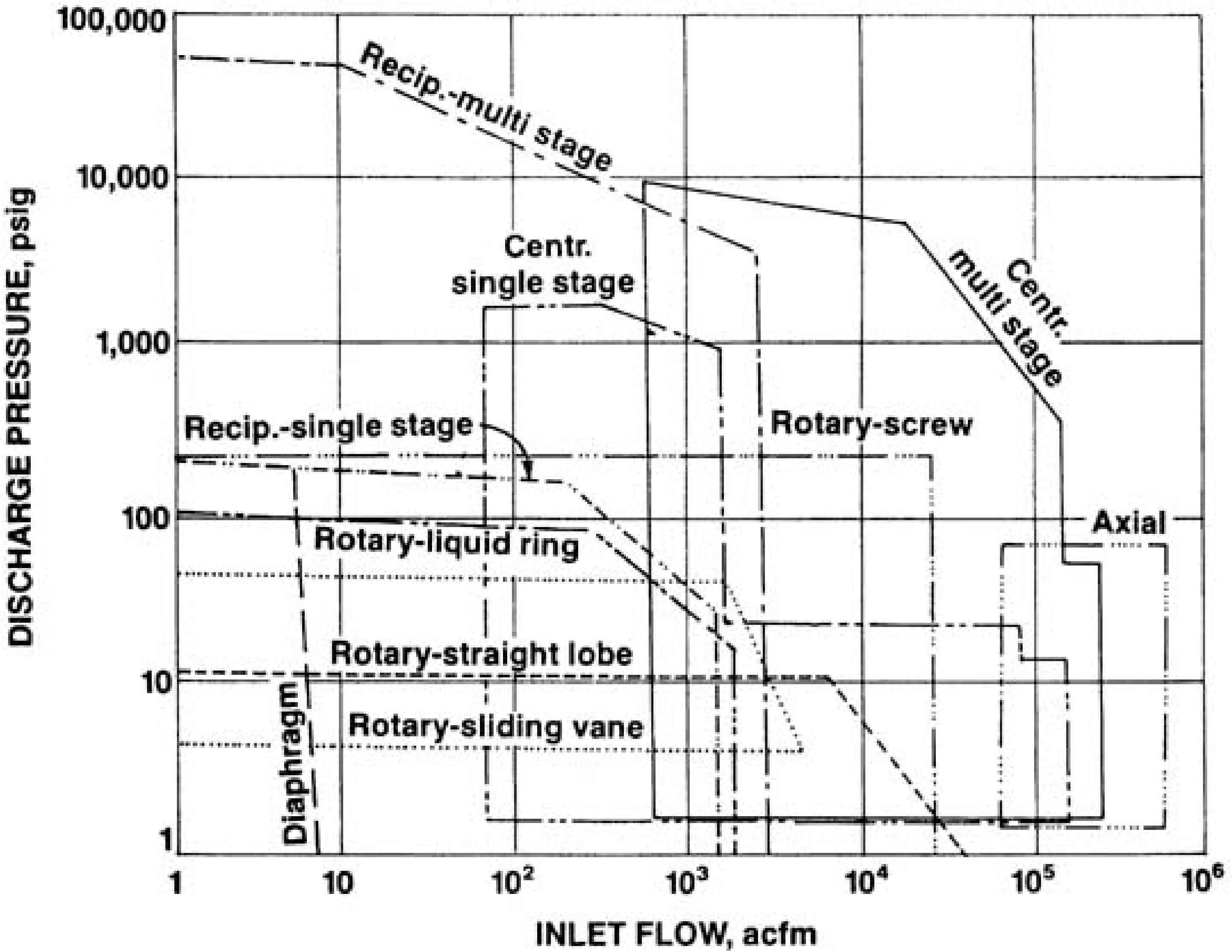


Normal contract specfication for gas to be transmitted through a high-pressure pipeline is a water content of 7 lb of water per million standard cubic feet of natural gas. This is approximately the water content of natural gas in equilibrium with water at the freezing point (32°F) when the gas is under a pressure of 1000 psia.

The solid desiccants offer much lower water contents for the dehydrated gas. With proper design and operation, molecular sieve dehydrators can satisfactorily prepare gas for total liquefaction at temperatures as low as -263°F (-165°C). Lower outlet water dew points can be obtained with triethylene glycol if the concentration of glycol going to the absorber is increased. This can be accomplished by using a vacuum in the reboiler or by introducing another material into the reboiler to lower the effective partial pressure of the water vapor above the glycol. Proper design and operation of such glycol systems can produce outlet water dew points in the -40°F (-40°C) range, or lower.

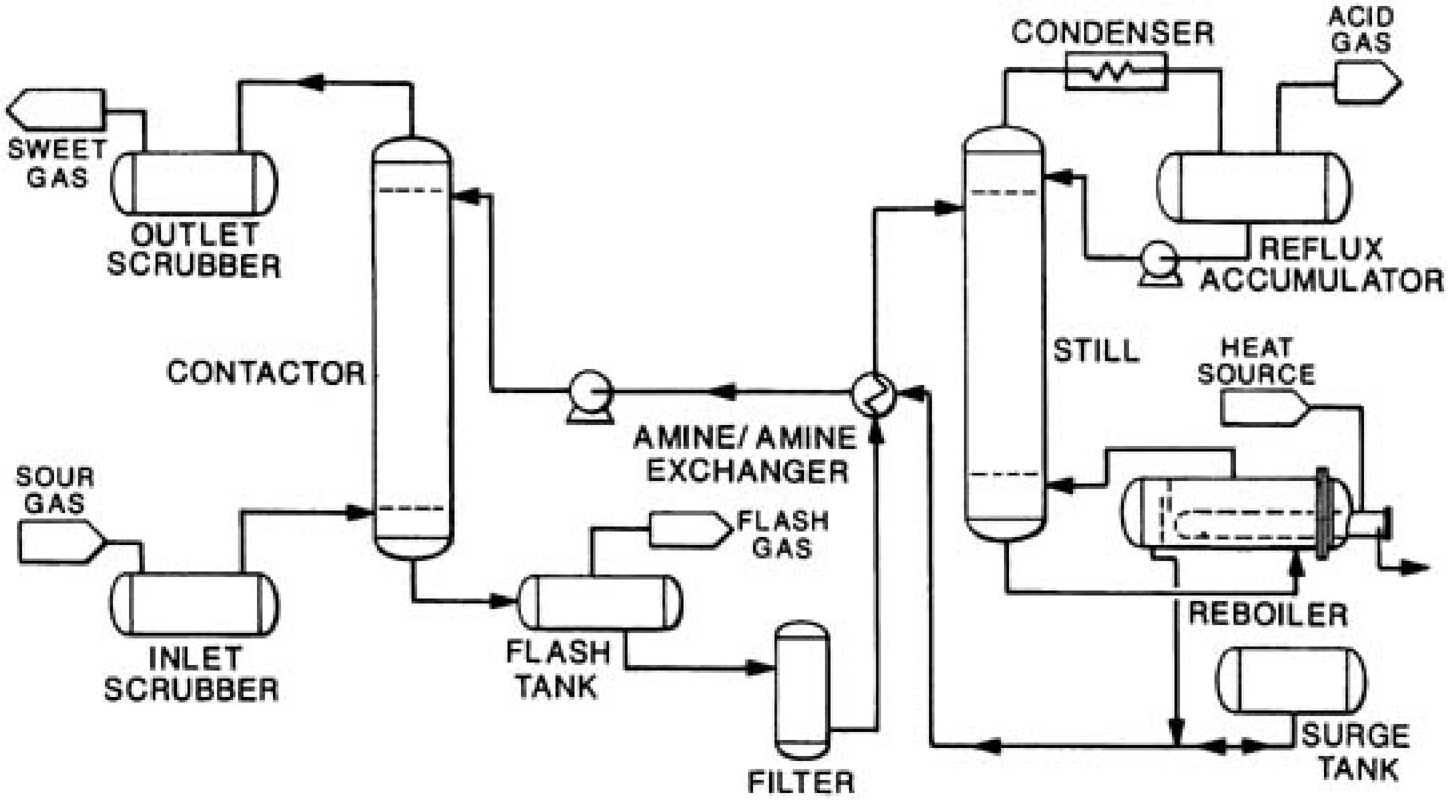
Natural gas pipelines normally operate at elevated pressures. In many instances the gas is available at low pressures so must be pumped or compressed to higher pressure. If the volume of the gas stream is small, reciprocating compressors must be used. For larger gas volumes, rotary compressors of various types are available. Figure 16.910 shows pressure and flow rate ranges in which different types of compressors currently available are used.

Fig. 16.9 Approximate range of application for compressors. (GPSA Engineering Data Book, *10th ed., Gas Processors Suppliers Association, Tulsa, OK. 1987*.)



If the natural gas stream contains unacceptable quantities of hydrogen sulfide and/or carbon dioxide, they must be removed in order to make the gas suitable for transmission and sale. The details of removal of H2S and CO2 from natural gas streams are beyond the scope of this chapter, but excellent discussions are available.10,12 There are many different processes available, depending upon the contaminants to be removed and their concentration in both the sour gas available and the sweetened gas to be produced. The dominant treating process is still the use of an alkanolamine. A typical flow diagram for an amine sweetening installation for removal of hydrogen sulfide and carbon dioxide from a natural gas stream is shown in Fig. 16.10.12

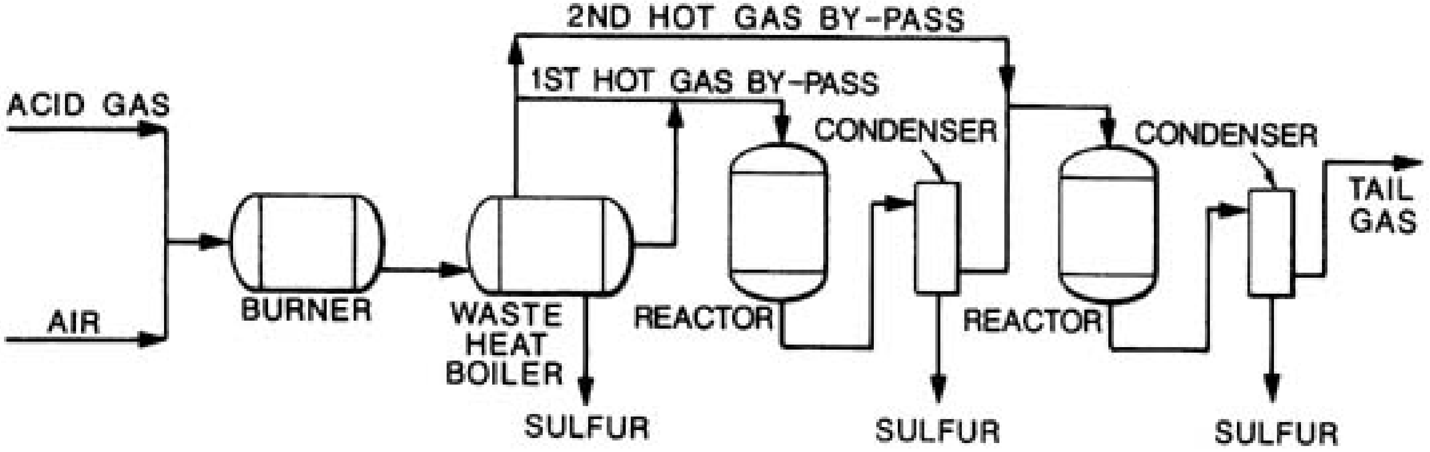
Fig. 16.10 Typical ethanolamine sweetening unit.



Environmental restrictions and controls normally will not permit the release of the hydrogen sulfide and/or its incineration to sulfur dioxide. As pointed out by Maddox,12 the least undesirable alternative is conversion of the hydrogen sulfide to sulfur. Again, there are many different processes available, several of which are discussed in the book by Maddox. The dominant method is by use of the Claus process, which involves partial combustion of the hydrogen sulfide to sulfur dioxide and catalytic combination of the hydrogen sulfide and sulfur dioxide to produce sulfur. Detailed discussions of the Claus technology as well as other techniques of converting hydrogen sulfide to sulfur are available.10,12

Figure 16.11 is a sketch of a "once-through" Claus process using a hot gas bypass for reheat. Sufficient air is used to burn one-third of the H2S and all hydrocarbons. Steam usually is generated in the waste heat boiler, where the combustion products are cooled, and sulfur is condensed. The gas must be reheated before introduction to the reactor to prevent sulfur condensation in the reactor bed. Tail gas usually is sent to a tail gas cleanup unit to minimize sulfur emissions.

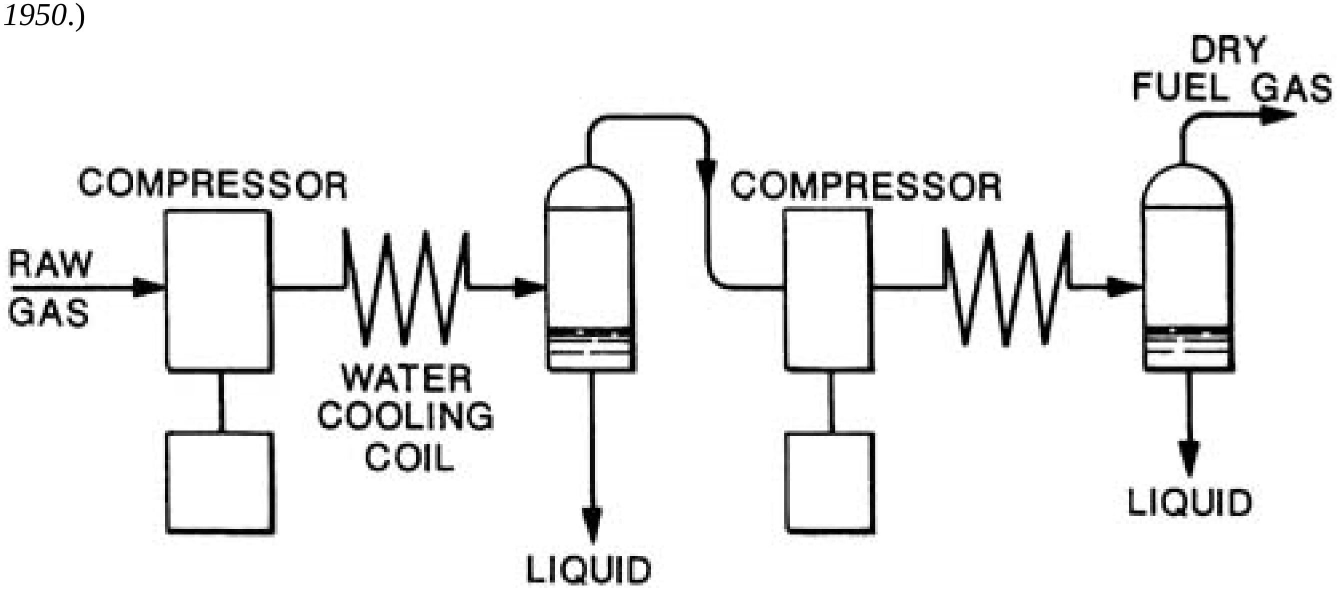
Fig. 16.11 Once-through Claus sulfur unit.



16.11 Processing for Liquids Recovery

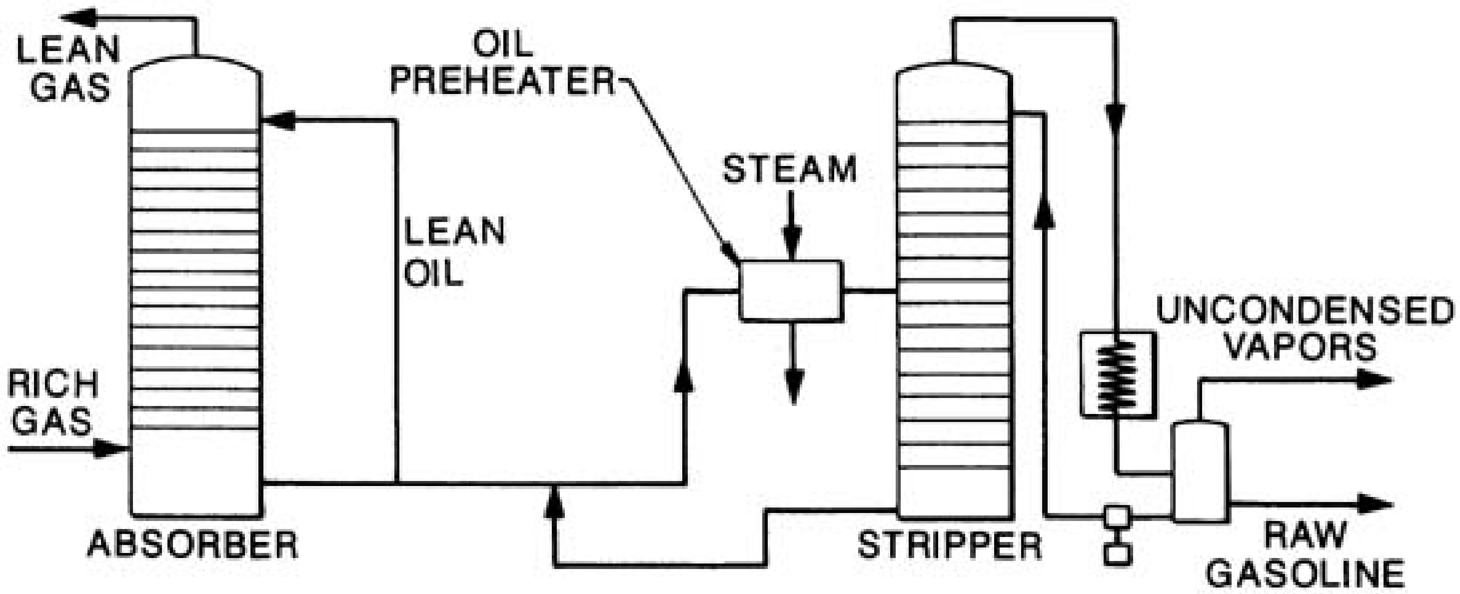
The first time that liquids were recovered from a natural gas stream probably happened more by accident than by design. Any time a gas stream from a liquidgas separator is either cooled or compressed, liquids will form. This simple statement still provides the basis for nearly all hydrocarbon liquid recovery systems from natural gas. Figure 16.1211 shows schematically the flow through a simple compression liquids-recovery facility. Replacing the water cooling after the second stage of compression with refrigeration would increase liquids recovery. If the gas had not been dehydrated, there would be danger of hydrate formation at this point unless the water had been removed.

Fig. 16.12 Liquid hydrocarbon recovery by compression and cooling. (*Adapted from Huntington, R. L.*, Natural Gas and Gasoline, *McGraw-Hill, New York, 1950*.)



As the natural gas industry developed and expanded, so did the demand for liquids recovered from natural gas. One way of recovering additional liquids is through the use of a heavy absorption oil that has good solubility characteristics for the propane and heavier hydrocarbon components in the gas stream. As in the case of dehydration, the solvent must be regenerated. Figure 16.1311 shows the absorption-stripping section of an absorption-type hydrocarbon liquids recovery facility. The absorbent is circulated to the top of the absorber and flows downward, absorbing liquefiable components from the gas stream. It then is heated in the stripper, with the stripped components being cooled and condensed while the stripped absorption oil is recirculated. The liquid hydrocarbon stream typically would flow through a fractionation train for separation into the individual components or mixtures that could be sold.

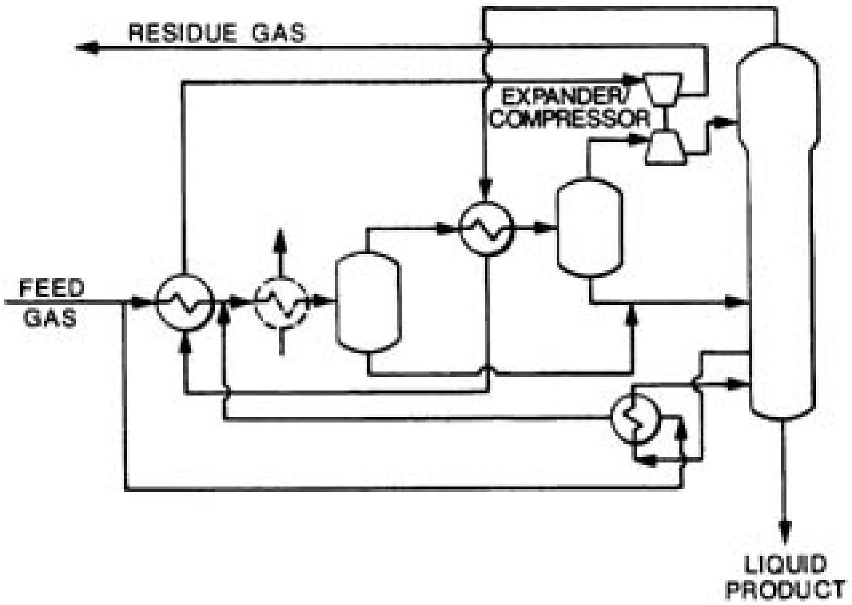
Fig. 16.13 Absorption-stripping unit. (*Adapted from Huntington, R. L.*, Natural Gas and Gasoline, *McGraw-Hill, New York, 1950*.)



Combining refrigeration with oil absorption allowed for additional liquid recoveries and greater economy of operation. Using temperatures as low as -40°F (-40°C), refrigerated oil absorption plants were capable of recovering 70 percent or more of the propane present in the natural gas stream. Recoveries of the butane and heavier constituents were essentially 100 percent.

Development of the turbo-expander process allowed the design and construction of plants for recovery of liquid ethane, as well as the heavier hydrocarbon components. The turbo-expander extracts useful work from the gas during expansion from a high pressure to a lower pressure. Because of the work extraction the gas is cooled more than in a simple pressure expansion, and, by means of suitable heat exchange, temperatures as low as -150°F (-100°C) can easily be achieved. The separation at low pressure gives higher relative values of the vaporliquid equilibrium constant for nitrogen and methane than for ethane and heavier hydrocarbons. The result is that the process shown in Fig. 16.1413 can recover as much as 75 percent or more of the ethane contained in the gas as the demethanized product. In some cases the process will be run to produce a de-ethanized product that contains essentially all the propane in the feed gas.

Fig. 16.14 Turbo-expander process flow.



The liquid product from the facility either can go to on-site fractionation into salable products or, more typically, be introduced into a pipeline and transported to a central fractionation facility near the point of ultimate use for the liquid product.

16.12 A New Potential Source for Natural Gas

Under proper conditions of temperature and pressure, water molecules can form ice-like structures in which there are cavities. If small molecules such as methane and nitrogen are present, they can be held in the cavities and there serve to stabilize the crystalline structure. The resulting crystalline structures are called clathrates.20,21 Although each crystalline form has a unique composition, there are no chemical bonds per se. Thermodynamically the structures are solid solutions. Although several different hydrate structures are known, the Structure 1 (S-1) hydrate typically dominates in natural gas systems. The general formula for type S-1 hydrates is 8X·46H2O, where X can be molecules of argon, krypton, xenon, nitrogen, oxygen, hydrogen sulfide, carbon dioxide, methane, ethane and/or propane. When gas containing more than one component is involved, the hydrate typically will contain more than a single gaseous constituent.

The stability of the hydrate structure depends heavily on the size of the entrapped guest molecules. Ultimately the molecules can become so large (pentane for example) that a stable hydrate cannot form. As noted earlier, many materials can interfere and prevent hydrate formation.

In recent years deposits of methane hydrate have been discovered in all parts of the world. The methane apparently forms in the decomposition of bio mass, primarily in watercontaining areas, and the stabilized hydrate accumulates.

The storage of methane as hydrates offers a potentially vast natural gas resource. As to the question of how much hydrate there is right now, there is no definitive answer. However, the worldwide amount of carbon bound in gas hydrates has been *estimated to total twice the amount of carbon to be found in all known fossil fuels originally on Earth*. Additionally, conventional gas resources appear to be trapped beneath methane hydrate layers in ocean sediments.22

In the United States, deposits have been confirmed on all areas of the continental shelf and under Alaskan permafrost. Similar deposits have been confirmed in many locations throughout the world. But little is known about the actual size and location of even most hydrates that are presumed to exist.

Estimates of the amount of gas sequestered in hydrates varies markedly. Today, most engineers and scientists estimate that at an absolute minimum there is approximately 100,000 tcf. They also estimate there is a maximum of 270,000,000 tcf, or more

The numbers for the United States are as follows. The U.S. Geological Survey (USGS) released a report in 1995 evaluating the U.S. hydrate resource base. It categorized estimates by the level of certainty that they exist.23

1. at a level of certainty of 95 percent, resources = 112,785 tcf
2. at a level of certainty of 50 percent, resources = 276,119 tcf
3. at a level of certainty of 5 percent, resources = 676,110 tcf
4. the mean of these estimates = 320,222 tcf

Obviously, hydrates have tremendous potential as a future source for natural gas. The concluding section of this chapter will serve as a summary of what is currently being used and of promising technology for future development.

16.13 Methane Conversion Processes

Methane conversion as treated here encompasses chemical transformation only, although in the broadest interpretation purification, compression, and liquefaction operations are sometimes included, especially in relation to methane consumption as a fuel.

Methane is the simplest, most abundant, and geographically most widely distributed hydrocarbon. It therefore receives constantly increasing attention as an alternate energy source to coal and petroleum from both the world fuels industry and from the science and engineering community to broaden its utility and enhance its transportability by energy-efficient conversion to liquid hydrocarbons and functional chemical raw materials.

Apart from combustion as fuel, the largest chemical conversion of methane is steam reforming to synthesis gas (syngas)mixtures of CO and H2 in ratios varied to fit the particular process. Syngas is the raw material base for methanol and its many derivatives including gasoline range hydrocarbons and methyl-t-butyl ether (MTBE), FischerTropsch hydrocarbons, hydroformylation (oxo reaction) products, formaldehyde, hydrogen source for ammonia, and a further wide spectrum of industrial chemicals.25 These are regarded as "indirect" methane conversion products since methane is consumed as the raw material for the final plant products and syngas accounts for over 80 percent of all methane conversion to commercial products.

The industrially important direct methane conversion processes comprise oxidative coupling, reductive coupling including pyrolysis reactions, partial oxidation, halogenation and oxyhalogenation,26 and ammoxidation. Other direct conversions include alkylation, electrophilic substitution, and CH bond activation over various complex and super acid catalysts. Several of these direct conversion technologies remain to be exploited to achieve their full commercial potentials.

Indirect Conversion via Syngas

Steam reforming of methaneand low to middle range hydrocarbonsis widely practiced to generate syngas for a number of basic, high-volume fuel stocks, such as gasoline range and middle distillate hydrocarbons, and also basic industrial chemicals such as methanol and ammonia.

The catalytic reaction of steam with methane at elevated temperatures (300 400°C) over various catalystscopper or nickel/molybdenum oxide/aluminacan be made to yield CO and H2 in desired ratios. The generalized reaction for hydrocarbons with steam is:

$$\mathrm{C}_n\mathrm{H}_m + n\mathrm{H_2O} \rightleftharpoons n\mathrm{CO} + [m/2 + n]\,\mathrm{H_2}$$

$$\mathrm{CO} + \mathrm{H_2O} \rightleftharpoons \mathrm{CO_2} + \mathrm{H_2}$$

With adjustment of the steam/methane ratio, the reactor can produce a synthesis gas with CO/H2 = 1/2, the stoichiometric proportions needed for methanol production. This mixture at approximately 200 atm pressure is fed to the methanol unit where the reaction then proceeds at 350°C. Per pass conversions range from 30 to 50 over the catalysttypically a supported copper oxide with a zinc, chromium, or manganese oxide promoter:3

$$\mathrm{CO} + 2\mathrm{H_2} \rightleftharpoons \mathrm{CH_3OH}$$

The methanol unit operating regime, process parameters, and sometimes the catalyst may be adjusted to yield a significant co-production of higher mixed alcohols typically ranging from C2 to C6.

Methanol to gasoline units following the Mobil fixed-bed design with a ZSM-5 molecular sieve catalyst are capable of producing high-grade gasoline with an iso-paraffinic hydrocarbon front end and a methlyated aromatic hydrocarbon back end. Dilution with other source alkylate to a lower aromatic content gives a premium grade mid-90s octane gasoline.

Alternatively, high-quality gasoline range distillates are produced directly from syngas via the FischerTropsch synthesis. The product from all the operating units is a mixture of paraffins and olefins having a straight chain structure and molecular weight distribution predicted by the SchulzFloryAnderson theory. Isomerization for improved octane rating yields superior automotive fuels. Recent reviews present a position of concern that GTL processes based in syngas intermediate unavoidably entail capital costs exceeding $25,00030,000 per daily barrel of capacity, and in most cases are not competitive with petroleum based liquid fuel processes. The 30+ percent consumption of originally contained energy in the methane consumed to power the syngas unitscompression and heat costsmust also be reckoned a process driving energy component.

Indirect Conversion via Nonsyngas Intermediates

The oxidative conversion of methane to reactive intermediates such as methyl sulfate has been reported recently.27 The reaction, catalyzed by Group V and VI metal complexes in concentrated sulfuric acid media, utilizes molecular oxygen (air or industrial gas source) selectively to break and substitute the CH bond in sequential oxidative scission and sulfate ester formation steps. Methanol may then be generated via hydrolysis with subsequent sulfuric acid recovery. While not yet industrially established as a methanol source for liquid fuels manufacture, this technology provides a methane CH bond-breaking technique that is restricted to one CH bond only and conserves much of the CH3 group energy content that is expended in driving the commercial syngas-based methanol process alternative.

Direct Methane Conversion to Hydrocarbons and Chemical Derivatives

The direct, one-step conversion of methane to higher hydrocarbons and chemical derivatives, after several decades, continues to be intensely researched worldwide as a possible source of liquid fuels and chemical raw materials based on natural gasindependent of petroleum and coal. For clarity and convenience the following "direct conversion" categories25 are recognized:

1. oxidative coupling to higher hydrocarbons
2. partial oxidation to chemical derivatives
3. pyrolysis or cracking
4. other direct conversion processes

Oxidative Coupling to Higher Hydrocarbons

The catalyzed gas phase reaction of methane with oxygen, wherein a hydrogen atom is abstracted from two methane molecules which are subsequently joined or "coupled" to form ethane with the coincidental formation of gaseous water, was reported28 in the early 1980s:

$$\mathrm{CH_4} + \tfrac{1}{2}\mathrm{O_2} \rightleftharpoons \mathrm{C_2H_6} + \mathrm{H_2O}$$

Immediate interest and continuing, increased research by additional investigators has shown that variations in reaction conditions (typically 500800°C, 15 bar, high space velocities) in combination with with homo- and heterogeneous, fixed and fluid bed catalysts, enables the formation of higher alkanes and olefins reaching to C6C8 chain length, although C2C3 dominates. With optimal methane/oxygen ratio, reaction temperature, and catalyst parameters, further oxydehydrogenation to ethlyene can occur:

$$\mathrm{CH_3{-}CH_3} + \tfrac{1}{2}\mathrm{O_2} \rightleftharpoons \mathrm{CH_2{=}CH_2} + \mathrm{H_2O}$$

Methane-based commercial production of ethylene via oxidative coupling has been investigated, but to date the lower per pass conversions required for acceptable ethylene selectivities combined with purified oxygen costs make this process noncompetitive with thermal cracking of ethane from natural gas liquids.

Oxidative coupling of methane nonetheless may be basically regarded as a polymerization reaction. Catalyzed oxidative coupling reactions are in commercial use today for high polymer production, for example, conversion of 2,6-dimethylphenol to poly (phenylene oxide) with molecular oxygen. However, the lower CH bond energy and resulting greater liability of the hydrogen atoms involved in the coupling reaction allow operation in the liquid phase at far lower temperatures (below 100°C).

Methane-coupling reaction conversions and yields less than 25 percent initially wereand still arebelow those acceptable for commercial fuel and chemical feedstock production. But worldwide research and development in more recent years continue to suggest that variations in process parameters, reactor design, and catalyst composition and structure may bridge this gap. Lower reaction temperaturesin the 300400°C range may be needed. Yields of 90 percent coupled with hydrocarbon products at 10 percent methane conversion have been reported at temperatures below 600°C where the activation energy for ethane and ethylene formation is below 90 kcal.29

The major obstacle to high-yield oxidative methane coupling remains the high energy necessary to dissociate the CH bond with its 94 kcal bonding energy. The rigorous reaction conditions and temperatures required to accomplish CH bond scission in the methane molecule are essentially the same as those favoring further oxidation and degradation of the methyl radical or related short-lived highly reactive intermediates to waste products such as CO and CO2. Research in recent years has identified catalysts, particularly including some based on transition metals, such as iridium and palladium complexes, which are able to dissociate the CH bond at much lower temperatures, even as low as below 200°C.

Industry experts today suggest conversions of 4050 percent and selectivities above 80 percent based on methane and oxygen as the minimum needed for commercial consideration after fixed and variable costs are added. Nonetheless, methane oxidative coupling holds the most promising combination of process simplicity, product slate versatility and low cost, and worldwide raw material availability not offered now by practiced fuel and chemical feedstock technologies.

An important aspect of direct methane oxidative coupling that merits further attention is optimization of the process to focus on the simplest products of the highest value and which are the easiest to obtain, namely the C-2 and C-3 products. Of these, the C-3 products, now commercially derived from light hydrocarbon and naphtha cracking or refinery streams, propane and particularly propylene, are in increasingly high demand and offer a versatility of existing end uses and potential utility not matched by the higher range C-4 and up aliphatic hydrocarbons, namely

1. Petrochemical raw materials for a multiplicity of plastics, films, fibers, elastomers and chemical industry intermediates.
2. Specialty (LPG) and commodity automotive fuelsthe research octane rating of 115 for propane compels industry interest;
3. High energy content, low vapor pressure (200 psi), and liquid state at ambient temperatures favors low-cost liquid pipeline transportation vs. highpressure compressed gas1000+ psig if methane conversion is done near the production site.

Methane oxidative coupling process simplifications and advancesespecially combinations of reactor design and catalyst features, product separation, and recycle economiesare areas marked for technological innovation with accompanying economic opportunity. C-3's handling and transportation advantages compared with remote natural gas, their potential as alternate fuel, and their demand and versatility as chemical feedstocks compared with higher liquid homologues should drive intensive methane-to-C-3's research and development for the next decade. Significant areas of exploration and development will focus on:

1. Reactor designfixed vs. fluid/moving bed variations.
2. Catalyst discovery researchmetal oxides and supports, shape selective and hetero metal substituted molecular sieves, pillared clays, biomimetic, methanotropic and other bio systems; and combinatorial catalytic screening techniques, liquid phase homogeneous systems.
3. Reactorcatalyst combinations with emphasis on porous substrates.
4. Unique absorptiondesorption product recovery systems.

Partial Oxidation to Chemical Derivatives

Oxygen-deficient cool flame partial oxidation of methane to methanol and/or formaldehyde has long been known,30 but methane-based selectivities of 71 percent methanol and 14 percent formaldehyde at 2 percent conversions require a recycle ratio of 200 to 1. Higher conversion levels led to flame temperature increase with markedly lower yields. Constraining factors were controlling the reaction stream adiabatic temperature rise within narrow ranges (25°C) to minimize waste product formation, and desirable narrow tube reactor design which raises the reactor pressure to undesirable levelsabove 60 barand also lowers yields. At 5 percent conversion, the adiabatic temperature increase was near 250°C.

An alternative approach to partial oxidation of methane is oxyhydrohalogenationusually oxyhydrochlorination.31 In this alternative chemistry modification, methane is converted first to methyl chloride in a reactor fed a mixture of HCl, steam, and methane. The intermediately formed methyl chloride then is converted in a second catalytic (zeolite) reactor to low-to-medium range hydrocarbons. The net reaction is:

$$nCH_4 + 2HCl + nO_2 \rightleftharpoons C_nH_{(2n+2)} + Cl_2 + nH_2O$$

In this process, the strategy of selective chlorination of methane to intermediate methyl chloride in effect provides for much lower temperatures and milder CH bond-breaking energetic regimes than those so far reported in the higher-temperature direct oxidative processes. The lower-temperature, halogen-based stoichiometry further sidesteps the problematic degradative oxidation of the extremely fragile methyl radical and derived intermediate species to waste products (CO and CO2) experienced in most reported direct oxidative coupling work. This innovative approach has received careful attention as an alternative to the other "direct" oxidative coupling processes. Conversely, the handling, recovery, and recycling of chlorine and hydrogen add complexity and possible cost to the technology.

Pyrolysis or Cracking

Acetylene production via high-temperature gas phase pyrolysis is the method of choice for the production of acetylene in large volume usage. It is essentially confined to on-site dedicated units to avoid the expense and hazards of transporting acetylene long distances by pipeline, truck, or rail. In the classic Wulff process and variations thereon, pure methane is fed to a pipe pyrolysis reactor operating slightly above atmospheric pressurenot more that a few barsat residence times of a few seconds and temperatures from near 1,000°C to much higher. The reactor effluent after heat exchanger cooling is passed to an absorber where acetylene is solvent scrubbed from the reactor stream comprising the major product hydrogen and unconverted/methane. Following hydrogen separation, methane is recycled. Small amounts of carbon may carry through to the absorber where they are removed from the solvent raffinate after acetylene recovery.

Methane high-temperature, in-situ cracking to carbon and easily disposed of gaseous by products is practiced in the carbon/graphite composites industry where additional layerings of amorphous or microstructured carbon deposits are fabricated on preformed scaffolds. These are utilized primarily in the aerospace industry. Thin diamond coatings are also fabricated by the low-vacuum, in-situ chemical vapor deposition from methane decomposition.

Other Direct Conversion Processes

Chlorination of methane to methyl chloride, methylene chloride, chloroform, and carbon tetrachloride is practiced industrially worldwide on a large scale. Bromination and iodination are similarly practiced on a smaller scale to prepare the corresponding bromine and iodine derivatives which are important pharmaceutical, synthetic, and fireproofing raw materials.

Methane Ammoxidation

Hydrogen cyanide (HCN) is a widely used high-volume raw and intermediate material in the world chemical industry. For the last several decades it has been increasingly supplied as a by-product of propylene ammoxidationby which acrylonitrile, a major chemical, plastics fiber, and rubber ingredient, is made. However, transportation of liquid HCN (boiling point 26°C) over significant distances is considered hazardous because of its high toxicity and volatility and on-site production for immediate use is usually desirable. Thus, methane catalytic oxidation with oxygen (air) in the presence of ammonia, or "ammoxidation," the Andrussow process,32 is the long-established most practiced industrial technology for HCN manufacture. The single-step reaction is straightforward and both conversions and yields are high:

$$CH_4 + NH_3 + \frac{11}{2}O_2 \rightleftharpoons HCN + 3H_2O$$

As can be seen from the above equation, formation of HCN is in reality a *hetero-bimolecular oxidative coupling* reaction of methane with ammonia. The ammoxidation reactor construction is a simple fixed-bed multi-tube and the catalyst is usually a platinum or sometimes a Group V or VI metal oxide on a silica or alumina support. The HCN product is recovered by condensation and fractionation. With the reaction simplicity and yield, and widespread availability of starting materials, in-situ HCN generation is an ideal industry solution to HCN supply. (See Chapter 29 for more details.)

Of coincidental interest, ammoxidation of propylene,33 itself an oxidative coupling product of methane as noted above, is the commercially practiced route to acrylonitrile, which is produced in high conversions and yields:

$$\begin{aligned}&CH_2{=}CH{-}CH_3 + \frac{11}{2}O_2\\ &\quad\rightleftharpoons CH_2{=}CH{-}CN + 3H_2O\end{aligned}$$

Acrylonitrile is a commodity monomer raw material for many fibers, plastics, rubbers, and chemical intermediates. Economically recoverable quantities of by-product hydrogen cyanide may be produced which are purchased by industrial consumers or, as necessary, disposed of by incineration.

Methane Reductive Nitrilization

Vapor phase co-pyrolysis of methaneammonia mixtures at temperatures in the 12001400°C range at modest pressures, the BMA Process,33 is also an industrial source of HCN and is carried out in several international locations where raw material costs are favorable. Another methane reductive nitrilization technology of considerable interest that has been carried out in several facilities internationally, is the Shawinigan fluohmic reactor process. Mixed methane and ammonia gases are reductively co-pyrolized to HCN and hydrogen in a fluidized bed of carbon particles maintained at pyrolytic temperatures in the 1400°C range. Propane is reported to be a preferred feedstock,34 but methane is sometimes the economic choice. The carbon particle bed temperature is achieved by the transmission of low-voltage electric alternating current from immersed electrodes through and between carbon particles with accompanying interparticle electric arcing. Reaction stoichiometry differs from the Andrussow given the absence of oxygen in the reactant stream, and of water in the reactor effluent stream.

Space restrictions do not allow a complete survey of the countless methane conversion examples available or in the published literature. This overview is intended to highlight those methane conversion technologies which are or have been used industrially and might impact conversion of methane from hydrate deposits.

The authors are firm in their belief that methane conversion(s) hold much greater promise for the future exploitation of huge hydrate reserves than does GTL technology.

16.14 References

Gas Processors Suppliers Association Engineering Data Book, 10th ed., 1987. The GPSA offices are at 6526 East 60th Street, Tulsa, OK 74145.
Gas Processors Association Standard 2145. Same address as GPSA.
Gas Processors Suppliers Association Engineering Data Book, 11th ed., Electronic FPS Version.
International Energy Annual, Energy Information Administration, U.S. Department of Energy.
U.S. Crude Oil, Natural Gas and Natural Gas Liquids Reserves, Energy Information Administration, U.S. Department of Energy.
Sales of Natural Gas Liquids and Liquefied Refinery Gases, American Petroleum Institute, Gas Processors' Association, and National Propane Gas Association.
Crockett, L. O., and Widgery, R. C., "Natural Gasoline Producers' Stake in Chemical Manufacturing," *Proceedings 34th Annual Convention*, Gas Processors' Association, 1955.
Campbell, J. M., Gas Conditioning and Processing, Vol. 1The Basic Principles, 6th ed., Campbell Petroleum Series, Norman, OK, 1984.
Campbell, J. M., Gas Conditioning and Processing, Vol. 2The Equipment Modules, 6th ed., Campbell Petroleum Series, Norman, OK, 1984.
GPSA Engineering Data Book, 10th ed., Gas Processors Suppliers Association, Tulsa, OK, 1987.
Huntington, R. L., Natural Gas and Natural Gasoline. McGraw-Hill, New York, 1950.
Maddox, R. N., Gas Conditioing and Processing, Vol. 4Gas and Liquid Sweetening, 3rd ed., Campbell Petroleum Series, Norman, OK, 1985.
Maddox, R. N., and Lilly, L. L., Gas Conditioning and Processing, Vol. 3Computer Applications for Production/Processing Facilities, 2nd ed., Campbell Petroleum Series, Norman, OK, 1990.
Segeler, C. G. (Editor-in-Chief), Gas Engineer's Handbook, 5th ed., 5th printing, Industrial Press Inc., New York, 1977.
Swindell, G., Personal Communication with R. N. Maddox, March 2002.
"North American Coalbed Methane Resource Map," gti e&p Services, Des Plaines, IL 60018.
Bakhtiari, A. M. Samsam, "Gas-to-Liquids: Much Smoke, Little Fire", *Hydrocarbon Processing* (Dec. 2001).
Romanovw, S., "Got Gas," *Hydrocarbon Processing* (Dec. 2001).
Energy Information Administration, *U.S. Crude Oil, Natural Gas and Natural Gas Liquids Reserves*, 2000. Annual Report.
Jeffrey, G. A., and McMullan, R. R., Progr. Inorg. Chem., 8, 43 (1967).
Ripmeester, J. A., Tse, J. S., Ratcliffe, C. I., and Powell, B. M., "A New Clathrate Hydrate Structure," Nature, 325 135 (1987).
Dillon, W., "Gas (Methane) HydratesA New Frontier," U.S. Geological Survey Fact Sheets, 1992.
Collett, T., and Kuuskraa, V., "Hydrates Contain Vast Store of World Gas Resources," *Oil & Gas*, May 11 (1998).
Wender, I., and Klier, K., "Review of Indirect Liquefaction," in (Ed.), Chapter 5, Coal LiquefactionA Research Needs Assessment, H.D. Schindler, DOE Contract DEAC01-87ER30110, 1989.
Kuo, J. C. W., and Ketlar, A. D., "Evaluation of Direct Methane Conversion To Higher Hydrocarbons and Oxygenates," Final Report DOE Contract DEAC2286PC90009, 1987.
Humphreys, G. C., and Dunster, M., "Engineering the World's Largest Methanol Plants," Design '79 Conference, Birmingham, 1979.
Olah, G. A., "Electrophilic Methane Conversion," Acc. Chem. Res., 20, 422428 (1987).
Keller, G. E., and Bhasin, M. M., J. Catal. 73, 9 (1982).
Rasko, J., Peirera, P., Somorjai, G. A., and Heinemann, H., Catal. Lett., 9, (1991).
Chemische Werke Huels, A. G., "Process for the Manufacture of Mixtures of Formaldehyde and Methanol by Partial Oxidation of Methane" GB Patent 1.410,709 (Oct. 22, 1975).
Fox, J. M., Chen, T. P., and Degen, B. D., "Direct Methane Conversion Process Evaluation," Final Report, U.S. Department of Energy Contract DE-AC22-87PC79814, August 1988.
Andrussow, U.S. Patent 1,934,839.
Idol, J. D., U.S. Patent 2,904,580, Sept. 15, 1959.
Encyclopedia of Chemical Technology, KirkOthmer, 6th ed., Vol. 7, pp. 756775.
Prensky, Stephen, "A Review of Gas Hydrates and Formation Evaluation of Hydrate-Bearing Reservoirs," Society of Professional Well Log Analysts 36th Annual Logging Symposium Transactions, paper GGG, 1995.
Trofimuk, A. A., Cheraski, N. V., and Tsaryov, V. P., "The Role of Continental Glaciation and Hydrate Formation on Petroleum Occurrence," in R. F. Meyer (Ed.), pp. 919926, The Future Supply of Nature-made Petroleum and Gas, New York, Pergamon Press, 1977.

17.1 Introduction

The United States contains about one-third of the world's coal reserves, and coal represents over 90 percent of U.S. proven reserves of fossil fuels. Recoverable reserves of U.S. coal are estimated to be 250 billion tons. Bituminous coals (with a heating value of 10,00015,000 BTU/lb) comprise nearly one-half of total U.S. coal reserves. Eastern U.S. coals are generally bituminous. Western and southwestern U.S. coals are mainly sub-bituminous (with a heating value of 900012,000 BTU/lb) and lignite (with a heating value of 800010,000 BTU/lb). Coal is a major source of energy for electric power production and process heat and can serve as a source of synthetic fuels and feedstock for the petrochemical industry.

Historically, interest in coal technology has been cyclic, depending primarily on the price of petroleum. Coal dominated the U.S. energy picture until the discovery of petroleum. Each subsequent threat to oil supplies spurred a renewed interest in coal, which rapidly faded as major new oil discoveries increased the supply of oil and lowered its price. The threat to oil supplies in 1973 produced the historic response, generating increased interest in converting coal to more convenient, cleaner fuels. Coal use for electric power generation has increased, but the economic conversion of coal to clean transportation fuels still stands as one of the key technical challenges of our time.

Bituminous and sub-bituminous coals account for the major share of all coal produced in the United States. In 2000, production of all types of coal totaled almost 1.1 billion short tons, of which about 95 percent was bituminous or sub-bituminous (Fig. 17.1). Despite its clean burning characteristics, anthracite, most of which is mined in northeastern Pennsylvania, accounts for a diminishing share of total coal production. Figure 17.2 presents data on coal consumption in the United States.

Fig. 17.1 U.S. coal production, 19492000. (Annual Energy Review, *Energy Information Agency, U.S. Department of Energy, DOE/EIA-0384(01).*)

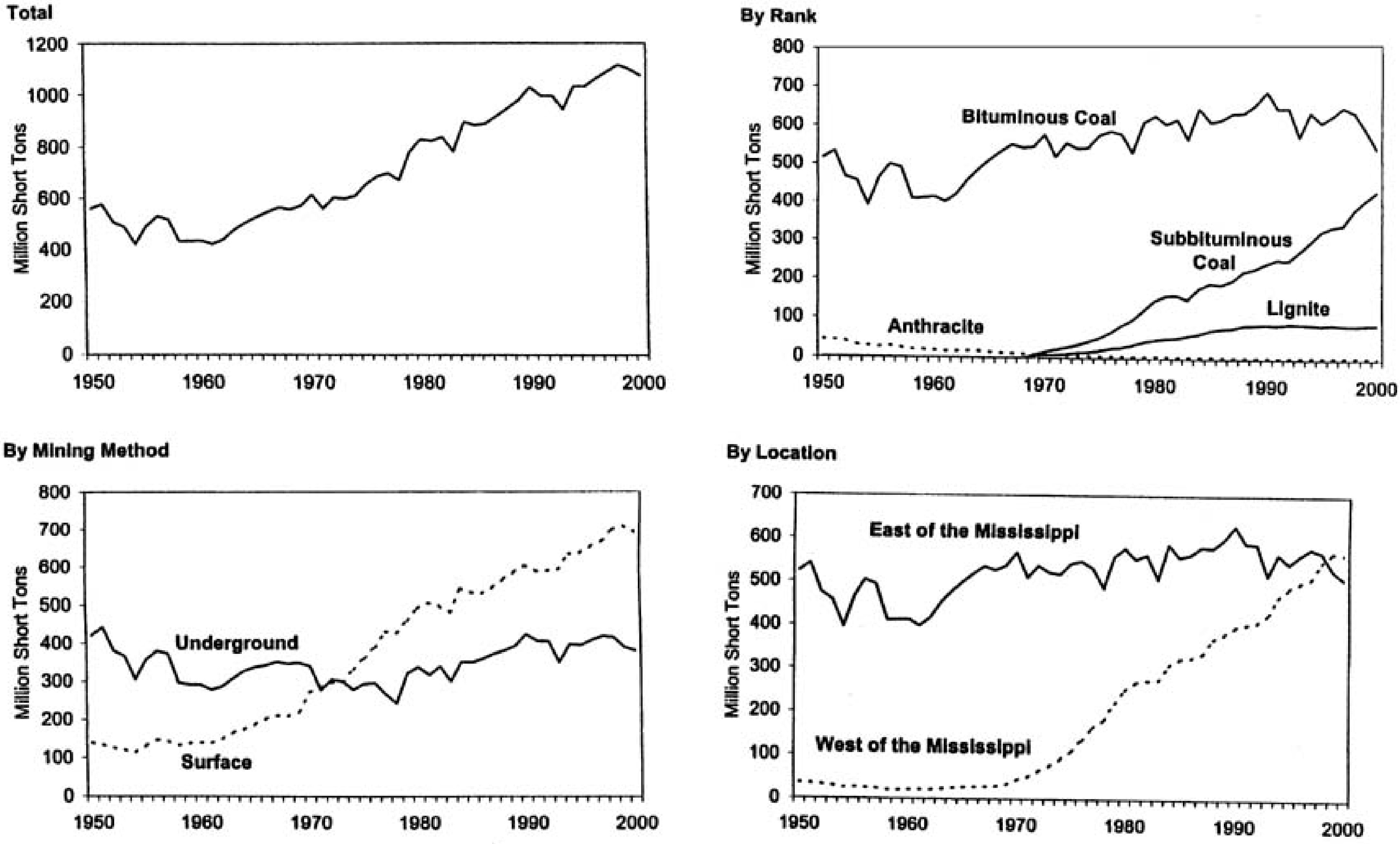
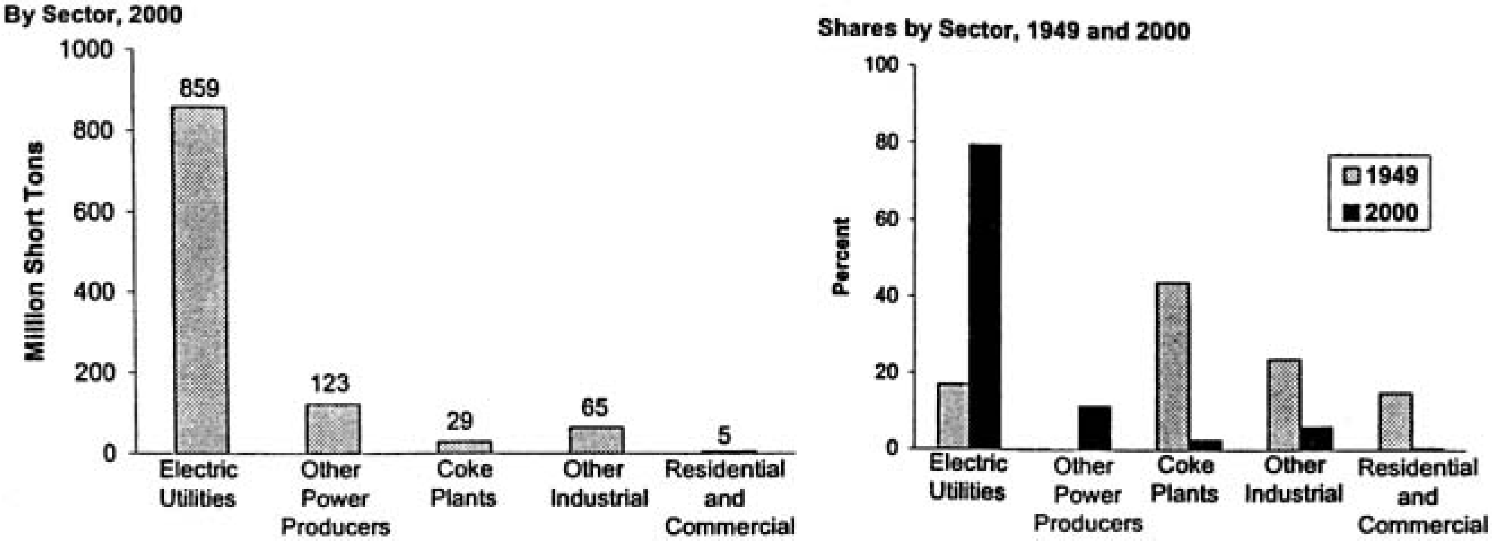


Fig. 17.2 U.S. coal consumption by sector. (Annual Energy Review, *Energy Information Agency, U.S. Department of Energy, DOE/EIA-0384(01).*)



Environmental concerns have led to an increased demand for low-sulfur coal, which is concentrated in the West. Also, surface mining, with its higher than average productivity, is much more prevalent in the West than in the East. Although more coal is mined east of the Mississippi than west, the share of western production has increased significantly since 1965, when production of western coal was only 27 million tons (5% of the total). By 1999, western production had increased to 571 million tons (52% of the total). In the period from 1990 to 1998, medium- and high-sulfur coal production decreased from 700 to 636 million tons.

Electric utilities are the dominant consumers of coal with consumption growing from 84 million tons in 1949 (a 17% share) to 983 million tons in 2000 (a 90% share). The percentage of total consumption in 2000 for all other sectors was lower than it was in 1949. The most dramatic declines occurred in the transportation sector (as railroads switched to diesel engines) and the residential and commercial sectors (as people switched from coal to oil and gas for heating). In 1949, residential and commercial use accounted for 187 million tons (39% of the total); but in 2000, consumption in these sectors totaled less than 6 million tons, less than 1 percent of U.S. coal usage. Consumption by the industrial sector trended downward after the mid-1960s, falling to about 65 million tons in 2000 from 205 million tons in 1966. As Table 17.1 shows, the United States is the second largest producer of coal in the world, right behind China.

TABLE 17.1 Top Coal-Producing Countries, 1998

Country	Production (Million Short Tons)
China	1,337.4
United States	1,117.6
India	322.2
Australia	313.7
Russia	256.7
South Africa	247.6
Germany	233.0
Poland	196.8
North Korea	93.6
Czech Republic	83.4
Ukraine	83.3
Canada	83.1
Turkey	74.3

Greece	67.1
Indonesia	66.5
Kazakhstan	66.0
World total	5,006.2

Average productivity of all types of mines in the United States has increased each year since 1949, reaching 3.7 tons per miner per hour in 1989. In 1998, average productivity reached 6 tons per miner per hour, largely due to increased production from large western surface mines and increased use of longwall mining in the eastern United States. Continued improvements in mine productivity (averaging 6.7% a year since 1978) are projected to cause falling real mine mouth prices. Higher electricity demand and lower prices, in turn, yield increasing coal demand. However, this demand is subject to a fixed sulfur emissions cap set by the Clean Air Act Amendments of 1990 (CAAA 90), which has led to progressively greater reliance on low-sulfur western coal. The use of western coals can result in up to 85 percent reduction in sulfur emissions compared with many eastern coals. As coal demand grows, however, new coal-fired generating capacity is required to use the best available control technologyscrubbers and other advanced techniques that can reduce sulfur emissions by 90 percent or more. Thus, there will still be a market for low-cost, higher-sulfur coal.

The literature on the subject of coal composition, structure, and use has grown significantly during the last two decades, as interest in coal research has continued. The future of coal science looks bright as researchers continue to make significant contributions to the elucidation of the structure, composition, and physicochemical behavior of coal. New analytical techniques have made an important contribution to these advances. The objective of this chapter is to provide a brief review of the state of the art of coal science and technology.

17.2 Origin and Classification of Coal

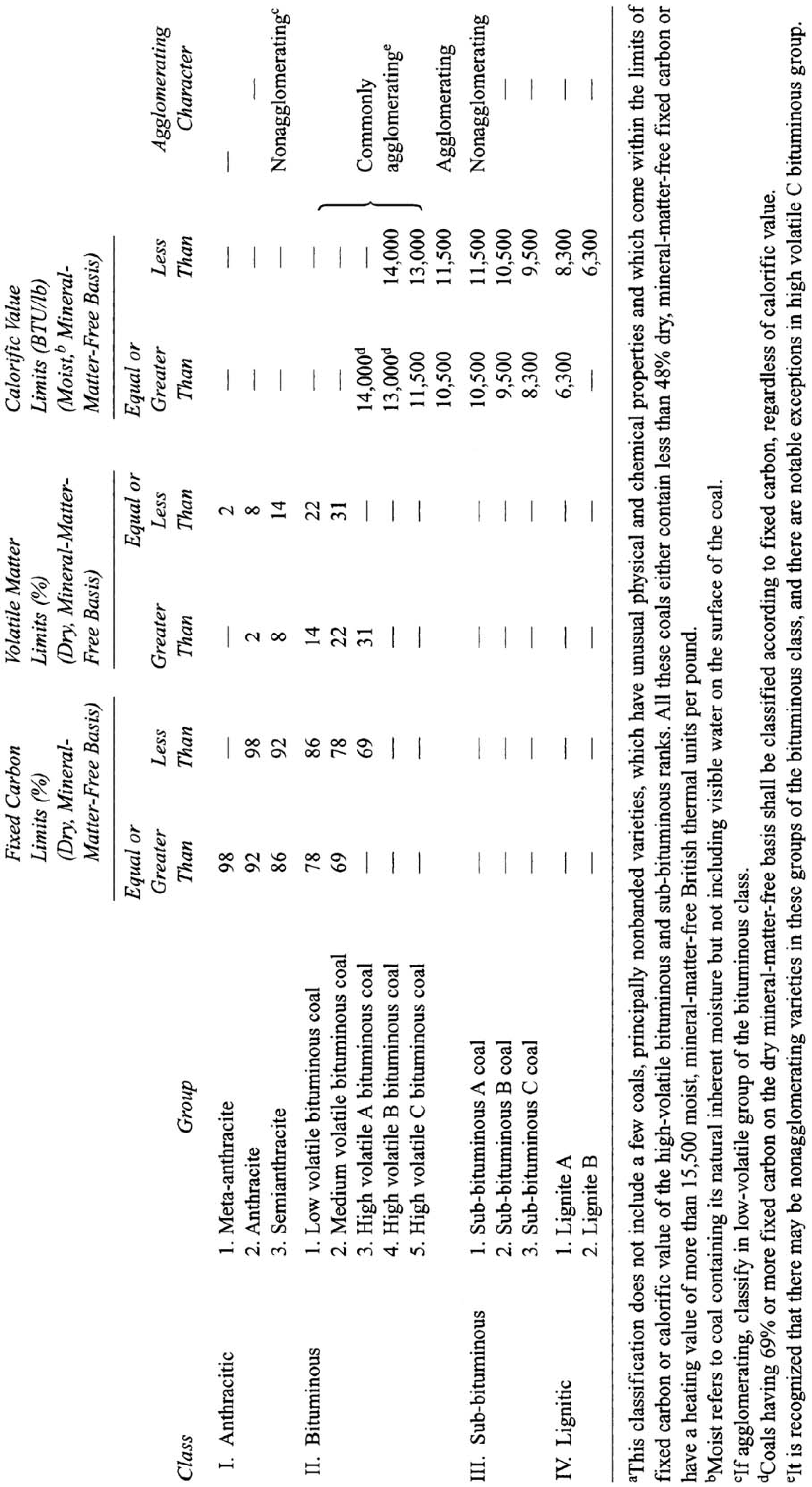
Coal is a brown to black carbonaceous sedimentary rock composed of macerals (organic material) and minerals (inorganic material). Coal originated from the remnants of various forms of plant life (mosses, ferns, shrubs, trees, etc.) that flourished in swamps and bogs millions of years ago during prolonged periods of warm temperatures and abundant rainfall. This accumulated organic matter, known as peat, was the precursor of coal. Although the peat may have been altered by bacterial and chemical action, overlying water excluded oxygen and retarded the rate of decay. This exclusion of oxygen from the peat was critical for preserving the organic matter. Over time, inorganic sediments accumulated on top of the deposits of plant debris. The gradual rising and lowering of the sea level over time led to deposits of peat alternating with deposits of sand and mud, resulting in the current situation of coal seams interlayered with rock layers.

As overlying sediments became thicker, the peat was compressed into a solid rock by the process of lithification. (Typically, 6 ft of peat will form 1 ft of coal.) This was followed by coalification, the complex progressive chemical changes that convert peat into coal. A major reaction occurring during coalification is the loss of oxygen (in the form of water) from the organic material. Thus, as coalification proceeds, the oxygen and hydrogen contents of the coal decrease, with a concomitant increase in carbon content. The length of time that these reactions proceed and the temperatures and pressures to which the organic molecules are exposed determine the degree of coalification (maturity) of the coal. Coal rank is a measure of the degree of coalification; as coalification proceeds, coal rank increases from lignite through sub-bituminous and bituminous to anthracite.

Coal is an inhomogeneous rock, since different types of plant material behave differently under coalification conditions, and mineral matter is not uniformly distributed, either by type or by amount. Microscopic examination reveals this inhomogeneity by showing that different areas of the coal structure have a different appearance. The organic components of coal are classified into six macerals: vitrinite, exinite, resinite, semifusinite, micrinite, and fusinite. The first three of these macerals are reactive (relatively easily hydrogenated or oxidized), and the other three are relatively inert. Most coals are banded, that is, show a banded structure. Besides banded coals, there are two types of dull, nonbanded coals: (1) cannels, which are rich in spores, and (2) bogheads, which contain abundant remains of algae.

Coals can be classified by rank and/or type. Table 17.2 presents a classification of coals in the United States according to rank. The factors determining rank (degree of coalification) are moisture, volatile matter (material that is volatilized when coal is heated at a certain temperature for a certain length of time), fixed carbon (the ash-free residue after the loss of moisture and volatile matter), heating value, caking tendency, and weathering properties. Vitrinite reflectance (a commonly used rank indicator for high-rank coals) is measured by comparing the intensity of a beam of light incident on a polished vitrinite surface with the light reflected directly back from that surface.

TABLE 17.2 Classification of Coals by Rank[a]



Class	Group	Fixed Carbon Limits (%) (Dry, Mineral-Matter-Free Basis)		Volatile Matter Limits (%) (Dry, Mineral-Matter-Free Basis)		Calorific Value Limits (BTU/lb) (Moist,[b] Mineral-Matter-Free Basis)		Agglomerating Character
		Equal or Greater Than	Less Than	Greater Than	Equal or Less Than	Equal or Greater Than	Less Than	
I. Anthracitic	1. Meta-anthracite	98	—	—	2	—	—	—
	2. Anthracite	92	98	2	8	—	—	—
	3. Semianthracite	86	92	8	14	—	—	Nonagglomerating[c]
II. Bituminous	1. Low volatile bituminous coal	78	86	14	22	—	—	Commonly agglomerating[e]
	2. Medium volatile bituminous coal	69	78	22	31	—	—	
	3. High volatile A bituminous coal	—	69	31	—	14,000[d]	—	
	4. High volatile B bituminous coal	—	—	—	—	13,000[d]	14,000	
	5. High volatile C bituminous coal	—	—	—	—	11,500	13,000	
						10,500	11,500	Agglomerating
III. Sub-bituminous	1. Sub-bituminous A coal	—	—	—	—	10,500	11,500	Nonagglomerating
	2. Sub-bituminous B coal	—	—	—	—	9,500	10,500	—
	3. Sub-bituminous C coal	—	—	—	—	8,300	9,500	—
IV. Lignitic	1. Lignite A	—	—	—	—	6,300	8,300	—
	2. Lignite B	—	—	—	—	—	6,300	—

---

[a] This classification does not include a few coals, principally nonbanded varieties, which have unusual physical and chemical properties and which come within the limits of fixed carbon or calorific value of the high-volatile bituminous and sub-bituminous ranks. All these coals either contain less than 48% dry, mineral-matter-free fixed carbon or have a heating value of more than 15,500 moist, mineral-matter-free British thermal units per pound.

[b] Moist refers to coal containing its natural inherent moisture but not including visible water on the surface of the coal.

[c] If agglomerating, classify in low-volatile group of the bituminous class.

[d] Coals having 69% or more fixed carbon on the dry mineral-matter-free basis shall be classified according to fixed carbon, regardless of calorific value.

[e] It is recognized that there may be nonagglomerating varieties in these groups of the bituminous class, and there are notable exceptions in high volatile C bituminous group.

An international coal classification (Table 17.3) classifies high-rank coals according to their volatile-matter content, calculated on a dry, ash-free (daf) basis. Because volatile matter is not a suitable parameter for classifying coals containing more than 33 percent volatile matter, the calorific value on a moist, ash-free basis is included as a parameter for such coals. The resulting classes of coal, based on volatile-matter content and calorific value, are grouped according to their plastic properties by employing either the free-swelling or the Roga test. They are then placed in subgroups according to plastic properties, by using either the AudibertArnu or the GrayKing test. A three-digit code number is used to identify a coal; the first digit indicates the class of the coal, the second digit indicates the group, and the third digit indicates the subgroup.

TABLE 17.3 International Classification of Coals by Type (*Courtesy Bureau of Mines, U.S. Department of the Interior*)

GROUPS (determined by caking properties)			CODE NUMBERS										SUBGROUPS (determined by coking properties)		
GROUP NUMBER	ALTERNATIVE GROUP PARAMETERS		The first figure of the code number indicates the class of the coal, determined by volatile-matter content up to 33% V. M. and by calorific parameter above 33% V. M. The second figure indicates the group of coal, determined by caking properties. The third figure indicates the subgroup, determined by coking properties.										SUBGROUP NUMBER	ALTERNATIVE SUBGROUP PARAMETERS	
	Free-swelling index (crucible-swelling number)	Roga index												Dilatometer	Gray-King
3	>4	>45					435	535	635				5	> 140	> G_8
						334	434	534	634				4	> 50-140	G_5-G_8
						333	433	533	633	733			3	> 0-50	G_1-G_4
						332 a 332 b	432	532	632	732	832		2	⩽ 0	E-G
2	2½-4	> 20-45				323	423	523	623	723	823		3	> 0-50	G_1-G_4
						322	422	522	622	722	822		2	⩽ 0	E-G
						321	421	521	621	721	821		1	Contraction only	B-D
1	1-2	> 5-20			212	312	412	512	612	712	812		2	⩽ 0	E-G
					211	311	411	511	611	711	811		1	Contraction only	B-D
0	0-½	0-5		100 A B	200	300	400	500	600	700	800	900	0	Nonsoftening	A
CLASS NUMBER →			0	1	2	3	4	5	6	7	8	9	As an indication, the following classes have an approximate volatile-matter content of: Class 6 33-41% volatile matter 7 33-44% " " 8 35-50% " " 9 42-50% " "		
CLASS PARAMETERS	Volatile matter (dry, ash-free) →		0-3	> 3-10 > 3-6.5 > 6.5-10	> 10-14	> 14-20	> 20-28	> 28-33	> 33	> 33	> 33	> 33			
	Calorific parameter a/ →		–	–	–	–	–	–	> 13,950	> 12,960-13,950	> 10,980-12,960	> 10,260-10,980			
CLASSES (Determined by volatile matter up to 33% V. M. and by calorific parameter above 33% V. M.)															

s calorific value on moist, ash-free basis (30 C., 96% relative humidity) BTU/lb.

(1) Where the ash content of coal is too high to allow classification according to the present systems, it must be reduced by laboratory float-and-sink method (or any other priate means). The specific gravity selected for flotation should allow a maximum yield of coal with 5–10% of ash. (2) 332a … > 14–16% V. M. 332b … > 16–20% V. M.

^{a}Gross c
Note: (1)
appropri

In assessing the suitability of a coal for a particular purpose, the above classification of coal types must be used with caution. The heterogeneous nature of the wide range of plant materials and minerals that were the genesis of coal guarantees an almost unlimited range of chemical constituents within the various types of coal; and these constituents determine the coal's reactivity under specific conditions.

Coal Structure

From the above discussion it should be clear that there is no such thing as pure coal. Because of variations in the organic and inorganic material that went into the formation of the coal, each seam will be different. Furthermore, there will be significant differences from point to point in the same seam. Composition and structure can vary dramatically between coals of different ranks, or even between coals of the same rank. However, some structural features are common to all coals.1

In general, at a molecular level, coals are composed of aromatic and hydroaromatic building blocks with a variety of substituent groups and some cross-linking between adjacent units. Some of these units are not strictly planar because of the presence of heteroatoms (oxygen, nitrogen, or sulfur) and hydroaromatic portions. Graphite consists of parallel, mutually oriented layers of carbon atoms separated by a distance of 0.335 µm. The parallel stacking of layers in coal is similar, but lacks mutual orientation between the layers, and the average spacing between layers is somewhat larger in coal than in graphite.

Low-rank coals are composed of small layers that are more or less randomly oriented and connected by cross links. Thus, the structure is highly porous. Medium-rank coking coals show a greater degree of orientation and a greater tendency toward parallel stacking. These coals have fewer cross links and fewer pores than lignites. High-rank coals (primarily anthracites) show a growth in the individual layers, a marked increase in the degree of orientation, and the development of a new type of pore that is elongated and parallel to the layers. These coals are the result of the highest degree of coalification.

The identification of the constituents of a complex mixture, such as coal, by molecular type may proceed in a variety of ways, but generally consists of three types of analyses: chemical, spectroscopic, and physical. Mathematical formulas have been developed that use the properties measured by these techniques to derive structural parameters. Several good reviews on this subject are available.2,3

Coal Composition and Analyses

Coal is composed of carbon, hydrogen, oxygen, nitrogen, sulfur, chlorine, other elements in trace amounts, and mineral matter. The percentage of these elements varies considerably, depending on rank and even within coals of the same rank. For coals of different ranks, the largest variation in composition occurs in the carbon and hydrogen contents, while for coals of the same rank, sulfur and mineral matter exhibit the greatest variation.

The two most common coal analyses are the proximate analysis and the ultimate analysis. In addition, there are a number of miscellaneous analyses. Many of these tests are empirical, requiring strict adherence to specified conditions. In the United States, the American Society for Testing and Materials (ASTM) standardizes these methods. International standards are developed by the International Organization for Standardization.

A proximate analysis includes the determination of moisture, ash, volatile matter, and fixed carbon (by difference). The proximate analysis, together with total sulfur and heating value, is usually sufficient to characterize coal burned for process heating and steam generation. An ultimate analysis includes the determination of carbon, hydrogen, nitrogen, total sulfur, chlorine (sometimes), ash, and oxygen (by difference). Other analyses that may be conducted include not only standard tests, such as heating value, free-swelling index, and ash composition, but also special tests to determine trace elements, mineral phases, maceral types, surface properties, conversion potential, and plastic properties.

Standard analyses are described in ASTM and Bureau of Mines publications, but nonstandard methods must be acquired from publications dealing with the specific area of interest. For example, the plastic properties of coal, which are important considerations in many process designs, can be characterized by various techniques, each having its own advantages and disadvantages (Table 17.4).

TABLE 17.4 Summary of Various Techniques to Characterize the Plastic Properties of Coal

	Description	Advantages	Disadvantages
Free-swelling index/GreyKing	Visual comparison of cokes carbonized under well-defined conditions, with standard conditions.	Inexpensive, rapid, reproducible.	Lack of flexibility; transition phenomena from coals to coke not known.
Hot-stage Microscopy	Morphological changes observed by optical microscopy.	Appropriate to supply the changes in macerals when heated.	Provides qualitative information "subjective."
Foxwell/gas flow Method	Measures the resistance of the gas flow through a bed of coal during continuous heating.	Indirectly measures the fluidity of the plastic mass by the pressure differences between the inlet and outlet ports.	Requires highly purified (oxygen-free) gas to avoid oxidation.
Plastometer			
(a) Constant torque	(a) Records the angular velocity of a rotating shaft through a bed of coal as a function of temperature.	Measures the changes in fluidity as a function oftemperature; quite sensitive for determining extent of oxidation or weathering.	Does not provide data on swelling properties, limited heating rates (-3°C/min).
(b) Variable torque	(b) Records the changes in torque required to rotate a cylindrical retort (when coal is carbonized) at constant speed as a function of temperature.		
Dilatometer	Expansion, contraction, and volume change on resolidification measured as a function of temperature when heated.	Provides data on dilation parameters and transition temperatures; wide range of heating rates (<100°C/min).	Does not provide quantitative information on changes in fluidity of coal.

Figure 17.3 shows typical heats of combustion, and Fig. 17.4 shows typical proximate analyses (reported on an ash-free basis) of various coal ranks. Analyses listed by coal bed, county, and state can be found in a number of publications, such as the data bank developed by the U.S. Department of Energy and the Pennsylvania State University Coal Bank, Bureau of Mines bulletins, Illinois State Geological Survey circulars, and the Argonne National Laboratory Coal Data Base (for a limited number of coal samples).

Fig. 17.3 Heat of combustion of U.S. coals selected to represent the various ranks. *(Courtesy Bureau of Mines)*

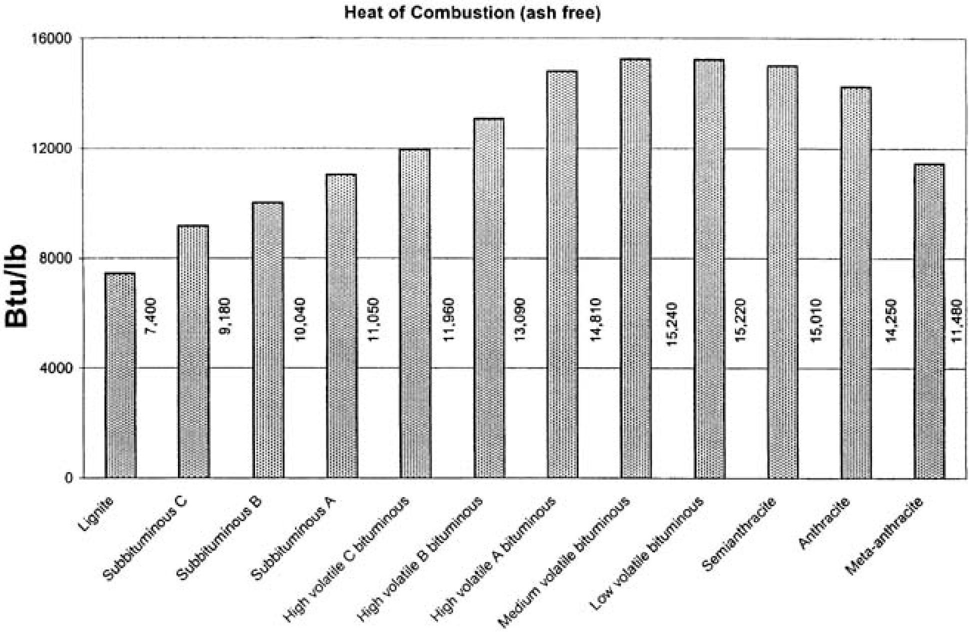
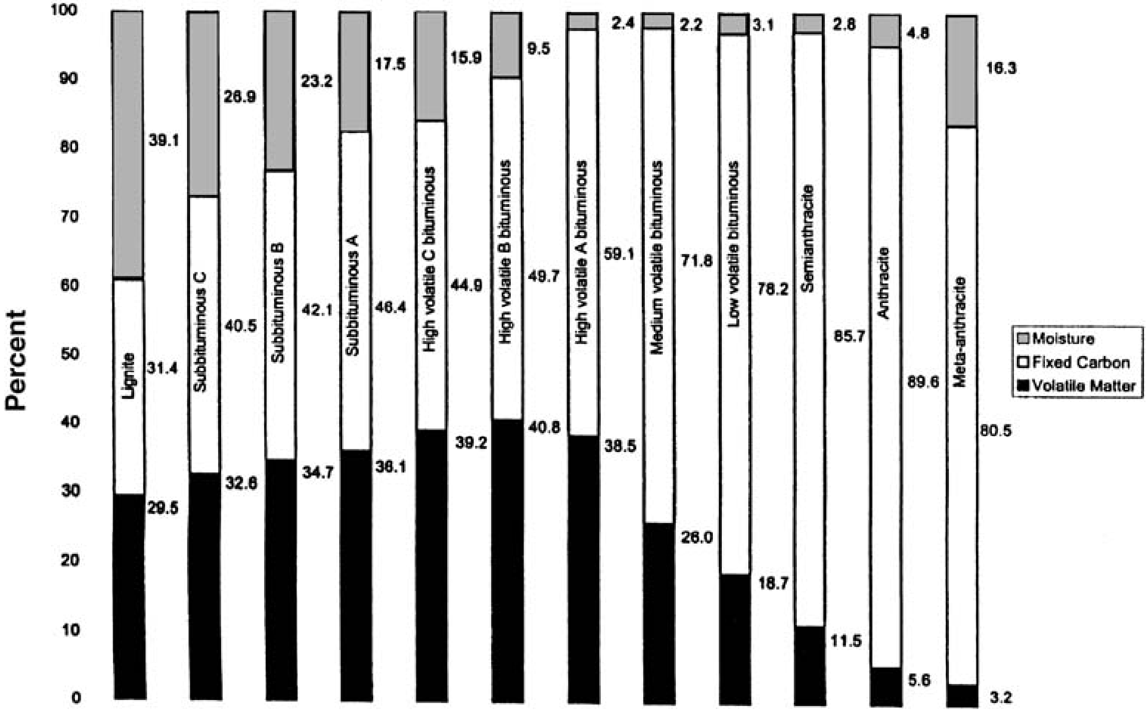


Fig. 17.4 Analysis of U.S. coals selected to represent the various ranks. *(Courtesy Bureau of Mines)*



The heating value (usually reported in BTU/lb) given in Fig. 17.3 is the higher-heating value (HHV), also referred to as the gross heating value. This quantity is defined as the heat produced by combustion of a unit quantity of coal at a constant volume, under specified conditions, with all water produced being condensed. The lower heating value (LHV), also referred to as the net heating value, is calculated from the HHV by deducting 1,030 BTU for each pound of water originally present or formed during combustion. The heating value can be estimated from a number of formulas based on the ultimate analysis for carbon (C), hydrogen (H), oxygen (O), and sulfur (S). Some correlations include additional parameters.

Nitrogen and sulfur are the two primary heteroatoms in coal and tend to form environmental pollutants when coal is processed. Therefore, the structure of sulfur and nitrogen compounds in coals, their evolution during thermal decomposition, and the distribution of these species in products have been extensively investigated. These studies suggest that nitrogen in coal is almost entirely contained in tightly bound ring structures. The nitrogen-containing structures evolve from coal without cleavage during pyrolysis. Nitrogen compounds in coal can interfere with synthetic fuels processing owing to the fact that most of these compounds are basic and are strongly adsorbed on the acidic catalysts used, thus poisoning them. Nitrogen present in coal is converted to NOx during combustion, the extent of the conversion depending on a variety of factors, including oxygen concentration, flame temperature, residence time at temperature, and the degree of fuelair mixing.

Unlike nitrogen that is almost entirely associated with the organic structure of coal, sulfur is present in both the organic and inorganic constituents. Inorganic sulfur is present in coal in several forms, such as pyrite (FeS_2) and sulfates (small amounts). Organic sulfur is mainly present in ring compounds. Pyrite decomposes at a relatively low temperature (approximately 400°C), producing FeS, H_2S, and organic sulfur compounds. Organic sulfur compounds can be classified into three groups, those with loosely bound sulfur, those with tightly bound sulfur, and an intermediate category. The tightly bound sulfur compounds decompose above 500°C, while the loosely bound sulfur compounds decompose at low temperatures (about 100°C). Some of the tightly bound organic sulfur compounds are formed by incorporating sulfur from pyrite decomposition into ring structures.

17.3 Coal Mining and Preparation

Coal mines fall into two general classificationssurface and underground. Production of coal by type of mining for the United States is shown in Fig. 17.1.

Surface Mining

Surface mining techniques are used when the coal is present near the surface, and the overlying strata (overburden) are thin enough to permit economic removal. Surface mining techniques include contour mining, strip mining, and auger mining.

Contour mining is used in hilly areas, where the slope of the surface will permit only a narrow bench cut around the side of a hill; the excavation is backfilled immediately after the removal of coal. It is the only method that can be used on slopes of 15 degrees or higher.

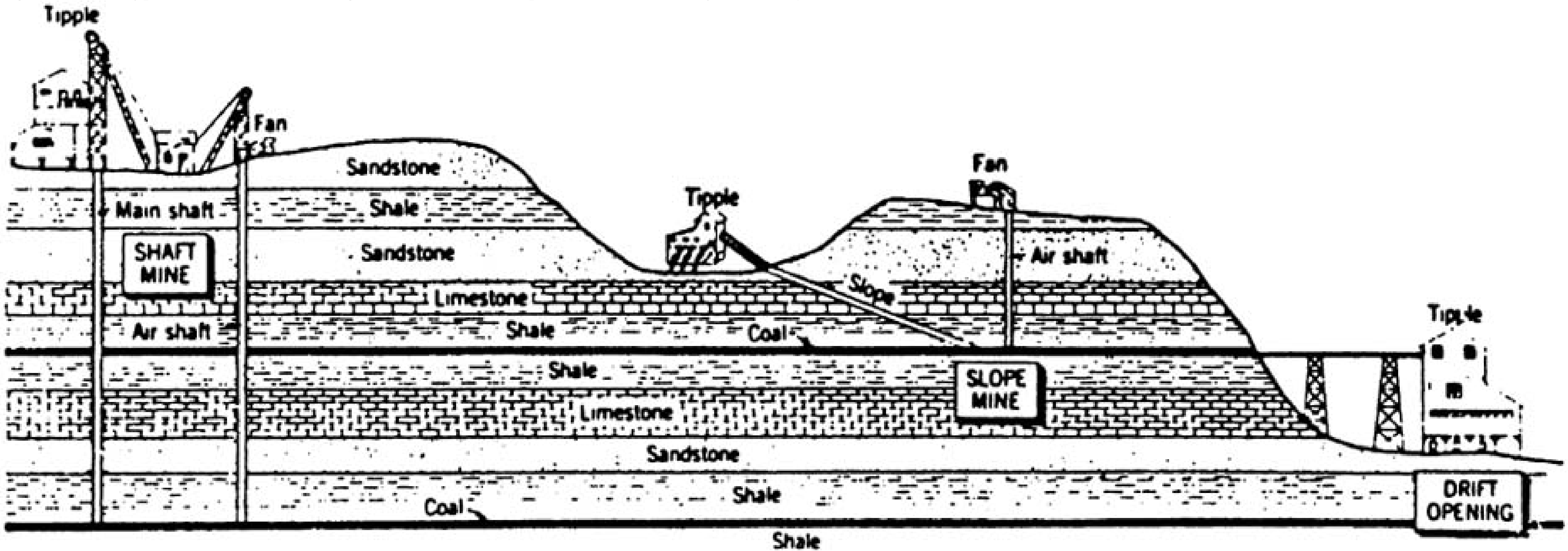
Strip mining is used in flat or gently rolling land, predominantly in the Midwest and West, where large, efficient equipment can be used. In this technique, the coal is exposed by removing the overburden. Blast holes are drilled, and explosives are loaded into these holes to shatter the rock cover; earth-moving equipment is used to remove the soil and the shattered rock. The coal is then collected with power shovels or other coal-digging machines and loaded directly into trucks. When the seam is near the top of a hill, the entire hilltop may be removed to expose the coal. Strip mining can be very efficient, especially in the West, where coal seams can be 100 ft thick or more. For example, the Wyodak seam averages nearly 65 ft thick.

Auger mining is a supplementary method used to reach coal in stripped areas where the overburden has become too thick to be removed economically. Large augers are operated from the floor of the surface mine and bore horizontally into the coal face to produce some reserves not otherwise minable. This technique is frequently used to supplement coal recovery from contour mining.

Underground Mining

Underground mining is somewhat more labor-intensive than surface mining and is used to remove coal located too deep for surface mining; but here too, machines are used in most instances to dig, load, and haul the coal. Access to the coal seam is through a drift (horizontal passage), a slope, or a shaft (Fig. 17.5), depending on the location of the coal seam.

Fig. 17.5 Three types of entrances to underground minesshaft, slope, and drift. *(Courtesy Bureau of Mines)*



A *drift mine* is one that enters a coal seam exposed at the surface on the side of a hill or mountain. The mine follows the coal horizontally.

A *slope mine* is one where an inclined tunnel is driven through the overlying rock to the coal, with the mined coal removed by conveyors or trucks.

A *shaft mine* is one where a vertical shaft is dug from the surface to reach the coal, which may be at great depth. The coal is then mined by horizontal entry into the seam, with the recovered coal hoisted to the surface through the vertical shaft.

Two general systems are used in underground mines: room-and-pillar mining and longwall mining.

In *room-and-pillar* mining, coal removal progresses in a nearly horizontal or low-angle direction by opening multiple stopes or rooms, leaving coal pillars to support the vertical load. This system recovers about 50 percent of the coal and leaves the mined area looking much like a checkerboard. It is used in areas where the overlying roof rock has geologic characteristics that provide good support. This system was used in old mines, where the coal was hand-dug. Two current methods for extracting the coal from the seam are the conventional method, where the coal is undercut and blasted free, and the continuous method, where a machine with a rotating cutter head moves along the coal face to extract the coal instead of blasting it loose. Roof control is the major problem for room-and-pillar mining. Ninety-eight percent of room-and-pillar mining is by the continuous method.

Longwall mining uses a machine that moves back and forth across the face of the coal seam in large rooms to remove up to 100 percent of the coal along faces of up to 1,000 ft or more across. As coal is removed, the mining machine moves forward, and the roof is allowed to collapse behind. Coal recovery using this method is greater than in room-and-pillar mining, and the method can be used where roof conditions are fair to poor. Strong roof rock, however, can be a problem, since roof collapse is desired in this type of mining. The seam should be over 42 in. thick to accommodate the large coal cutter or plow that is used, and a large reserve is necessary. The use of longwall mining has increased significantly in the past several years to the point where it accounts for 3040 percent of underground production. Despite the increase in the use of longwall mining, subsidence remains a significant problem. Surface structures can suffer significant damage, and wells and streams can be disrupted.

A modification of this method, using a continuous mining machine on faces up to 150 ft long, is known as *shortwall mining*. It uses the roof support system of self-advancing chocks developed for longwall operations.

Coal Preparation

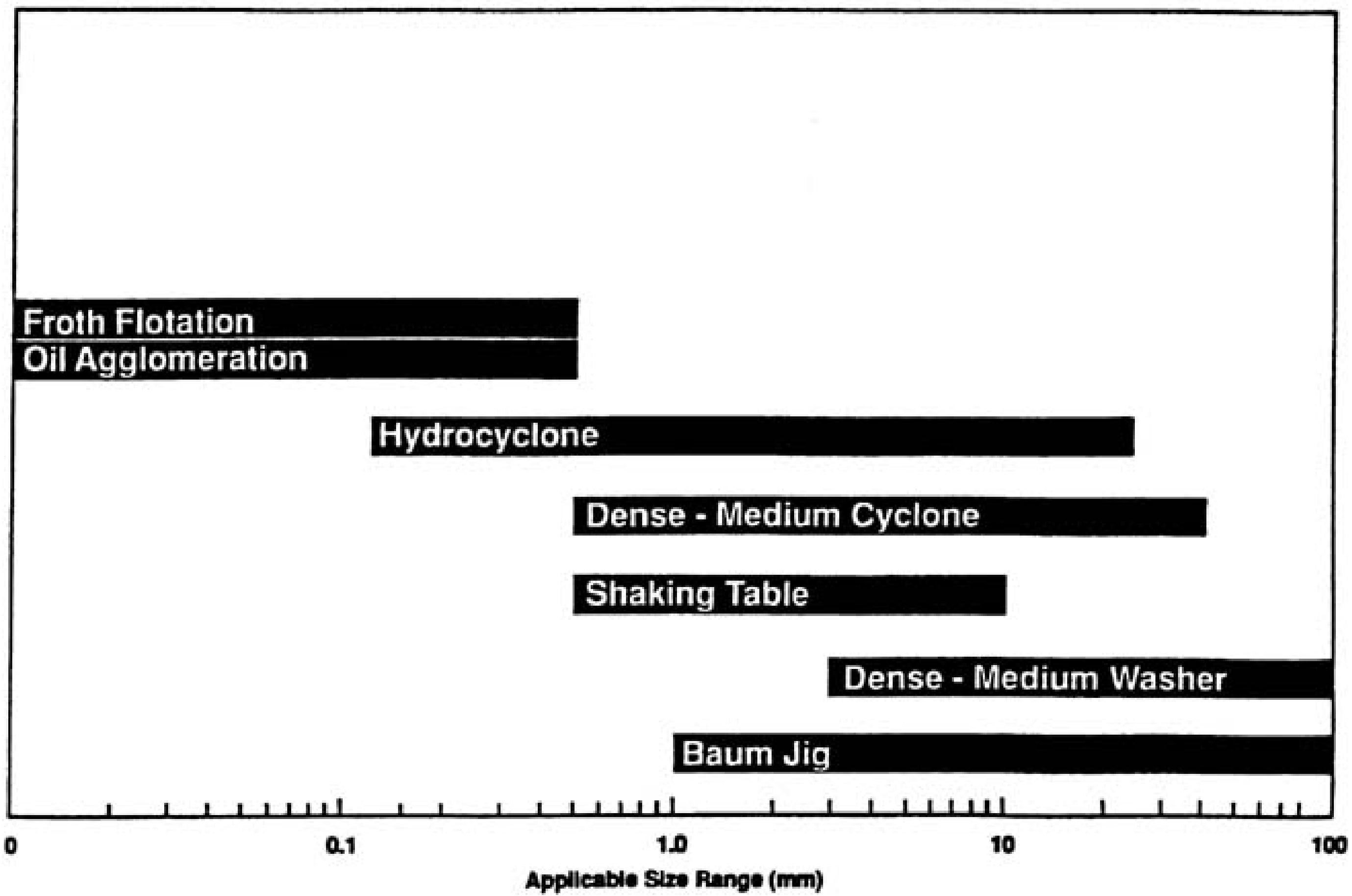
Coal preparation is a term applied to upgrading coal to make it suitable for a particular use; this includes blending and homogenization, size reduction, and beneficiation or cleaning. It is this last aspect, as well as the degree to which it is required, that most significantly governs the cost of coal preparation. Figure 17.6 shows levels of cleaning in terms of broad categories. Only about one-third of the 3.3 billion metric tons of coal produced every year is at present cleaned by breaking, crushing, screening, and wet and dry concentrating processes.

Fig. 17.6 Different levels of coal cleaning and the effect on coal recovery and economics.

The purpose of coal cleaning[4] is to remove mineral matter, consisting of rock, slate, pyrite, and other impurities, and to increase the organic content. Coal cleaning typically involves size reduction, particle sizing, cleaning, and dewatering. In general, cleaning processes utilize differences in density or surface properties between the organic and mineral matter in the coal. Depending upon the size of the coal particles being cleaned, cleaning techniques can be divided into three areas: coarse, medium, and fine.

The vast majority of the coal cleaned commercially is separated by processes that rely on the density differential of the various components. Technologies that make separations on this basis are particularly effective for treating the coarse to intermediate sizes of coal (i.e. particle sizes ranging from greater than 10 cm to about 0.10.5 mm), with efficiency decreasing as particle size decreases. Centrifugal force is often applied to improve separation, particularly for finer particles. Density-based processes are particularly effective in removing pyrite from coal. With very few exceptions, coal cleaning is done using aqueous media. Figure 17.7 shows various coal-cleaning methods and the corresponding size ranges of the coal treated.

Fig. 17.7 Coal-cleaning equipment in common use in the coal industry with respect to the coal size processed.



The first step in most coal-cleaning operations is size reduction, the main objective being to liberate mineral matter from organic matter. Typically, coal is crushed to 24 in. top size, and no grinding is performed. Size reduction equipment ranges from heavy-duty crushers and breakers, capable of crushing lumps up to a meter in size, to coal pulverization equipment, capable of milling coal to a fine powder. Grinding the coal finer liberates more mineral matter, but fine coal is more difficult and expensive to process. The ease with which coal can be ground (coal hardness) is measured in terms of the Hardgrove Grindability Index (HGI). A high HGI (>80) indicates soft, easy to grind coal, and a low HGI (•35) indicates difficult-to-grind coal. Details on size reduction can be found in various books on coal preparation.5,6

After the coal is crushed, it is generally screened to separate the raw coal into various sizes for cleaning operations. Screening of coal particles above $\frac{1}{2}$-in. size usually is carried out dry. Double-deck vibrating screens commonly are used for this purpose. For sizing below $\frac{1}{2}$ in., wet screening is used, employing either high-frequency vibrating screens or the Sieve Bend. Classifying cyclones are used to size coals finer than 100 mesh.

Coarse-Coal Cleaning

Coarse coal, which is usually the largest weight fraction in the coal feed to the cleaning plant, typically has a size greater than $\frac{3}{4}$ in. Jigging and dense-medium separation are the most frequently used coarse-coal cleaning techniques. In jigging, the separation of coal from mineral matter is accomplished via a fluidized bed created by a pulsating column of water, which produces a stratifying effect on the raw coal. The lighter coal particles rise to the top, overflow at the end of the jig, and are removed as clean product. The denser mineral matter settles and is removed as refuse. Dense-medium separations include processes that clean raw coal by immersing it in a fluid with a density intermediate between the density of clean coal and that of reject mineral matter. Most dense-medium washers use a suspension of fine magnetite in water to achieve a fluid with the desired density. Dense-medium systems are more expensive to install and operate than water-only systems, but their superior performance makes them particularly well suited for cleaning difficult-to-clean coals.

Medium-Coal Cleaning

Medium-size coal, which is usually the second largest weight fraction, includes coal ranging in size from $\frac{3}{4}$ in. to 28 mesh (0.5 mm). The principal techniques used for medium-coal cleaning are dense-medium cyclones, hydrocyclones, spirals, and wet concentrating tables. All of these equipment types are widely used in the coal industry, with the dense-medium cyclone being the most efficient and prevalent. In a dense-medium cyclone, a mixture of raw coal and a dense medium (magnetite suspension) enters the cyclone tangentially near the top, producing free-vortex flow. The refuse is discharged through the underflow orifice, and the clean coal is discharged from an axial outlet located at the top of the cyclone. When only water is used in a cyclone for cleaning coal, it is called a hydrocyclone. The spiral separator is usually 810 ft in height and consists of a trough going downward in a spiral. The coal slurry is fed in at the top; and, as the slurry follows the spiral down, centrifugal force separates the coal from the denser mineral matter. Wet concentrating tables, also known as shaking tables, work much like the classical miner's pan.

Fine-Coal Cleaning

Coal below 0.5 mm in size, classified as fine coal, typically makes up 520 percent of the feed to the cleaning plant. It is generally processed by methods such as froth flotation, which depend on surface properties. In froth flotation, the fine coal slurry, to which a small amount of flotation agent (usually fuel oil and a short-chain alcohol, such as methyl isobutyl carbinol) is added, is processed through a flotation cell. In the cell, fine bubbles are generated by using either forced air or suction. The coal, being hydrophobic, attaches to the air bubbles and rises to the top, where it is removed as froth. The refuse, being hydrophilic, remains in the water and is removed from the bottom. The process is very effective in recovering high-grade coal at moderate cost.

Conventional mechanical flotation cells have been widely used in the industry for many years and remain the primary means of cleaning coal particles smaller than 100 mesh. These units consist of a number of cells (usually 46) connected in series, each cell containing a mechanical agitator and aerator. The feed slurry, pretreated with frother and, usually, collector reagents, passes from one cell to the next with the most floatable coal being recovered in the first cell and the least floatable coal in the last cell.

A newly developed alternative to mechanical flotation cells is column flotation. Column flotation differs from conventional flotation in that columns have a much greater height-to-width ratio and do not require mechanical agitation to induce particle-bubble attachment. Columns are claimed to be more effective than conventional cells for beneficiating finer particles due to better control of bubble size (smaller is better), improved particlebubble interaction, and froth washing capabilities. There are a number of similar column processes, with each possessing one or more unique technological advances. Among the most significant of these are Microcel™, Kenflote™, Flotaire, and the packed column. The Jameson Cell differs from other columns in that particlebubble contact is achieved in a downcomer tube, where air and feed mixing occurs in a venturi-type system, rather than in the slurry column itself. The Microcel™ has been applied commercially for the recovery of fines from both freshly mined coal and refuse ponds and has also realized significant applications in the minerals industry.

In addition, improved separation of fine particles has resulted from recent advances in density-based separation techniques. This includes the development of several new water-only devices, such as various concentrating spirals, the Kelsey Jig, the Multi-Gravity Separator, the Falcon Concentrator, and the Knelson Concentrator. The development of advanced ultrafine dense-medium systems, such as Carefree Coal™ and MicroMag, also have the potential to improve cleaning of coal fines. However, none of these processes has yet achieved any significant degree of commercialization.

An oil agglomeration process that utilizes oil, or a similar hydrocarbon, to agglomerate coal and leave mineral matter in aqueous suspension has been tested on pilot and commercial scales. Other processes (high-gradient magnetic and electrostatic cleaning) utilize differences in magnetic and electrical charge properties of the mineral matter present in coal. None of these processes has achieved significant commercial success.

Conventional mechanical dewatering devices, such as vibrating screens and vibrating basket centrifuges, are effective in dewatering coarse and medium coal to moisture levels of less than 5 percent.

Vibrating and high-speed vibrating screens are typically used to dewater coal larger than about $\frac{3}{8}$ in. For coal in the $\frac{3}{8}$ in. × 30-mesh size range, vibrating basket centrifuges have proven to be highly effective and are widely used throughout industry.

One of the biggest problems in processing fine coal is dewatering and drying. Fine coal dewatering is one of the most costly operations in the coal-cleaning plant; and, therefore, a significant number of operators bypass this operation by discarding coal fines directly to their waste slurry impoundments. Although these impoundments create a tremendous environmental liability and represent a loss of valuable coal resources, they still are widely used with an estimated 3050 million tons of coal being added to ponds each year. Abandoned and active ponds in the United States are estimated to contain 23 billion tons of coal.

Vacuum filtration and screen-bowl centrifuges are the principal unit operations used for mechanical dewatering of fine coal. Pressure filtration (plate and frame filter presses and continuous belt filter presses) is used to a much lesser extent to dewater fine coal and refuse. Sedimentation devices (thickeners, including conventional, high-rate, and multiplate units) are widely used to provide clarified water for recycle back to the plant and to preconcentrate dilute slurries (usually refuse fines) prior to dewatering or disposal. Thickener operation almost always requires the use of flocculants (primarily synthetic polymers or, to a lesser extent, modified starches) to promote particle settling.

The use of screen-bowl centrifuges has increased significantly in popularity to the point where now they are the predominant method for dewatering fine coal. Compared with vacuum filters, these units provide lower operating costs and increased moisture reductions; however, typically 1015 percent of the coal entering the unit is lost to the effluent. These units typically produce filter cakes in the range of 1520 percent moisture. Solid-bowl centrifuges occasionally are used on very fine coal, where coal losses using screen-bowl centrifuges would be excessive.

The primary form of vacuum filtration is the rotary-disk vacuum filter, which consists of multiple vertical disks, connected by a horizontal hollow shaft, that are partially submerged in open tanks containing the slurry to be dewatered. A vacuum is applied to the portion of the disk that is submerged to draw slurry toward the filter medium supported by the disk to form a filter cake on the surface of the filter as water passes through. As the disk rotates out of the slurry, the filter cake is dried by the continuing vacuum action, and a final product is produced by pressurized air blowback and mechanical scrapers. Flocculating agents may be added to the slurry prior to filtration to improve dewatering performance. These systems typically produce final product moistures in the range of 2030 percent depending on the percentage of minus 200-mesh material in the feed.

Filtration devices that utilize high-pressure forces are capable of reducing moisture in the filter cake to about 20 percent, but capital and operating costs are high for these devices. Other newly developed techniques that have been tested on a pilot scale include an ultrahigh-*g* centrifuge, which generates forces up to 4000*g*, and an electro-acoustic technique that utilizes the synergistic effect of electric, ultrasonic, mechanical, and surface chemical forces to remove moisture from the cake.

Chemical Coal Cleaning

Chemical cleaning processes utilize alkali or acid to leach out the impurities present in coal. Chemical methods are effective in removing both organic and pyritic sulfur. For example, the Molten Caustic Leaching Process is quite effective in its ability to remove both organic and pyritic sulfur, mineral matter, and trace elements from coal.7 The process consists of treating coal with molten caustic (usually a eutectic mixture of sodium and potassium hydroxides) at 350400°C for up to 4 hr. The treated coal is washed with water and dilute acid to give a low-ash, low-sulfur product. The major drawbacks of chemical cleaning processes are economics and the need to use corrosive materials. The optimum coal desulfurization scheme may include both physical and chemical treatment.

17.4 Coal Utilization

The vast majority of the coal produced is burned to generate heat and power. The next most important use is for the production of metallurgical coke. Other uses are for the production of synthetic fuels and as a raw material for the production of chemicals. This section discusses the technologies involved in these uses of coal.8

Environmental Concerns Related to Coal Use

As discussed above, coal contains many of the elements in the periodic table, at least in trace amounts; and, during processing and/or combustion, these elements or their compounds can be released into the environment. Examples of harmful pollutants that can result from coal use are sulfur dioxide, sulfur trioxide, nitrogen oxides, particulates, hydrogen chloride, mercury vapor, and a wide variety of trace metals. In addition, most uses of coal result in the production of carbon dioxide, the most important greenhouse gas from the point of view of the quantity in the atmosphere. Since coal has the lowest H/C ratio of the fossil fuels, coal combustion releases more CO_2 per unit of heat released than combustion of oil or gas. Because of these problems, considerable effort is being expended on developing technologies to permit the use of coal with minimum environmental degradation. The discussions on the various coal utilization processes in the following sections include techniques to mitigate environmental concerns.

17.5 Combustion

In the United States, approximately 90 percent of the coal consumed annually is burned in boilers to generate steam911 to run turbines for electric power production. The burning of coal in small-scale combustion devices for direct heating is more significant outside the United States, so that, on a worldwide basis, about 55 percent of the coal produced is used for electricity generation.

Combustion Equipment

A summary of the most important direct combustion processes for coal is given in Table 17.5. Options for firing industrial and electric utility boilers are stokers, pulverized-coal-fired burners, and cyclone burners, the choice depending on the kind of coal and the amount of steam needed. Table 17.6 lists the types of coal that can be burned with the various combustion systems. Although good results usually will be obtained with the indicated coal and firing equipment, this table should be used only as a rough guide, and final equipment selection should be based on a sound engineering analysis.

TABLE 17.5 Summary of Direct Combustion Coal Process (Smoot)

Process Type	Description	Extent of Coal Use In U.S. (% of Total Used)	Commercial Use	Scale Size (TPD)a	Coal Types	Coal Size
Power station	Commercial electricity production	7880%				
Pulverized	Rapid burning of finely grained goal		Common	100010000	All	0.010.025 mm
Fluidized-bedb	Well-stirred combustion		Pilot plant	20008000	All	0.150.6 cm
Stoker	Mechanically fed fixed bed		Small	100	Noncaking	15 cm
MHD	Combustion energy capture by magnetic fields		Laboratory	8004000		
Coal/oil mixture (COM)	Burning coal/oil mixtures in oil furnaces		Demonstration			
Industrial heat/steam	Industrial plant power	811%				
Pulverized			Small	1100		
Fluidized-bed			Pilot	1100	Same as above	
Stoker			Common	1100		
COM			Demonstration	1100		
Domestic/commercial	Hand-stoked space heating	1%		0.0050.05	Noncaking	310 cm
Transportation	Fuel for railroads	0.010.02%		0.11	Noncaking	

aTPD = tons/day; bVery adaptable, exhibits high heat transfer in bed as well as low-level pollutant products.

Source: Excerpted from Smoot, L. D. in *Fossil Fuel Combustion*, William Bartok and Adel F. Sarafin (Eds.). John Wiley & Sons, New York, 1991. Copyright © 1991 by John Wiley & Sons, Inc. and used by permission of the copyright owner.

TABLE 17.6 Suitability of Various Combustion Systems for Various Coals

Fuel	Under Feed	Stokers Traveling or Chain Gate	Spreader	Pulverized-Coal Burner	Cyclone Burner
Anthracite		X		X	
Bituminous					
1725% volatile	X		X	X	X
2535% volatile				X	X
strongly coking	X	X	X	X	X
weakly coking		X	X	X	X
Lignite		X	X	X	X

Electric utility and large industrial plants favor pulverized-coal-fired and cyclone-fired furnaces because of their inherent capability to effectively burn a range of coal types with varying qualities, their comparatively high availability, their quick response to load changes, and their extremely large steamgenerating capacity. Large units can generate as much as 9.5 million pounds of steam per hour at 3,500 psig and 1,000°F. Smaller industrial installations typically favor stoker firing.

Stoker Firing

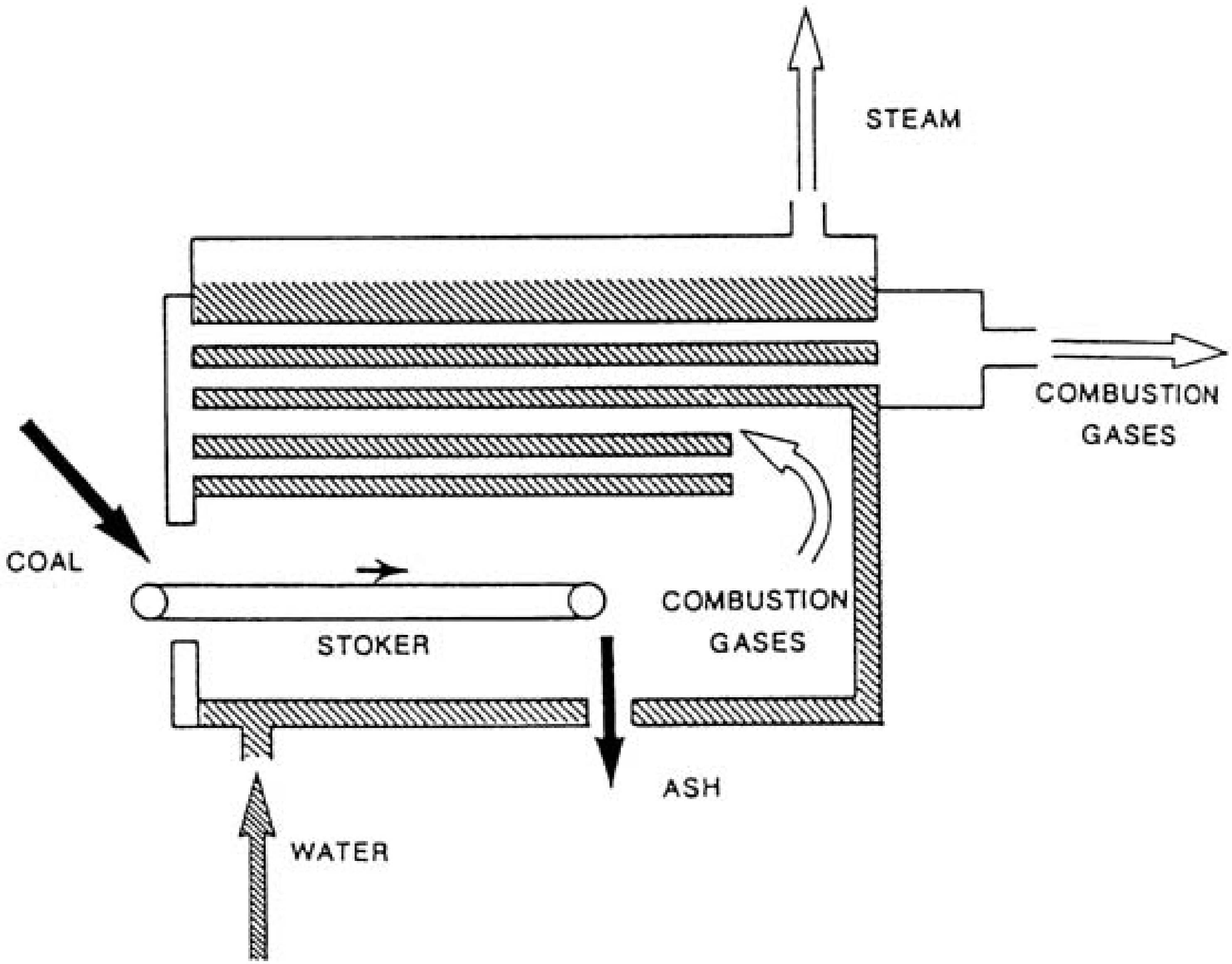
Stokers have been in operation for more than 150 years and are still used to fire small-capacity boilers, generally for steam production. There are several types of stokers, which are classified by the method of coal feeding and the relative flow directions of the fuel and air. The three main types are underfeed, crossfeed, and overfeed. Both fuel and air flow in the same direction in retort stokers; this arrangement is called underfeed burning. The fuel moves across the air direction in chain- or traveling-grate stokers; this design is called crossfeed burning. The spreader stoker approximates overfeed burning, the incoming fuel moving toward the air. Except for certain types of coal gasifiers, in which lump coal moves downward toward a grate against air (or oxygen and steam) coming through the grate, no conventional combustion system operates purely in the overfeed mode. Table 17.7 gives the approximate capacity range for each type of stoker.

TABLE 17.7 Approximate Capacity Range of Stokers

Type	Steam Production (1000 lb/hr)	Grate Heat Release (1000 BTU/hr/ft2 (max))
Single retort	550	200
Multiple retort	40300	300
Traveling or chain grate	10300	300
Spreader	10300	1000

An example of a crossfeed chain-grate stoker being used to heat a fire-tube boiler is given in Fig. 17.8. This stoker consists of an endless track of grate bars that pass over rotating sprockets at each end, similar to a conveyor belt. Coal is fed onto one end of the grate, using a hopper, to a depth of about 4 in. The coal burns as it moves along the grate, and, at the far end of the grate, only ash remains. The ignition of the coal is assisted by radiant heat from a refractory lining.

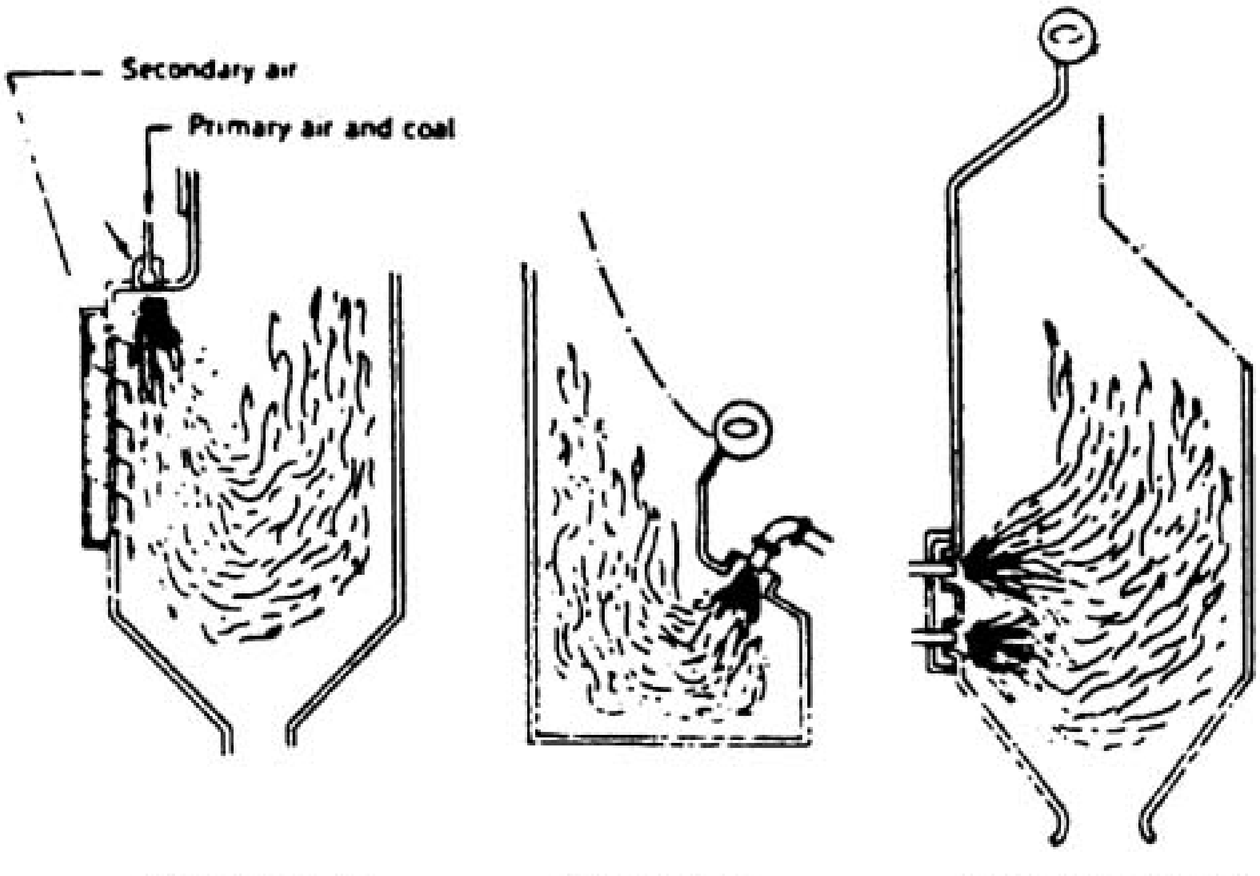
Fig. 17.8 Example of a chain-grate stoker equipped with a fire-tube boiler. (*Reprinted with permission from Schobert, H.H.,* Coal: The Energy Source of the Past and the Future, *American Chemical Society, Washington D.C., Copyright 1987, American Chemical Society*.)



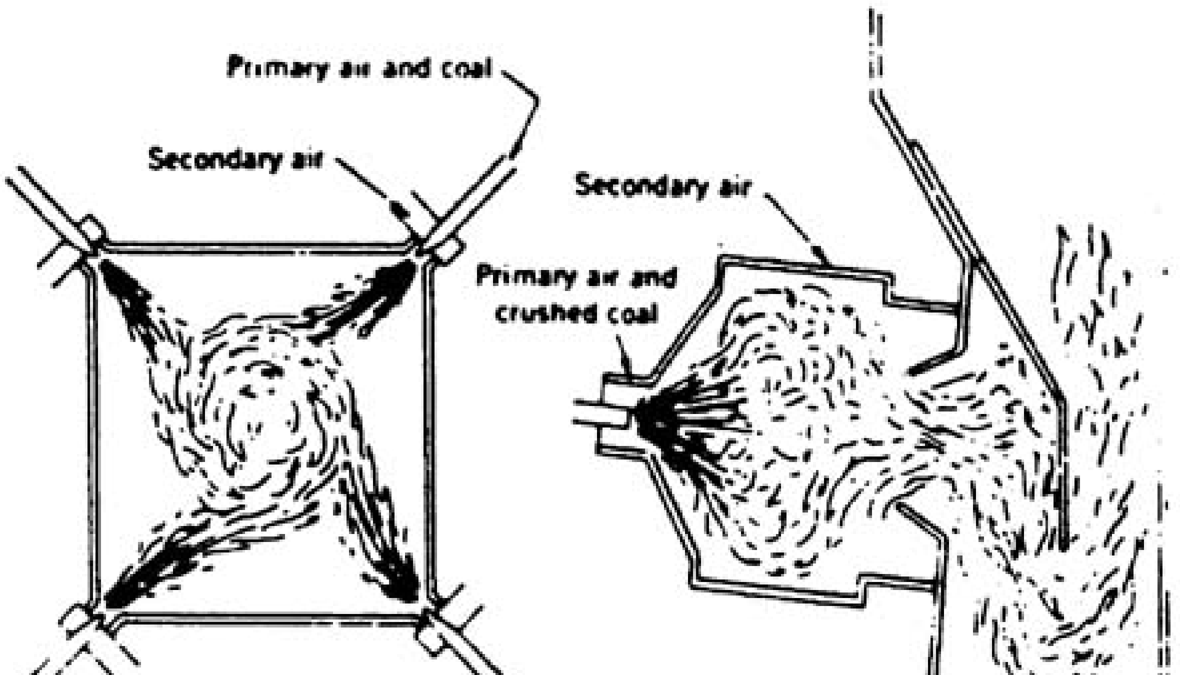
Pulverized-Coal Firing

The burner and furnace configurations for the main types of pulverized-coal firing (often called suspension firing) are shown in Fig. 17.9. There are some design variations among various vertical, impact, and horizontal suspension firing systems, but these schematic drawings serve to illustrate the general principles.

Fig. 17.9 Methods of firing pulverized and crushed coal. (*Courtesy Bureau of Mines*)



VERTICAL FIRING IMPACT FIRING HORIZONTAL FIRING



TANGENTIAL FIRING FROM CORNERS OF FURNACE WALLS

CYCLONE FIRING

Coal-fired boilers are further differentiated by the operating mode, either dry-bottom or wet-bottom (slagging), depending on whether the ash leaves the boiler as a solid or as a liquid. Most modern pulverized-coal-fired boilers are of the dry-bottom type, in which the ash is cooled before contacting any surface.12 Wet-bottom firing provides a more convenient method of ash removal and disposal but has several disadvantages, including less flexibility in fuel selection, higher incidence of fouling and corrosion, and higher levels of NOx formation, that have led to its virtual elimination in the United States.13

The first suspension-fired furnace in the United States was designed like the one shown for vertical firing in Fig. 17.9. Pulverized coal (about 70% through a 200-mesh screen) is transported to the burner with primary air, amounting to about 20 percent of the total air needed for complete combustion. The balance of the air, known as secondary air, is admitted through openings in the furnace wall. Because a large percentage of the total combustion air is withheld from the fuel stream until it projects well down into the furnace, ignition stability is good. This type of firing is well suited for coals that are difficult to ignite, such as those with less than 15 percent volatile matter. Although no longer used in central-station power plants, this design, with delayed admission of secondary air, may find favor again if low-volatile chars from various coal-conversion processes become available for burning for heat and power.

The other types of suspension firing use burners in which the primary air, coal, and secondary air are mixed just before or immediately after entering the furnace. With tangential firing, the burners are arranged in vertical banks at each corner of a square, or nearly square, furnace and directed tangentially toward an imaginary circle in the center of the furnace. This produces a vortex with its axis along the vertical centerline. The burners consist of an arrangement of slots, one above the other, admitting through alternate slots the primary airfuel mixture and secondary air. The burners can be tilted upward or downward 30 degrees from the horizontal plane, enabling the operator to control superheat and to permit selective utilization of furnace heat-absorbing surfaces. In tangential firing, the turbulence needed for mixing the fuel and air is generated in the furnace instead of in the burners.

The other major type of pulverizedcoal-fired boiler is the wall-fired boiler. In wall-fired units, instead of being mounted in the corners, the burners are mounted on the walls of the furnace. They may be mounted on only one wall, or they may be mounted on opposing walls. The burners typically are mounted in a grid pattern.

Cyclone Firing

Cyclone firing is a form of two-stage wet-bottom firing in which the coal is not pulverized, as for suspension firing, but is crushed to 4-mesh size and admitted tangentially with primary air to a water-cooled cylindrical chamber called a cyclone furnace, as shown in Fig. 17.9. The finer particles burn in suspension, while the coarser ones are thrown by centrifugal force to the furnace wall. The wall, having a sticky coating of molten slag, retains the coal particles until they complete combustion. Secondary air, which is admitted tangentially along the top of the furnace, completes the combustion of the coarse particles. Slag drains continuously into the main furnace and through a tap hole into a water-filled slag tank, where it solidifies. The main advantage of cyclone firing is the ability to use larger coal particles. However, being a wet-bottom technique, it suffers from the disadvantages discussed above. In particular, the high temperature required to melt the slag results in high NOx production.

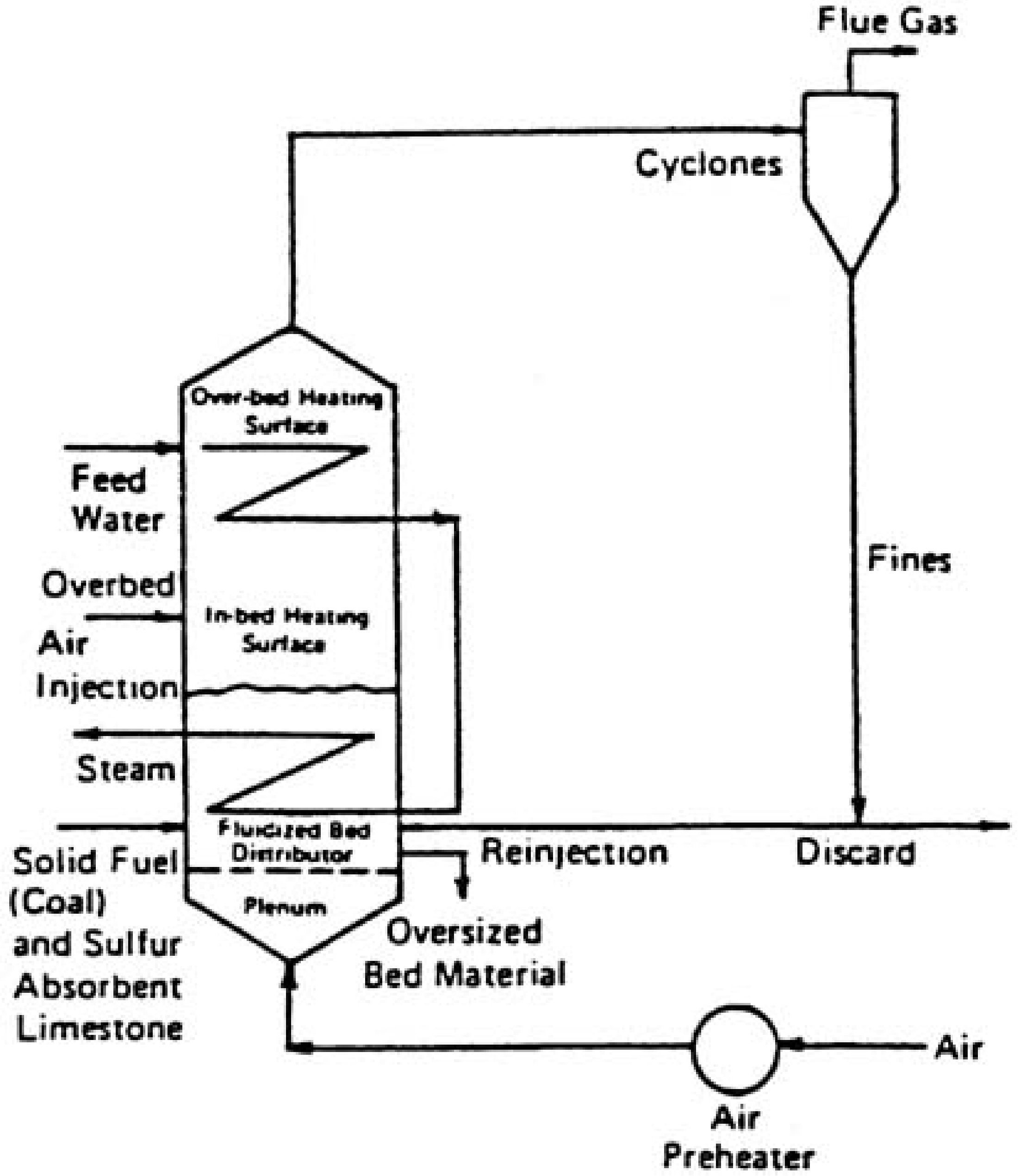
Fluidized-Bed Combustion

In the past 20 years, fluidized-bed combustion (FBC) has advanced from the laboratory and pilot-plant scale to commercial status. A fluidized-bed combustor consists of a vessel containing a granular material, such as sand, coal ash, or limestone. This bed is fluidized by blowing air up through it. Fuel particles are injected into the bed and burn as they contact the oxygen in the air. As ash particles agglomerate, they sink to the bottom of the bed and are removed. Fluidized-bed combustors are very flexible with respect to the fuel that can be burned, and almost any carbonaceous material that can be injected into the bed can be used. Steam tubes are generally immersed in the bed, where the heat transfer coefficient is high.

Although this technology is not yet widely used in the electrical power industry, the potential advantages are clear. (1) The bed temperature is low, about 1550°F, which means less formation of NOx and the retention of some of the sulfur in the ash of certain coals. (Adding dolomite or limestone to the bed greatly improves sulfur retention.) Of equal importance is less volatilization of sodium and potassium in the coal; consequently, there are fewer deposits on, and corrosion of, the furnace, superheater, and reheater tubes. (2) Heat-transfer rates from the fluidized bed to the immersed heat-transfer surfaces are relatively high, as much as 100 BTU (hr ft2 °F). (3) Temperature control is easier, because rapid mixing in the bed results in a relatively uniform bed temperature. (4) FBC has the capability to burn many kinds of coal and is less sensitive to changes in coal heating value or ash characteristics than other combustion methods. (5) FBC is adaptable to several kinds of power generation cycles (e.g. co-generation or combined cycle), can operate at atmospheric or elevated pressure, and can use bubbling or circulating beds with in-bed or external heat exchangers. This versatility allows engineers to develop FBC systems that are tailored to specific applications. This can result in significant cost savings when compared with other technologies.

The application of fluidized bed boilers to electric utility stations should reduce operating costs and increase efficiency compared with conventional pulverized-coal-fired plants of the same capacity equipped with emission controls. The principle of the fluidized-bed combustor is illustrated schematically in Fig. 17.10.

Fig. 17.10 Basic components of a fluidized bed combustor. (Shang, J.Y., Department of Energy, Morgantown, WV.)



Interest is increasing in running units at elevated pressures (615 atm) in order to reduce equipment size, lower emissions, reduce sorbent requirements, and increase the thermal efficiency by using a gas turbine in addition to a steam turbine in a combined cycle. There is also a trend away from bubbling beds and toward circulating beds, which use finer solids and higher velocities and are able to use sorbents more efficiently.

Several large demonstrations of FBC technology for electrical power generation have proven the viability of this technology. These include the 110-MW(e) atmospheric pressure circulating fluidized-bed boiler at the ColoradoUte Electric Association's Nucla Station, which began operation in August 1988. A second project is the American Electric Power Tidd 70-MW(e) pressurized fluidized-bed combustor/combined-cycle demonstration, which began generating power in December 1990. Both projects were partly funded by DOE's Clean Coal Technology Program.14

Boiler Types

Industrial and utility boilers are broadly classified as fire-tube or water-tube. In fire-tube boilers, the hot combustion gases pass through tubes, and heat is transferred to water outside the tubes. Most steam locomotives had this type of boiler. The most common and least expensive boiler of this type is the horizontal return tubular (HRT) boiler. However, because of the design and construction of fire-tube boilers, there is a definite limitation to their size and the pressure that they can tolerate.

Water-tube boilers may be broadly classified as straight-tube and bent-tube types; the latter have several variations in design and are preferred for applications where higher capacities and steam pressures are required. In both types, heat is transferred by radiation or convection to the outside of the tubes, and water flows inside the tubes as a result of thermal circulation, or in the case of certain bent-tube boilers, as the result of forced circulation. A comparatively new version of the forced-circulation, bent-tube boiler for central station power plants is the "once-through" type. The feed-water passes progressively through the heating, evaporation, and superheater sections; no drum is used for separating steam from unvaporized water as in other boilers, so the ratio of water circulated in the boiler to steam generated is unity. Table 17.8 presents approximate ranges of steam capacity and pressure for the principal types of boilers.

TABLE 17.8 Approximate Range of Capacities of Various Types of Industrial and Utility Boilers

Type	Capacity (lb steam/hr)	Maximum Design Pressure (psig)
Fire tube (HRT)	1,00015,000	250
Water tube		
Straight	15,000150,000	2,000
Bent, three-drum, low-head	1,00035,000	400
Two-drum, vertical	1,000350,000	1,000
Electric utility	Up to 9,000,000	3,500

Pollution Controls

Coal combustion is drawing increasing scrutiny because of concerns over coal's contribution to several environmental problems: acid rain (from oxides of sulfur and nitrogen), global warming (from CO2), and particulates and trace elements (from ash).15 Despite increasingly stringent environmental regulations, coal is likely to remain an important fuel for power plants in the United States and many other countries.16,32 The amount of coal burned is likely to increase steadily in developing countries like China and India. The combustion of coal is also important in a number of eastern European economies, such as Poland. Consequently, there is a strong impetus for developing highly efficient clean coal technologies. Achieving significant reduction of SO2 and NOx emissions from coal combustion is a major goal of the 1990 Clean Air Act Amendments. The achievement of these goals has been assisted by a U.S. DOE effort known as the Clean Coal Technology Program, which has provided government funding for a range of advanced pre- and post-combustion technologies to mitigate environmental problems.

Emissions of sulfur oxides, nitrogen oxides, and particulates from coal combustion are problems of increasing concern and regulation. Coal combustion contributes about 25 percent of the particulate matter, 25 percent of the sulfur oxides, and 5 percent of the nitrogen oxides emitted to the atmosphere. Much of the particulates are derived from the mineral content of the coal, but some particulates also result from sulfur and nitrogen oxides that react to form various sulfate and nitrate salts. A major concern about particulate matter is that the smallest particles are respirable and may pose a health hazard. Particulate matter is recovered in most power plants by the use of electrostatic precipitators, which have been developed to very high efficiencies (>99%). Other methods of particulate removal include baghouses and cyclone collectors.

An important technology for removal of SO2 is Flue Gas Desulfurization (FGD), carried out in units known as scrubbers. Most scrubbers contact the flue gas with a slurry of lime or limestone to capture the sulfur oxides and produce a sludge containing calcium sulfite and calcium sulfate. However, disposal of sludge is another environmental problem, and some scrubbers include oxidation to convert all the calcium sulfite to sulfate (gypsum), which can be used for wallboard manufacture. Fluidized-bed combustion units add a sulfur capture agent, such as limestone, to the bed, and Integrated Gasification Combined Cycle (IGCC) systems convert the sulfur in the coal to hydrogen sulfide, which can be removed easily from the fuel gas stream before it is burned.

Most of the NOx emitted from coal combustion is in the form of nitric oxide (NO). Some of the NOx originates from oxidation of nitrogen in the coal, and the rest comes from the reaction of nitrogen and oxygen in the combustion air. Consequently, partial control of NOx can be achieved by changing the conditions of the combustion process. This usually involves the use of staged combustion, in which coal is first burned at fuel-rich conditions, followed by addition of secondary air to complete combustion at fuel-lean conditions. Staged combustion decreases NOx by lowering both oxygen concentration and temperature during combustion. So-called low-NOx burners have been developed that can be retrofitted to most utility boilers to reduce NOx emissions. Other techniques that have been developed to reduce NOx emissions involve post-combustion operations, such as reburning, selective noncatalytic reduction (SNCR), and selective catalytic reduction (SCR). In reburning, additional fuel (natural gas, oil, or pulverized coal) is added to the combustion gases downstream of the main combustion to result in fuel-rich, reducing conditions, which convert NOx into nitrogen and water. Overfire air is then added to complete combustion, but because of the lower temperature, additional NOx is not formed. In SNCR, ammonia or urea is injected into the furnace to react with NOx and form N2. For SNCR to be effective, the flue gas, after urea or ammonia injection, must spend a minimum residence time within a specific temperature range. In SCR, ammonia is injected into the flue gas, which then passes over a catalyst where reduction of NOx to N2 occurs. SCR is used extensively in Japan and Germany; and, because of decreasing costs, its use in the United States is steadily increasing. SCR is attractive because it can remove upwards of 90 percent of the NOx. Fluidized-bed combustors, in which combustion occurs at a lower temperature than in PC fired furnaces, generally produce less NOx.

Another pollutant which is coming under increasing scrutiny is mercury.17,18 Coal contains trace amounts of mercury (an average of about 0.1 ppm) which, because of its high volatility, is released into the atmosphere when the coal is burned. Approximately 50 tons/year of mercury are released by power plants. Although coal cleaning can reduce the mercury content of coal, it will prove difficult to achieve complete removal. Therefore, post-combustion technology is the most probable approach to controlling mercury emissions. Possible technologies include adsorbents and scrubbers. Scrubbers are already used at many power plants to remove SO2 from the flue gas, and many of these scrubbers have been found to remove a significant fraction of the mercury. One adsorbent that has been used effectively is activated carbon. The EPA is under mandate to develop regulations that may require removal of up to 90 percent of mercury emissions.

Advances in Combustion Technology

Because of improvements in steam-generating equipment and turbines, remarkable advances have been made in the energy efficiency of electric utility plants. One method of expressing the efficiency of the conversion of coal to electricity is the heat rate, which is the amount of heat energy required to generate one kilowatt-hour of electricity. The heat rate of a modern, well-maintained plant in 1990 was about 8750 BTU/kWhr, compared with an average value of 17,500 BTU/kWhr in 1937 and 10,000 BTU/kWhr in 1968. Some new designs project heat rates close to 7500 BTU/kWhr, corresponding to an overall thermal efficiency of 45 percent.19

Further improvements in thermal efficiency may be achieved with new power generation systems now being developed. One promising technology is IGCC, in which both a gas turbine and a steam turbine are used to generate electricity. In this concept, the coal is first gasified to produce fuel gas (consisting mainly of CO and H2, with small amounts of methane and other gases), which is cleaned of H2S, NH3, HCl, and particulates and then burned in a gas turbine. The hot exhaust gases from the gas turbine are used to generate steam that drives a steam turbine. An important advantage is that the thermal efficiency of a combined-cycle plant exceeds 40 percent, whereas that of a conventional pulverized-coal-fired boiler equipped with flue gas desulfurization is about 35 percent. The development of more advanced gas turbines, capable of operating at higher gas inlet temperatures, will significantly increase the efficiency advantage of combined-cycle plants. IGCC is discussed in more detail later in this chapter in the section "Gasification."

Combined-cycle plants require clean fuel gas for the gas turbine. To maximize efficiency, gas cleanup should be carried out at high temperature. Removing trace amounts of impurities from the gasifier effluent without cooling the gas stream presents a technical challenge. An advantage of combined cycle plants is that, because fuel gas volume is much less than stack gas volume, pollutant removal equipment can be much smaller. Also, the sulfur and nitrogen are in forms that are more easily removed. Another advantage is the modular nature of combined cycle plants, which offers flexibility in construction and the ability to add capacity in increments. The feasibility of combined-cycle plants is being demonstrated in two of DOE's Clean Coal Technology Programs, the Tampa Electric Company IGCC Project and the Wabash River Coal Gasification Repowering Project. These plants are recognized as being two of the cleanest coal-fired power plants in the world.

The use of coal slurries, either coaloil mixtures (COM) or coalwater mixtures (CWM), has been the subject of considerable research because of the potential to replace fuel oil in boilers. Recent work has shifted in favor of CWM over COM, as the cost savings for the latter were not enough to offset conversion costs. CWM is typically composed of 6075 percent coal, 2439 percent water, and 1 percent chemical additives. The advantages of CWM compared with dry pulverized fuel are the ease and safety of handling, transporting (e.g. via pipeline), and injecting the fuel. It is possible that the manufacture of CWM can be integrated into coal beneficiation processes, such as froth flotation. Potential problems from replacing oil with CWM are: (1) limited flame stability and turndown; (2) incomplete combustion; and (3) slagging and fouling caused by inorganic constituents. This technology is still in the development stage. Several demonstration projects have been completed or are under way, with the primary focus being on utilization of waste coal fines from coal-cleaning operations.

17.6 Coke Production

The production of metals frequently requires the reduction of oxide-containing ores, the most important being production of iron from various iron oxide ores. Carbon in the form of coke is often used as the reducing agent in a blast furnace, which is basically a vertical tubular vessel to the top of which is fed alternate layers of iron ore, coke, and limestone. Coke is the porous residual solid remaining after a coal's volatile matter is released at high temperature. Coal cannot be fed directly at the top of a blast furnace, because it does not have the structural strength to support the column of iron ore and limestone in the furnace while maintaining sufficient porosity for the air blast to pass upward through the furnace.

Not all coals can produce coke that is suitable for use in a blast furnace. The property that distinguishes coking coals is their caking ability. Various tests such as maceral analysis, free swelling index, chemical analysis, etc. are performed to identify suitable coals. Frequently, to achieve the necessary coal properties to produce a suitable coke, several coals are blended to generate the feed to the coke plant. Commercial cokemaking processes can be divided into two categories: nonrecovery cokemaking and by-product cokemaking.

Nonrecovery Cokemaking

In nonrecovery coke plants, the volatile components released during cokemaking are not recovered, but rather are burned to produce heat for the coke oven and for auxiliary power production. One of the earliest nonrecovery units was the beehive oven, which for many years produced most of the coke used by the iron and steel industry. With these ovens, none of the by-products produced during coking were recovered.

A beehive oven consists of an old-fashioned beehive-shaped brick chamber. It has a charging hole in the top of the dome and a discharging hole in the oven's wall near the bottom. Coal is charged through the opening in the top and spreads over the oven floor. Retained heat from the previous charge is sufficient to start devolatilization. Evolved gases mix with air entering through the discharge door and burn, with the combustion products exiting through the charging port in the top. The heat generated by the burning gases is sufficient to complete the coking process. The finished coke is removed through the discharge door. Because of their low efficiency and pollution problems, beehive ovens are no longer used in the United States.

The general principle of operation of modern nonrecovery coke plants is the same as for the beehive oven, except that the units are designed to prevent the escape of volatile organic compounds (VOCs) and other pollutants to the atmosphere. Coking takes place in large rectangular chambers that are heated from the top by radiant heat transfer and from the bottom by conduction through the floor. Primary air for the combustion of evolved volatiles is introduced through several ports located above the charge level. Combustion gases exit the chamber through downcomers in the oven walls and enter the floor flue, thereby heating the floor of the oven. Combustion gases from all the chambers collect in a common tunnel and exit via a stack that creates a natural draft for the oven. To improve efficiency, a waste heat boiler can be added before the stack to recover waste heat and generate steam for power production.

At the completion of the coking process, the doors of the chamber are opened, and a ram pushes the hot coke (approximately 2000°F) into a quench car, where it is typically cooled by spraying it with water. The coke is then screened and transported to the blast furnace.

By-Product Coke Production

The majority of coke produced in the United States comes from wet-charge, by-product coke oven batteries. By-product cokemaking consists of the following operations: (1) selected coals are blended, pulverized, and oiled for bulk density control; (2) the blended coal is charged to a number of slot-type ovens, each oven sharing a common heating flue with the adjacent oven; (3) the coal is carbonized in a reducing atmosphere with the off-gas being collected and sent to the by-product plant for by-product recovery; and (4) the hot coke is discharged, quenched, and shipped to the blast furnace.

After the coke oven is charged with coal, heat is transferred from the heated brick walls to the coal charge. In the temperature range from 375 to 475°C the coal decomposes to form a plastic layer near the walls. From 475 to 600°C there is marked evolution of aromatic hydrocarbons and tar, followed by resolidification into semicoke. At 6001100°C coke stabilization occurs, characterized by contraction of the coke mass, structural deve-lopment of coke, and final hydrogen evolution. As time progresses, the plastic phase moves from the walls to the center of the oven. Some gas is trapped in the plastic mass, giving the coke its porous character. When coking is complete, the incandescent coke mass is pushed from the oven and wet or dry quenched prior to being sent to the blast furnace. Modern coke ovens trap the emissions released during coke pushing and quenching, so that air pollution is at a minimum.

By-Product Recovery

Coke-oven gas is released during the coking process. A typical composition for raw coke-oven gas is given in Table 17.9. In addition to the components shown in the table, raw coke-oven gas also contains various contaminants, such as tar vapors, light oil vapors (mainly benzene, toluene, and xylene), naphthalene, ammonia, hydrogen sulfide, and hydrogen cyanide. The by-product plant removes these contaminants so that the gas can be used as fuel. The volatiles emitted during the coking process are recovered as four major by-products: clean coke-oven gas, coal tar, ammonium sulfate, and light oil. The following description represents a typical coke-oven gas cleanup scheme, but considerable variations are possible from plant to plant.

TABLE 17.9 Composition of Coke-Oven Gas

Component	Composition (vol. % (dry))
Hydrogen	54
Methane	25
Nitrogen	10
Carbon monoxide	6
Carbon dioxide	3
Higher hydrocarbons (ethane, propane, etc.)	2

Tar is the first by-product recovered. Hot coke-oven gas at a temperature of about 1100°C is quenched to about 80°C by contact with an aqueous spray (flushing liquor) in the primary cooler. The cooled gas, along with excess flushing liquor, is collected in the coke-oven battery gas-collecting main. The condensed tars and water are transported to a tar decanter, where the heavier tars are separated out, and the water is recycled to the primary cooler. The gas passes through a compressor, called the exhauster, which sucks the gas from the ovens and provides sufficient pressure to allow the gas to flow through the rest of the by-product train. During cooling of the coke-oven gas, some of the tar forms an aerosol, which is carried along with the gas. Since these tar droplets would contaminate and foul downstream equipment, the tar mist is removed in an electrostatic precipitator. The recovered tar is pumped to the tar decanter.

Ammonia removal is the next step in the gas cleanup process. A typical approach is to contact the coke-oven gas with a solution of sul-furic acid to form ammonium sulfate, which is then recovered by crystallization, dried, and sold as fertilizer. More modern processes for ammonia removal include the water wash process, in which the coke-oven gas is scrubbed with water that dissolves the ammonia along with some H_2S and HCN. The scrubber solution is sent to a fractionator to recover pure ammonia, which can be sold or catalytically decomposed. Another option is the PHOSAM Process, developed by U.S. Steel, which absorbs ammonia in a solution of mono-ammonium phosphate and produces anhydrous ammonia for sale.

The purpose of the final cooler is to remove the heat of compression added by the exhauster and to cool the gas to its final temperature so that downstream absorbers will operate more efficiently. Final cooling is typically achieved by direct contact with the cooling medium, either water or oil. An important function of final cooling is removal of naphthalene. In final coolers using wash oil, the naphthalene dissolves in the oil, and a side stream is steam stripped to remove the naphthalene. If water cooling is used, the condensed naphthalene must be absorbed by contacting the water with tar. The tar is continuously exchanged with fresh tar to prevent naphthalene buildup.

The cooled coke-oven gas then passes to the light oil absorber, where benzene, toluene, xylene, and higher aromatics are scrubbed from the gas using a wash oil. The light oil is recovered from the wash oil in a fractionator. The crude light oil can either be sold, or it can be refined to recover individual aromatics, such as benzene. In the past, coke-oven gas was an important source of aromatic compounds, but now they can be more cheaply obtained from petroleum.

The final step is desulfurization. A number of processes have been developed which are suitable for removing H_2S from coke-oven gas. These include using solutions of potassium carbonate, monoethanolamine (MEA), or ammonia to absorb the H_2S. If ammonia solution is used as the absorbent, desulfurization is frequently combined with the ammonia removal step. Recovered H_2S can be converted to elemental sulfur or sulfuric acid. The product remaining after all the above steps is cleaned coke-oven gas, some of which is used to heat the coke ovens and produce more coke with the rest going to the boiler house and/or the blast furnace for direct injection.

The tar and liquor plant handles the flushing liquor that circulates between the by-product plant and the primary cooler. It also processes the wastewater that results from recovery of the coal moisture and chemically bound water in the coal. The flushing liquor flows into tar decanters, where the tar separates from the water and is pumped to storage for later sale. Heavier solid particles separate from the tar and are removed as tar decanter sludge. After withdrawal of a bleed stream that contains ammonia, the aqueous liquor is pumped back to the primary cooler. Following removal of tar particles, the wastewater stream is steam stripped. An alkali, such as sodium hydroxide, is added to decompose ammonia compounds dissolved in the liquor. The ammonia vapor is recovered and combined with the rest of the ammonia recovered from the coke-oven gas.

In the past, many products valuable to industry and agriculture were produced as by-products of coke production, but today most of these materials can be made more cheaply by other techniques. Therefore, the main emphasis of modern coke by-product plants is to treat the coke-oven gas sufficiently so that it can be used as a clean, environmentally friendly fuel. Although coke-oven gas is now generally used in the coke plant or a nearby steel plant, during the

1950s and 1960s a significant contribution to domestic energy use was made by coke-oven gas, which is roughly similar to the gas generated by a moving-bed gasifier.

Direct Coal Utilization in the Steel Industry

Although coke is still required for blast furnace operation, the amount of coke used per ton of hot metal can be significantly reduced by injecting granular or pulverized coal into the furnace along with the hot blast near the bottom of the blast furnace. The coal reacts with oxygen to form H_2 and CO, which act as reducing agents as they rise through the furnace. As long as sufficient coke is used to preserve bed porosity, a portion of the coke can be displaced by injected coal on a roughly pound per pound basis. Reducing coke requirement not only reduces costs, because coal is cheaper than coke, but also reduces the environmental problems associated with coke production. Furthermore, furnace capacity is increased, because the smaller volume of coke charged means a larger volume of iron ore can be charged.

A wide variety of noncoking coals have been successfully tested for direct injection, ranging from lignite to anthracites. The choice depends on price and availability rather than on attaining the highest injection rates. Among coals of similar cost and availability, important factors are ash, volatile matter, moisture content, and grindability. Also, sulfur and alkali contents should be low. Because ash lowers temperature and affects grindabilty, ash should be less than 10 percent. When injecting pulverized coal, volatile matter should be as high as possible, since low volatile coals tend not to burn completely and contribute particulate matter to the off-gas. Moisture should be minimized, as water absorbs heat in the blast furnace, causes difficulties during grinding, and can cause plugging problems in storage vessels and injection equipment. In addition to coal, natural gas, oil, and tar are also injected into blast furnaces.

Another approach receiving increased interest is direct reduction of iron ore. In this process, coal is added directly to the smelting vessel and is the source of the reducing gases and thermal energy. Thus, easily crushed and handled, inexpensive noncoking coals of high calorific value can be used. Particle size requirements are variable, but less than 1 mm is most often used; and, in some cases, there is an effort to restrict moisture content to below 6 percent.

Mild Gasification

Mild gasification is actually a pyrolysis-based process, closely related to coking. It is designed to produce the most profitable slate of alternative fuels and chemicals by decomposing coal at relatively mild conditions of temperature and pressure.20,21 Mild gasification differs from coking in that in mild gasification the goal is the production of liquid and gaseous fuels with the accompanying char being the by-product, whereas in coking the coke is the main product, and the liquids and gases are the by-products. A major problem with this technology is the low quality of the liquid products, which are hydrogen deficient and require expensive upgrading to convert them into gasoline, diesel fuel, and other products. Pyrolysis-based processing for the production of liquids is discussed in more detail later in this chapter under Pyrolysis-Based Processes in the section "Coal Liquefaction."

17.7 Gasification

Coal was first gasified in England by William Murdock in 1792, and the world's first coal gas company was chartered in England in 1812. Coal gas was first produced in America in 1816 in Baltimore, and by 1850, over 55 commercial coal gasification plants in the United States were generating gas for lighting and heating. During the late 1800s and early 1900s, a large number of coal gasifiers operated commercially in the United States and Europe to produce industrial and residential fuel gas. In the early 1930s, there were over 11,000 coal gasifiers operating in the United States, consuming about 13 million tons of coal per year and producing 500 million BTU/day of coal gas.

Most of the early gasifiers were moving bed units, charged with sized coal and blown with steam and air to generate "producer gas" (150 BTU/scf). Operation of the moving-bed gasifiers in a cyclic mode (blowing first with air to heat the coal, followed by contact with steam to produce "water gas") increased the heating value of the product gas to 300 BTU/scf. The heating value of the fuel gas was further increased to about 500 BTU/scf by co-feeding oil with steam in the moving bed process to produce "carburetted water gas," which contained hydrocarbons in addition to H_2 and CO. Coke oven gas was also used as "town gas." An early gasification process, still in use today, was that developed by Lurgi. This gasifier operates at elevated pressure and uses oxygen, instead of air, thus increasing the heating value of the product gas.

By the mid-1950s, the availability of inexpensive natural gas and petroleum had led to the abandonment of most coal gasification units. However, even as oil and natural gas use increased, coal gasification plants continued to be built on a limited scale in certain areas of the world, particularly in countries with a limited supply of petroleum.

Extensive process development was carried out in the United States in the late 1940s to mid-1950s, prompted by a concern that natural gas reserves were limited. This work included (1) development of a 100 ton/day Texaco entrained-flow gasifier at Morgantown, West Virginia, by Texaco and Olin Matheson; (2) demonstration of a 360 ton/day Babcock and Wilcox entrained-flow gasifier at Belle, West Virginia; and (3) development of a 24 ton/day KoppersTotzek entrained-flow gasifier at Louisiana, Missouri. Further work has resulted in the development of several additional gasifiers, including the Shell, KRW, E-Gas™, British Gas/Lurgi, Kilngas, and Dow gasifiers. Recent interest in coal gasification has been driven by the potential of IGCC facilities to increase the efficiency of power production and reduce emissions. Development of improved coal gasification processes increases the potential for implementation of coal gasification technology for a variety of applications.[22]

In modern gasification systems,[23] a bed of coal or other carbonaceous material is blown with a mixture of steam and air or oxygen under pressure to produce a low to medium heating value gas whose main constituents are CO and H_2 with smaller amounts of CO_2, CH_4, and other components. (A mixture of H_2 and CO is called synthesis gas or syngas. Chapter 22 discusses some of the chemicals that can be made from syngas.) Most of the mineral matter (ash) in the feed leaves the bottom of the gasifier either as a dry solid or as a liquid, which solidifies to form an inert glassy slag, suitable for sale or disposal. Only a small fraction of the ash is entrained in the syngas and requires removal downstream. During gasification, sulfur in the feed coal is reduced to H_2S, and nitrogen is reduced to NH_3. These materials, along with HCl and entrained particulates, are removed in the downstream gas cleanup system. A typical gasifier raw gas composition is shown in Table 17.10.

TABLE 17.10 Typical Coal Gasifier Raw Gas Composition

Component	Concentration (vol. %)
Hydrogen	2530
Carbon monoxide	3060
Carbon dioxide	515
Water	230
Methane	05
Hydrogen sulfide	0.21
Carbonyl sulfide	00.1
Nitrogen	0.54
Argon	0.21
Ammonia/hydrogen cyanide	00.3

Chemistry of Coal Gasification

The initial step in coal gasification involves grinding and/or pretreatment of the coal to put it into a form suitable for injection into the gasifier. In the gasifier, the coal is heated in the presence of a reactive gas whose composition depends on the product desired. Walker[24] has presented an extensive review of gasification reaction fundamentals.

Reactions occurring during gasification of coal can be divided into three groups: (1) pyrolysis reactions (thermal decomposition of the coal), (2) gasification reactions (gassolid reactions), and (3) gasgas reactions. The major reactions taking place are listed in Table 17.11.

TABLE 17.11 Reactions Occurring during Coal Gasification

	ΔHr kcal/gmol	Name	Number
Pyrolysis reactions			
$CH_x = C + CH_4$		Pyrolysis	(17.1)
$CH_x + H_2 = C + CH_4$		Hydropyrolysis	(17.2)
Gassolid reactions			
$C + 0.5O_2 = CO$	-26.45	Combustion	(17.3)
$C + O_2 = CO_2$	-94.1	Combustion	(17.4)
$C + H_2O = CO + H_2$	41.9	Steam/carbon	(17.5)
$C + CO_2 = 2CO$	41.25	Reserve boudouard	(17.6)
$C + 2H_2 = CH_4$	-17.9	Hydrogasification	(17.7)
Gasgas reactions			
$CO + H_2O = H_2 + CO_2$	-0.7	Watergas shift	(17.8)
$CO + 3H_2 = CH_4 + H_2O$	-59.8	Methanation	(17.9)

The first reaction that occurs is pyrolysis or devolatilization (Reaction 17.1). Reaction 17.2 represents the conversion of gaseous pyrolysis products into predominantly methane and coke, the extent of this reaction depending upon gasifier design. Gasification is related to pyrolysis/coking in that, when coal is gasified by reacting it at a high temperature with steam and an oxidant (air or oxygen), pyrolysis is the first reaction as the coal is heated to reaction temperature. Depending on the type of gasifier, condensable hydrocarbons may be collected as a by-product or may be completely destroyed.

Combustion reactions, Reactions 17.3 and 17.4, are the fastest gasification reactions and are highly exothermic. The oxygen can come from air, enriched air, or pure oxygen. In practice, the oxidation reactions occur rapidly with essentially complete oxygen consumption, so that most of the reactor volume is utilized for the other reactions listed in Table 17.11. The primary combustion products are CO and CO_2. The equilibrium constants for both Reactions 17.3 and 17.4 are greater than 1 up to 3,500°C; and, as a result, both reactions are essentially irreversible at typical gasifier temperatures. The CO/CO_2 ratio should increase with lower pressure and higher temperature, because pressure does not influence the equilibrium of Reaction 17.4, and Reaction 17.3 is not favored under pressure.

Reactions 17.5, 17.6, and 17.7 illustrate the gasification of char by reaction with various gases. The carbonsteam Reaction 17.5 is an endothermic reversible reaction. Steam undergoes a side reaction, Reaction 17.8, called the watergas shift reaction. This reaction, which is very rapid, is catalyzed by various impurities and surfaces. The carbonCO_2 reaction, Reaction 17.6, is favored at high temperatures and low pressures, whereas the carbonH_2 reaction, Reaction 17.7, is favored at low temperatures and high pressure. Since only three of Reactions 17.517.9 are independent, if the equilibrium constants for Reactions 17.6, 17.7, and 17.8 are known, the equilibrium constants for the other reactions can be calculated.

The reactions in Table 17.11 are useful for examining the effect of changes in gasifier operating conditions on product gas composition. As temperature is increased, endothermic reactions are favored over exothermic reactions, which means that methane production will decrease and CO formation will increase. An increase in pressure favors reactions in which there are fewer moles of products than moles of reactants. Thus, higher pressure favors methane and CO_2 formation. Coal char gasification is generally equilibrium controlled. For design purposes, it is common to consider char as thermodynamically equivalent to graphite, although this is not strictly true.

Even when gasifier operating conditions are known, it is not possible to calculate the composition of the product gas exactly. First, the pyrolysis reactions and those of the pyrolysis products are highly dependent on coal pro-perties and gasifier operating mode and vary significantly from case to case. Second, the thermodynamic equations treat coal char as if it were pure carbon, which is not the case. Finally, the system may not reach equilibrium. Thus, although useful predictions of gas composition can be made, exact calculations are not possible. Nevertheless, knowledge of stoichiometry, equilibrium constants, and reaction rates provides a reasonable basis for the modeling and design of gasification systems.

Pore structure can markedly affect char reactivity. Coals in general are highly porous with a polymodal pore size distribution. Pores normally are classified into macropores (>500 Å in diameter), transitional pores (20500 Å in diameter), and micropores (<20 Å in diameter). Upon pyrolysis, the pores in the coal open up, but the, char still contains microporosity. Coal chars in general, and lignitic chars in particular, retain coal's polymodal pore distribution. Surface areas of coal chars are in the range 100800 m2/g. Most of this surface area and, therefore, most of the active surface reside inside the pores, so accessibility of reactive gases to active sites is very important.

The fundamental reactions occurring during gasification can be described by the LangmuirHinshelwood and RidealEley mechanisms. The LangmuirHinshelwood mechanism involves three steps: (1) adsorption of the gas onto the solid surface; (2) surface migration/reaction; and (3) desorption of the products from the solid surface. In the RidealEley mechanism, the basic steps are (1) reaction between gas molecules and surface atoms by direct collision and (2) desorption of products.

Because coal chars are highly microporous, most of the gasification reactions take place inside the char particles. Therefore, diffusion of gas into, and products out of, porous particles is required. The overall diffusion process can be described by the following steps: (1) diffusion of the reactant from the bulk gas to the solid surface (film diffusion); (2) diffusion of the reactant from the particle's surface to its interior (internal diffusion); (3) diffusion of the product from the interior to the particle's surface (internal diffusion); and (4) diffusion of the product from the surface to the bulk gas (film diffusion).

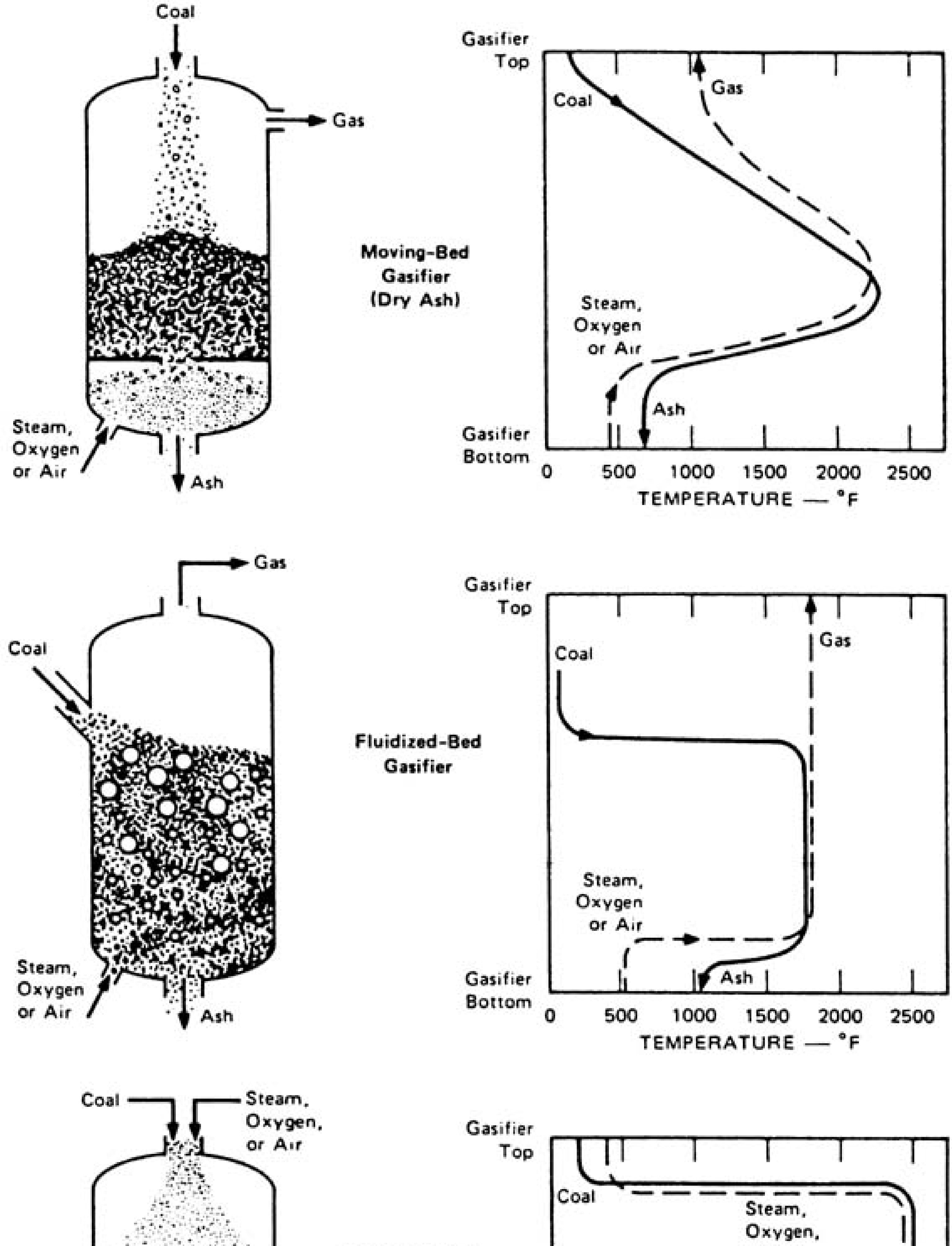
If the pore diameter and process conditions are well defined, the rates of internal and film diffusion can be calculated. The temperature dependency of the rate can be presented in the form of an Arrhenius plot, that is, log rate vs. reciprocal absolute temperature. Gasification rates can be divided into three zones, depending on whether reaction rate is controlling, diffusion is controlling, or both reaction rate and diffusion are important. To determine the overall gasification rate, it is necessary to model the actual gasifier.

Because of the reducing nature of the product gas, heteroatoms (sulfur, nitrogen, and chlorine) appear in reduced form, that is, sulfur appears as hydrogen sulfide, nitrogen as ammonia, and chlorine as hydrogen chloride. In most cases, these materials are scrubbed from the product gas before it is burned. Ammonia and HCl are very water soluble and are easily removed by a water wash. A number of processes have been developed for H_2S removal; many of these process are based on absorption in solutions of amines, such as monoethanolamine (MEA).

Types of Coal Gasifiers

Gasification processes can be separated into three major types: (1) moving-bed (countercurrent flow) reactors; (2) fluidized-bed (back-mixed) reactors; and (3) entrained-flow (not back-mixed) reactors. Figure 17.11 shows the types of gasification reactors together with temperature profiles and locations of feed and product streams. Table 17.12 summarizes the important characteristics of each type of gasifier, and Table 17.13 presents the performance characteristics of selected gasifiers.

Fig. 17.11 Generic coal gasification reactors. (*Source: Electric Power Research Institute and Synthetic Fuels Associates, Inc.*)



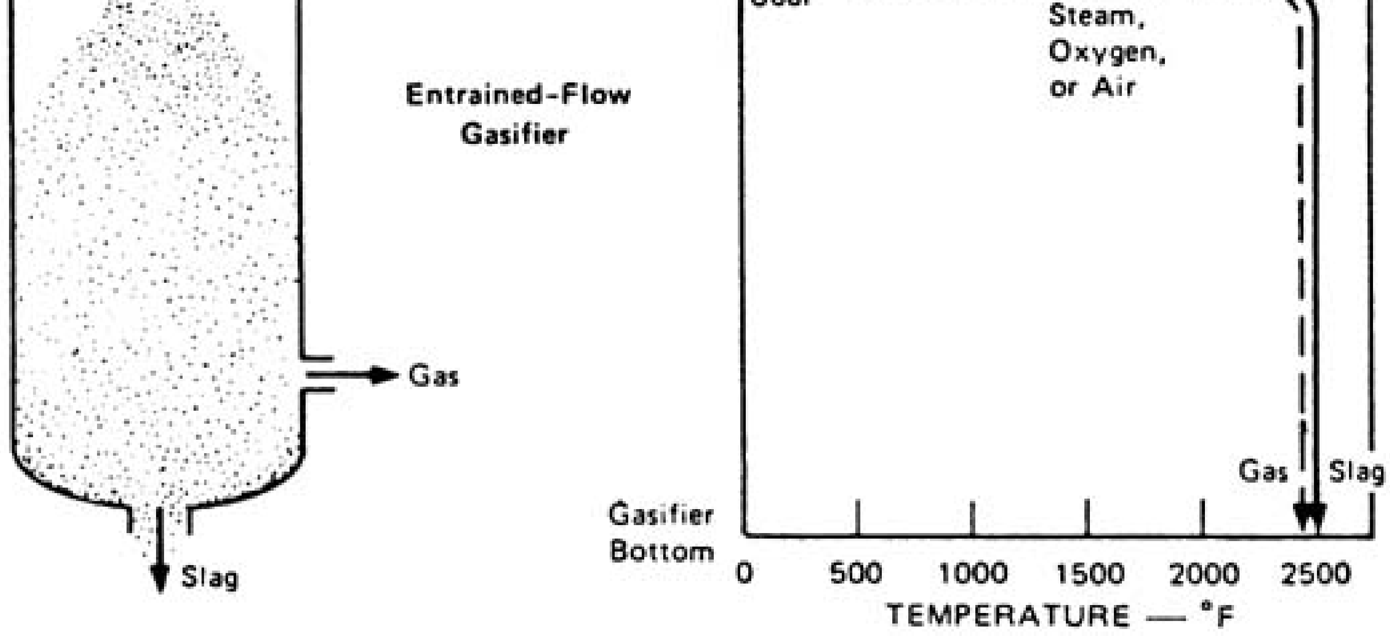


TABLE 17.12 Features of Fixed-Bed, Fluidized-Bed, and Entrained-Bed Gasifiers

Parameters

Moving bed	*Fluidized bed*	*Entrained bed*
Bed of coal descending under the influence of gravity	One or more fluidized bed zones	Upflow or downflow suspension gasification
Mechanical grates/distributors	Uniform temperature and composition throughout each fluidized zone Moderate process response	High temperaturehigh rate process Fast process response
Discrete zones:	1. Preheating-drying devolatilization 2. Gasification 3. Combustion	

Advantages

Moving bed	*Fluidized bed*	*Entrained bed*
High carbon conversion efficiency	High degree of process uniformity	Handles all types of coal without pretreatment
Low ash carryover	Excellent solids/gas contact	Low steam consumption
Lower temperature operation	Lower residence time than fixed bed gasifier	Excellent solidsgas contact
Lowest air/oxygen requirement	Higher coal throughput per unit volume of reactor	No tar formation No phenol formation Ability to slag ash High capacity per unit volume of reactor Produces inert slagged ash

Disadvantages

Moving bed	*Fluidized bed*	*Entrained bed*
Sized coal required	Sized coal required	Requires finely crushed coal
Coal fines must be briquetted	Dry coal required for feeding	Small surge capacity requiring close control
Low off-gas temperature	Requires complicated gas distributor	

TABLE 17.13 Performance Characteristics of Selected Gasifiers

	Lurgi Dry Ash	Lurgi Slagger	Texaco	Shell	KRW PDU	KRW Design Estimate	U-Gas
Parameters	Illinois No. 6 coal		Pittsburgh No. 8 coal				West Kent Coal
T, °F	1600	2700	2700	3000	1793	1850	1831
p, psia	315	300	600	365	230	600	15
H_2Ocoal, molemole	2.42	0.293	0.027	0	13	0.29	1.34
O_2coal, molemole	0.286	0.259	0.463	0.435	0.44	0.30	0.45
C conversion, %	99.3	99.E	99	99.3	76.0	89.6	93.3
Gas composition (dry)							
CO	15.36	58.05	51.69	61.46	43.38	51.5	26.1
CO_2	31.1	1.94	10.6	1.65	35.30	9.3	23.4
H_2	42.9	30.41	35.1	30.6	18.04	25.9	37.4
CH_4	3.78	7.76	0.09	0.04	1.33	10.4	2.8

Source: Shinner. R., Fourth Technical Meeting of COGARN, presented in "Coal Gasification: Direct Applications and Synthesis of Chemicals and Fuels: A Research Needs Assessment," by Penner, S. S., and others for US DOE.24

Moving-Bed Gasifiers

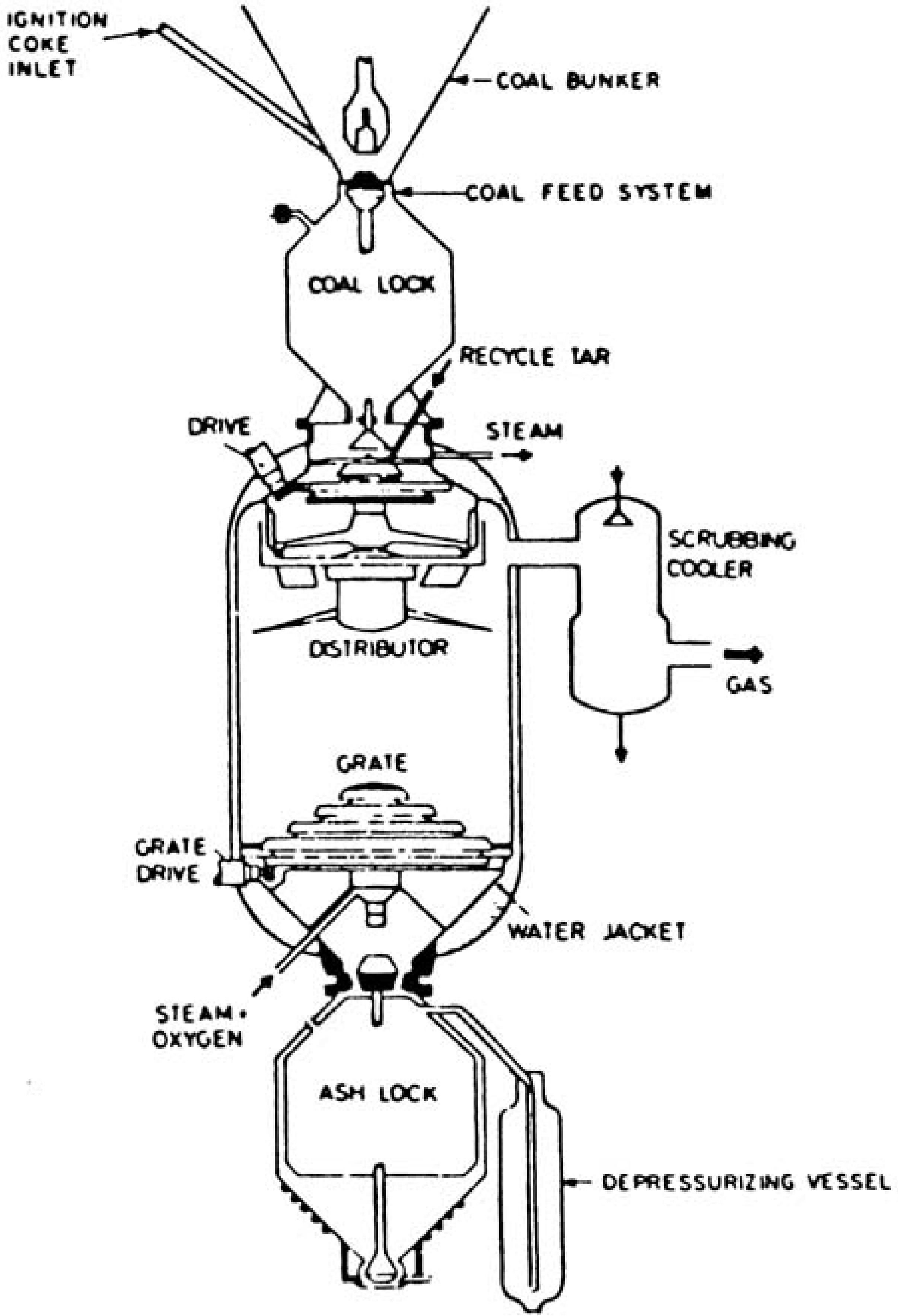
Moving-bed gasifiers consist of a downward-moving bed of coal contacted by a countercurrent flow of gases moving upward through the bed. In moving down the bed, coal sequentially undergoes drying, devolatilization, gasification, combustion, and ash cooling. Moving-bed gasifiers can be operated at atmospheric pressure or can be pressurized with either air or oxygen as the oxidant, with either dry ash or molten ash removal, and with or without stirrers to prevent agglomeration.

In the drying zone at the top of the bed, the hot upward-flowing gases remove moisture from the coal. As the coal moves down the bed, its temperature increases, and at about 600900°F pyrolysis occurs, liberating gases, oils, and tars. As the devolatilized coal (char) further descends, it is gasified by reaction with steam, carbon dioxide, and hydrogen to produce a mixture of carbon monoxide, hydrogen, and methane, containing unreacted steam and carbon dioxide. Below this, char is combusted by reaction with oxygen. The final composition of the gas produced and the rates of reaction are strongly influenced by gasifier temperature and pressure.

The ratio of steam to oxygen (air) in the gasifier controls the peak temperature in the combustion zone. If a nonslagging operation is desired, sufficient steam is added to offset exothermic oxidation reactions with endothermic steamcarbon reactions to stay safely below the ash-fusion temperature. Slagging gasifiers operate at a higher temperature, and ash is removed in a molten state and then quenched in a water bath.

Moving-bed gasifiers require sized coal for proper operation; typically, coal between $\frac{1}{4}$ in. and 2 in. in diameter is used. The countercurrent flow of gases and solids leads to high thermal efficiency, with exit gas temperature usually in the range of 5001,000°F. The raw product gas contains fines and tar/oil produced during gasification. These materials are removed prior to downstream processing. Figure 17.15 illustrates the mechanical configuration for a Lurgi dry bottom gasifier.

Fig. 17.15 Lurgi dry bottom gasifier. *("Comparative Study of Coal Gasification ProcessesKoppersTotzek, Lurgi, and Winkler," Koppers Co., Pittsburgh, PA.)*



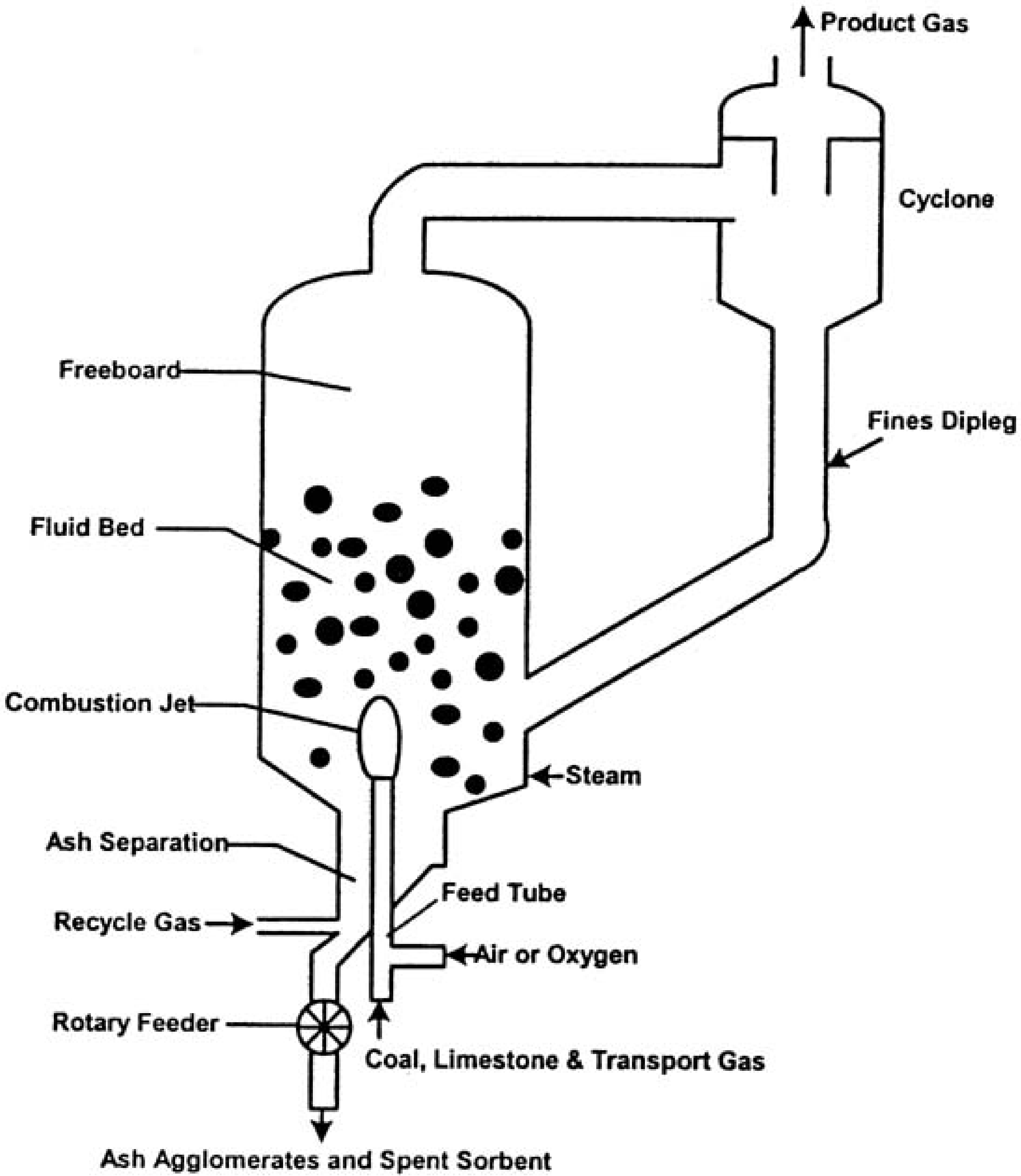
Fluidized-Bed Gasifiers

In a fluidized bed gasifier, reactant gases are introduced through a distributor at the bottom of the bed at a velocity sufficient to suspend the feed particles. The result is a bed of highly mixed solids in intimate contact with the gas phase. The agitation leads to a uniform temperature throughout the bed. The high degree of mixing results in reaction rates that are typically higher than those experienced in moving-bed gasifiers, although fluidized-bed gasifiers generally do not achieve as high a thermal efficiency as moving-bed gasifiers.

The exit gas temperature for a fluidized-bed gasifier is higher than that for a moving-bed gasifier. The higher temperature results in further reaction of the pyrolysis products, and, therefore, the product gas contains a much lower concentration of tar/oil. However, particle attrition generally leads to higher levels of fines/dust carryover. Unconverted char and ash are removed as dry solids. Pretreatment is required if strongly caking coals are used.

Fluidized-bed gasifiers can be operated at atmospheric or higher pressure. Fluidizing gases can be mixtures of steam with either air or oxygen and, in some cases, mixtures of hydrogen with other gases. The KRW gasifier is illustrated in Fig. 17.18.

Fig. 17.18 KRW gasifier.

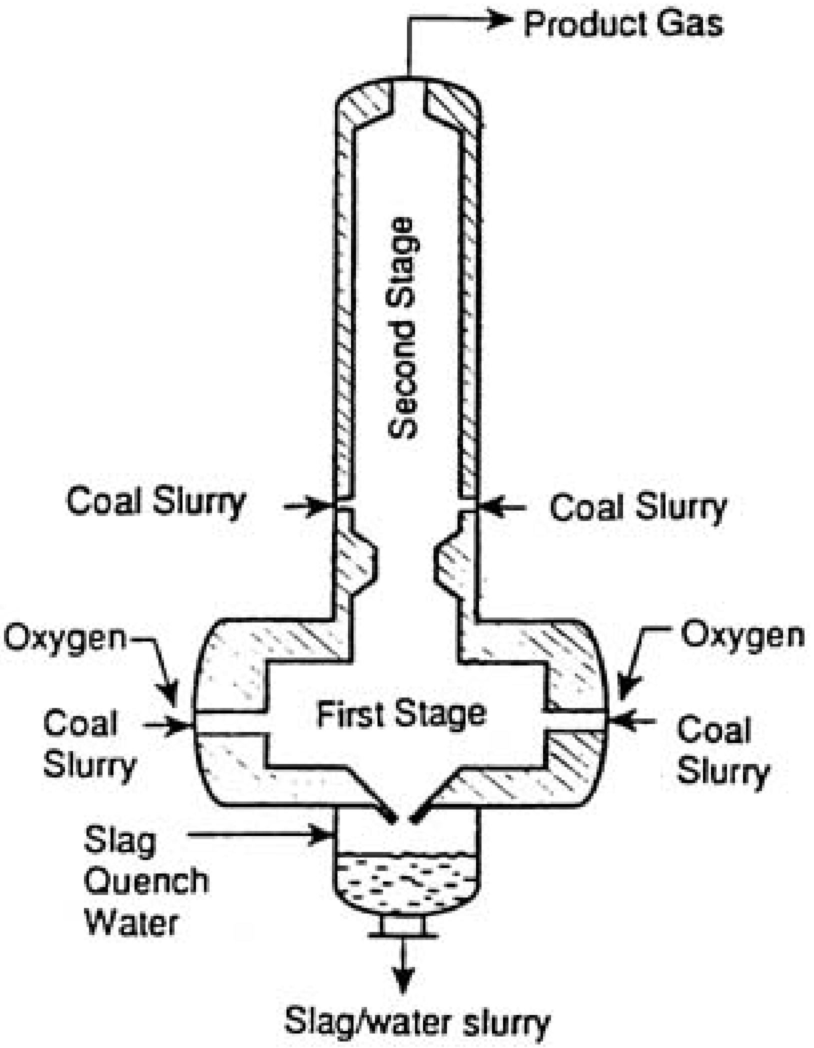


Entrained-Flow Gasifiers

In an entrained-flow gasifier, a mixture of finely ground coal entrained in a reactant gas flows co-currently through the reactor, with little or no back-mixing. This type of gasifier may be either single-stage or two-stage.

In general, high temperatures (22003200°F) are used to achieve complete gasification of the coal in a mixture with steam and oxygen or air. Because of the high temperatures employed, gasification rates are considerably higher than those achieved with either fixed-bed or fluidized-bed gasifiers. Entrained-flow gasifiers can handle all coals, including strongly caking coals, without pretreatment. The high temperature of operation produces a gas devoid of both methane and tars/oils. In two-stage gasifiers, the incoming coal is first entrained with reactant gases to produce gas; the resultant char is gasified further in a second stage, which may or may not be entrained. As is the case with fluidized-bed gasifiers, a staged operation achieves better overall thermal efficiency without sacrificing higher throughput, because the more reactive incoming coal can be gasified at a lower temperature than the less reactive char. Entrained-flow gasifiers can be operated at atmospheric pressure or pressurized, and ash may be removed either dry or molten. Figure 17.17 illustrates the two-stage entrained-flow E-Gas™ gasifier.

Fig. 17.17 E-Gas™ gasifier. (*Source*: Dastec Energy.)

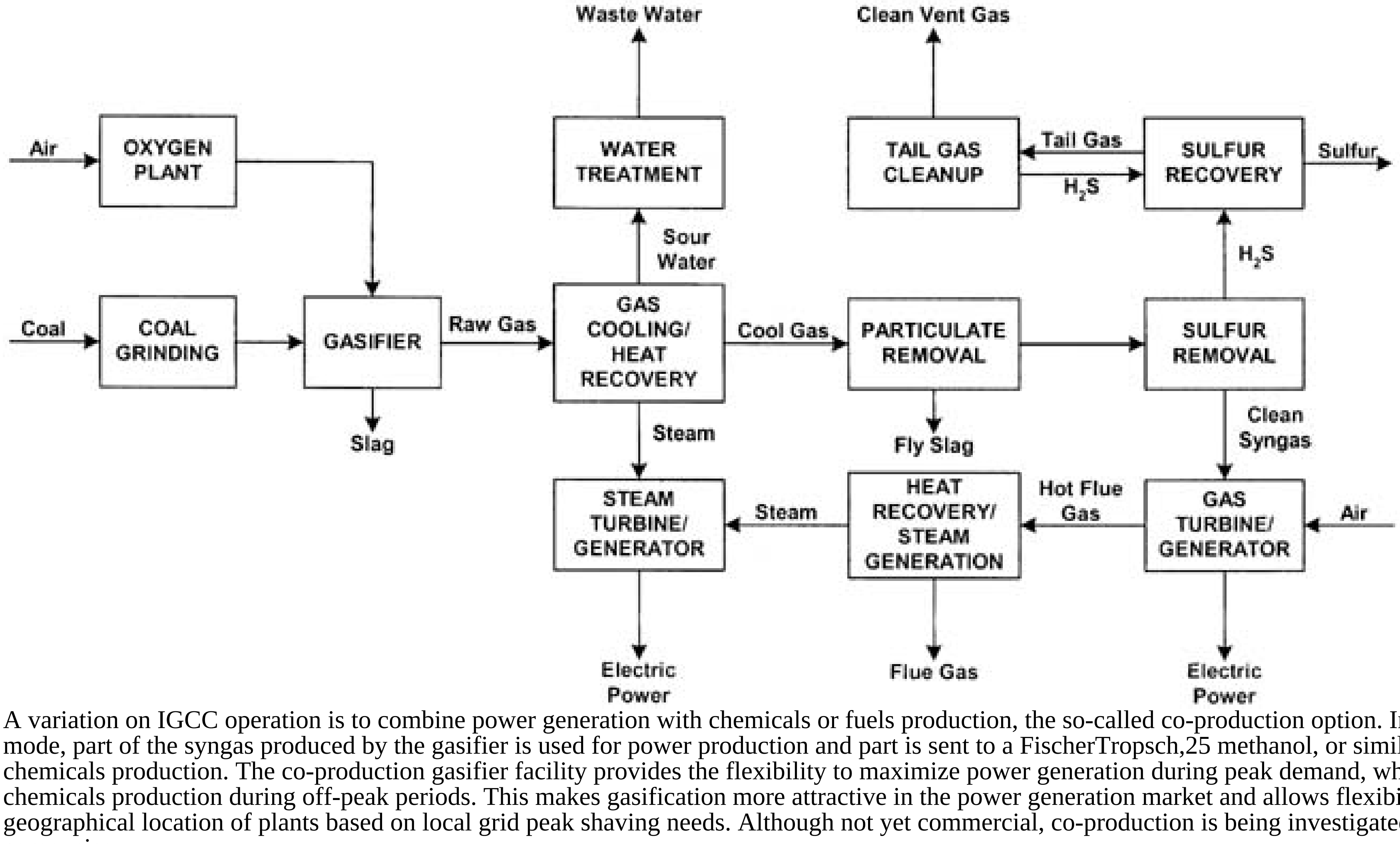


Gasification for Power Generation

High reliability, acceptable capital and operating costs, and minimal environmental impact are requirements for gasification systems proposed for utility applications. Operating costs can be minimized by using a gasifier that is capable of operating on run-of-mine coal, including caking coals and fines. The potential also exists to feed other materials such as petroleum coke, lignite, sewage sludge, refuse derived fuel (RDF), biomass, etc. to the gasifier. Use of such opportunity fuels can lower electricity costs.

Numerous studies confirm that gasifiers coupled with gas turbinesteam turbine combined cycles represent one of the most promising technologies for future coal-based power generation systems. IGCC technologies offer the potential for high efficiencies with low pollutant emissions. High efficiencies are achieved in IGCC operation by combining efficient combustion turbines with steam turbine bottoming cycles. The gas turbine converts high-temperature (3300°F) heat efficiently, while the steam turbine utilizes lower-temperature heat (1050°F) in the form of steam generated by recovering heat from the gas turbine exhaust. A typical IGCC facility is illustrated in Fig. 17.12.

Fig. 17.12 Block flow diagram of a typical integrated gasification combined cycle power plant.



A variation on IGCC operation is to combine power generation with chemicals or fuels production, the so-called co-production option. In this operating mode, part of the syngas produced by the gasifier is used for power production and part is sent to a FischerTropsch,25 methanol, or similar facility for fuels/ chemicals production. The co-production gasifier facility provides the flexibility to maximize power generation during peak demand, while maximizing fuels/ chemicals production during off-peak periods. This makes gasification more attractive in the power generation market and allows flexibility in the geographical location of plants based on local grid peak shaving needs. Although not yet commercial, co-production is being investigated by several energy companies.

A major goal of power production is minimal environmental impact. Since the product gas from IGCC systems is purified before combustion,26 burning this clean fuel results in low pollutant emission levels. Another potential environmental problem is eliminated by designing and operating the gasifier so that tars are not formed. Ash leaving the system is usually in the form of molten slag, which is water quenched to form benign vitreous material suitable for use or disposal. On balance, coal gasification systems are environmentally superior to other alternative coal utilization technologies and can meet rigorous environmental standards for SO_2, NO_x, and particulates. Furthermore, because of their increased efficiency, IGCC plants emit less CO_2, the major greenhouse gas, per unit of electricity generated.

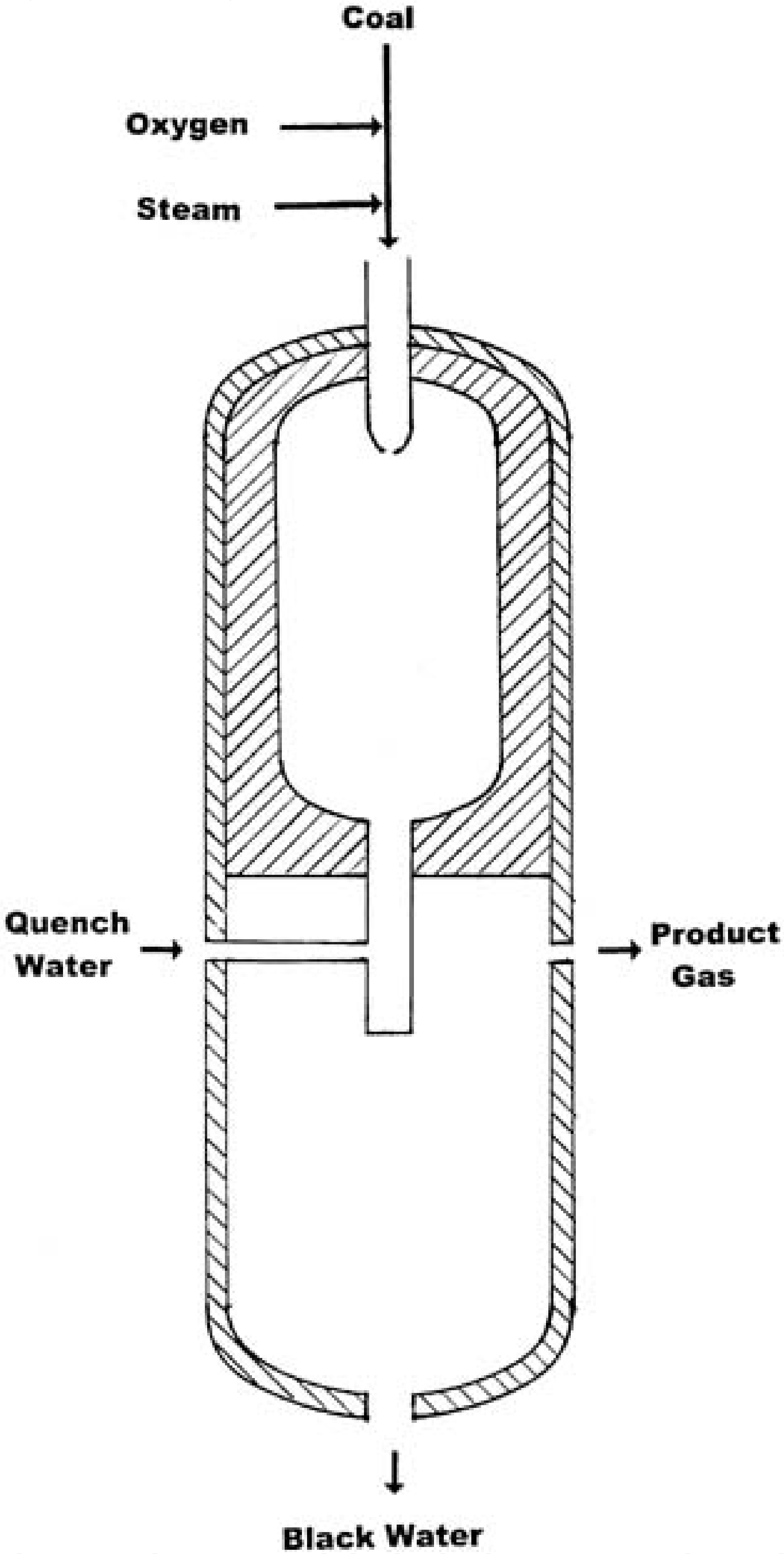
Descriptions of Selected Gasification Processes

Many different gasifiers have been developed, at least through the demonstration stage, by a variety of organizations, both public and private. However, not all of these gasifiers have achieved commercial success, and improved processes now supercede some technologies that were widely used in the past. In 1999, the Texaco, Shell, and Lurgi (dry ash) processes accounted for over 75 percent of the installed and planned coal gasification capacity, with Texaco being the leader with almost 40 percent of installed capacity. It is beyond the scope of this chapter to describe all existing gasifiers. The following processes cover a wide variety of gasifier types and designs.

Texaco Coal Gasification Process

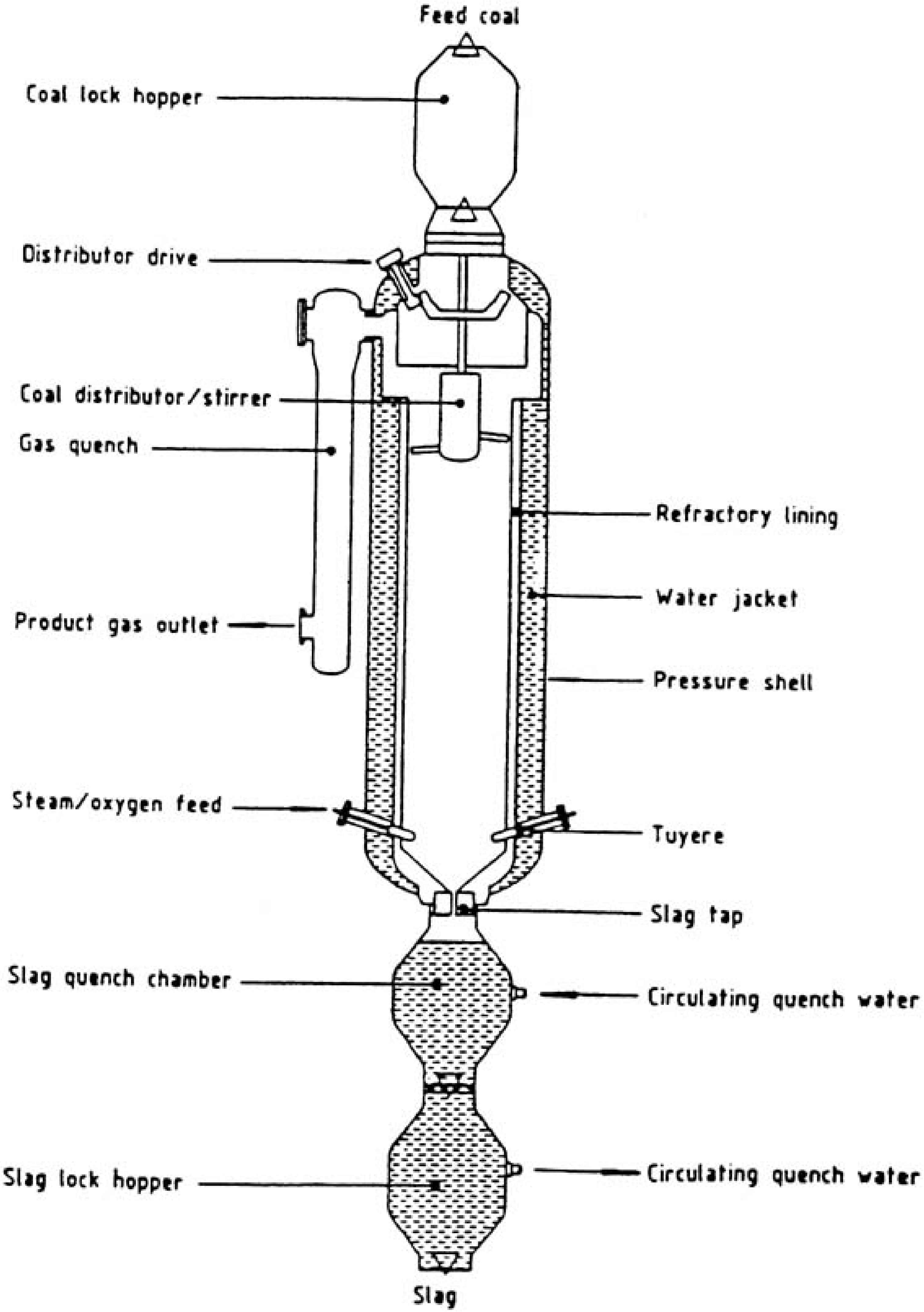
The Texaco Coal Gasification Process (TCGP) offers the following design features: a down-fed, pressurized, entrained-flow, slagging reactor; air- or oxygen-blown; high operating temperature; flexible feedstocks and products; and system coupling for cogeneration. The TCGP is designed to operate at pressures between 20 and 30 atm and temperatures between 1200 and 1500°C. Figure 17.13 shows a direct quench gas cooling configuration for the TCGP. The coal is wet-ground and mixed with water to make a slurry containing 6070 wt. percent solids. The slurry is mixed with oxygen or air in the gasifier burner. By properly adjusting the oxygen/slurry ratio, gasifier temperature is maintained above the ash fusion temperature.

Fig. 17.13 Texaco gasifier (direct quench mode).



The raw product gas is cooled either by direct contact with quench water, as in Fig. 17.16, or by passage through a radiative cooler followed by a convective cooler. The cooled gases go to a scrubber to remove entrained char and ash. The product gas leaving the scrubber contains H2, CO, CO2, H2O, and traces of Ar, N2, CH4, H2S, and COS. Spencer and others have reported on the performance of the TCGP.27,28

Fig. 17.16 British Gas/Lurgi slagging gasifier. (Source: British Gas.)



Various products have been prepared in commercial plants employing Texaco gasifiers, including H2, NH3, fuel gas for electricity generation, and syngas for production of methanol and other oxo products. The direct water quench cooling method is preferred when the desired final product is NH3 or H2, as the required shift reactor is easily integrated into the quench train.

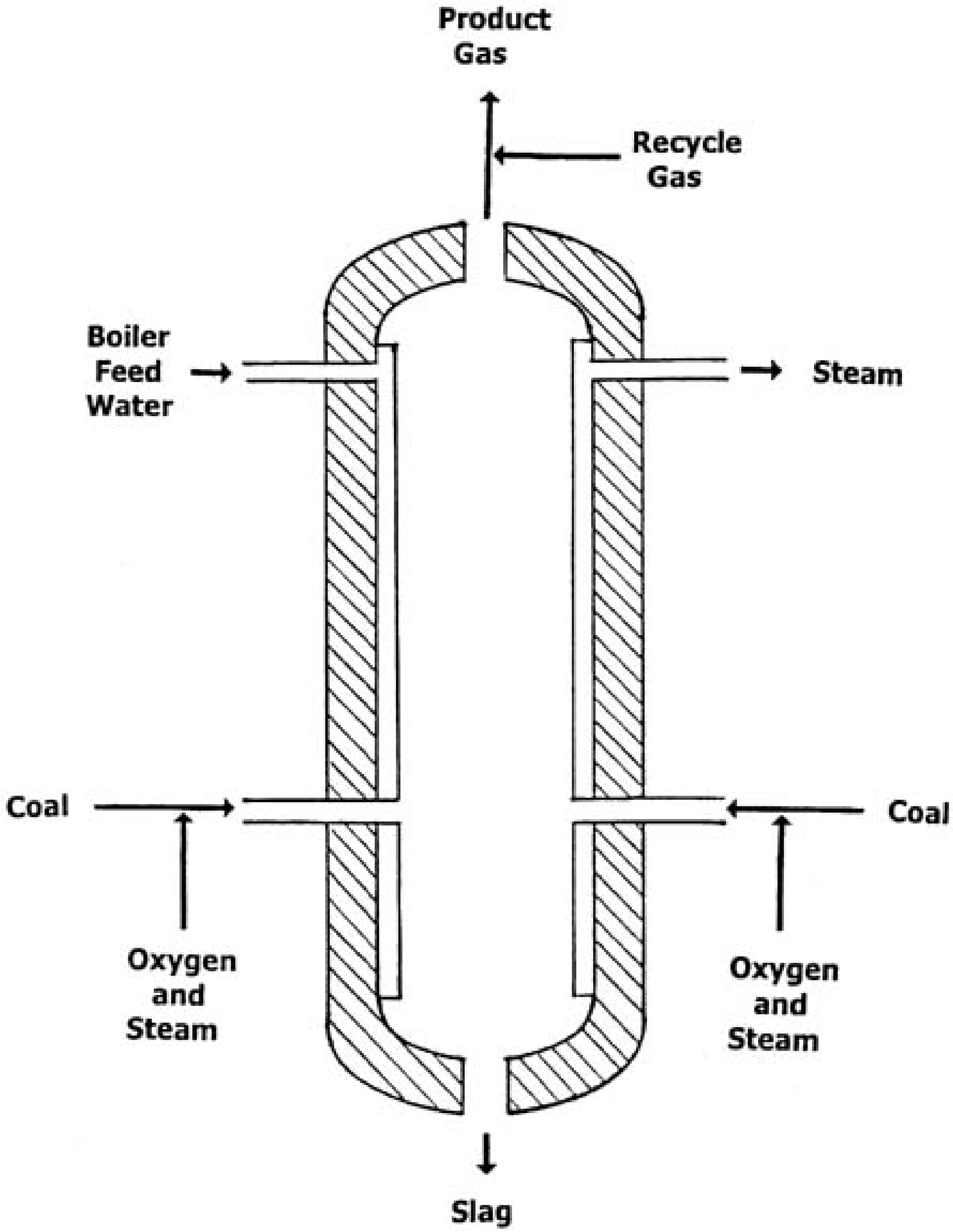
Any carbonaceous material that can be processed into a concentrated, pumpable slurry can be gasified in a TCGP. For example, petroleum coke is converted to syngas in a 30-ton/day plant in Ube City, Japan, which has been operating since 1982. In 1991, a mixture of coal and sewage sludge was successfully gasified. The ability of the Texaco gasifier to handle a wide range of feedstocks without the need for additional equipment or changes in operating procedures ensures flexibility in selecting the least expensive feed alternative while maintaining high availability and reliability.

Shell Gasification Process

Shell's experience with coal gasification started with operation of a 6-ton/day pilot plant in Amsterdam. This was followed by a 150-ton/day unit in Hamburg, Germany, and a 250400-ton/day unit in Houston, TX. This led to the successful operation of a 2000-ton/day commercial unit.

The Shell gasifier (Fig. 17.14) consists of a pressure vessel containing a gasification chamber whose inner wall consists of closely spaced tubes. The wall temperature is controlled by circulating water through these tubes to generate saturated steam. Coal is ground (90% less than 100 µm) and dried in an inert atmosphere and pneumatically transported to the coal pressurization and feeding system. Coal, oxygen, and steam are injected into the gasifier through pairs of opposed burners. Oxygen purity is typically 95 percent, but other purities can be used, depending on the application. Gasifier pressure is 2040 bar, and temperature is above 1500°C.

Fig. 17.14 Shell gasifier.



Most of the mineral content of the feed coal leaves the gasification zone in the form of molten slag. The high operating temperature ensures that the molten slag flows freely down the tube walls and into a water bath at the bottom of the gasifier, where it solidifies into dense glassy granules that are washed, depressured, and sent to storage or disposal. Greater than 99 percent carbon conversion is obtained. Because of the high temperature, no hydrocarbons heavier than methane exist in the raw gas. The insulation provided by the molten slag layer minimizes heat loss.

The raw gas leaving the gasifier is quenched with recycled cooled product gas to solidify any entrained slag droplets. The gas then enters a syngas cooler that generates high-pressure steam. The cooled gas flows through filters or cyclones for fly slag removal. The recovered fly slag can be recycled via the coal feed system to enhance gasification efficiency. The syngas goes to a scrubber for removal of remaining solids and water-soluble contaminates and then to an acid gas removal system. Recovered H2S is converted to elemental sulfur. Overall efficiency (including high pressure steam production) is very high.

The Shell process can handle a wide variety of coals, varying from lignite to bituminous, in an environmentally acceptable way and produce a high-purity, medium heating value gas for use as a fuel for power generation, as a chemical feedstock, or as a source of hydrogen.

Lurgi Dry Ash Gasifier

The Lurgi Dry Ash Coal Gasification Process (Fig. 17.15) is among the earliest commercially successful gasifiers. The first commercial plant was built in 1936. Since then numerous plants have been built. In particular, Lurgi gasifiers provide the syngas for the Sasol indirect coal liquefaction plants in South Africa and also for the Dakota Gasification synthetic natural gas plant.

Sized coal enters the top of the bed by way of a lock hopper and moves down through the bed. Ash is withdrawn by a rotating grate and is removed from the gasifier through another lock hopper. Steam and oxygen enter the bottom of the gasifier through the grate and react with the coal as the gas moves up through the bed. As the coal passes down through the gasifier vessel, it undergoes in turn drying, devolatilization, pyrolysis, gasification, and combustion. This countercurrent operation results in a temperature drop across the reactor. Temperature in the combustion zone near the bottom is in the range of 2000°F, whereas gas temperature leaving the drying and devolatilization zone near the top is typically 5001000°F. The main factor affecting exit temperature is the moisture content of the feed coal. For feed coals high in moisture, such as lignite, the gasifier exit temperature is at the lower end of the range, and for a low moisture bituminous coal it is at the upper end.

The raw gas is quenched directly with recycle water to condense the tar it contains. The gasifier vessel is cooled by water jacketing, which generates a portion of the steam that is used in the gasifier. Sufficient steam is injected to keep the temperature at the bottom of the gasifier below the melting temperature of the ash. The quenched gas is sent to cleanup for removal of NH3, H2S, and other impurities. After cleanup, the gas has a higher methane and hydrogen content than the gas from other gasifiers. Because of the lower exit temperature, oxygen requirement is reduced. However, recovery and disposal of tar can be a problem. High-reactivity, low-rank coal with a high ash softening temperature and minimal fines is favored.

British Gas/Lurgi (BGL) Slagging Gasifier

The BGL slagging gasifier technology (Fig. 17.16) is very similar to the conventional dry ash Lurgi gasifier, the key difference being that the BGL removes ash as molten slag. Advantages of slagging the ash include a significantly lower steam requirement and improved utilization of high-rank coals and coal fines. Slagging operation also eliminates liquid by-products by recycling them.

Like the Lurgi dry ash gasifier, the BGL gasifier is a high-pressure, moving-bed, oxygen-blown unit. Coal with a size distribution of 2 in. × 0 in. (the amount of minus $\frac{1}{4}$-in. material allowed depends on the coal's swelling characteristics) and/or briquettes are fed to the top of the bed via a lock hopper system. A rotating internal fuel distributor rakes and levels the top of the bed. The coal is gasified as it passes down through the reactor. Ash is removed from the bottom of the gasifier as molten slag through a slag tap, then quenched in water and removed by a lock hopper. Oxygen and steam are injected through tuyeres near the bottom of the bed.

As with the Lurgi dry ash gasifier, there is a large temperature difference between the top and bottom, the magnitude of the difference depending mainly of the moisture content of the feed coal. For bituminous coals, this temperature difference is about 10001200°F. The raw gas leaving the gasifier is directly quenched to remove solids and condensable liquids. The solids entrained in the raw product gas, along with hydrocarbon by-products, such as tars and oils, naphtha, and phenols, are recycled to the top of the gasifier and/or reinjected into the gasifier at the tuyeres, where they are gasified. Coal fines can also be fed through the tuyeres, either pneumatically as dried solids or slurried with wastewater.

E-Gas™ Process

The first large-scale experience with the E-Gas™ gasifier (formerly the Destec gasifier) was gained in more than 34,000 hr of operation at Louisiana Gasification Technology, Inc. from April 1987 through November 1995. This was followed by construction of an E-Gas™ gasifier at PSI Energy's Wabash River Generating Station in Indiana as part of DOE's Clean Coal Technology program. This plant started operations in 1995 and has the capacity to gasify about 100 ton/hr of bituminous coal.

The E-Gas™ Gasifier consists of two stages: an entrained-flow, slagging first stage, and a nonslagging second stage, as shown in Fig. 17.17. The first stage is a horizontal, refractory-lined vessel in which carbonaceous fuel is partially combusted with oxygen at an elevated temperature and pressure (2600°F and 400 psia). The gasifier is initially heated with natural gas and then switched to the feed slurry, produced by wet grinding of coal. Oxygen and preheated slurry are fed to two opposed mixing nozzles, one on each end of the horizontal section of the gasifier. The oxygen feed rate is controlled to maintain the gasification temperature above the ash fusion point. Under gasifier conditions, the feed is almost completely converted to a gas consisting primarily of H2, CO, CO2, and H2O.

Coal ash and any added flux form a molten slag that flows continuously through a tap hole in the floor of the horizontal section into a quench water bath, where the slag solidifies. A continuous removal system transfers the slag to a dewatering bin, where the bulk of the slag settles out. The overflow goes to a settler for removal of the remaining fines. To improve carbon utilization, the fine particles are recycled to the slurry preparation area.

Raw syngas from the gasifier's first stage flows up into the vertical, refractory-lined, second stage, where additional slurry, but no additional oxygen, is injected. The injected fuel undergoes devolatilization and pyrolysis that both cools the syngas (because these reactions are endothermic) and increases its heating value (because of the nature of the products produced). In addition, water reacts with some of the carbon to produce CO and H2. Unreacted char is carried overhead with the syngas that leaves the gasifier at 1900°F.

The hot syngas is cooled to approximately 700°F in the syngas cooler, a vertical firetube heat recovery boiler with the hot syngas on the tube side, that generates 1600 psia steam. The cooled syngas is sent to the particulate removal system that removes over 99.9 percent of the particulates. The particulate-free syngas is further cooled and then water scrubbed to remove chlorides and volatile trace metals. The syngas is then cooled to 100°F before being sent to the acid gas removal system. During cooling, water condenses and dissolves most of the NH3, along with some CO2 and H2S. This water is sent to the sour water treatment unit. The essentially sulfur-free syngas is moisturized and superheated and sent to the combustion turbine.

KRW Fluidized-Bed Gasifier

Development of this process was started by Westinghouse. Later, Kellogg Rust purchased an interest, and the technology was renamed the Kellogg-Rust Westinghouse (KRW) process. A 15-ton/day unit was successfully operated at Waltz Mills, Pennsylvania. Based on this success, a KRW gasifier was included as part of a 100-MW(e) IGCC Clean Coal Technology project near Reno, Nevada. Owing to problems, mainly with the hot gas cleanup system, this unit only operated for short periods of time.

The KRW fluidized-bed gasifier consists of a vertical vessel with a smaller diameter at the bottom than at the top (Fig. 17.18). It is fitted with a central feed tube through which coal, sorbent (limestone), oxidant (air or oxygen), and steam are introduced. Solids are fed through a series of bins, which raise them from atmospheric to gasifier operating pressure. Coal and limestone are then pneumatically transported to the gasifier central feed tube. Oxidant is added to the feed tube, and the streams merge to form a central jet. The coal quickly devolatilizes, and the remaining char and limestone enter the gasifier bed. Combustion of char and gas occurs within the jet to provide the heat necessary for the endothermic devolatilization, gasification, and desulfurization chemical reactions. Extraction steam from the steam turbine is also fed to the gasifier. CO and H2 are the major combustible constituents of the product gas. Methane and other hydrocarbons are produced in small quantities during devolatilization of the coal. The gasifier temperature is high enough to crack the tars and oils produced. Because of the reducing atmosphere present during gasification, sulfur in the coal is released as predominantly H2S. At conditions in the gasifier, limestone is rapidly calcined to form lime that reacts with the H2S to form CaS. Since chemical equilibrium limits sulfur capture, remaining H2S in the product gas is captured in a downstream desulfurizer.

As the carbon in the coal is gasified, the particles become enriched in ash. These ash particles tend to agglomerate and, along with dense calcium sulfide/oxide particles, separate from the char bed because of their higher density and different fluidization characteristics. This separation occurs primarily in the region surrounding the central feed tube at the bottom of the gasifier. These solids are cooled in the gasifier annulus by a countercurrent stream of recycle gas. The spent solids (unconverted calcined limestone, sulfided limestone, and ash) leaving the gasifier are transported to the ash collection system, consisting of a series of bins designed to reduce pressure back to atmospheric.

Gas exiting the gasifier flows through a cyclone for removal of entrained solids (char, ash, and sorbent) and then to the product gas cooler, where it is cooled to 1000°F while generating 1075 psia steam. Solids collected in the cyclone are returned to the gasifier, while the gas goes to sulfur removal before being burned in the gas turbine.

17.8 Coal Liquefaction

The primary purpose of any coal liquefaction process is to produce a liquid product with an increased hydrogen-to-carbon ratio, a reduced heteroatom (sulfur, nitrogen, and oxygen) content, and a minimal inorganic content. This has been accomplished by a variety of methods, which can generally be grouped into three major categories: pyrolysis, direct liquefaction, and indirect liquefaction. In pyrolysis, coal is heated to generate gases and liquids/tars while leaving a char behind. The liquids/tars are condensed and further treated to produce conventional liquid fuels. A process variation (hydropyrolysis) produces higher-quality products by carrying out the pyrolysis in a hydrogen atmosphere. In direct coal liquefaction, powdered coal is mixed with a solvent and, usually, a catalyst, and heated under pressure in a hydrogen atmosphere to produce gaseous and liquid products. The liquids are separated from ash and unconverted coal and refined to useful fuels. In indirect liquefaction, the coal is first gasified to produce syngas, which is then converted in the presence of a catalyst at an appropriate temperature and pressure to liquid products. Each of these approaches is discussed in more detail below.

Pyrolysis-Based Processes

When coal is heated, it undergoes thermal decomposition, which results in the evolution of gases and liquids/tars from the char that remains. In all coal conversion and utilization processes (including combustion, gasification, and liquefaction), some form of coal pyrolysis occurs. Because coal pyrolysis is an important process for making metallurgical coke, a vast amount of literature exists on the topic. By carefully varying pyrolysis conditions and optionally adding a reactive atmosphere (such as hydrogen), one can optimize production of a desired product (solid, liquid, or gas).

Coal pyrolysis is a very old technique (dating back to the eighteenth century), based on relatively inelegant technology. Most pyrolysis systems in use in the late 1800s and early 1900s were in Europe and had as their objective the production of smokeless fuel (char) for domestic use. However, within a short time it was realized that the coal tar fraction contained valuable chemical products. However, as inexpensive petroleum appeared on the scene, interest in coal by-products faded. A detailed review has been published.[29]

Interest in pyrolysis was rekindled by the oil embargo of 1973, and emphasis was placed on developing pyrolysis processes that would achieve a maximum yield of liquid products capable of displacing petroleum as a source of distillate fuels. Several processes were developed to the pilot-plant stage in the 1970s, but development was discontinued when oil prices softened in the early 1980s, and none of these processes achieved commercial success.

Pyrolysis conditions determine the extent to which the coal is carbonized and the nature of the liquid and gaseous products. Coal pyrolysis processes are generally classified as low temperature (<700°C), medium temperature (700900°C), or high temperature (>900°C). A number of reviews on advanced pyrolysis processes are available.[30,31] Two highly developed pyrolysis processes were the COED process, developed by FMC Corporation, which used a series of fluidized beds operating at successively higher temperatures, and the TOSCOAL process, which used a horizontal rotating kiln.

Factors Affecting Coal Pyrolysis

During pyrolysis, the yield of gaseous and liquid products can vary from 25 to 70 percent by weight, depending on a number of variables, such as coal type, type and composition of the atmosphere present, final pyrolysis temperature, timetemperature path, and pressure. Although certain operating conditions may lead to increased product yield, achieving these conditions may result in increased costs. Many excellent reviews (Howard,[32] Gavalas,[3] and Solomon and Serio[33]) have been published that discuss the factors affecting coal pyrolysis and product composition. In the following sections, major conclusions presented in the literature are discussed (as summarized in Table 17.14).

TABLE 17.14 Summary of Effects of Process Parameters on Coal Pyrolysis

Heating rate
1. A rapid heating rate increases liquid/gas yield and reduces char yield
2. A rapid heating rate produces poorer quality (lower H/C ratio) tar than that obtained at a slower heating rate[19,20]
3. A rapid heating rate (in a reactive gas atmosphere) produces char with increased open structure and reactivity
4. A rapid heating rate increases the thermoplastic (softening and swelling) behavior of coal
5. Achieving a rapid heating rate requires a sophisticated (often expensive) system

Temperature
1. Low-temperature operation (500700°C) improves liquid yield
2. Temperature affects heteroatom distribution among char, liquid, and gas
3. At elevated temperatures (>1300°C), inorganics are removed as slag
4. Lower temperatures require a longer residence time for complete reaction

Pressure

Inert gas atmosphere:
1. Higher pressure operation reduces reactor size (i.e. increases throughput)
2. Higher pressure reduces tar yield
3. Coal feeding and product separation are more difficult at high pressure
4. Higher pressure improves gassolid heat transfer

H_2 atmosphere:
1. Improves the yields of liquid and lighter products
2. Requires a sophisticated pressure control system
3. May increase undesirable agglomerating properties of coal[34]
4. H_2 cost must be compared to the increased value of the product

Other atmospheres (H_2O, CO_2, CO, CH_4, CS_2):
1. Probably improve liquidgas yield
2. Little information available

Vacuum:
1. Plastic behavior of coal is reduced[23]
2. Increases liquid/gas yield[23]
3. Difficult to achieve gassolid heat transfer (solidsolid heat transfer feasible)
4. Little information available

Particle size
1. Smaller particle size improves gasliquid yield
2. Smaller particle size reduces secondary reactions
3. Grinding cost increases with the reduction in size

Coal rank
1. High Volatile A (HVA) bituminous coals produce the largest quantity of tar
2. Lignites are rich in oxygen functional groups that lead to overall reduction in the calorific value of the product
3. The type of sulfur (pyritic versus organic) present influences sulfur distribution among the products

Temperature and Heating Rate

Coal undergoes many physical and chemical changes when heated gradually from ambient temperature to approximately 1000°C. Some changes, such as carboncarbon bond scission, are observed before the onset of thermal decomposition that occurs above 350°C. When heated to approximately 100°C, physically sorbed moisture is liberated. Heating low-rank coals, such as lignites, that contain appreciable carboxylic functional groups will liberate carbon dioxide by thermal decarboxylation. Over 50 percent of the carboxylic acid functional groups are lost in the temperature range 100250°C.

As the temperature of thermal treatment increases to the range 200400°C, coal produces a number of lower molecular weight organic species (especially aliphatic compounds), which are believed to arise from components that are loosely bound to the more thermally stable part of the coal structure. At a higher temperature (375700°C, depending on the heating rate), thermal destruction of the coal structure occurs, as reflected by the formation of a variety of hydrocarbons, including methane, other alkanes, polycyclic aromatics, phenols, and nitrogen-containing compounds. In this temperature range, bituminous coals soften and become plastic (thermoplastic) to varying degrees.

At still higher temperatures (600800°C, depending on the heating rate), the plastic mass undergoes repolymerization, forming semicoke (solid coke containing significant volatile matter). At temperatures exceeding 600°C, semicoke hardens to form coke with the evolution of methane, hydrogen, and traces of carbon oxides. Pyrolysis of coal is essentially complete at approximately 1000°C. The temperature at which the maximum devolatilization rate occurs depends on the heating rate. For a slow heating rate (about 5°C/s) the maximum rate occurs at around 400°C, whereas for a rapid heating rate (about 10°C/s) the maximum rate might not occur until 900°C.[35] Maintaining the coal at the final temperature for an extended period normally increases the yield of volatile material, because char decomposition is a relatively slow process.

Since pyrolysis reactions do not occur at sharply defined temperatures, the heating rate has a marked effect on the nature and distribution of pyrolysis products, as summarized in Table 17.14. Solomon and coworkers conducted extensive work on the kinetics of coal devolatilization, and many reviews are available.[36]

TABLE 17.16 Distillate Yields from Various Direct Liquefaction Processes

					Heteroatoms (wt. %)		
Process	Configuration	Distillate (wt. % Maf. Coal)	Yield (bbl/ton Maf. Coal)	Distillate Quality (gravity °API)	N	O	S
SRC-II (1982)	One-stage Noncatalytic	41	2.4	12.3	1.0	2.33	0.33
H-Coal (1982)	One-stage Noncatalytic	52	3.3	20.2	0.50	1.0	0.20
Wilsonville (1985)	ITSL Thermal-catalytic	62	3.8	20.2	0.25	1.9	0.23
Wilsonville (1986)	CC-ITSL Catalytic-catalytic	70	4.5	26.8	0.16	<1	0.11

Coal Rank

The type of coal strongly influences pyrolysis behavior. Low-rank coals, such as lignite, contain oxygen functional groups that evolve water and carbon oxides upon pyrolysis. Higher-rank bituminous coals contain less oxygen; consequently, these coals produce significantly less water and carbon oxides when pyrolized. The nature of the tar produced is also dependent on coal rank. Bituminous coal tars tend to be more aromatic (and relatively more thermally stable) than the tars generated from lignites. When heated, bituminous coals soften, become plastic, and swell to varying degrees, whereas lower-rank coals generally do not become plastic. However, at a rapid heating rate (about 10°C/min) or elevated pressure, certain lower-rank coals may melt and demonstrate some plastic and swelling characteristics.

Other Factors

Several other factors influence, at least to some extent, the course of the pyrolysis process. These include particle size, bed configuration, pressure/vacuum during pyrolysis, nature of the coal ash, secondary reactions, etc.[37] It is beyond the scope of this chapter to consider these items, but the interested reader can find additional information in the literature, including reports on pressure effects,[21,38] effect of vacuum,[23] effect of inorganics,[26,39] and effect of a reactive atmosphere.[23,40]

Utilization and Characterization of Pyrolysis Products

Efficient utilization of all the products, solid, liquid, and gaseous, is essential if favorable economics for a pyrolysis process are to be achieved. Products may require varying degrees of treatment before they are usable. The stream exiting the pyrolyzer requires separation of gas, liquid, and particulates, similar to the situation for a fixed-bed gasifier (i.e. cold gas cleanup). Possible operations include hot dust removal cyclones, quench/ particulate scrubber towers, and venturi scrubbers to remove tar mist. In general, gas stream sulfur removal and wastewater treatment are also required. Finally, environmental factors, such as toxicity,[41] carcinogenicity, and mutagenicity of the coal pyrolysis liquids (CPL), need to be considered. For example, the mutagenicity of CPL is strongly dependent on the conditions of pyrolysis (temperature, coal type, and atmosphere during pyrolysis).

Liquid fuels markets tend to have product specifications that do not vary widely. However, the characteristics of CPL can vary greatly, depending on coal type and operating conditions. In general, the H/C atomic ratio for CPL is lower than that for petroleum products, which is in the range 1.81.9 for No. 2 fuel oil and 1.71.8 for No. 6 fuel oil. In contrast, the H/C ratio for CPL is in the range 0.91.5, indicating that CPL is significantly more aromatic (5070%) than petroleum-derived products.

The uses of CPL can be divided into two broad categories: (1) direct combustion, requiring little or no upgrading, and (2) transportation fuels and chemicals, requiring extensive upgrading. The high aromaticity of CPL indicates that conversion to gasoline is preferred to conversion to diesel fuel. Much of the attention in low-temperature tar processing has been devoted to hydroprocessing techniques, such as hydrotreating and hydrocracking, with the primary objectives of reducing viscosity, reducing polynuclear aromatics, and removing heteroatoms (sulfur, nitrogen, and oxygen) to produce usable fuels and chemicals.

The cost of H_2 is still the primary impediment to tar upgrading. Although the tar fraction can be used as a source for chemicals, such as phenolics, road tars, preservatives, and carbon binders, these uses do not constitute a large enough market to support a major industry. Likewise, the char has a range of potential applications, but the char must provide an economic return for pyrolysis processes to operate profitably. One obvious application is on-site combustion to generate heat for the pyrolysis process. Alternatively, the char can be combusted in an electric power plant. Char could also be gasified to produce hydrogen for use in hydropyrolysis or tar hydrotreatment. Other applications include production of activated carbon and molecular sieves.

Direct Coal Liquefaction

Direct coal liquefaction (DCL) refers to the process of converting coal to liquid products by mixing ground coal with a recycled process solvent and/or petroleum-derived residual oil and reacting the slurry in a hydrogen atmosphere at 750850°F (400450°C) and 10002500 psig (717 MPa). Under these conditions, the coal structure breaks down into a variety of products, some of which are in the gas phase and some soluble in the liquid phase. The highly aromatic liquid products are separated from ash and unconverted coal by solvent deashing or vacuum distillation and upgraded by conventional petroleum refining techniques to gasoline and distillates.

The success of any DCL process is highly dependent on the type of coal used. Coal rank influences both overall conversion and product distribution. Lower-rank coals have been reported to give both higher and lower conversions than bituminous coals,[42] with most evidence supporting the latter. Furthermore, the liquid products from lower-rank coals are generally more volatile and of lower molecular weight[43] than those from bituminous coals.

Germany demonstrated commercial-scale operation of DCL during World War II, employing the Bergius process, which they had developed. After World War II, the United States constructed a 200300 bbl/day DCL plant at Louisiana, Missouri, based on the German technology. This plant was operated by the Bureau of Mines from 1949 to 1954. However, as the extent of petroleum deposits in the Middle East and other locations became apparent and fears of an oil shortage subsided, interest in DCL diminished.

The 1973 oil embargo created a new surge of interest in DCL as one of a number of routes to produce petroleum substitutes from indigenous resources. From this renewed effort in the early 1970s, three major DCL processes emerged: Solvent Refined Coal (SRC-II), Exxon Donor Solvent (EDS), and H-Coal. These processes converted coal to liquids in a single reaction stage and, by the late 1970s, had been developed at pilot-plant scale, with designs underway for demonstration plants. However, in the early 1980s, these plans were cancelled due to projected unfavorable economics. Interest then shifted to developing a two-stage process, which promised higher yields and better quality product with a concomitant improvement in economics. For information on the early history of DCL, see Wu and Storch,[44] Donath and Hoering,[45] Ministry of Fuel and Power,[46] and *Perry's Chemical Engineers' Handbook*.[47]

Chemistry of Direct Coal Liquefaction

As coal is heated in a suitable solvent, it is rather rapidly (15 min) converted into soluble entities that have a high molecular weight and an average elemental composition relatively little changed from that of the moisture and ash free (maf) feed coal. As residence time is extended, hydrocracking converts the dissolved entities to lower-boiling liquids with reduced heteroatom content.48 The exact product slate depends on many factors, including operating conditions, coal properties, and the solvent used. The dissolution reactions produce free radicals, which may undergo retrograde reactions to form insoluble coke or, if there is sufficient hydrogenation activity, be stabilized by reaction with hydrogen atoms. The stabilized free radicals can then undergo hydrocracking reactions to form a wide range of products from methane through heavy oil. Some hydrogen comes from the gas phase, but most is transferred from the solvent. The most effective hydrogen transfer agents (hydrogen donor solvents) are hydroaromatic compounds, which can lose hydrogen and revert to polynuclear aromatics and later be rehydrogenated to hydroaromatics. The mechanism of these and other DCL reactions is not fully understood, and since every coal molecule is unique, it is unlikely that precise kinetics will ever be possible.

Because of the high concentration of ring structures in coal, coal liquids are highly aromatic, particularly the heavier fractions. This can be a problem, since some polynuclear aromatics are carcinogenic. Even though the carcinogens can be eliminated from the final product by limiting the endpoint of the distillate to about 700°F and recycling heavier components to extinction, care must be taken to prevent exposure of plant workers. Another problem is that the current trend is to limit the aromatics content of transportation fuels for environmental reasons. Therefore, production of specification fuels from coal liquids can require extensive (and potentially expensive) hydroprocessing.

The first DCL processes were single-stage processes, in which both coal dissolution and hydrocracking take place in the same reactor. Later, it became apparent that these two steps could not both be optimized at the same operating conditions. This led to the development of a two-stage process, in which a short contact time (SCT) thermal liquefaction reactor was followed by a catalytic upgrading reactor to optimize conditions for both reactions, thus leading to higher yields, more efficient use of hydrogen, and a better product distri-bution. Other variations of two-stage processing, such as the Integrated Two-Stage Liquefaction (ITSL) process, were developed later. Tables 17.15 and 17.16 present typical compositions and yields, respectively, of distillates from various DCL processes.

TABLE 17.15 Heteroatom Content of Products from Direct Liquefaction Processes

	Wt. % in naphtha (C5390°F)				Wt. % in distillate (400975°F)			
Process	O	N	S	H	O	N	S	H
H-Coal	0.6	0.3	0.07	13.19	1.1	0.5	0.08	9.52
EDS	2.42	0.23	0.28	12.03	2.0	0.6	0.3	8.97
SRC-II	3.51	0.42	0.26	10.46	3.63	0.99	0.25	9.27
ITSL	0.41	0.14	0.01	13.78	0.56	0.30	0.015	10.21

Direct Coal Liquefaction Processes

All DCL processes consist of three basic steps: (1) slurrying powdered coal in a solvent, (2) dissolving the coal under high pressure and temperature in a hydrogen-containing atmosphere, and (3) separating the products from ash and unconverted coal. However, the specific configuration of equipment and the processing conditions used vary considerably among different processes. It is beyond the scope of this chapter to review all DCL processes, but some of the different approaches that have been used are discussed below.

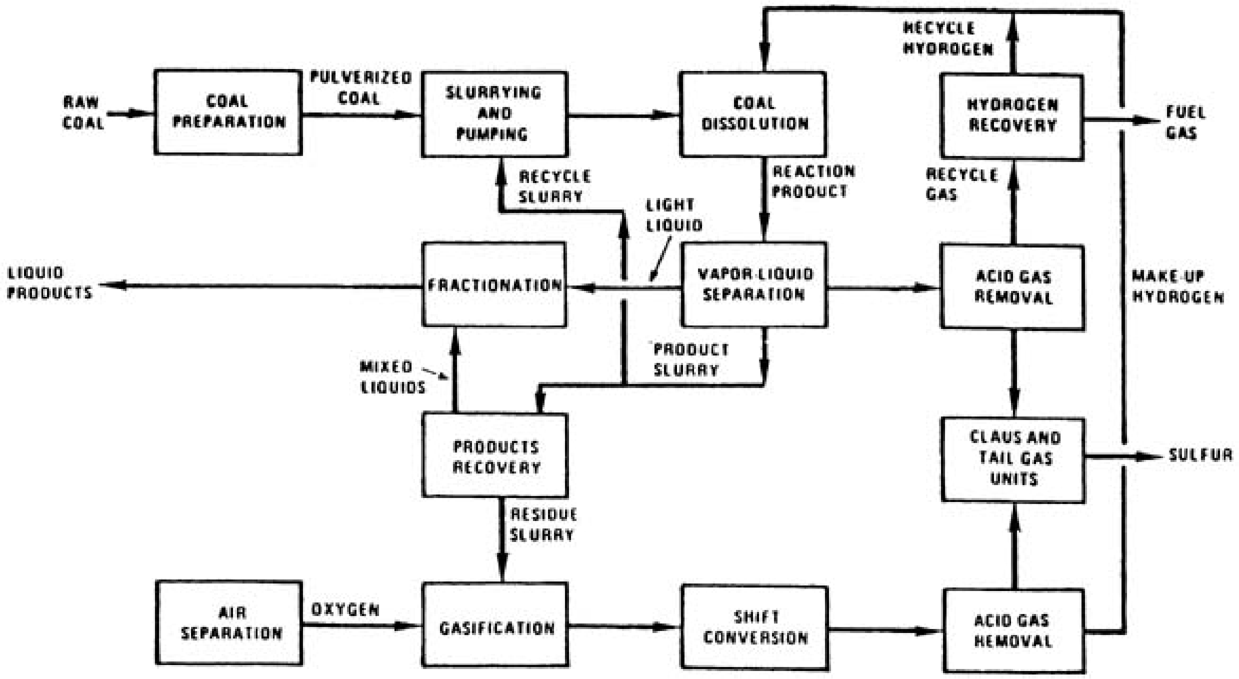
The Solvent Refined Coal (SRC-I and SRC-II) Processes

The SRC-II process began as a study of the feasibility of coal deashing by Spencer Chemical (later purchased by Gulf Oil) in 1962.49 The aim was to produce a very low-ash, low-sulfur solid fuel that could substitute for the coal burned in boilers. Feed coal is crushed, mixed with a recycled solvent and hydrogen, preheated, and fed to the dissolver, which operates at 840870°F (450465°C). Dissolver effluent is flashed, subjected to ash removal and solvent recovery, and then solidified by cooling. The product is a low-sulfur, low-ash, solid boiler fuel with a melting point of approximately 300°F (150°C), a heat of combustion of 16,000 BTU/lb, and a H/C ratio approximately the same as that of the feed coal. A 50-ton/day SRC plant, located in Fort Lewis, Washington, began operating in October 1974. Ash removal was initially accomplished using a filter, which was later replaced by a critical solvent deashing (CSD) unit. Operation of the plant revealed two problems: (1) it was difficult to produce enough solvent to stay in solvent balance; and (2) the economics of producing a coal substitute were not attractive.

To overcome these problems, the plant was modified in late 1976 to increase residence time in the reactor and to recycle some of the ash-containing bottoms. These modifications resulted in greatly enhanced distillate production and changed the objective of the process from producing a solid fuel (SCR-I) to producing a liquid product (SCR-II). The plant was shut down in 1981 and later dismantled.

Although SRC-II was basically a thermal liquefaction process, it was most successful using bituminous coals with a high native pyrite content. Iron sulfides are well known to have catalytic properties for coal liquefaction. Recycling part of the ash-minerals-containing bottoms had two beneficial effects: (1) it increased the pyrite concentration in the reactor feed, and (2) it increased the residence time for heavy components, thus giving them more time to hydrocrack to distillate products. A block flow diagram of the SRC-II process is shown in Fig. 17.19.

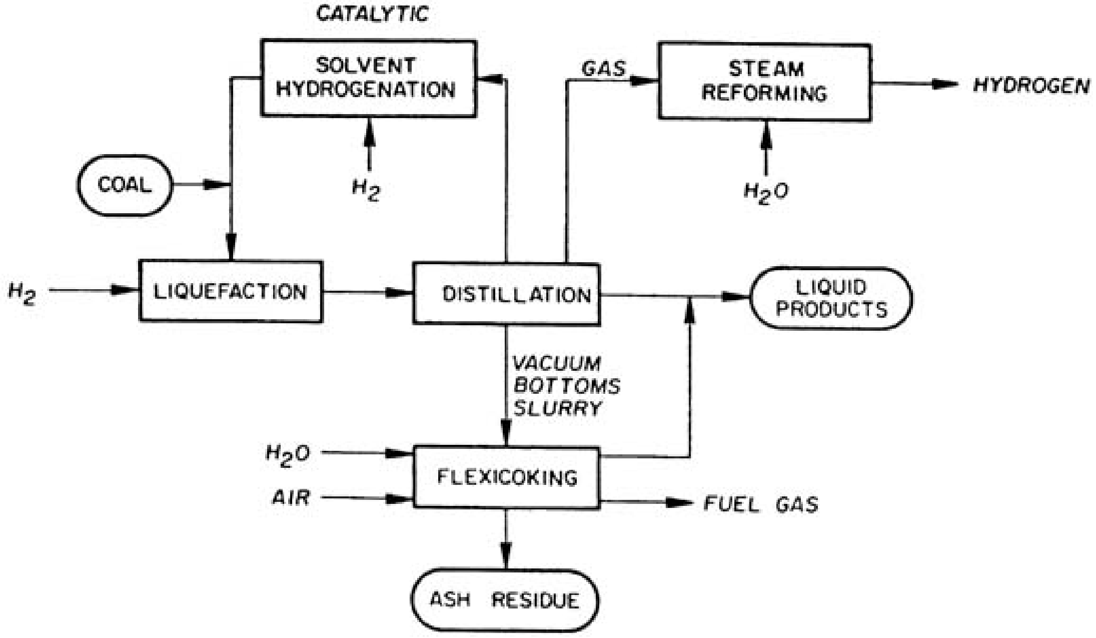
Fig. 17.19 Block flow diagram of SRC-II process for solvent refined coal. *(Courtesy The Pace Company, Denver, CO)*



The Exxon Donor Solvent Process

Research on the Exxon Donor Solvent (EDS) process began at Exxon in 1966. A schematic of the EDS process is shown in Fig. 17.20.37,50 Coal is ground, mixed with hydrogenated distillate recycle solvent, and preheated before entering a slurry-bed liquefaction reactor, which operates at 800880°F (425470°C) and 15002000 psig (1014 MPa). Hydrogen is added to the reactor, which operates without addition of an external catalyst. The solvent transfers hydrogen to dissolved coal fragments in the liquefaction reactor with some additional hydrogen being supplied from the gas phase. Distillate production is achieved by thermal cracking and hydrogenation. The total reactor product is sent to a distillation unit where it is separated into gas, naphtha, distillate, and vacuum bottoms containing ash and undissolved coal. The 400850°F (204455°C) distillate fraction is sent to a fixed-bed catalytic hydrogenation unit to regenerate the hydrogen donor (H-donor) solvent.

Fig. 17.20 Schematic of Exxon Donor Solvent process. *(Courtesy The Pace Company, Denver, CO)*



The use of an external reactor to hydrogenate the recycle solvent distinguished the EDS process from other single-stage processes.51 (Although this process uses a separate solvent hydrogenation unit, it is considered to be a single-stage process, because the coal itself is only treated in one reactor.) A 250-ton/day pilot plant operated at Baytown, Texas, from 1980 through 1982.38

Operating conditions for the EDS process are similar to those for the SRC-II process; and, like SRC-II, EDS works best with bituminous coals. The distillate yield is about 2.53 barrels of liquid per ton of coal. This somewhat low yield is probably attributable to the absence of a catalyst in the liquefaction reactor.

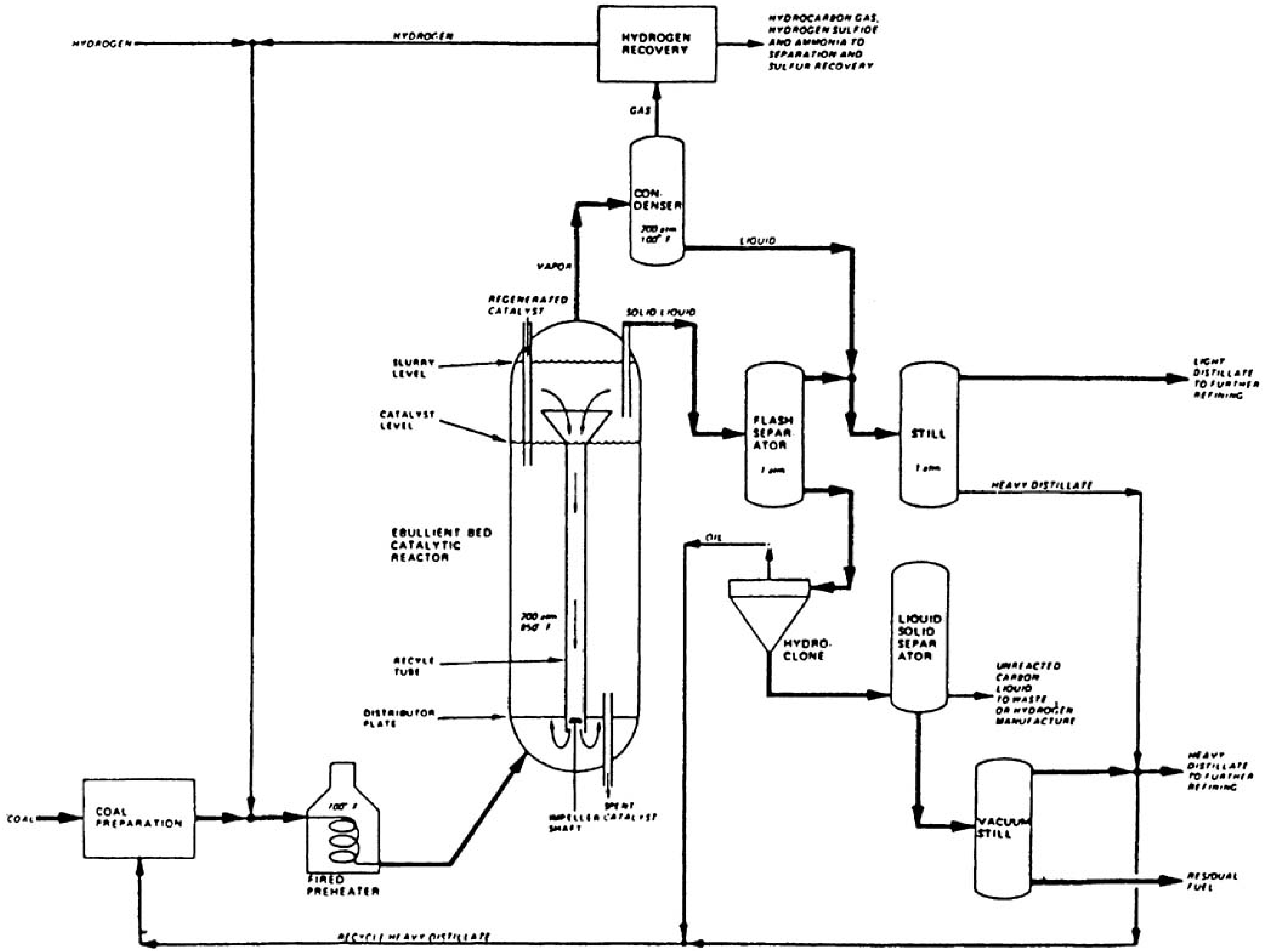
Initial operation of the EDS process showed that coal could be adequately hydrogenated in the absence of added catalyst. However, a later modification, which used vacuum bottoms recycle (similar to the bottoms recycle demonstrated in the SRC-II process) in addition to donor solvent recycle, displayed improved operating characteristics and higher distillate yield, particularly for low-rank coals. The increased residence time in the reactor of the heavy, high-boiling-point material increased the yield of naphtha.52 In addition, the ash in the bottoms recycle may have provided a catalytic effect, resulting in increased conversion for lignite and sub-bituminous coals.53 The recognition of these factors was important to the later development of the two-stage liquefaction concept.

The H-Coal Process

The H-Coal process is an adaptation of the H-Oil process, which uses a catalytic ebullated bed reactor to convert petroleum residues to lower-boiling-point products.54 Research on the H-Coal process began in 1964 at Hydrocarbon Research, Inc. (HRI) on a bench-scale unit. Construction of a 200600-ton/day pilot plant adjacent to the Ashland Oil refinery in Cattletsburg, Kentucky, was completed in early 1980 and operated through November 1982.

Figure 17.21 shows a schematic of the H-Coal process, which employs a single catalytic stage to produce a synthetic crude oil.37,38 Coal is crushed, dried, and mixed with recycle oil and hydrogen before being preheated to approximately 850°F (454°C). The preheater effluent is fed to the bottom of an ebullated bed reactor. During operation, fresh catalyst (a cobaltmolybdenum extrudate) is fed to the top of the reactor, while spent catalyst is removed from the bottom to maintain constant reactivity and inventory. The upward flow of the coal slurry and hydrogen causes the catalyst to remain in suspension in the reactor; however, there is a small volume at the top of the bed which remains catalyst free. Products are removed from the top of the reactor, and the various fractions are separated by conventional techniques.

Fig. 17.21 Schematic of H-Coal Process. *(Courtesy The Pace Company, Denver, CO)*



The H-Coal product has a high H/C ratio and a low heteroatom content, as shown in Table 17.15. However, the use of high temperatures in the reactor causes excessive thermal cracking, resulting in high gas yields and correspondingly poor hydrogen utilization efficiency. Typically, between 14,000 and 20,000 standard cubic feet of hydrogen are required for each ton of coal processed.

The H-Coal process produces approximately 33.5 barrels of liquid product for each ton of coal.42 Tests showed that the process is best suited for high-volatile bituminous coal; the use of low-rank coals significantly reduced throughput and distillate yields. The successful performance of the ebullated bed reactor in the H-Coal process led to its later use in two-stage liquefaction systems.

Two-Stage Process

As DCL research progressed, it became apparent that both coal dissolution and hydrocracking could not be optimized at the same operating conditions and that two-stage operation offered the potential for process improvements. This led to the use of a short contact time (SCT) thermal liquefaction reactor followed by a catalytic upgrading reactor (hydrotreater) to provide optimum conditions for both reactions. This arrangement resulted in higher yields, more efficient use of hydrogen, and a better product distribution. The development of the two-stage process involved contributions from a number of companies, including Lummus, KerrMcGee, and HRI (which later became HTI), and proceeded in an evolutionary manner through a series of steps.

The first approach to separation of the two stages was the nonintegrated two-stage liquefaction (NTSL) process,55 in which a second-stage upgrading reactor was placed downstream of an SRC-I type dissolution reactor. Nonintegrated means that the second-stage catalytic hydrotreater did not contribute solvent to the first-stage dissolver. To avoid plugging and rapid catalyst deactivation that could result from use of a fixed bed, an ebullated-bed reactor was used as the second stage. Although improved, distillate yields were still relatively low.56

To further improve yields, the integrated two-stage liquefaction (ITSL) process was developed. In ITSL hydrogenated resid from the second stage is recycled to the first stage to serve as solvent. In addition, solvent deashing (SDA) was used to remove ash from the resid. The SDA unit was originally located between the first and second stages, but later, in the reconfigured two-stage liquefaction (RITSL) process, the SDA unit was relocated after the second-stage. This latter configuration showed that ash in the first-stage reactor effluent did not deactivate the catalyst in the second stage. The liquid product from the ITSL process was of better quality, as shown in Table 17.16, and more easily upgraded by conventional petroleum-refining techniques than the product from the single-stage process.

The next advance was the close-coupled integrated two-stage liquefaction (CC-ITSL) process, in which the residence time between reactors was minimized by having the first-stage effluent pass directly to the second-stage reactor, thereby improving thermal efficiency, reducing the potential for retrograde reactions, eliminating the need for pressure letdown between stages, and producing a higher-quality solvent.37,57,58 Distillate yields from the CC-ITSL of up to 64 percent (maf coal) were reported for Illinois No. 6 bituminous coal.45 Several versions of the CC-ITSL process have been investigated, including thermal/catalytic (T/C CC-ITSL), in which the first stage is thermal and the second stage catalytic; catalytic/catalytic (C/C CC-ITSL), in which both first- and second-stage reactors are ebullated beds; and use of a disposable slurry catalyst in place of ebullated catalyst.

The most advanced two-stage process involves close coupled operation with a disposable slurry catalyst (typically iron and/or molybdenum) added with the feed coal and an integrated hydrotreater to treat the distillate condensed from the gaseous phase from the high-pressure separator. This arrangement produces a high yield of distillate products with a very low heteroatom content.

Direct Coal Liquefaction and the Future

Future research will be primarily driven by economic considerations. The use of large-scale commercial DCL will be feasible only when the costs of coal-derived products are comparable with those of petroleum-derived products. Special circumstances, such as those present in Germany during World War II, may occasionally make DCL economically feasible, but so far these situations have proven to be only temporary. Two studies37,59 have reported that, at an equivalent crude petroleum cost of around $35/bbl, coal liquefaction would be economically feasible. However, crude oil prices have tended to fluctuate, and even if they rose to $35/bbl, the viability of a large-scale DCL operation would be tenuous. Investors would be reluctant to commit the large capital expenditure required for fear that crude prices would drop and make their facility uneconomic. Several approaches to reducing costs are being pursued that may ultimately make DCL economically competitive.

Decreasing the cost of hydrogen will greatly improve the economics of any coal liquefaction scheme. This may be done either by lowering the cost of hydrogen production or by improving the efficiency of hydrogen use, that is, minimizing the yield of light hydrocarbons (methane through butanes).

Coal cleaning, which involves the separation of some of the ash from the coal prior to liquefaction, can yield several benefits. Primarily, it reduces the load on the solidliquid separator, thereby reducing the amount of organic material rejected with the ash. Other benefits include the reduction of wear on plant components and better utilization of reactor volume. However, removal of mineral matter prior to reaction may eliminate beneficial catalytic effects that coal minerals, particularly iron pyrite, provide.

Improved catalysts, in both stages of a C/C CC-ITSL system, offer a significant potential benefit to process economics.60 Coal conversion reactions are generally slow, and any improvement in reaction kinetics would result in a large economic return through reduced reactor costs. In addition to supported catalysts, research is being devoted to developing dispersed slurry phase catalysts that are more active, selective, and resistant to deactivation in the presence of high-boiling-point components.

The use of inexpensive western U.S. sub-bituminous coals as feedstock could further improve process economics. It has been shown that these coals can produce lower-boiling and higher-quality liquids31,32,41 and slower deactivation of the second-stage catalyst.47 The main problem is their lower reactivity, which causes inefficient use of reactor volume and the loss of product on solids separation. The different behavior of low-rank coals is believed to relate to their thermal sensitivity and high propensity for cross-linking. Dissolution promoted by an effective slurry-phase catalyst may be able to counter this tendency by increasing the effectiveness of hydrogenation.

An economically viable DCL process may result from the incorporation of one or a combination of the above techniques. There is always the possibility of discovering an approach that represents a radical departure from conventional techniques.

Indirect Coal Liquefaction

In indirect coal liquefaction, synthesis gas is first produced in a suitable gasifier by reaction of the coal with oxygen and steam. Before being fed to the synthesis reactor, the syngas must be cooled and cleaned, which involves removal of particulates, H_2S, NH_3, and other impurities. Then, depending on the catalyst being used, it may be necessary to adjust the H_2/CO ratio. Modern coal gasifiers typically produce syngas with a H_2/CO ratio in the range 0.450.7, which is lower than the stoichiometric ratio of about 2.0 for hydrocarbon production. Some catalysts, particularly iron-based catalysts, possess water gas shift conversion activity and permit operation with a low H_2/CO ratio syngas. Other catalysts may require shift conversion before the syngas is sent to the synthesis reactor. Because many syngas conversion catalysts are very sulfur sensitive, a guard chamber before the reactor is sometimes used to remove the last traces of H_2S.

Any of the gasifiers discussed earlier in this chapter in the section "Gasification" can be used to produce the syngas intermediate. Thus, the feature that distinguishes one indirect liquefaction process from another is the technology used to produce liquid products from the syngas, rather than the technology used to make the syngas. The liquefaction process usually involves some variation of FischerTropsch chemistry to produce naphtha and distillate fuels, but methanol, dimethyl ether (DME), and other fuels and chemicals may also be produced.61

FischerTropsch

The process most frequently considered for indirect coal liquefaction is the FischerTropsch (FT) synthesis, developed in 1925 by German chemists Franz Fischer and Hans Tropsch. In the FT process, synthesis gas is reacted over a catalyst, typically iron or cobalt based, at 130 atm and 200350°C to produce a wide range of mainly aliphatic hydrocarbons, including gas, LPG, gasoline, jet fuel, diesel oil, middle distillates, heavy oil, and waxes. Germany used this technology during World War II to produce nearly 15,000 barrels/day of military fuels.

The FT synthesis is basically a polymerization reaction in which carbon monoxide molecules are integrated one at a time into a growing chain, followed by hydrogenation. Depending on the hydrogenation activity of the catalyst, the product may be predominantly paraffinic, or it may contain appreciable amounts of olefins and alcohols. The basic reactions in the FT synthesis are:

Paraffins:

$$(2n+1)\mathrm{H}_2 + n\mathrm{CO} = \mathrm{C}_n\mathrm{H}_{2n+2} + n\mathrm{H}_2\mathrm{O}$$

Olefins:

$$2n\mathrm{H}_2 + n\mathrm{CO} = \mathrm{C}_n\mathrm{H}_{2n} + n\mathrm{H}_2\mathrm{O}$$

Alcohols:

$$2n\mathrm{H}_2 + n\mathrm{CO} = \mathrm{C}_n\mathrm{H}_{2n+1}\mathrm{OH} + (n-1)\mathrm{H}_2\mathrm{O}$$

If the catalyst has water gas shift activity, then the following reaction may also occur.

$$\mathrm{CO} + \mathrm{H}_2\mathrm{O} = \mathrm{H}_2 + \mathrm{CO}_2$$

This reaction converts CO into H_2 and allows operation with syngas having a H_2/CO ratio below the required ratio of about 2.0.

The FT synthesis typically follows polymerization kinetics. The AndersonSchulzFlory equation describes the product distribution:

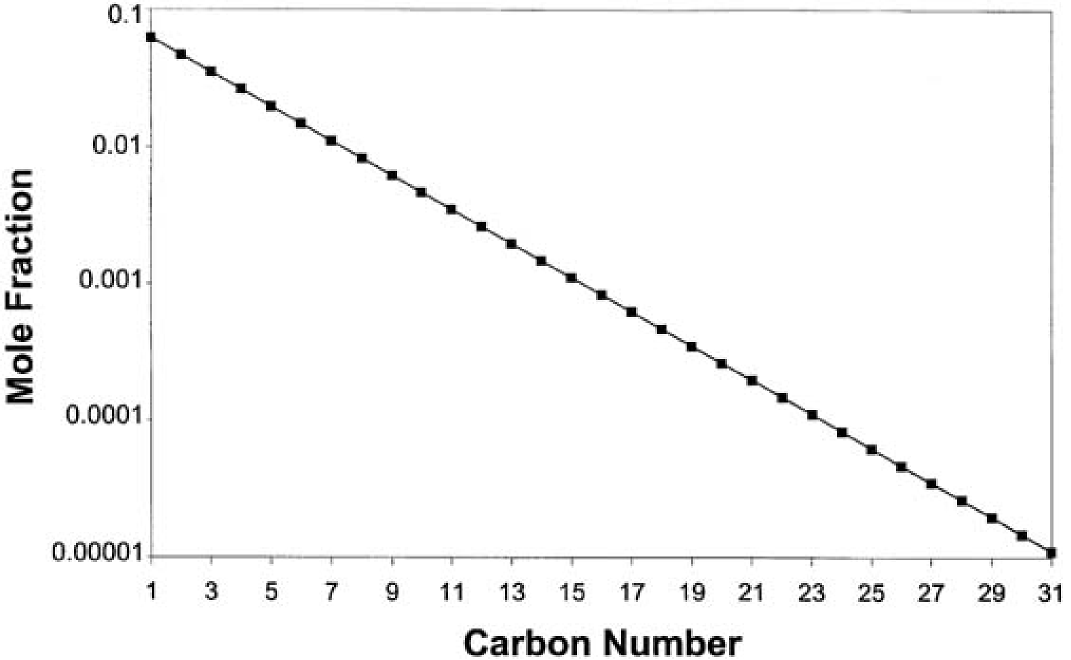
$$W_n = n(1-\alpha)^2\alpha^{n-1}$$

where W_n is the weight fraction of products with carbon number n, and α is the chain growth probability, sometimes referred to as the SchulzFlory alpha, the value of which depends on the catalyst and operating conditions used. When α equals 0, the product is pure methane, and as α approaches 1, the product becomes predominantly high molecular weight wax. In logarithmic form, the distribution becomes:

$$\log(W_n/n) = n\log\alpha + \log[(1-\alpha)^2/\alpha)]$$

A plot of $\log(W_n/n)$ vs. n is a straight line, with a slope depending upon the value of α (see Fig. 17.22).

Fig. 17.22 Typical SchulzFlory plot.



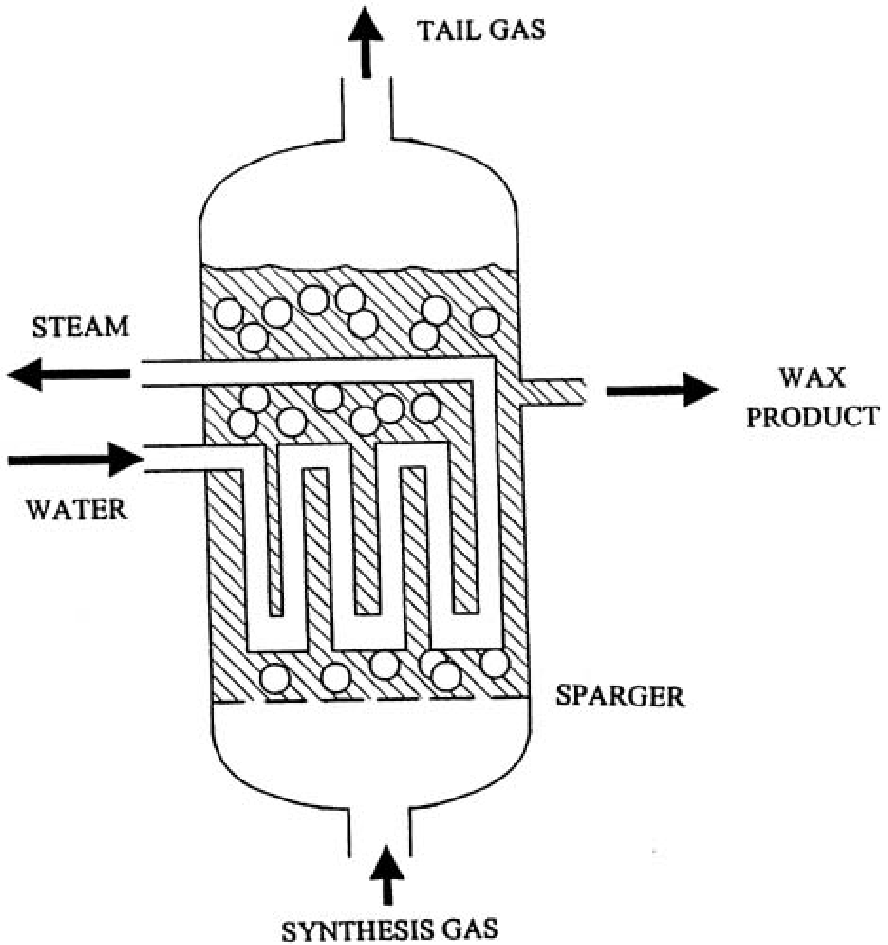
Catalysts of commercial significance are either iron-based or cobalt-based. Iron-based catalysts are typically not supported, whereas cobalt-based catalysts are usually supported on alumina, silica, or a similar material. Cobalt catalysts do not have water gas shift activity; therefore, the syngas used must have a H_2/CO ratio close to the stoichiometric ratio. Iron catalysts, on the other hand, have water gas shift activity and can operate on syngas with a low H_2/CO ratio, which typically occurs when the syngas is derived from coal gasification. Cobalt catalysts have more hydrogenation activity than iron catalysts, with the result that the product from cobalt catalysis tends to have a lower olefin and alcohol content and a higher methane content. Also, cobalt catalysts tend to be more active and operate at somewhat lower temperatures.

Because the FT reaction is highly exothermic, high heat removal and good temperature control are vital for successful process operation. Three types of reactors have been used in FT plants: tubular fixed-bed reactors, fluidized-bed reactors, and slurry bubble column reactors (SBCR). A tubular fixed-bed reactor resembles a vertical tubular heat exchanger. Catalyst is loaded into the tubes, and a cooling medium, typically water, is circulated in the shell for heat removal and temperature control. To prevent the temperature in the center of the tube from getting too high, reactor tube diameter is limited to about 2 in. Also, catalyst loading and unloading can be a problem.

Two types of fluidized-bed units have been used. So-called fixed fluidized beds, in which the catalyst remains in the reactor and the syngas passes up through it, and transport reactors, in which the fluidized catalyst is carried along with the syngas, separated from the product gas outside the reactor, and recirculated to the reactor inlet. These reactors normally are operated in a high-temperature mode (about 340°C), which produces lower-boiling-point olefins and other compounds in the gasoline boiling range. Heavier products would remain on the catalyst, thus destroying fluidization characteristics in the reactor.

Because of their excellent heat-removal capabilities and temperature control, SBCRs are currently receiving the most interest. In 1980, due to the limitations of other reactors, the DOE identified an opportunity to develop SBCRs for liquid phase processes to substantially improve overall economics. An SBCR is a vertical vessel containing heat transfer tubes for heat removal (Fig. 17.23). Finely divided catalyst suspended in an inert liquid fills the reactor. Syngas is fed through a sparger at the bottom. The upward flow of this gas keeps the catalyst in suspension and provides backmixing. The combination of high heat and mass transfer permits high syngas conversion, even making once-through operation feasible. Furthermore, the relative simplicity of the reactor design makes for attractive economics.

Fig. 17.23 Schematic of a slurry-phase reactor.



The early thrust of the DOE program was aimed at developing a slurry phase process for the production of methanol. Development took place in a 5 tons of methanol per day plant located in LaPorte, Texas. Air Products' Liquid Phase Methanol (LPMEOHTH) process was later scaled up to 260 tons/day at Eastman Chemical Company's Kingsport, Tennessee, facility. In addition to methanol production, SBCRs are ideally suited to the production of other fuels from syngas, including FT liquids, DME, and various other products.

In 1990, Sasol operated a 1-m diameter SBCR FT demonstration unit. This was then scaled up to a 2,500 barrels/day commercial unit with a 5-m diameter by 22-m high reactor that was commissioned in 1993.

The Sasol Complex

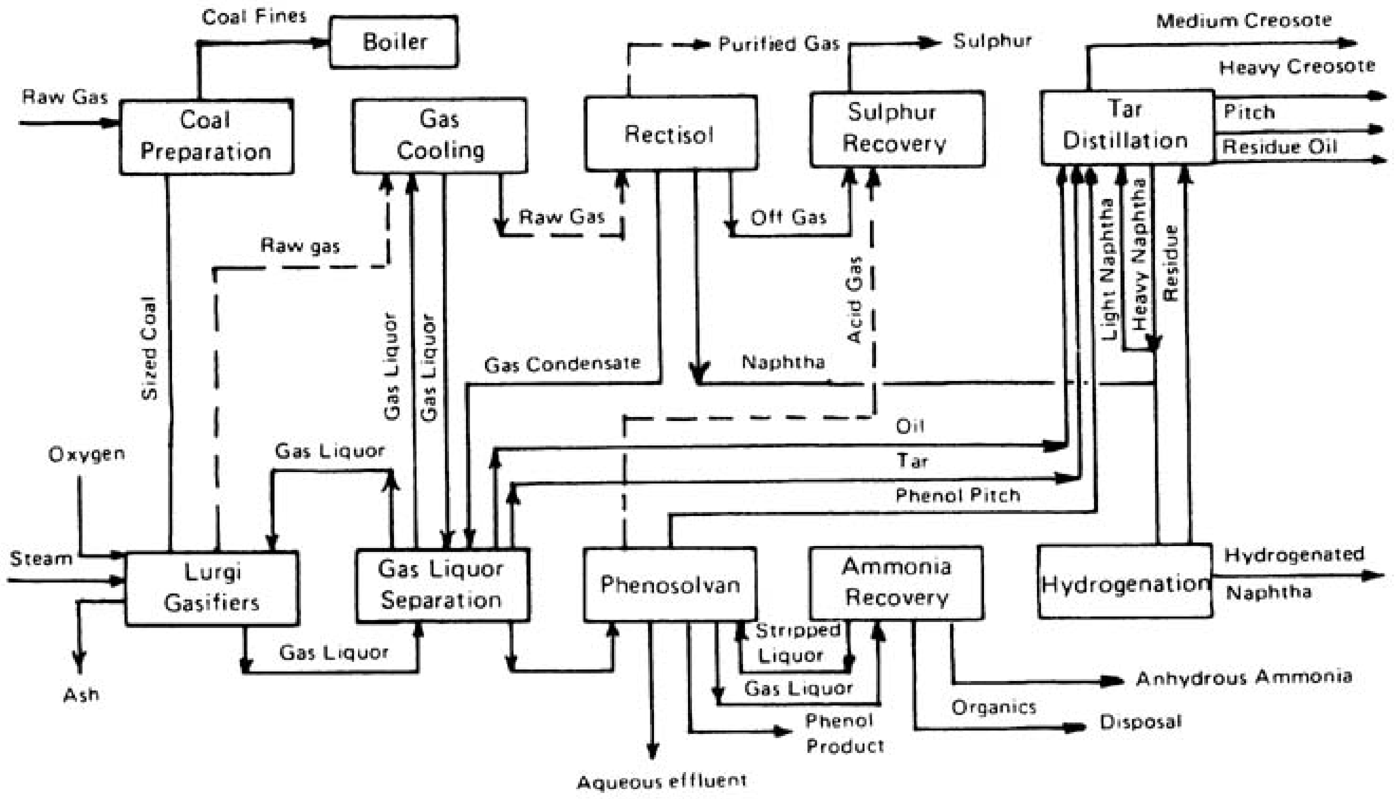
South Africa's Sasol installation is the world's largest synfuels plant based on coal. It converts an inferior, high-ash (35%) coal into a broad spectrum of products, which include ethylene, propylene, butylene, gasoline, and long straight-chain alcohols and hydrocarbons, SASOL I, which went into commercial operation in 1955, is part of a versatile chemical complex that supplies much of South Africa's needs for materials such as nitrogen fertilizers, plastics, and pipeline gas. Five million tons of coal per year are gasified in Lurgi gasifiers with steam and oxygen under pressures of about 20 atm. The principal components of the syngas are H_2, CO, CO_2, and CH_4, with the key materials being H_2 and CO. The H_2/CO ratio can be adjusted by changing the amounts of H_2O and O_2 used in the gasification.

Other constituents present in the raw syngas include NH_3, H_2S, other sulfur-containing compounds, and tars. The raw product is purified by cooling the gas, condensing water and phenolics, and passing the remaining gas through three absorption trains using very cold methanol. This essentially removes all the gaseous constituents except H_2, CO, and CH_4. Sulfur in the purified gas amounts to less than 0.01 ppm, thus virtually eliminating sulfur poisoning of catalysts and ensuring a low sulfur final product. A flow diagram is shown in Fig. 17.24.

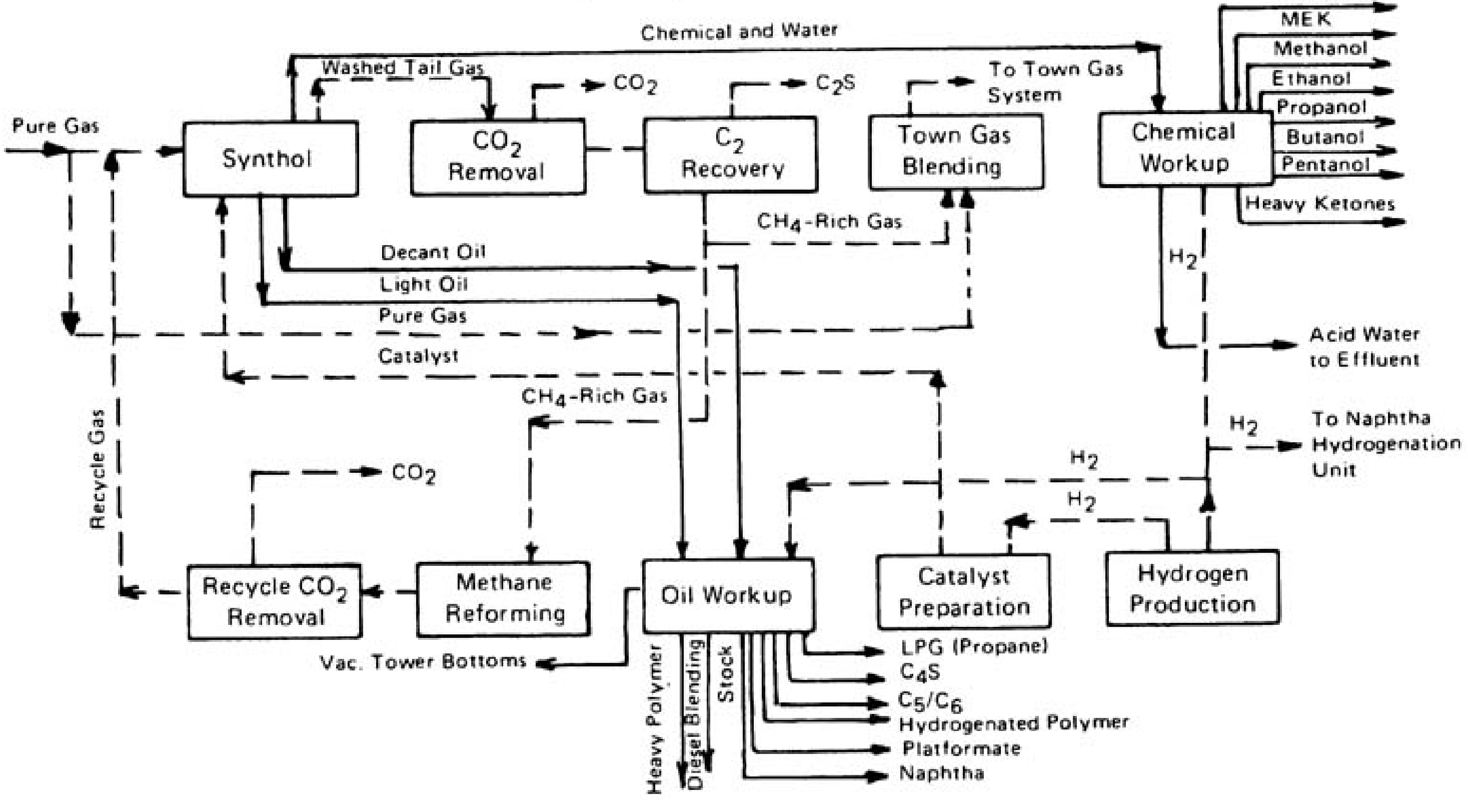
Fig. 17.24 Block flow diagram of Sasol's coal-to-oil process. (*Reprinted from* Encyclopedia of Chemical Processing Design, *pp. 310, 320, by courtesy of Marcel Dekker, Inc., 1979.*)

(a) Sasol's Two Main Coal-Into-Oil Processes

Coal gasification and purification flow diagram



(b) Fischer-Tropsch Synthesis Flow Diagram



Sasol II, with an output of 50,000 barrels per day of products, came on stream in 1980. This was followed in 1983 by Sasol III. These two new plants were designed for zero emissions of liquid wastes. With the addition of these projects, which cost over $7 billion, South Africa had the capacity to provide approximately 40 percent of its liquid fuels and chemicals needs from coal.

In 1980, Sasol introduced a second generation of its fluidized-bed FT technology (the Synthol process), and in 1989 a third generation. Each new generation has utilized improved engineering design and has enhanced the utilization of the chemicals generated. At Sasol, the process is operated to produce primarily gasoline and diesel fuels, but it can be modified by using different catalysts and reaction conditions to enhance the yield of oxygenated products that can be recovered for chemical uses.62,63 Some of the oxygenates available from the process are methanol, ethanol, propanol, butanone, and higher alcohols and ketones. FT olefins can be separated and used for production of oxo chemicals, synthetic rubber, and alkylbenzene sulfonate detergents.

Hydrocarbon waxes produced in a fixed-bed reactor, which has operated since 1955, have found a variety of uses. Also, by-products from the Sasol Lurgi coal gasifiers are recovered for chemical and solvent applications. These products include phenol, cresols, toluene, xylenes, ammonia, and sulfur. An addition to the spectrum of chemical products from Sasol is polypropylene. Also, ethane is being cracked to supplement ethylene production for sale to polyethylene producers. Additional work is in progress to evaluate the recovery of organic acids from aqueous waste streams.

Methanol

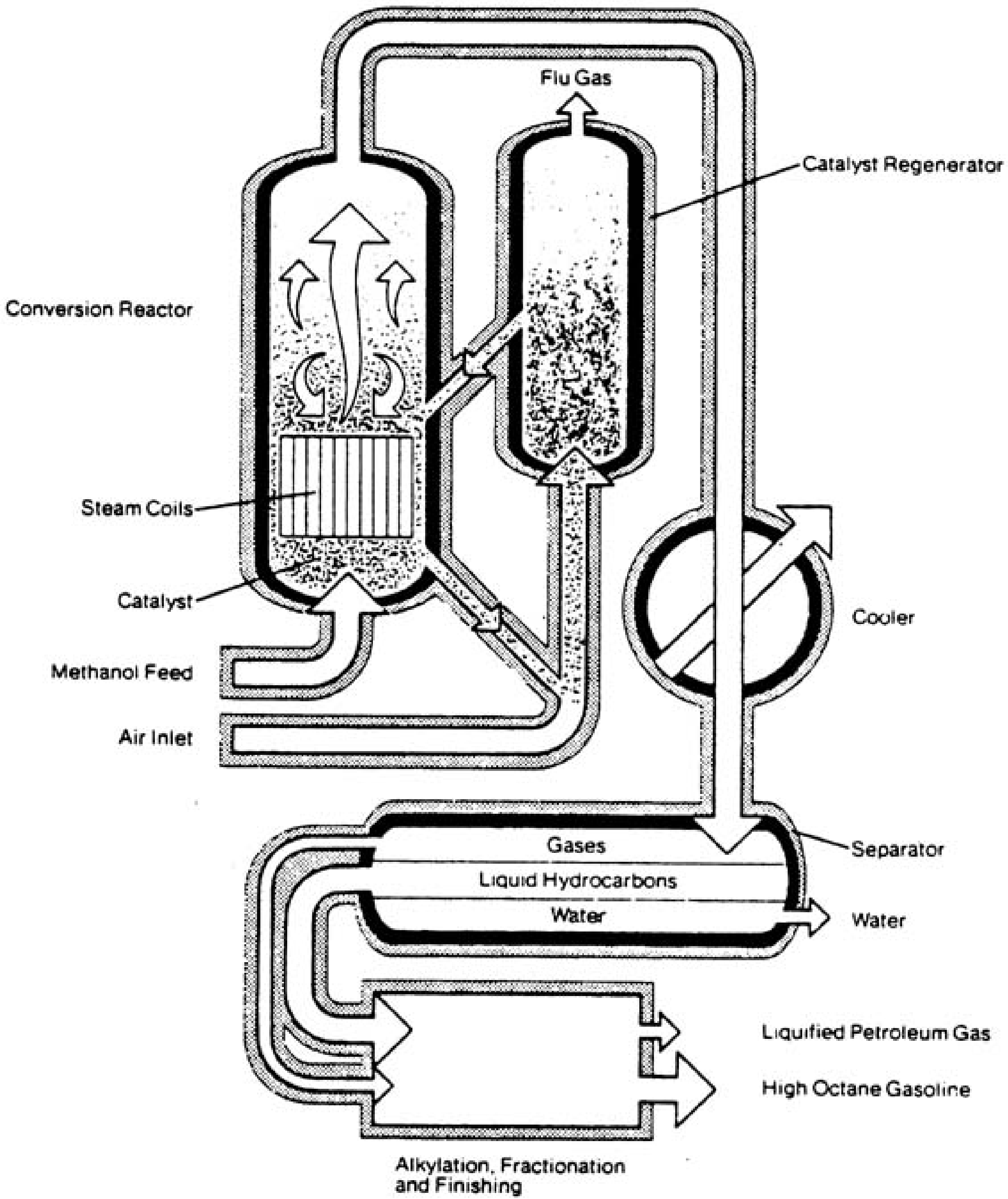
Methanol is a water-soluble, low molecular weight alcohol that may be of increasing importance as a low-sulfur fuel, a chemical feedstock, and perhaps as a gasoline additive or an intermediate in gasoline production. The synthesis of methanol is accomplished by the catalytic conversion of synthesis gas containing two moles of hydrogen for each mole of carbon monoxide. Methanol synthesis is widely practiced in industry on a commercial scale. See Chapter 22 for a discussion of methanol manufacture.

Methanol to GasolineThe Mobil Process

Mobil Research and Development Corporation developed a process that catalytically dehydrates and polymerizes methanol to produce a high octane unleaded gasoline. The catalyst is one of a new family of synthetic zeolites designated ZSM-5 by Mobil. These new zeolites have a unique channel structure, different from previously known wide-pore (910 Å in diameter) and other narrow-pore (5 Å) zeolites. One of these new zeolites converts methanol into a mixture of hydrocarbons corresponding to high-quality gasoline. The mass balance shows 0.44 lb of gasoline and 0.56 lb of water from each pound of methanol; however, the gasoline produced contains 95 percent of the heating value in the methanol feed.

A schematic diagram of the process is shown in Fig. 17.25. Crude methanol is vaporized at the bottom of the reactor and passes through a dense fluidized bed at 775°F and 25 psig. The methanol is converted to hydrocarbons and water. The catalyst is separated from reaction products in a disengager section at the top of the reactor, the reactor effluent is condensed, and the water and hydrocarbon products are separated. To make additional gasoline, propene and butenes can be alkylated with isobutane by conventional petroleum technology (see Chapter 15).

Fig. 17.25 Schematic of Mobil's fluid-bed MTG process, which uses a unique zeolite catalyst to convert methanol to high octane, unleaded gasoline. (*Courtesy The Pace Company, Denver, CO*)



Portions of the powdered catalyst are periodically removed from the reactor, regenerated with air, and returned to the reactor. Small amounts of carbon monoxide, carbon dioxide, and coke are formed as by-products.

17.9 Petrochemical Feedstocks

Feedstocks for the petrochemical industry are produced mainly from crude oil (Chapter 15) and natural gas (Chapter 16). About 90 percent of all organic chemicals are manufactured from just six feedstocks: synthesis gas, ethylene, propylene, butadiene, benzene, and *p*-xylene, with synthesis gas accounting for over half of these feedstocks. Although a lot of syngas for the manufacture of ammonia, methanol, and other chemicals is produced by steam reforming of natural gas or light oils, as discussed above syngas can also be produced by coal gasification.

Many references discuss and evaluate processes for the production of chemicals from coal.4,6466 Most basic chemicals currently produced from oil or natural gas can be produced from coal using demonstrated technology. The broadest range of synthesized products is available by conversion of coal to synthesis gas and utilization of existing processes for production of chemicals, such as ammonia, methanol, and acetyl derivatives. Advances in the chemistry of synthesis gas are rapidly expanding the array of chemicals obtainable from coal by this route.6769 Many other chemicals can be separated from the by-products of gasification and pyrolysis processes. In fact, the foundation of the modern chemical industry was built on processes utilizing coal hydrogenation and pyrolysis to obtain a variety of chemicals, including dyes, solvents, fuels, and pharmaceuticals.70 However, in the 1950s the wide availability and low cost of crude oil made petroleum-derived materials the preferred source for chemicals production.

Today, the choice of feedstock for chemicals production depends on complex technical, economic, environmental, and political factors. Clearly, not all chemicals are suitable for production from coal with current technology. Some factors to be considered in the evaluation of the appropriate feedstock for a particular chemical product are: (1) the relationship between the carbon/hydrogen ratio in the chemical product and the feedstock, (2) the delivered cost of alternative raw materials, (3) capital costs, (4) environmental protection, and (5) the reliability of supply. Recently, except for special situations, such as that for Sasol in South Africa, the manufacture of chemicals from coal at coal prices relative to the prices of petroleum and natural gas has not been attractive economically.

Chemicals from Coal

The use of coal as a raw material for the production of chemicals started with the discovery of the first coal tar dye by Sir William Perkins in 1856. Prior to that time, the liquid by-products of coking were regarded as wastes. Perkins' discovery revealed that valuable materials could be isolated or made from coal tar and ultimately led to the establishment of the modern chemical industry. Coke ovens rapidly came to be important as sources of chemicals, such as benzene, naphthalene, ammonia, and hydrogen. Coal tar became an important source for aromatic and heterocyclic chemicals for the dye and pharmaceutical industries.

Coke production remains the most important demonstrated technology associated with the direct production of chemicals from coal. Industrial chemicals currently obtained in significant amounts as coke by-products include benzene, toluene, xylene, naphthalene, anthracene, phenanthrene, phenol, ammonia, ammonium sulfate, sulfur, and carbon dioxide. The vast majority of aromatics production from coal occurs in Eastern Europe, India, and Japan.75

Chemical by-products from coke manufacture amount to roughly 4 percent of the coal feed. Although this yield initially supplied adequate quantities of chemicals for developing end uses, the growth of the industry eventually required additional sources of chemicals. The advent of technology for producing calcium carbide and acetylene in 1892 provided a means for conversion of coke to chemical products.71 Also, this technology stimulated the growth of an independent chemical industry by allowing the manufacture of chemicals to be pursued apart from the manufacture of coke for steel production. More recently, independent production of chemicals from coal has gone through the synthesis gas route.

Nevertheless, chemicals from coal typically have been obtained as by-products from processes primarily designed to produce other products, such as motor fuel, town gas, or coke. Except under certain favorable conditions, economics still suggest that the production of chemicals must be coupled with the manufacture of other end products. The combination of chemical and fuel production results in synergism that can benefit the economics of both products. Thus, some chemicals that are not economic as primary products from coal may be obtained as revenue-generating co-products that enhance overall plant economics. This concept has prompted efforts at Great Plains in North Dakota and Sasol to increase the recovery of chemical by-products. At Great Plains, ammonia and sulfur are recovered, but facilities are being added to recover phenol from gasification liquids and a kryptonxenon mixture from the air separation plant. Research and development efforts are in progress to enable profitable recovery of other by-products.

In recent years several commercial plants have been constructed for conversion of coal to synthesis gas for chemical manufacturing. These include the Eastman Chemical's acetic anhydride plant, the Ube (Japan) ammonia plant, and the SAR (Germany) oxo chemicals plant. Of these, the Eastman plant is the only one still operating exclusively on coal. Feedstock changes at the other plants illustrate the vulnerability of coal conversion processes to a changing economic climate. The fact that the Eastman process remains competitive under changing conditions is due to a set of special circumstances that favor a coal-based process. The success of the Eastman chemicals from coal complex demonstrates that synthesis gas from coal is a viable feedstock for some industrial chemicals under certain conditions.

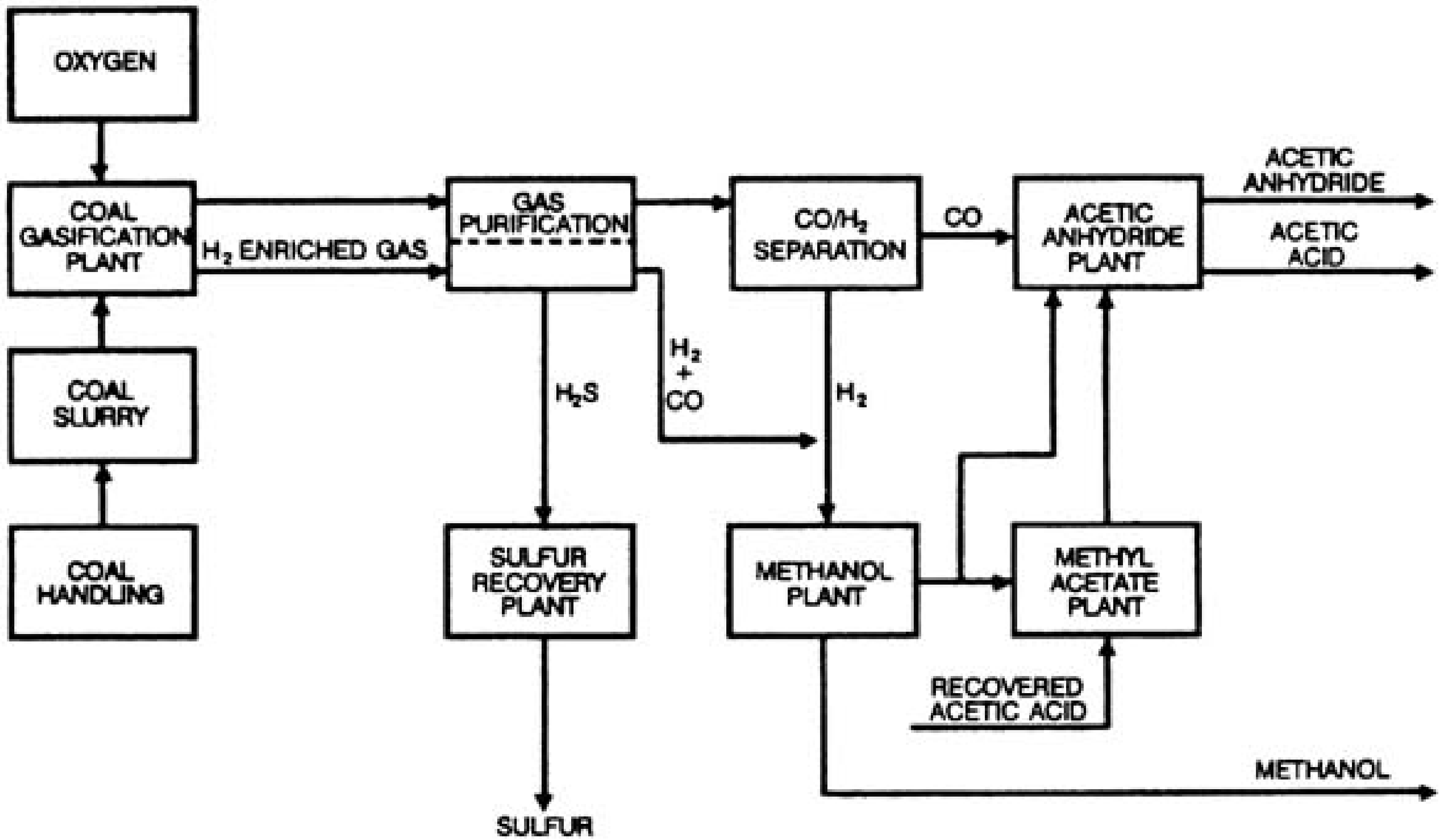
Clearly, the number of chemicals produced from coal-derived synthesis gas can expand as new technologies are developed and favorable economics exist. The most likely such chemicals are those for which processes have been demonstrated but which presently are uneconomic. Relatively small improvements in technology, shifts in feedstock availability and/or cost, decreased capital costs, or political factors could enhance the viability of coal-based processes for the production of methanol, ethanol, and higher alcohols, vinyl acetate, ethylene glycol, carboxylic acids, and light olefins.

New technologies are being developed for the partial oxidation of coal to organic acids, for plasma pyrolysis of coal to produce acetylene and other small molecules,72 and for the manufacture of additional chemicals such as acetaldehyde from synthesis gas.73 Also, the use of coal tar for the manufacture of high-temperature engineering plastics74 and bioconversion of coal to aromatics, organic acids, alcohols, and methane75 have been reported. Innovative strategies are being considered that offer the potential for reduced costs; for example, by co-gasifying coal with waste, waste disposal credits might be obtained.76

Examples of Chemicals Production from Coal

The first U.S. manufacturer to produce a slate of industrial chemicals from coal was Eastman Chemical Company (Division of Eastman Kodak Company). Commercial facilities include a coal gasification plant for synthesis gas manufacture, raw gas cleanup and separation facilities, a sulfur recovery unit, a coal-fired steam plant, and chemical plants to produce methanol, methyl acetate, acetic acid, and acetic anhydride. A flow diagram of Eastman's chemicals from coal plant is shown in Fig. 17.26.

Fig. 17.26 Overall block flow diagram for coal gasificationacetic anhydride complex.



A Texaco coal gasifier provides synthesis gas at elevated pressure from local coal. Medium- and low-pressure steam for use elsewhere in the complex are produced by recovering waste heat. After the gasifier product is scrubbed with water to cool the gas and remove ash particles, a portion of the syngas is sent to a watergas shift reactor to increase its hydrogen content. A Rectisol unit, using a cold methanol wash, then removes hydrogen sulfide and carbon dioxide from the product gas streams. Cryogenic separation in a Linde "cold box" provides a carbon monoxide stream for the acetic anhydride plant and a hydrogen-rich stream for methanol production. The recovered hydrogen sulfide is converted to elemental sulfur in a Claus unit followed by a SCOT tail gas treating unit.

Methanol is produced from carbon monoxide, carbon dioxide, and hydrogen. The proper feed composition is achieved by combining the hydrogen-enriched syngas from the shift reactor and the hydrogencarbon monoxide stream from the gas separation unit. The feed stream goes to both an energy-efficient, low-pressure, gas-phase, catalytic methanol process and a liquid-phase process. Methanol is reacted with acetic acid to form methyl acetate. In the final step of the process, purified carbon monoxide from the gas separation plant is reacted with methyl acetate to form acetic anhydride. Part of the acetic anhydride is reacted with methanol to co-produce acetic acid. The methyl acetate resulting from this reaction is recarbonylated in the acetic anhydride reactor. The acetic anhydride is used to make cellulose acetate. The reaction produces by-product acetic acid that is recycled to the methyl acetate plant to react with methanol.

Another example of large-scale production of chemicals from coal is the Sasol facility in South Africa. Sasol produces many products from coal-derived syngas, including ethylene, propylene, α-olefins, alcohols, and ketones. They have also increased their production of methanol, synthetic lubricants, detergent alcohols, acrylic acid and acrylates, oxoalcohols, styrene and polystyrene, propylene oxide, and propylene glycol.

As the above examples illustrate, the potential exists for producing larger quantities of commodity chemicals from coal. Some factors that could facilitate the increased use of coal for this purpose include price increases for petroleum and increased use of coal gasifiers for power production. As IGCC plants are built, there will be an opportunity to incorporate in the plant design the co-production of high-value chemicals.

17.10 References

Dryden, I. G. C., " Chemical Constitution and Reactions of Coal," in Chemistry of Coal Utilization, Supplementary Volume, H. H. Lowry (Ed.), pp. 223295, John Wiley & Sons, New York, 1963.

Wender, I., Heredy, L. A., Neuworth, M. B., and Dryden, I. G. C., "Chemical Reactions and the Constitution of Coal," in Chemistry of Coal Utilization, M. A. Elliott (Ed.), Second Supplementary Volume, pp. 425522, John Wiley & Sons, New York, 1981.

Gavalas, G. R., "Coal Pyrolysis," in Coal Science and Technology, Vol. 4, Elsevier, 1982.

Lin, Y. A. (Ed.), Physical Cleaning of CoalPresent and Developing Methods, Marcel Dekker, New York, 1982.

Leonard, J. W. (Ed.), Coal Preparation, 5th ed., American Institute of Mining, Metallurgical and Petroleum Engineers, Inc., New York, 1992.

Osborne, D. G., Coal Preparation Technology, Graham & Trotman, 1989.

Gala, H. B., Srivastava, R. D., Rhee, K. H., and Hucko, R. E., "Coal Preparation," Vol. 7, No. 1, July 1, 1989.

Sakkestad, B. A. (Ed.), The Proceedings of the 23rd International Technical Conference on Coal utilization & Fuel Systems, Coal & Slurry Technology Association, Washington, DC, 1998.

Stultz, S. C., and Kitto, J. B. (Eds.), SteamIts Generation and Use, The Babcock & Wilcox Company, Barberton, Ohio, 1992.

Singer, J. G. (Ed.), Combustion Fossil PowerA Reference Book on Burning and Steam Generation, Combustion Engineering, Inc., Windsor, Connecticut, 1991.

Bartok, W., and Sarofim, A. F. (Eds.), Fossil Fuel CombustionA Source Book, John Wiley & Sons, New York, 1991.

Bartok, W., Lyon, R. K., McIntyre, A. D., Ruth, L. A., and Sommerlad, R. E., "Combustors: Applications and Design Considerations," Chem. Eng. Progr., 84 (3), 54 (1988).

Ceely, F. J., and Daman, E. L., "Combustion Process Technology," in Chemistry of Coal Utilization, M. A. Elliott (Ed.), WileyInterscience, New York, 1981.

Department of Energy, Clean Coal Technology Compendium, www.lanl.gov/projects/cctc

Corcoran, E., "Cleaning Up Coal," Sci. Am., 264(5), 106 (1991).

Haggin, J., "Coal," Chem. Eng. News, p. 32 (June 17, 1991).

Environmental Protection Agency, "Mercury Study Report to Congress Volume VIIAn Evaluation of Mercury Control Technologies and Costs," EPA-452/R-97-010 (Dec. 1997).

Brown, T., Smith, D., Hargis, R., and O'Dowd, W., "Mercury Measurement and Its Control: What We Know, Have Learned, and need to Further Investigate;" J. Air Waste Management Assn., 197 (June 1999).

Schobert, H. H., Coal: The Energy Source of the Past and the Future, American Chemical Society, Washington, DC, 1987.

Haggin, J., Chem. Eng. News, 69(21), 33 (1990).

Sondreal, E. A., Ness, R. O., Timpe, R. C., Knudson, C. L., Jha, M. C., Berggren, M. H., and Sinor, J. E., Proc. Sixth Annual Int. Pittsburgh Coal Conf., Vol. 2, pp. 785794, Sept. 2529, 1989.

Synthetic Fuels Assn., Inc., "Coal Gasification Systems: A Guide to Status, Applications and Economics," EPRI AP-3109, Project 2207 (June 1983).

SFA Pacific, Inc., "Coal Gasification Guidebook: Status, Applications, and Technologies," Research Project 2221-39, prepared for EPRI, Final Report (Dec. 1993).

Shinner, R., Fourth Technical Meeting of COGARN, presented in "Coal Gasification: Direct Applications and Syntheses of Chemicals and Fuels; A Research Needs Assessment," by Penner, S. S., and others for US DOE, Office of Energy Research, DOE Contract No. DE-AC01-85ER30076, NTIS-PR-360 (1987).

Anderson, R. B., The FischerTropsch Synthesis, Academic Press, London, 1984.

Newman, S. A. (Ed.), Acid and Sour Gas Treating Processes, Gulf Publishing Co., Houston, 1985.

Spencer, D. F., Alpert, S. B., and Gluckman, M. J., "Integrated Coal Gasification Combined Cycles (IGCC) an Emerging Commercial Option for the Power Industry," paper presented at 1985 ACS Meeting, CA, 1985.

Spencer, D. F., 1986, The Commercial Implications of the Cool Water Project for the Electric Power Industry, EPRI, Palo Alto, CA, 1986.

Khan, M. R., and Kurata, T., in "The Feasibility of Mild Gasification of Coal: Research Needs," DOE/METC-85/4019, NTRS/DE85013625, 1985.

Probstein, R. F., and Hicks, R. E., Synthetic Fuels, McGraw-Hill, New York, 1982.

Seglin, L., and Bresler, S. A., "Low-Temperature Pyrolysis Technology," in Chemistry of Coal Utilization, M. A. Elliott (Ed.), Second Supplementary Volume, pp. 785846, John Wiley & Sons, New York, 1981.

Howard, J. B., "Fundamentals of Coal Pyrolysis and Hydropyrolysis," in Chemistry of Coal Utilization, M. A. Elliott (Ed.), Second Supplementary Volume, pp. 665784, John Wiley & Sons, New York, 1981.

Solomon, P. R., and Serio, M. A., "Evaluation of Coal Pyrolysis Kinetics," in Fundamentals of Physical Chemistry of Pulverized Coal Combustion, J. Lahaye and G. Prado (Eds.), Martinus Nijhoff Publishers, 1987.

Khan, M. R., "Thermoplastic Properties of Coal Pyrolysis at Elevated Pressures: Effects of Experimental Variables, Inorganic Additives, and Preoxidation," Ph.D. thesis, The Pennsylvania State University, pp. 333 (1985).

Anthony, D. B., and Howard, J. B., "Coal Devolatilization and Hydrogasification," AlChE J., 24(2), 625656 (1976).

Solomon, P. R., Serio, M. A., and Suuberg, E. M., "Coal Pyrolysis: Experiments, Kinetics Rates and Mechanisms," Progress in Energy and Combustion Science (1992), Vol. 18, No. 2, pp. 133220.

Arthur, J. R., and Wadsworth, K. D., "The Effect of Inorganic Compounds on the Behavior of Coals and Cokes," BCURA Progress in Coal Science, pp. 396410 (1950).

Gryaznov, N. S., and Petrov, V. K., Coke and Chemistry, pp. 2732, USSR (Coal Tar RA Translation), 1960.

Given, P. H., and Yarzab, R. F., "Analysis of the Organic Substance of Coals: Problems Posed by the Presence of Mineral Matter," in Analytical Methods for Coal and Coal Products, Vol. 11, C. Karr, Jr. (Ed.), pp. 341, Academic Press, Inc., New York, 1978.

Walker, P. L., Jr., Davis, A., Verina, S. K., Rivera-Utrilia, J., and Khan, M. R., "Interaction of Gases, Vapors, and Liquids with Coals and Minerals at Ambient TemperaturesTheir Use to Characterize Porosity," The Pennsylvania State University, DOE-30013-19, Under Contract No. DE-AC22-80PC30013 (1984).

Gray, R. H., Drucker, H., and Massey, M. J. (Eds.), Toxicology of Coal Conversion Processing, John Wiley & Sons, New York, 1988.

Neavel, R. C., "Liquefaction of Coal in Hydrogen-Donor and Non-donor Vehicles," Fuel, 55, 237242 (1976).

Derbyshire, F., and Stansberry, P., "Comments on the Reactivity of Low-Rank Coals in Liquefaction," Fuel, 66, 17411742 (1987).

Wu, W. R. K., and Storch, H. H., "Hydrogenation of Coal Tar," U.S. Bureau of Mines Bulletin No. 633 (1968).

Donath, E. E., and Hoering, M., "Early Coal Hydrogenation Catalysis," Fuel Proc. Tech., 1, 320 (1977).

Ministry of Fuel and Power, Report on the Petroleum and Synthetic Oil Industry of Germany, Ministry of Fuel and Power, His Majesty's Stationery Office, London, 1947.

Green, D. W. (Ed.), Perry's Chemical Engineers' Handbook, 7th ed., McGraw-Hill, New York, 1997.

Derbyshire, F., "Catalysis in Coal Liquefaction: New Directions for Research," IEACR/08, IEA Coal Research, London, 1968.

DOE, "Coal Liquefaction: A Research and Development Needs Assessment," DOE Coal Liquefaction Research Needs (COLIRN) Panel Assessment, Final Report, Vol. 2, DOE/ER-0400 (Mar. 1989).

Derbyshire, F., "Coal Liquefaction," in Ullmann's Encyclopedia of Industrial Chemistry, Vol. A7, VCH Publishers, New York, 1986.

Maa, P. S., Trachte, K. L., and Williams, R. D., "Solvent Effects in Exxon Donor Solvent Coal Liquefaction," presented at the ACS National Meeting, New York, Aug. 2328.

Neavel, R. C., "Exxon Donor Solvent Liquefaction Process," Phil. Trans. R. Soc. Lond. A, 300, 141156 (1981).

Tomlinson, G. C., Gray, D., Neuworth, M. B., and Talib, A., "The Impact of Rank-Related Coal Properties on the Response of Coals to Continuous Direct Liquefaction Processes," Sandia Contractors Report, SAND85-7238 (Oct. 1985).

Holmes, D. F., Jr., Mitchell, J., and Farthing, S., "A Final Report of Kentucky's Role in the H-Coal Pilot Plant Project," Kentucky Energy Cabinet (Aug. 1983).

EPRI, "Two-Stage Coal Liquefaction Integrated Configurations: The Advanced Coal Liquefaction R&D Facility, Wilsonville, Alabama," EPRI GS-7293 (May 1991).

Gollakota, S. V., Lee, J. M., and Davies, O. L., "Process Optimization of Close-Coupled Integrated Two-stage Liquefaction by the Use of Cleaned Coals," Fuel Proc. Technol., 22, 205216 (1989).

Nalitham, R. V., Lee, J. M., Lamb, C. W., and Johnson, T. W., "Two-Stage Coal Liquefaction Process Performance with Close-Coupled Reactors," Fuel Proc. Technol., 17, 1327 (1987).

Rosenthal, J. W., Dahlberg, A. J., Kuehler, C. W., Cash, D. R., and Freedman, W., "The Chevron Coal Liquefaction Process (CCLP)," Fuel, 61, 10451049

(1982).
El Sawy, A., Gray, D., Talib, A., and Tomlinson, G., "A Techno-Economic Assessment of Recent Advances in Direct Coal Liquefaction," Sandia Contractor Report SAND86-7103 (June 1986).
Derbyshire, F. J., "Role of Catalysis in Coal Liquefaction Research and Development," Energy and Fuels, 3, 273277 (1989).
Wender, I., "Reactions of Synthesis Gas," Fuel Proc. Technol., 48(3) (September 1996).
Schultz, H., Pure Appl. Chem., 51, 22252241 (1979).
Dry, M. E., in Chemicals From Coal: New Processes, K. R. Payne (Ed.), Critical Reports on Applied Chemistry, Vol. 14, Chapter 4, John Wiley & Sons, Chichester, 1987.
Oara, J. B., et al., "Petrochemical Feedstocks from Coal," Chem. Eng. Prog., 6472 (June 1977).
O'Hara, J. B., et al., "Project POGOA Coal Refinery," Chem. Eng. Prog., 4663 (Aug. 1978).
Conference on Chemical Feedstocks Alternatives, American Institute of Chemical Engineers and National Science Foundation, Houston, Texas, October 25, 1977.
Wender, I., in Chemicals from Coal: New Processes, K. R. Payne (Ed.), Critical Reports on Applied Chemistry, Vol. 14, Chapter 5, John Wiley & Sons, Chichester, 1987.
Cornils, B., in Chemicals from Coal: New Processes, K. R. Payne (Ed.), Critical Reports on Applied Chemistry, Vol. 14, Chapter 5, John Wiley & Sons, Chichester, 1987.
Pruett, R. L., Science, 211, 1116 (1981).
Shreve, R. N., Chemical Process Industries, Chapter 5, McGraw-Hill, New York, 1967.
Spitz, P. H., CHEMTECH, 19, 92100 (1989).
Anderson, L. L., and Tillman, D. A., Synthetic Fuels from Coal, WileyInterscience, New York, 1979.
Moloy, K. G., and Wegman, R. W., Organometallics, 8, 2883 (1989).
Jedlinski, Z. J., and Gaik, U., Polimery, 29, 424 (1984).
Haggin, J., Chem. Eng. News, 67(7), 28 (1989).
Coal and Synfuels Technology, 11(16), 1 (1990).

18.1 Introduction

The word "rubber" immediately brings to mind materials that are highly flexible and will snap back to their original shape after being stretched. In this chapter a variety of materials are discussed that possess the characteristics describing this class of materials. There will also be a discussion on the mechanism of this "elastic retractive force." Originally, rubber meant the gum collected from a tree growing in Brazil. The term "rubber" was coined for this material by the English chemist Joseph Priestley, who noted that it was effective for removing pencil marks from paper. Today, in addition to Priestley's natural product, many synthetic materials are made that possess these characteristics and many other properties. The common features of these materials are that they are made up of long-chain molecules that are amorphous (not crystalline), and the chains are above their glass transition temperature at room temperature.

Rubber products appear everywhere in modern society from tires to biomedical products. The development of synthetic rubber began out of the need for countries to establish independence from natural products that grew only in tropical climates. In times of conflict the natural product might not be available, and its loss would seriously threaten national security. Synthetic rubber, then, became a strategic concern during World Wars I and II.1 Beyond the security issue, the need for materials with better performance also provided a strong impetus for the development of new rubbery materials. In particular, improvements in oil resistance, high-temperature stability, and oxidation and ozone resistance were needed.2 Research today is driven to develop materials with even better performance in these areas. In the 1980s and 1990s tires with lower rolling resistance were demanded by car manufacturers to improve fuel economy. This was accomplished, in part, by developing functional tread polymers which chemically bond to fillers resulting in a dramatic reduction in the hysteresis (energy loss) of the tire tread. These new functional polymers will be discussed in this chapter.

The worldwide demand for rubber was estimated to be 15.27 million metric tons for 2000,3 excluding latex materials. Of this demand, 55 percent is synthetic rubber of various kinds. Rubber has been classified by use into general purpose and specialty.

The major general purpose rubbers are natural rubber, styrenebutadiene rubber, butadiene rubber, isoprene rubber, and ethylenepropylene rubber. These rubbers are used in tires, mechanical goods, and similar applications. Specialty elastomers provide unique properties such as oil resistance or extreme heat stability. Although this differentiation is rather arbitrary, it tends also to classify the polymers according to volumes used. Styrenebutadiene rubber, butadiene rubber, and ethylenepropylene rubber account for 74 percent of all synthetic rubber consumed.

The 1999 synthetic rubber capacity by type is presented in Table 18.1, and Table 18.2 lists the U.S. rubber consumption forecast by type for 2000.

TABLE 18.1 Synthetic Rubber Production by Type1999

Type	Description	Capacity (Metric Tons)
SBR-solid	Styrenebutadiene rubber	2,970,000
BR	Butadiene rubber	1,885,000
IR	Isoprene rubber	221,000
CR	Chloroprene (or neoprene)	243,000
EPDM	Ethylenepropylene terpolymer	853,000
NBR	Nitrile or acrylonitrilebutadiene rubber	287,000
Others		790,000
Total		7,249,000

Source: *Worldwide Rubber Statistics 2000*, International Institute of Synthetic Rubber Products, Inc.

TABLE 18.2 U.S. Rubber Consumption Forecast, 2000 (Metric Tons)

Total new rubber	3,711,000
Natural	1,253,000
Total synthetic	2,458,000
SBR	909,000
Nitrile	88,000
Polybutadiene	580,000
EPDM	350,000
Other synthetics	382,000
% Natural	33.8
% Synthetic	66.2

Source: *Worldwide Rubber Statistics 2000*, International Institute of Synthetic Rubber Producers, Inc.

18.2 Rubber Concepts

Several key principles (outlined below) greatly help in understanding the performance of elastomeric materials. This outline should present these concepts well enough for use in the following discussions of specific polymer types. However, the reader should consult the references for a more complete understanding of these principles. The concepts are classified as those that relate to polymer structure, those that relate directly to physical properties, and those that relate to the use of the material, as listed below.

1. Polymer structure:
    1. Macrostructure
        1. molecular weight
        2. molecular weight distribution
        3. branching
    2. Microstructure
    3. Network structure
2. Rubber properties:
    1. Elasticitythe retractive force
    2. Glass transition temperature
    3. Crystallinity
3. Rubber use:
    1. Compounding
    2. Processing

18.3 Polymer Structure

Macrostructure

Molecular Weight

The single most important property of any polymer is the size or length of the molecule. The polymerization process consists of combining many of the simple monomer molecules into the polymer chain. Most of the monomers used to produce rubbers are either gases or low-viscosity liquids under normal conditions; upon polymerization they form liquids whose viscosity increases to extremely high values as the chain length is increased. At very low chain length, this increase is linear with molecular weight until the chains are long enough to become entangled. Above the entanglement molecular weight, the viscosity increases to the 3.43.5th power of molecular weight increase.4 In addition to viscosity, a great many other physical properties of any polymer depend upon the molecular weight.5

Molecular Weight Distribution

A given polymer sample is composed of many polymer chains, which in most cases are not of the same length. This variability can be a result of the synthesis process or of possible random scission and cross-linking that can occur upon processing. For economic reasons, it is not possible to separate the various polymer chains by length prior to use; so it is important to characterize this distribution in order to describe the polymer and understand its performance. As with any distribution, no single number is a totally satisfactory descriptor.

The commonly used molecular weight parameters are the number, weight, and *z* average molecular weight, which are defined, respectively, as:6

$$M_n = \frac{\sum N_i M_i}{\sum N_i}$$

$$M_w = \frac{\sum N_i M_i^2}{\sum N_i M_i}$$

$$M_z = \frac{\sum N_i M_i^3}{\sum N_i M_i^2}$$

where *N*i is the number of moles of species *i*, and *M*i is the molecular weight of species *i*.

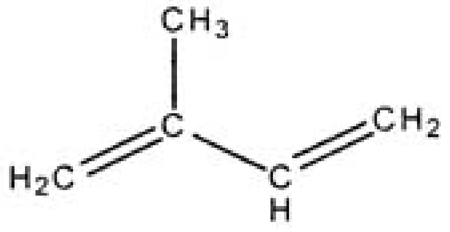
Although there are many different statistical ways to describe any population, the above parameters have been widely used because they are readily understood in physical terms, and they can be measured directly in the laboratory.7 A fourth parameter, the dispersion index, frequently is used to characterize the breadth of the distribution. This parameter is simply the ratio of the weight to the number average molecular weight, with 1.0 being the lowest possible number (i.e. all chains of exactly the same length). Typical values for commercial polymers are in the 25 range, with those under 2 considered relatively narrow and those over 2.5 considered broad in distribution. The measurement of these molecular weight averages once was a time-consuming task, but with the development of gel permeation chromatography (GPC), also referred to as size exclusion chromatography, the measurement of these distributions has become commonplace.8 Units are even available that automatically sample polymerization reactors, process the sample, and perform all necessary calculations to provide data for process control. These units can have multiple detectors, thereby providing compositional distribution as a function of molecular weight.

Branching

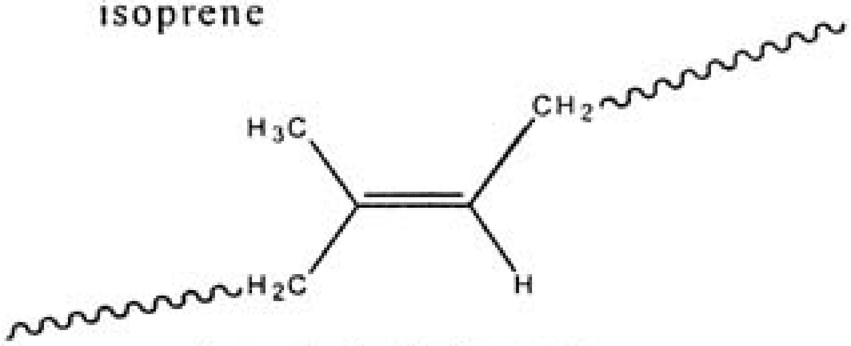
The concept of a polymer chain implies two ends per chain. However, because of the nature of the process used to form the polymer, the chain may contain one or more branch points, resulting in multiple ends per chain. These chain ends can have an adverse effect on polymer performance. Branching, molecular weight, and molecular weight distribution have been shown to affect processability as well.9 The optimum macrostructure often represents a compromise between processing and ultimate performance. Branching can also be measured using the GPC technique with special detectors.

Microstructure

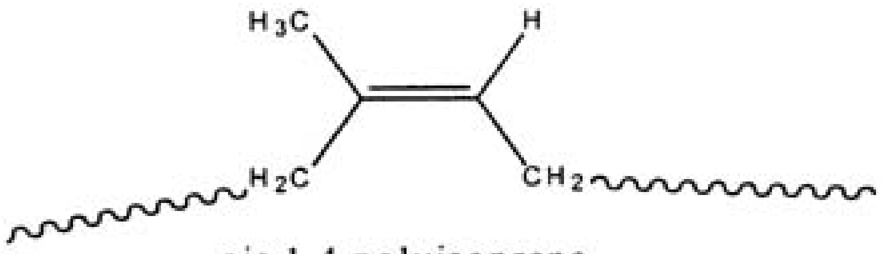
In the formation of elastomers from diolefin monomers such as butadiene or isoprene, there are a number of possible structures. Since the control of these structures is critical in obtaining optimum properties, this area has received great attention from the synthesis chemist. The possible polyisoprene structures are:



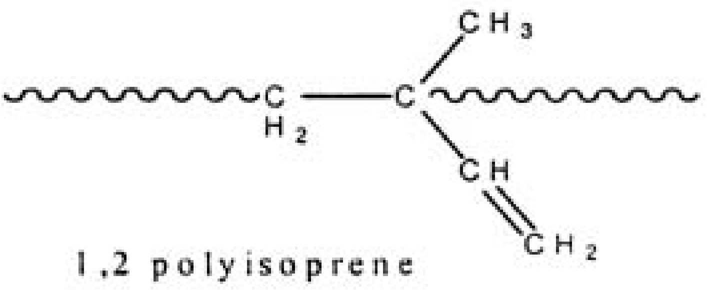
isoprene



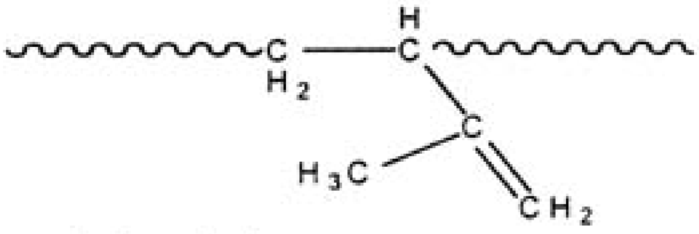
trans-1,4 polyisoprene



cis-1,4 polyisoprene



1,2 polyisoprene



3,4 polyisoprene

For butadiene (no methyl group) the 3,4 form does not exist. The 1,2 addition is referred to as *vinyl addition*.

The polymer in natural rubber (from the *Hevea brasiliensis* tree) is pure *cis* polyisoprene; gutta percha and balata are composed of the *trans* isomer.

Many of the commercial synthetic elastomers are synthesized from more than one monomer, such as styrenebutadiene and ethylenepropylene rubbers. The properties of the resultant polymer depend on the ratio of the two monomers in the polymer and upon the distribution of the monomers within the chain.

If the monomers are uniformly distributed within the polymer chain, the ratio of monomers will define the flexibility of the polymer chain.10 Because many properties depend on this chain mobility, polymer composition is carefully controlled. In addition to chain mobility, polymer composition also defines the solubility parameter of the polymer, which is a critical property relative to the type of solvents in which the polymer is soluble, the ability of the polymer to accept and hold oil, and the relative compatibility of the polymer with other polymers.11 Basically, these properties all involve polymersolvent interaction, with the difference being the increasing molecular weight of the solvent.

In addition to the relative ratio of the monomers, the arrangement of the units in the chain is important. This arrangement is referred to as the copolymer *sequence distribution*. In the previous discussion, the assumption was made that the comonomer units were well mixed in the polymer chain. If this is not the case, parts of the chain can reflect properties of the corresponding homopolymer. It is thus possible to produce polymers that have significantly different properties in different parts of the polymer chain. A most dramatic example of this can be found in styrenebutadienestyrene or styreneisoprenestyrene thermoplastic elastomers. The properties of these unique materials will be discussed in the section "Thermoplastic Elastomers."

Network Structure

A critical requirement for obtaining engineering properties from a rubbery material is its existence in a network structure. Charles Goodyear's discovery of vulcanization changed natural rubber from a material that became sticky when hot and brittle when cold into a material that could be used over a wide range of conditions. Basically, he had found a way to chemically connect the individual polymer chains into a three-dimensional network. Chains that previously could flow past one another under stress now had only limited extensibility, which allowed for the support of considerable stress and retraction upon release of the stress. The terms "vulcanization," "rubber cure," and "cross-linking" all refer to the same general phenomenon.

For most rubber systems the network is formed after the polymer is compounded and molded into the desired final shape. Once cross-linked, the material no longer can be processed. If cross-linking occurs prior to compounding or molding, the material is referred to as *gelled*, and it cannot be used. Most rubber is used in a compounded and cured form. There is an optimum *cross-link density* for many failure properties such as tensile strength and tear which will be discussed in the next section.

The one general class of polymers that fall outside this concept is the thermoplastic elastomers, which will also be discussed later.

18.4 Rubber Properties

ElasticityThe Retractive Force

The fact that cross-linked rubber materials can be extended to several times their original length and return to that original length when released is certainly their most striking feature. This is in contrast to crystalline solids and glasses, which cannot normally be extended to more than a fraction of their original length and also to ductile metals which can be extended to large deformations but do not return to the original length after the stress is removed.

There have been both statistical and thermodynamic approaches to solving the problem of rubber elasticity leading to a phenomenological treatment; however, these methods are beyond the scope of this chapter. The important, and most interesting, result of these theories is that rubber elasticity arises from changes in entropy of the network. Rubber molecules are capable of geometric isomerization, examples of which are *cis* and *trans*, just like any other unsaturated organic molecule. Upon stretching the network of chains, no "configurational" changes take place. That is, *cis* is not changed to *trans* and thus there is no configurational contribution to entropy. Elasticity *does not* arise from stretching or deforming covalent bonds either. What does take place are rotations about single bonds in the chain backbone during stretching. It is these "conformational" changes which give rise to the entropy decrease upon stretching. The chains do not like being in a stretched state because there are many conformational states not available to them until the stress is released. The chain ends are held apart at a statistically unfavorable distance, which gives rise to the entropy decrease. The following expression, for extensions > 10%, shows that the elastic force, f, is directly proportional to the absolute temperature, or the elastic response of the rubber is entirely governed by the decrease in entropy (S) which it undergoes upon extension.[12] The term δL is the change in length of the sample:

$$f = -T\left[\frac{\partial S}{\partial L}\right]_{T,V}$$

This expression was derived for constant temperature and volume experiments.

The shear modulus of the rubber network is related to the molecular weight between cross-link points or M_c. The lower the molecular weight of chains between cross-links (network chains), the higher the cross-link density and the higher the modulus. This is shown in the following expression:

$$G_0 = \frac{\rho RT}{M_c}$$

where G_0 is the elastic shear modulus, ρ is the density, R is the gas constant, and T is absolute temperature. There is an optimum cross-link density for ultimate strength properties, above which the highly cross-linked network no longer can dissipate strain energy in the form of heat (hysteresis) so all the energy goes into breaking network chains and the material becomes brittle. Below this optimum cross-link level the material simply has too much viscous flow and pulls apart easily. Therefore, the rubber chemist must optimize the state of cure if high strength is desired.

Glass Transition Temperature

In order for a polymer to behave as a rubbery material, it is necessary for the chain to have great mobility. As the temperature is lowered, the ability of the chain segments to move decreases until a temperature is reached where any large-scale motion is prevented. This temperature is referred to as the glass transition temperature (T_g). Below this temperature the rubber becomes a glassy materialhard and brittle. Above this temperature amorphous plastics, such as polystyrene, can exhibit a rubbery character if the molecular weight is sufficiently high. All rubbery materials, then, must have glass transition temperatures below room temperature. For good low-temperature properties, it obviously follows that a low glass transition temperature polymer is required. The control of the glass transition temperature of the polymer is critical for many properties other than low-temperature use. For example, the wet traction and wear of a passenger tire have been shown to greatly depend directly on the polymer glass transition temperature.[13]

Control of this property is possible by controlling the structure of the polymer chain. Monomers with bulky side groups restrict chain mobility and thus raise the glass transition temperature. The composition of copolymers and the ratio of polymer blends often are determined by the desired glass transition temperature of the final product.

The glass transition temperature is usually measured using thermal methods such as differential scanning calorimetry (DSC) which looks at the change in heat capacity when a material goes through its glass transition. Another very useful method is to use dynamic mechanical properties where the polymer is subjected to a temperature sweep in a dynamic mechanical spectrometer from very low temperature (-120°C) to well above room temperature (+100°C). If the modulus is plotted vs. temperature, there will be a sharp decrease as the polymer warms to above its T_g. There will also be a peak in the energy loss property known as tan delta. The tan delta vs. temperature plot is very useful for predicting tire properties such as rolling resistance and wet traction.

Crystallinity

Polymer chains can exist in spatial arrangements that are orderly enough to allow the chains to form crystalline structures. The existence of strong interchain interactions via polar forces, hydrogen bonding, or ionic groups can facilitate crystallization. The existence of crystallization is very important for many plastics and fibers, but crystallinity cannot be appreciably present in rubbery materials, as the corresponding restriction in chain mobility could preclude the very chain mobility needed for rubbery properties. Rubbery materials must have both their melting temperature (if they have a crystalline point) and their glass transition temperature below the use temperature.

Crystallinity can be reduced by disruption of the order in the chain by copolymerization.[14] For example, both polyethylene and polypropylene are crystalline plastics, whereas ethylenepropylene rubber produced at about a 50:50 ratio is an amorphous elastomer. Compositional excursions much outside this range lead to crystalline materials.[15] For some materials, such as natural rubber, that are close to crystallizing, stretching the chains can align them sufficiently for crystallization to occur. Such polymers can exhibit excellent gum properties and improved strength in the uncured state that greatly facilitate processing. Attempts have been made with some success to produce other polymers that exhibit this property of natural rubber. Although the melting temperature can be matched by appropriately disrupting the crystallizable structure through controlled introduction of another monomer, an exact match is not possible because the extent of crystallinity and the kinetics of crystallization will differ.

Crystallinity can be measured using the same thermal and dynamic mechanical methods described for measuring T_g however, the melting transition is much sharper than the T_g because it is a first-order transition compared with second-order for the T_g.

18.5 Rubber Use

Compounding

The rubber industry began when Charles Goodyear developed the first useful rubber compound: natural rubber plus sulfur. The concept of mixing materials into rubber to improve performance is still of primary importance today. Without compounding, few rubbers would be of any commercial value. Any given rubber application will have a long list of necessary criteria in addition to cost, encompassing appearance, processing, mechanical, electrical, chemical, and thermal properties. Developing such compounds requires a broad knowledge of material science and chemistry combined with experience. The use of designed experiments can greatly facilitate selecting the optimum compound formulation.

The major components in a compound are curatives, reinforcing agents, fillers, plasticizers, and antidegradants.

Curatives

The function of curatives is to cross-link the polymer chains into a network; the most common ones are the sulfur type for unsaturated rubber and peroxides for saturated polymers. Chemicals called accelerators may be added to control the cure rate in the sulfur system; these materials generally are complex organic chemicals containing sulfur and nitrogen atoms. Stearic acid and zinc oxide usually are added to activate these accelerators. Metal oxides are used to cure halogenated polymers such as polychloroprene or chlorosulfonated polyethylene.

Reinforcing Agents

Carbon black and silica are the most common reinforcing agents. These materials improve properties such as tensile strength and tear strength; also, they increase hardness, stiffness, and density and reduce cost. Almost all rubbers require reinforcement to obtain acceptable use properties. The size of the particles, how they may be interconnected (structure), and the chemical activity of the surface are all critical properties for reinforcing agents. In tire applications, new polymers are currently being developed which contain functional groups that directly interact with carbon black and silica, improving many properties.

Fillers

Fillers are added to reduce cost, increase hardness, and color the compound. Generally they do not provide the dramatic improvement in properties seen with reinforcing agents, but they may have some reinforcing capability. Typical fillers are clays, calcium carbonate, and titanium dioxide.

Plasticizers

These materials are added to reduce the hardness of the compound and can reduce the viscosity of the uncured compound to facilitate processes such as mixing and extruding. The most common materials are petroleum-based oils, esters, and fatty acids. Critical properties of these materials are their compatibility with the rubber and their viscosity. Failure to obtain sufficient compatibility will cause the plasticizer to diffuse out of the compound. The oils are classified as aromatic, naphthenic, or paraffinic according to their components. Aromatic oils will be more compatible with styrenebutadiene rubber than paraffinic oils, whereas the inverse will be true for butyl rubber. The aromatic oils are dark colored and thus cannot be used where color is critical, as in the white sidewall of a tire. The naphthenic and paraffinic oils can be colorless and are referred to as nonstaining.

Antidegradants

This group of chemicals is added to prevent undesirable chemical reactions with the polymer network. The most important are the antioxidants, which trap free radicals and prevent chain scission and cross-linking. Antiozonants are added to prevent ozone attack on the rubber, which can lead to the formation and growth of cracks. Antiozonants function by diffusion of the material to the surface of the rubber, thereby providing a protective film. Certain antioxidants have this characteristic, and waxes also are used for this purpose.

Processing

A wide range of processes are used to convert a bale of rubber into a rubber product such as a tire. The first process generally will be compounding. Typical compounding ingredients were discussed previously. In many compounds more than one rubber may be needed to obtain the performance required. Uncured rubber can be considered as a very highviscosity liquid; it really is a *viscoelastic* material possessing both liquid and elastic properties. Mixing materials into rubber requires high shear, and the simplest method is a double roll mill in which the rubber is shear-mixed along with the other compounding ingredients in the bite of the mill. Large-scale mixing is most commonly done with a high-shear internal mixer called a Banbury. This mixing is a batch process, although continuous internal mixers also are used.

The compounded rubber stock will be further processed for use. The process could be injection or transfer molding into a hot mold where it is cured. Tire curing bladders are made in this fashion. Extrusion of the rubber stock is used to make hose or tire treads and sidewalls. Another common process is calendaring, in which a fabric is passed through rolls where rubber is squeezed into the fabric to make fabric-reinforced rubber sheets for roofing membranes or body plies for tires. The actual construction of the final product can be quite complex. For example, a tire contains many different rubber components some of which are cord or fabric reinforced. All of the components must be assembled with high precision so that the final cured product can operate smoothly at high speeds and last over 50,000 miles.

18.6 Natural Rubber

More than 500 years ago, the people of Central and South America were using a product that they collected from certain trees to make balls and to coat fabric to make it waterproof. This material they called cauchuc, which means "weeping wood." Today we know the tree as the *Hevea brasiliensis* and the material as natural rubber. Although a number of plants produce rubber, the only significant commercial source is the *Hevea* tree. Natural rubber initially had only limited applications because it flowed when hot and had poor strength. In 1839 Charles Goodyear found that when combined with sulfur and heated, the material changed into cured rubber with properties much as we know them today. The development of the pneumatic tire in 1845 combined with the dramatic growth of the automotive industry led to a rapid increase in the demand for natural rubber.

Although the tree was indigenous to Brazil, seeds were taken to England where they were germinated, and the plants were sent to the Far East. Rubber plantations were in existence by the late 1800s, and in the 1920s plantations were begun in West Africa. Because of a leaf blight disease, essentially all natural rubber now comes from plantations in Africa and the Far East rather than Central and South America.16

The production process starts with the trees. Over the years considerable biological research has been done to produce trees that grow faster, produce more latex, and are resistant to wind and disease damage. Once such an improved tree has been identified, buds are grafted from the tree onto root stock. All such trees are referred to as clones and will have the same characteristics. It typically takes 67 years of growth before a tree is ready for rubber recovery. Peak rubber production is reached at 1215 years of age. Another major development in improving tree performance has been the use of tree stimulants, which has resulted in an overall yield increase of 30 percent without adverse effects on the trees.

The production process of natural rubber in the tree is not yet fully understood. However, it involves a long series of complex biochemical reactions that do not involve isoprene as a monomer, even though the resulting polymer is 100 percent *cis* 1,4 polyisoprene. Because the tree makes the product, the rubber production process is really one of recovery.

The recovery process starts with tapping of the tree, which involves manually removing thin sections of bark at an angle so that as the latex is exuded from the damaged living latex cells, it will flow down to be collected in a cup. The depth of the cut is critical, as a tool shallow cut will not allow optimum latex yield, and too deep a cut will damage the tree. The damaged latex vessels will seal off after several hours. A preservative is added to the collection cup to prevent coagulation of the latex. The tapper then collects the latex and takes it to a collection station, from which it is shipped to the rubber factory. Additional preservative is added at the collection station to assure stable latex. The latex contains around 3045 percent rubber. Approximately 3 percent of the solid material is nonrubber materials, consisting primarily of proteins, resins, and sugars.

The latex is processed in one of several ways, depending on the desired final product. If it is going to be used in the final product application in the latex form (such as for dipped goods and adhesives), it will be concentrated to 60 percent or higher solids. The most common process is centrifugation. The latex separates into the high-solids product and a low-solids skim material that contains a much higher percentage of the nonrubbery components; rubber produced from skim is generally of a lower quality because of these impurities. A variety of chemicals can be added to the latex to provide the necessary preservation and mechanical stability. As the final use of the latex will involve destabilizing the rubber particles, care must be taken not to overstabilize the latex concentrate. Being a natural product, the latex tends to change upon aging, a factor that also must be compensated for in the process.

Dry rubber is produced from the latex first by dilution, then by coagulation with organic acids, and finally by formation into sheets or crepe. Rubber smoked sheets are made by working the coagulated sheets between rolls to remove as much of the nonrubbery components as possible, followed by drying for up to a week in a smokehouse. The smoke serves as a fungicide that prevents biological attack of the rubber. To provide a more well-defined uniform product, Technically Specified Rubber (TSR) processes have been developed. These processes involve converting the coagulated rubber into rubber crumb, which is further washed, dried, and baled. Constant-viscosity grades of natural rubber have been developed by chemically reacting the aldehyde groups, which otherwise would lead to cross-linking upon storage. In the pale crepe process the latex is carefully selected for colored bodies (from carotene) and treated with sodium bisulfite to stop enzyme activity. The rubber is extensively washed through rollers to remove serum, as this can lead to yellowing. The wet crepe is dried under carefully controlled temperatures and in the absence of light to assure optimum properties. The previously described processes involve considerable investment and are practical only for plantation operations. Small shareholders may allow their latex to coagulate naturally and sell it to processors. Such a product will vary greatly, depending on the specific history of each rubber slab.17

Uses

The largest use of natural rubber is in the manufacture of tires. Over 70 percent of its consumption is in this area. The next largest use is as latex in dipped goods, adhesives, rubber thread, and foam. These uses account for approximately another 10 percent. The remainder is used in a variety of applications such as conveyor belts, hoses, gaskets, footwear, and antivibration devices such as engine mounts.

Because of the high stereoregularity of natural rubber, the units in the polymer chain can form very orderly arrangements, which result in crystallization upon storage at low temperatures or upon stretching. Stored crystallized rubber may be converted to its original amorphous state by heating. Several crystalline forms are reported, with melting points varying from 14°C to 36°C.18 Although crystallization upon storage can be a problem to users, the ability to crystallize reversibly upon stretching accounts for many of the unique properties of natural rubber. Specifically, the ability of natural rubber to be used as a gum polymer (unfilled vulcanizate) depends on this property. The crystallites that form act both as filler and as temporary cross-links, providing high tensile properties.19 In tire fabrication this property is reflected in natural rubber stocks possessing high green (uncured) strength and excellent building tack. In the final product, strain-induced crystallization provides tear and cracking (cut growth) resistance.

Although many other polymers can crystallize, only natural rubber has been found to have the necessary combination of rate of crystallization, degree of crystallization, and melting point to provide all the properties discussed above. Natural rubber has good flexibility and resilience. In truck and bus tires, it is used extensively with blends of polybutadiene to give the low heat buildup needed as well as wear and cut growth resistance. In passenger tires, natural rubber is used in the sidewalls and carcass areas. These areas require the building tack, ply adhesion, and hot strength properties that it imparts. Synthetic rubber is used almost exclusively in passenger tire treads.

The largest use of natural rubber latex is in the dipped goods area. Products include balloons, surgical and examination gloves, and prophylactics. The rapid spread of AIDS has led to a dramatic increase in the surgical and examination glove market. Latex also is used to make adhesives, rubber thread, and foams. However, natural rubber is being replaced in these two latex areas by urethanes and synthetic rubber latex. Some of the replacement has been driven by skin allergy problems which are caused by the natural proteins in the latex. Although natural rubber latex is a mature product, research continues on improving its uniformity, stability, and performance.20

18.7 Polyisoprene

Faraday discovered in 1826 that natural rubber was composed of a hydrocarbon with a ratio of five carbons to eight hydrogen atoms; and in 1860, G. Williams isolated isoprene by collecting the distillate from the heating of natural rubber. By 1887, scientists in France, England, and Germany had converted isoprene back into a rubbery material. Because this offered a potential for manufactured "natural" rubber, research was undertaken to find ways to obtain isoprene from sources other than rubber itself.21

Monomer Production

The primary source of isoprene today is as a by-product in the production of ethylene via naphtha cracking. A solvent extraction process is employed. Much less isoprene is produced in the crackers than butadiene, so the availability of isoprene is much more limited. Isoprene also may be produced by the catalytic dehydrogenation of amylenes, which are available in C-5 refinery streams. It also can be produced from propylene by a dimerization process, followed by isomerization and steam cracking. A third route involves the use of acetone and acetylene, produced from coal via calcium carbide. The resulting 3-methyl-butyne-3-ol is hydrogenated to methyl butanol and subsequently dehydrogenated to give isoprene. The plants that were built on these last two processes have been shut down, evidently because of the relatively low cost of the extraction route.

Polymer Production Process

The free-radical catalysts were found to produce a product that did not have the tack, green strength, or gum tensile of natural rubber. Whereas natural rubber is an essentially pure *cis*-1,4 structure, the emulsion product was of mixed microstructure. This precluded the ability of the latter to undergo strain-induced crystallization, which is required to obtain many of the desired natural rubber properties.

In 1955 investigators from the Firestone Tire and Rubber Company and the B. F. Goodrich Company announced the synthesis of polyisoprene with over 90 percent *cis*-1,4 structure. The work at Firestone was based on lithium metal catalysts, whereas the work at Goodrich was the result of using ZieglerNatta type coordination catalysts.22,23

Use

Although considerable interest was generated by these discoveries, their commercial success has been rather limited. The lithium-based polymers were found to produce up to 94 percent *cis*, which still was not high enough to provide the properties of natural rubber. Polymers made with the coordination catalysts have *cis* contents of up to 98 percent, providing products that can more closely serve as replacements for natural rubber than the lithium-based polymers. In comparison with natural rubber, they offer the advantage of a more highly pure rubber (no nonrubber material) and excellent uniformity. For economic reasons, polyisoprene has seen only limited success. Several of the plants built to produce polyisoprene have been either shut down or converted for use to produce other polymers. In terms of synthetic rubber production in 1999, only 221,000 metric tons were produced (excluding centrally planned economy countries, CPEC), which represented only 3 percent of total synthetic rubber production. In comparison, in 1999 natural rubber usage was 6,708,000 metric tons.24 Evidently because of its strategic importance, the Former Soviet Union (FSU) continues to rely heavily on polyisoprene.

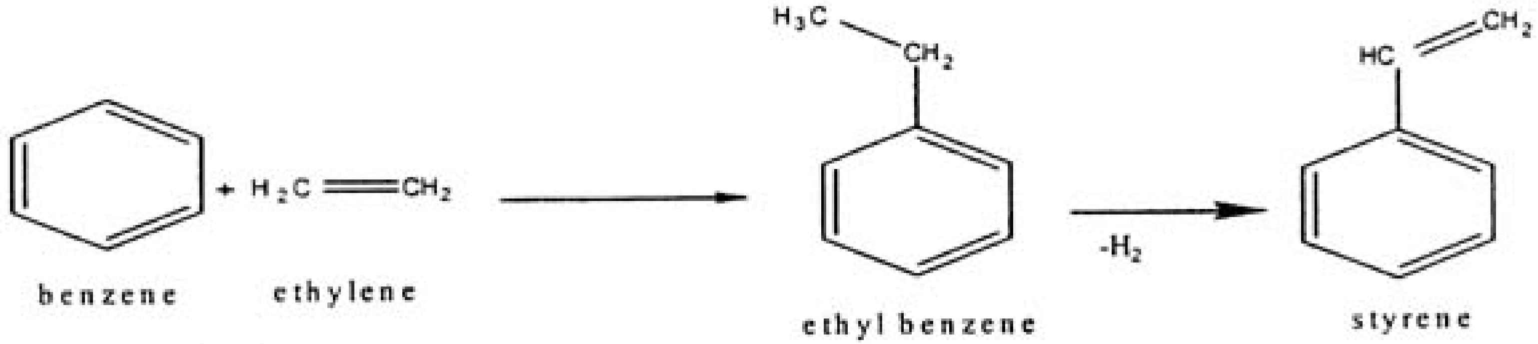
18.8 StyreneButadiene Rubber

The largest-volume synthetic rubber consumed is styrenebutadiene rubber (SBR). In 1999, SBR solid rubber accounted for 39 percent of all synthetic rubber. If SBR latex and carboxylated SBR latex are included, its share increases to 55 percent. The major application of solid SBR is in the automotive and tire industry, accounting for approximately 70 percent of the use. Therefore, SBR has been tightly tied to the tire business.25

Initially, SBR was developed as a general purpose alternate material to natural rubber. In the United States the thrust came early in World War II when the U. S. supply of natural rubber was cut off. The basic technology was developed in Germany in the late 1920s, and by 1939 Germany had 175,000 metric tons of capacity in place. The first U.S. production was 230 metric tons in 1941, but by 1945 there were more than 850,000 metric tons of capacity. Basically, in a period of five years the emulsion SBR business as we know it today was put in place. By 1973 U.S. capacity had increased to almost 1,400,000 metric tons, but in 1989 it had contracted to 881,000 tons close to the 1945 capacity.26

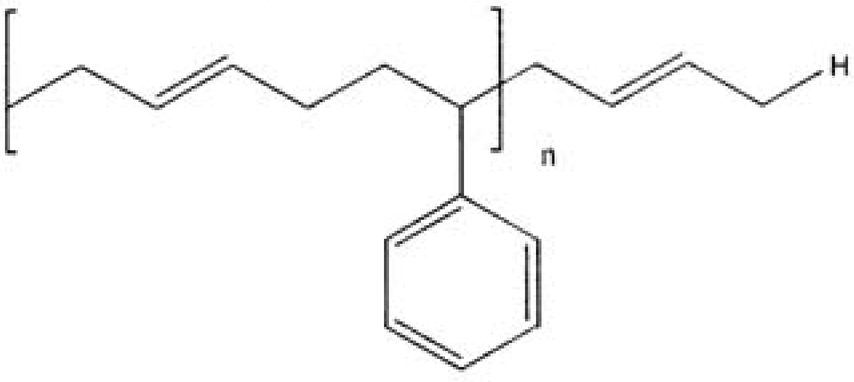
Monomer Production

The production of butadiene monomer is discussed below in polybutadiene section "Polybutadiene." The largest volume of styrene is produced by the alkylation of benzene with ethylene to give ethyl benzene, which is then dehydrogenated to give styrene.27



Polymer Production Process

SBR is produced by two different processes: emulsion and solution. The emulsion process involves a free-radical mechanism, whereas solution SBR is based on alkyllithium catalysis.



copolymer of butadiene and styrene

Emulsion Process

The formula developed to provide SBR during World War II was standardized, with all rubber plants owned by the U.S. government. The standard recipe is listed below:28

Component	Parts by Weight
Butadiene	75
Styrene	25
n-Dodecyl mercaptan	0.5
Potassium peroxydisulfate	0.3
Soap flakes	5.0
Water	180

Initiation occurs through reaction of the persulfate with the mercaptan, as shown below:

$$K_2S_2O_8 + 2RSH \rightarrow 2RS\bullet + KHSO_4$$

$$RS\bullet + M \rightarrow RSM\bullet$$

Chain propagation occurs by the growing chain free radical attacking either the butadiene or styrene monomer. The active radical chain can react with mercaptan to form a new mercaptyl radical and a terminated chain. The mercaptyl radical then can initiate an additional chain. The molecular weight of the chain P can be controlled by the concentration of mercaptan via this chain transfer mechanism.

$$P\bullet + RSH \rightarrow PH + RS\bullet$$

$$RS\bullet + M \rightarrow RSM\bullet$$

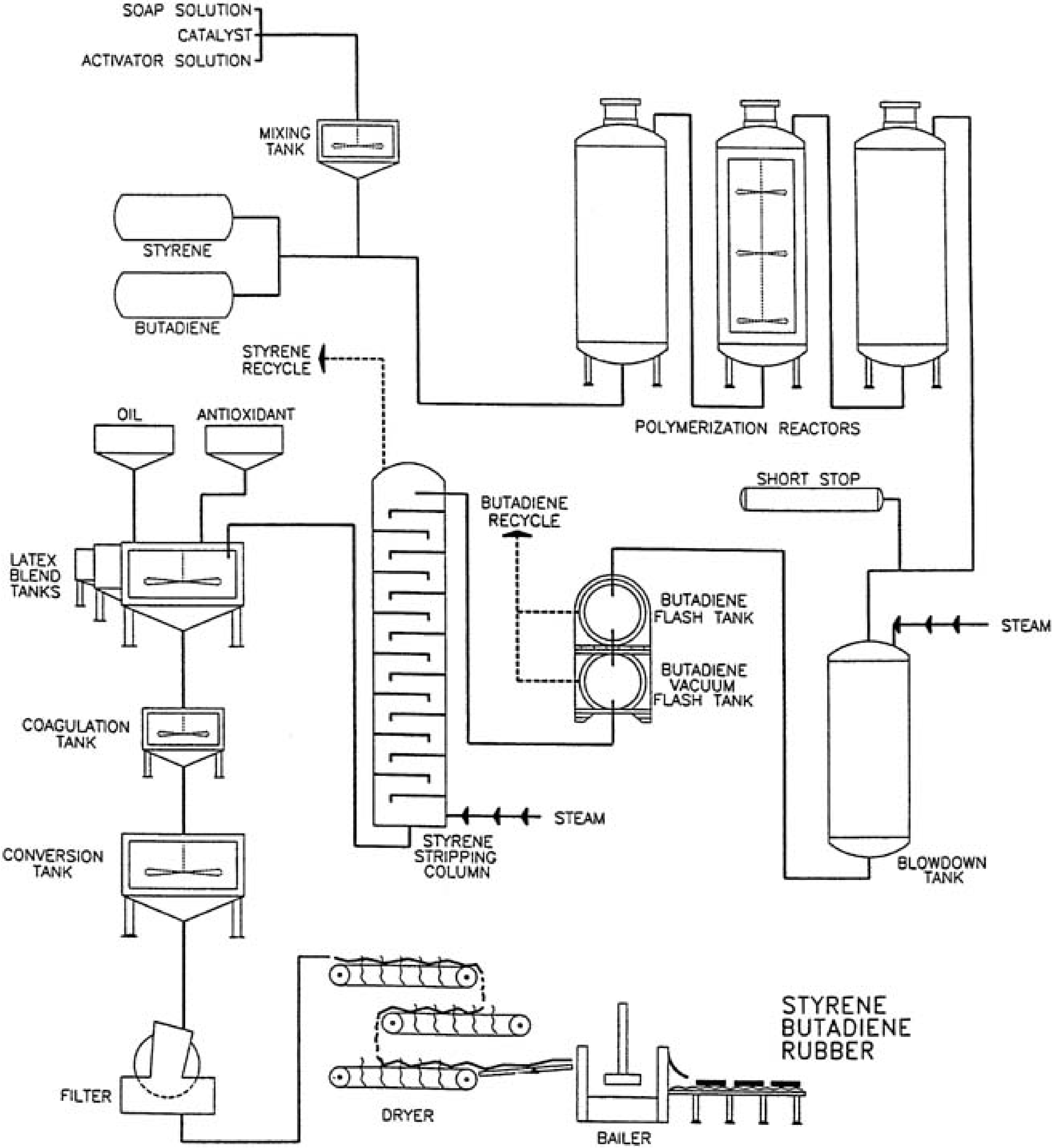
Termination also can occur by the reaction of two free radicals, through either combination or disproportionation reactions.

$$\begin{aligned} P\bullet + P\bullet \quad &\rightarrow \quad \text{P-P combination (bimolecular coupling)} \\ P\bullet + P\bullet \quad &\rightarrow \quad \text{P-CH}{=}\text{CH}_2 \\ &\qquad + \text{PH disproportionation} \\ &\qquad \text{(hydrogen-free radical transfer)} \end{aligned}$$

The mercaptyl radical also can react with growing chains, to lead to termination.29

Polymerization is initially carried out at 50°C until conversion of 7075 percent is reached, at which time the polymerization is terminated by the addition of a freeradical scavenger such as hydroquinone. Polymerization beyond this point results in excessive free-radical attack on the polymer chains. Products made under such conditions have poor properties due to excessive branching and gelation. Unreacted butadiene and styrene are removed by flashing and steam stripping. Antioxidant is added to the latex, followed by coagulation with the addition of polyelectrolytes and salt-acid. The coagulated crumb then is washed, dried, baled, wrapped, and packaged for shipment. Because of the soap and other chemicals in the formulation, most emulsion polymers will contain about 7 percent of nonrubber residues. The emulsion process flow sheet is shown in Fig. 18.1.

Fig. 18.1 The production of styrene-butadiene rubber. (*Modified from a drawing in the Vanderbilt Rubber Handbook, 13th ed.*)



It was soon discovered that polymers made at lower temperatures had significantly better properties, especially in tire treads. This was mainly because they contained fewer low molecular weight species. A lower temperature process, using "redox" chemistry, was eventually developed. It used peroxides or hydroperoxides with a reducing agent such as a water-soluble transition metal salt which were active even at 0°C.

Mercaptans are also used as chain transfer agents to provide a mechanism for molecular weight control. Commercially these types of polymerization are carried out at 5°C and are referred to as "cold" polymerizations to differentiate them from the previously discussed "hot" systems. A typical formula is listed below.30

Component	Parts by Weight
Butadiene	71
Styrene	29
tert-Dodecyl mercaptan	0.18
p-Methane hydroperoxide	0.08
Ferrous sulfate heptahydrate	0.03
Trisodium phosphate decahydrate	0.50
Tetrasodium ethylenediaminetetraacetate	0.035
Sodium formaldehyde sulfoxylate	0.08
Rosin acid soap	4.5
Water	200

The improved tire wear of cold polymerization SBR led to the very rapid replacement of hot SBR for most applications. This change was relatively easy to make, as all the equipment could be used with the only modification required being the addition of reactor cooling, which is achieved with either the reactor jacket, internal coils, or both.

It later was found that even more improvements could be realized by polymerizing to very high molecular weights and then adding petroleum-based oils to the latex prior to coagulation. The oil is absorbed by the rubber which, upon coagulation, produces oil-extended polymers. For tread applica-tions, oils of higher aromatic content were preferred because of their excellent compatibility with the rubber. Typically 37.5 parts of oil are added, although grades containing up to 50 parts have been produced. Very high molecular weight polymers thus can be processed without requiring excessive energy to mix them. The oil also allows these tough polymers to be processed without excessive degradation. Carbon black masterbatches also are produced. In this process carbon black is added to the latex prior to coagulation, and the black, along with oil, is incorporated into the latex in the coagulation step. These products offer the user the advantage of not having to handle free black in their mixing operation, and can provide additional compounding volume for manufacturers with limited mixing capacity.

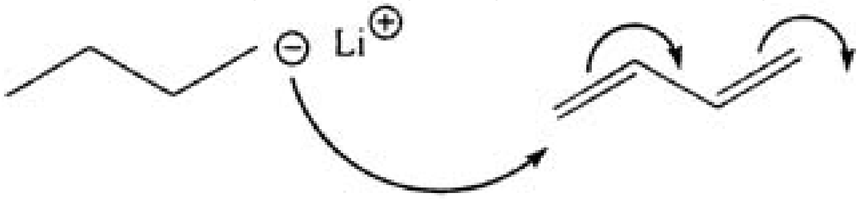
Not all emulsion SBR is converted to dry rubber for use. There is a variety of applications where the latex can be used directly in the final fabrication process. This technology logically grew out of the latex technology developed for natural rubber. For latex applications the particle size distribution can be critical because of its effect on viscosity and performance variables, as when used to provide impact strength in plastics such as ABS.31 Careful control of the mechanical stability of the latex also is critical, as these systems must destabilize under relatively mild conditions such as those in a coating operation. A number of processes have been developed to control particle size via partial destabilization of the latex. Among the commercial methods are: careful control of a freezethaw cycle, controlled shear agitators, high-pressure colloid mills, and the addition of chemicals such as hydrocarbons or glycols.

A special variation of SBR latex containing terpolymerized vinyl pyridine is used in the tire industry to provide adhesion of organic fiber tire cords to rubber stock. The vinyl pyridine SBR latex is combined with resins and coated on the fiber by a dipping process. The adhesive is set by a controlled temperature and tension process to control the shrinkage properties of the cord.

Solution Process

The discovery of the ability of lithium-based catalysts to polymerize isoprene to give a high *cis* 1,4 polyisoprene was rapidly followed by the development of alkyllithium-based poly-butadiene. The first commercial plant was built by the Firestone Tire and Rubber Company in 1960. Within a few years the technology was expanded to butadienestyrene copolymers, with commercial production under way toward the end of the 1960s.

The copolymerization with alkyllithium to produce uniformly random copolymers is more complex for the solution process than for emulsion because of the tendency for the styrene to form blocks. Because of the extremely high rate of reaction of the styryllithium anion with butadiene, the polymerization very heavily favors the incorporation of butadiene units as long as reasonable concentrations of butadiene are present. This observation initially was somewhat confusing because the homopolymerization rate of styrene is seven times that for butadiene. However, the cross-propagation rate is orders of magnitude faster than either, and it therefore dominates the system. For a 30 mole percent styrene charge the initial polymer will be almost pure butadiene until most of the butadiene is polymerized. Typically two-thirds of the styrene charged will be found as a block of polystyrene at the tail end of the polymer chain:



anionic initiation of butadiene by butyllithium

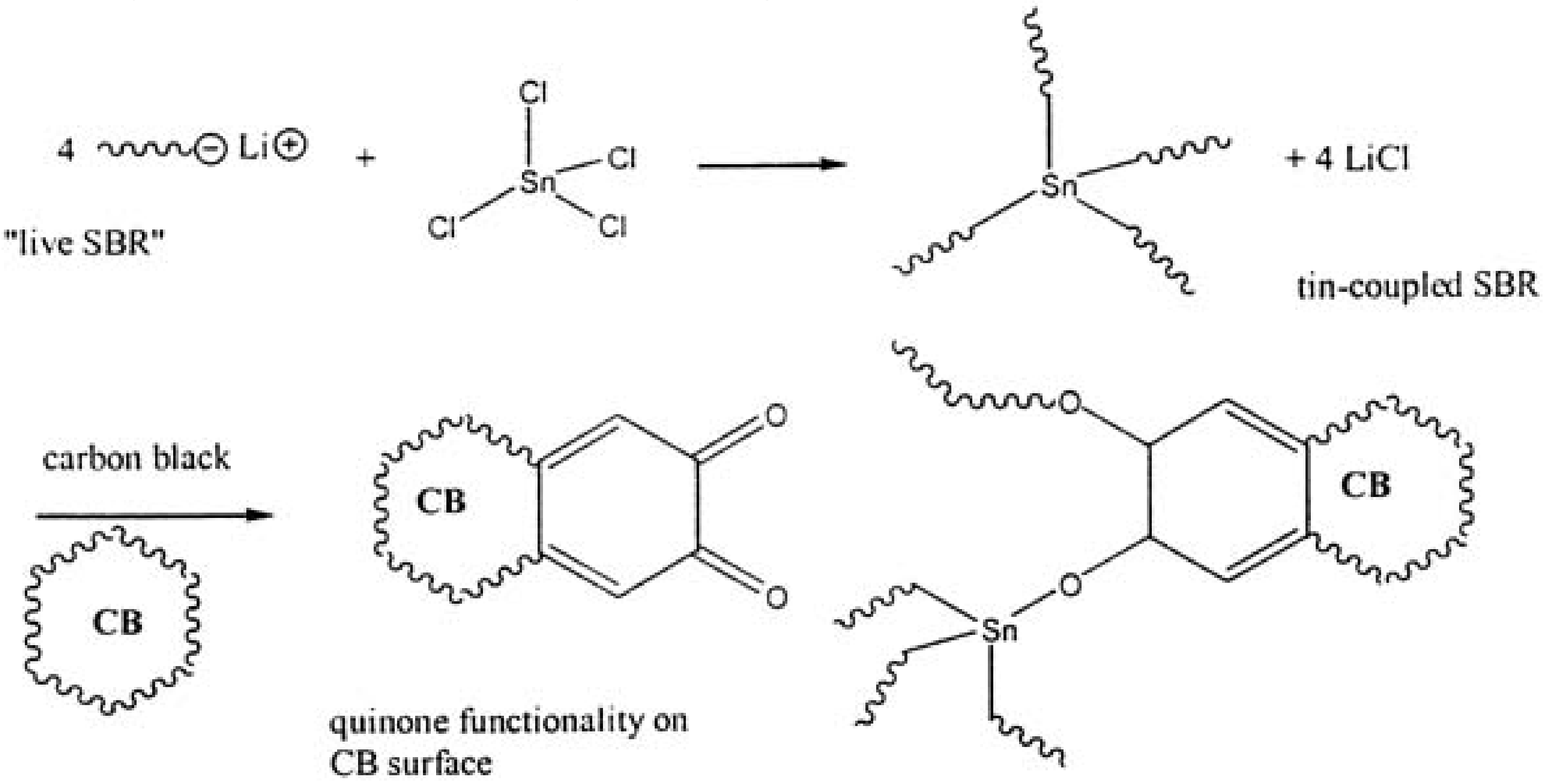
Several methods have been proposed to overcome this problem. In one, the styrene and part of the butadiene are charged initially with butadiene metered at a rate equivalent to its incorporation into the chain. A second approach involves adding both monomers at a relatively slow rate so that the equilibrium monomer concentration reaches a pseudosteady state that will produce polymer at the desired composition.32 This process can be done in either a batch or a continuous mode.33

In addition to these reaction engineering approaches to produce uniformly random copolymer, the chemistry may be changed by the addition of polar agents such as amines and ethers. This action results in bringing the reactivity rates much closer together. The change in chemistry also is reflected in the microstructure of the butadiene portion of the polymer. Whereas in the nonpolar system the vinyl content of the butadiene portion is around 10 percent, in polar systems vinyl contents of 3040 percent typically are obtained when a 20 percent styrene polymer is randomized. Higher styrene contents require higher modifier levels, resulting in even higher vinyl contents. An added complication with polar modifiers is their ability to react with the growing chain, resulting in undesired termination.

In spite of these complications, all recent U.S. expansions or announced plants for SBR have been for solution polymers. The ability to better design the polymer structure and produce special functional polymers (described below) accounts for most of this shift.

Functional Solution SBR

The driving force toward functional solu-tion SBR is its improved hysteresis properties for passenger tire treads. The fact that anionic SBR has a much more narrow molecular weight distribution compared with emulsion gives it lower hysteresis. However, the big advantage is the relatively stable growing chain ends which can be chemically modified to improve interaction with carbon black and silica in tire compounds.34 This modification can lead to a dramatic reduction in rolling resistance, which is critical for automotive manufacturers who must meet government-mandated fuel economy targets. The most active functional end-groups contain either organotin or certain amines. Termination with tin tetrachloride is the easiest and most popular method which generates a four-armed star polymer. The polymertin bonds break down during mixing of the compound and both lower the compound viscosity and create active sites for reaction with carbon black surfaces. The creation of this "carbon-bound rubber" effectively prevents the carbon black from agglomerating on a microscale to form hysteretic, three-dimensional networks. The breaking of these networks during the deformation of a tire tread is a major source of rolling resistance. Polymers with amine end-groups also show good activity with carbon black. Termination with silane esters is usually used to obtain interaction with silica fillers producing the same decrease in hysteresis. There have recently been efforts to make functional anionic initiators, some of which could be used to make low hysteresis rubber.35 However, the main use for such polymers have been in adhesives and other nontire applications. One study comparing an emulsion polymer, its solution counterpart, and a chemically modified version of the solution polymer showed a 23 percent hysteresis reduction in going from emulsion to solution and an additional 15 percent reduction for the chemically modified polymer, to provide an overall reduction of 38 percent.36 To date, this type of chemical modification is only possible using anionic techniques.



18.9 Polybutadiene (BR)

Next to SBR, polybutadiene is the largest volume synthetic rubber produced. Consumption was approximately 1,953,000 metric tons in 1999.37

Monomer Production

Butadiene monomer can be produced by a number of different processes. The dominant method of production is as a by-product from the steam cracking of naphtha to produce ethylene. The butadiene is recovered from the C-4 fractions by extractive distillation.38 "On-purpose" butadiene is generally produced by dehydrogenation or oxidative dehydrogenation of four-carbon hydrocarbons.39,40

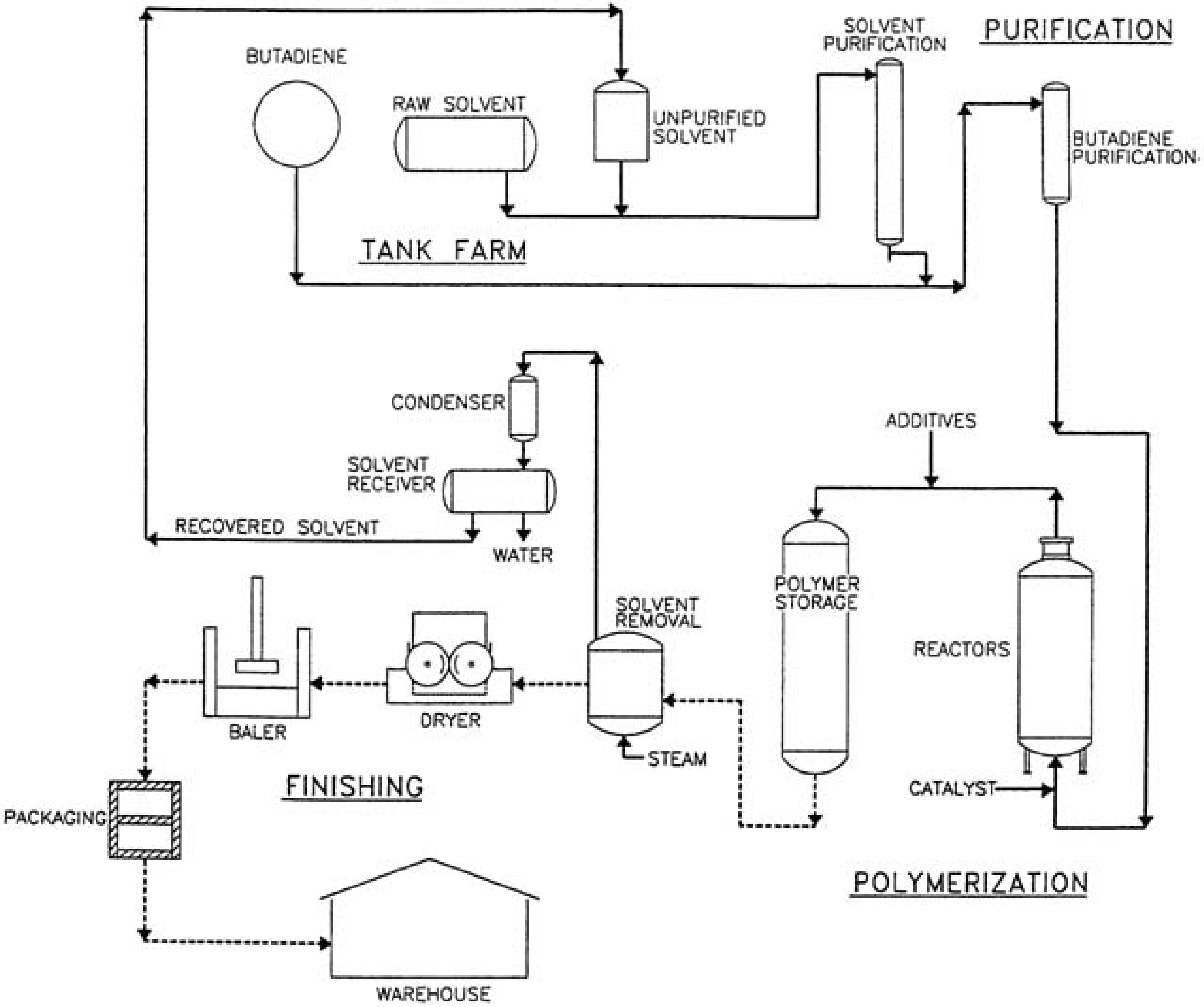
Polymer Production Process

Polybutadiene is usually produced by alkali metal, and transition metal coordination solution processes. Most production is based on the solution processes because of the ability to obtain preferred microstructures by these routes.

Alkali-metal-based polymerization (usually organolithium) produces a product with about 36 percent *cis*, 54 percent *trans*, and 10 percent vinyl. The polymerization process is conducted in an aliphatic hydrocarbon under an inert atmosphere in either a batch or a continuous mode. Because of the characteristics of this polymerization system, polymers of extremely narrow molecular weight distribution and low gel can be produced.41 The narrowest distribution is produced via batch polymerization. Coupled (star branched) and end-functional polybutadienes are possible using organolithium technology due to the living anion on the chain end which is available for further reactions. Upon the addition of polar agents, such as ethers or amines, the organolithium initiators can produce polybutadienes with vinyl contents up to 100 percent.42 The vinyl content can be controlled by the ratio of modifier to catalyst and the polymerization temperature, with lower temperatures favoring increased vinyl formation. Even with high vinyl contents such polymers do not crystallize because of the atactic nature of the vinyl units. High *cis* polybutadiene is produced via solution processes using ZieglerNatta type transition metal catalysts. The major commercial catalysts of this type are based on titanium, cobalt, nickel, and neodymium.43 Typically the transition metal is used in the form of a soluble metal salt, which can react with an organoaluminum or organoaluminum halide as a reducing agent to give the active species. Because of the active nature of transition metals, the polymer solutions are treated to deactivate or remove such materials from the final product. All of these catalysts produce products with 90 percent or higher *cis* content. The neodymium system is reported to produce the highest *cis* (9899%) with the most linear chain structure.44 The highest branched *cis* BR is produced with the Co system with Ni giving intermediate branching.

All the solution processes require high efficiency in recovering the solvent. The most widely used process consists of termination of the polymerization and the addition of antioxidant to the polymer solution. The solution may be treated to remove catalyst residue and then transferred into an agitated steam stripping vessel in which unreacted monomer and solvent are flashed off, leaving the rubber as a crumb slurry in water. The watercrumb slurry then is dewatered and dried. The recovered monomer/solvent is recirculated to a series of distillation columns to recover monomer and purify the solvent. As both the anionic and the coordination catalyst systems are highly sensitive to impurities such as water, the purification system is very critical for satisfactory process control (Fig. 18.2).

Fig. 18.2 Flow diagram for a typical solution process for the manufacture of polybutadiene. (*Courtesy of Firestone Polymers, Akron, Ohio*)



Uses

The major use for polybutadiene is in tires, with over 70 percent of the polymer produced used by the tire industry. Cured polybutadiene has excellent low-temperature properties, high resiliency, and good abrasion resistance due to its low glass transition temperature. However, this same fundamental property also leads to very poor wet skid resistance. For this reason, polybutadiene is blended with other polymers such as natural rubber and SBR for use in tread compounds. In general, polybutadiene is a poorer-processing polymer than SBR, but this is generally not a problem as it is blended with other polymers in use. The very high *cis* polymers have the potential for strain-induced crystallization, which can lead to improved green strength and increased cut growth resistance in the cured product. High *cis* polybutadiene is reported to have a melting point of 6°C.45

The other major use for polybutadiene is as an impact modifier in plastics, in particular high impact polystyrene (HIPS) and acrylonitrilebutadienestyrene resin (ABS). In the HIPS application the rubber is dissolved in the styrene monomer, which is then polymerized via a free-radical mechanism. A complex series of phase changes occurs, resulting in small rubber particles containing even smaller polystyrene particles being incorporated into a polystyrene matrix. The rubber is added to increase impact strength. Because of the unique morphology that is formed, low levels of rubber (typically around 7%) provide rubbery particles having a volume fraction of 3040 percent. This morphology leads to high impact at very low rubber levels, providing good stiffness and hardness.46

There is also a fairly large market for high *cis* BR in solid core golf balls. In this application, the polymer is compounded with zinc acrylate and the mixture is cured with peroxide.47 This produces an ionically cross-linked compound that has outstanding resilience. The covers are also ionomers with superior cut resistance. In the last few years the golf ball market has been shifting away from the traditional wound ball to these new solid core balls that use polybutadiene.

18.10 EthylenePropylene Rubber

There are two general types of polymers based on ethylene and propylene: ethylene propylene rubber (EPM) and ethylenepropylene terpolymer (EPDM). EPM accounts for approximately 20 percent of the polyolefin rubber produced. Comprising a totally saturated polymer, these materials require free-radical sources to cross-link. EPDM was developed to overcome this cure limitation. For EPDM a small amount (less than 15%) of a nonconjugated diene is terpolymerized into the polymer. One of the olefinic groups is incorporated into the chain, leaving its other unsaturated site free for vulcanization. This ensures that the polymer backbone remains saturated, with corresponding stability, while still providing the reactive side group necessary for conventional cure systems. The nonconjugated dienes used commercially are ethylidene norbornene, 1,4 hexadiene, and dicyclopentadiene. The selection of the termonomer is made on the basis of the reactivity of the termonomer, both in polymerization and in vulcanization. The estimated 1999 worldwide consumption (excluding CPEQ) was 842,000 metric tons.

Monomer Production

Ethylene and propylene are produced primarily by the cracking of naphtha. They also are available from the fractionation of natural gas. Ethylidene norbornene is produced by reacting butadiene with cyclopentadiene. 1,4 Hexadiene is produced from butadiene and ethylene. Dicyclopentadiene is obtained as a by-product from the cracking of heavy feedstocks to produce ethylene.

Polymer Production

There are two processes used to produce EPM/EPDM: solution and suspension. In either case a ZieglerNatta type catalyst is used (aluminum alkyl or aluminum alkyl chlorides and a transition metal salt). The most generally used transition metal is vanadium in the form of the tetrachloride or the oxytrichloride.48 The solution process is similar to that used for other solution polymers. The polymer cement can be finished by stream stripping and drying of the resulting crumb.49

In the suspension process, the polymer is suspended in the monomer propylene. This process offers the advantages of being able to operate at higher solids owing to the lower viscosity of a suspension compared with a solution at comparable solids. Other advantages are simple heat removal by the evaporative cooling of the propylene, more uniform reactor temperature profile, and ease of production of high molecular weight or semicrystalline polymers.50

A specially developed titanium-based catalyst has been used in the suspension process for EPM and EPDM where the termonomer is low-boiling. The advantages claimed, in addition to those characteristic of the suspension process, are better structural control and high catalyst efficiency, resulting in a high-purity product without requiring catalyst removal.51

The polymer composition for both EPM and EPDM is usually in the 40/60 to 60/40 ethylene/propylene ratio. Outside these ranges, the polymer will start to crystallize because of either polyethylene or polypropylene blocks.

Use

EPM/EPDM polymers exhibit outstanding resistance to heat, ozone, oxidation, weathering, and aging due to the saturated backbone. They have low density, are miscible with aliphatic and naphthenic oils, and maintain acceptable properties at high filler loadings. They are used in single-ply roofing, wire and cable, automotive parts, impact modification of polypropylene, and viscosity index additives for automotive oils. They also can be used in producing thermoplastic olefin elastomers by blending with polypropylene, which may be partially grafted or cross-linked by dynamic vulcanization. These "polymer alloys" will be discussed in the section "Thermoplastic Elastomers." Although at one time EPDM was expected to become the major polymer for tires, this market has not materialized for a variety of processing and performance reasons.52

18.11 Butyl Rubber

Butyl rubber is one of the older synthetic rubbers, having been developed in 1937. Because of the saturated nature of a polyolefin elastomer, the commercial polymer is actually a copolymer of isobutylene and isoprene. The isoprene is added to provide cure sites. In addition, halogenated (bromo or chloro) derivatives are available. The halogenated products improve the mixing and cure compatibility with the more common unsaturated rubbers such as natural or styrenebutadiene rubber.

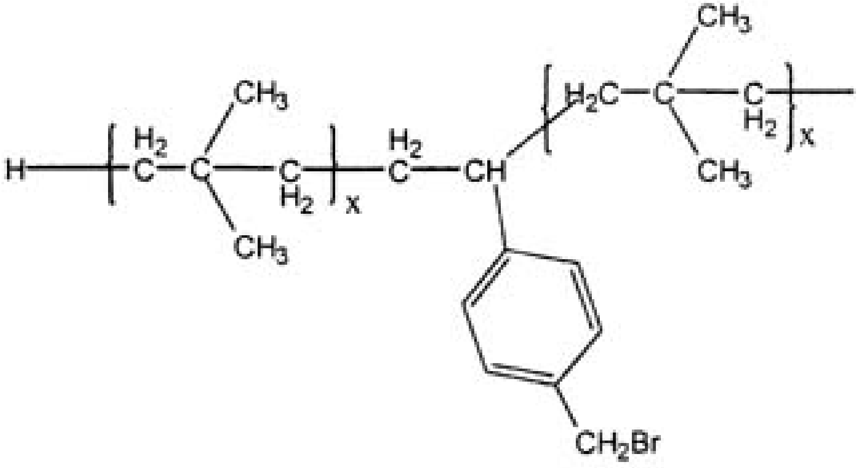
Monomer Production

Isobutylene is obtained as a by-product from petroleum and natural gas plants. The monomer must be highly purified to assure high molecular weight.

Production Process

Butyl rubber is produced at very low temperature (below -90°C) to control the rapid exotherm, and to provide high molecular weight. The process consists of charging isobutylene along with isoprene (24%) with an inert diluent such as methyl chloride to a reactor to which a FriedelCrafts catalyst is added. The polymerization is very rapid, and the polymer forms in a crumb or slurry in the diluent. Heat is removed via the reactor jacket. The slurry is steam-stripped to remove all volatiles. The catalyst is neutralized, and antioxidants are added to the slurry prior to drying.53 The halogenated derivatives are produced by the direct addition of the halogen to a solution of the isobutyleneisoprene polymer.

During the last 10 years another type of butyl rubber was developed which is derived from a copolymer of isobutylene and *p*-methylstyrene.54 They are subsequently brominated to varying degrees producing different grades of the elastomer. Bromination occurs selectively on the methyl group of the *p*-methylstyrene providing reactive benzylic bromine functionality, which can be used for grafting and curing reactions.



brominated poly(-isobutylene-*p*-methylstyrene)

Properties and Use

The most important characteristics of butyl rubber are its low permeability to air and its thermal stability. These properties account for its major uses in inner tubes, tire inner liners, and tire curing bladders. Because of the poor compatibility of butyl with other rubbers (with respect to both solubility and cure), the halobutyls are preferred. The brominated *p*-methylstyrene-containing butyl rubbers are used in a number of grafting reactions for tire applications and adhesives. Other uses for butyl rubber are automotive mechanical parts (due to the high damping characteristics of butyl), mastics, and sealants.55

18.12 Nitrile Rubber

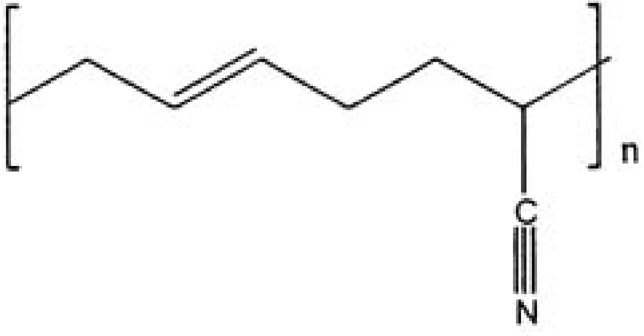
Nitrile rubber was invented at about the same time as SBR in the German program to find substitutes for natural rubber.56 These rubbers are copolymers of acrylonitrilebutadiene, containing from 15 to 40 percent acrylonitrile. The major applications for this material are in areas requiring oil and solvent resistance. The estimated worldwide consumption in 1999 was 322,000 metric tons.57

Monomer Production

The production of butadiene is discussed in the diene section "Polybutadiene." Although several routes have been developed to produce acrylonitrile, almost all now is produced by the catalytic fluidized-bed ammoxidation of propylene.

Polymer Production

The polymerization process parallels the emulsion process used for styrenebutadiene rubber. Either a hot or a cold process can be used, with the cold polymerization providing the same improved processing and vulcanizate properties as seen in SBR. Polymerizations are carried to 7080 percent conversion and terminated to avoid gel formation. The latex must be stripped to remove unreacted butadiene and acrylonitrile.



butadiene-acrylonitrile copolymer

Properties and Use

As the acrylonitrile content increases in the polymer chain, the properties change predictably. The glass transition temperature increases approximately 1.5°C for each percent increase in acrylonitrile. Properties such as hysteresis loss, resilience, and low-temperature flexibility will correspondingly change. The oil resistance increases with increased acrylonitrile content, as does the compatibility with polar plastics such as PVC. The major market for nitrile rubber is in the automotive area because of its solvent and oil resistance. Major end uses are for hoses, fuel lines, O-rings, gaskets, and seals. In blends with PVC and ABS, nitrile rubber acts as an impact modifier. Some nitrile rubber is sold in latex form for the production of grease-resistant tapes, gasketing material, and abrasive papers. Latex also is used to produce solvent-resistant gloves.58

18.13 Hydrogenated Nitrile Rubber

During the last 15 years several companies have developed hydrogenated grades of nitrile rubber to both improve its thermal stability and solvent resistance. Although the hydrogenation of a polydiene backbone was done as early as the 1920s, real commercial products with acrylonitrile were not introduced until the mid-1980s.59

Hydrogenated NBR (HNBR) is produced by first making an emulsion-polymerized NBR using standard techniques. It then must be dissolved in a solvent and hydrogenated using a noble metal catalyst at a precise temperature and pressure.60 Almost all the butadiene units become saturated to produce an ethylenebutadieneacrylonitrile terpolymer. These "post-polymerization" reactions are very expensive so HNBRs usually command a premium price. HNBR is usually cured with peroxides, similar to ethylenepropylene elastomers, because it has no unsaturation for a conventional sulfur cure system.

Uses

HNBR has many uses in the oil-field, including down hole packers and blow-out preventers, because of its outstanding oil resistance and thermal stability. For the same reasons, it has also found uses in various automotive seals, O-rings, timing belts, and gaskets. Resistance to gasoline and aging make HNBR ideal for fuel-line hose, fuel-pump and fuel-injection components, diaphragms, as well as emission-control systems.

18.14 Chloroprene Rubber

Chloroprene rubber (Neoprenetrade name of DuPont) was one of the earliest synthetic rubbers, first commercialized in 1932. It has a wide range of useful properties but has not become a true general purpose synthetic rubber, probably because of its cost. It does possess properties superior to those of a number of general purpose polymers, such as oil, ozone, and heat resistance; but for these properties other specialized polymers excel. Polychloroprene thus is positioned between the general purpose elastomers and the specialty rubbers.

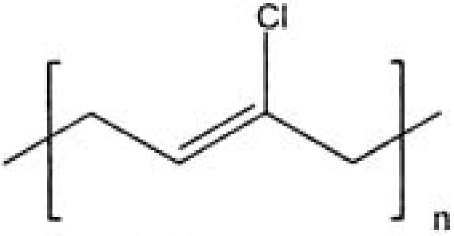
Monomer Production

Chloroprene monomer production starts with the catalytic conversion of acetylene to monovinylacetylene, which is purified and subsequently reacts with aqueous hydrogen chloride solution containing cuprous chloride and ammonium chloride to give Chloroprene.61

Production Process

Polychloroprene is produced by using an emulsion process. Two general types of processes are used: sulfur modified and unmodified. In the sulfur modified process, sulfur is dissolved in monomer and is incorporated into the polymer chain. Upon the addition of thiuram disulfide-type materials and under alkaline conditions, some of the sulfur bonds are evidently cleaved to give the soluble polymer. In the unmodified process chain transfer agents are used. If neither the sulfur modified nor the chain transfer system is used, the resulting polymer is a gelled tough material. Typical polymerization systems consist of rosin acid soap emulsifier and persulfate catalyst. Conversions of 8090 percent are obtained. Polymerizations are run at around 40°C. For the modified polymer, the thiuram disulfide is added after polymerization, and the latex is aged to allow the peptization (chain scission) reaction to occur. Acidification stops the peptization reaction. The latex is vacuum-stripped and coagulated using a cold drum dryer process. The coagulated rubber is washed and dried.

The polymerization produces primarily *trans*-1,4-polychloroprene. The *trans* content can be increased somewhat by lowering the polymerization temperature:62,63



polychloroprene

Properties and Uses

Polychloroprene is stable to oxidation and ozone. It also is flame resistant, and its oil resistance is better than that of general purpose rubbers. Its major disadvantage, other than cost, is relatively poor low-temperature properties. Because of the high stereoregularity, polychloroprene will strain crystallize, giving good tensile to unfilled stocks. At low temperatures the polymer can crystallize, making processing more difficult. Polymer made at lower temperatures will show higher unfilled tensile properties and more rapid crystallization due to the higher *trans* content (i.e. less disruption of the crystal structure). The major end uses are conveyor belts, V-belts, hoses, and mechanical goods such as wire insulation, O-rings, and gaskets. It also has found use in single-ply roofing and adhesives.64

18.15 Silicone Elastomers

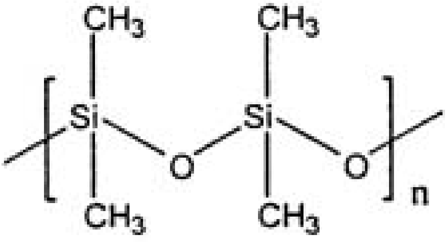
Silicone Rubber elastomers represent a rather unique group of polymers in that they consist of alternating siliconoxygen bonds to form the polymer chain backbone. Side groups off the silicon atoms are selected to provide very specific properties that differentiate one type from another. The most common side group is the dimethyl structure. Replacement of small amounts of the methyl group with vinyl provides sites for cross-linking. Phenyl groups are used to improve low-temperature properties. Fluorosilicones are produced by replacing the methyl with trifluoro-propyl units. The addition of bulky phenyl side groups leads to an increase in the glass transition temperature. However, the disruption caused by such groups leads to the desired reduction or elimination of crystallization, which is critical for low-temperature properties. Such polymers have glass transition temperatures around -110°C, which is the range of the lowest *T*g carbon-backbone polymers.

Monomer Production

The actual polymerization process involves a ring-opening reaction of dimethyl-substituted cyclic siloxanes. The preparation of the cyclic materials starts with the production of pure silicon via the reduction of quartz with coke in an electric arc furnace. The silicon metal then reacts with methyl chloride to give a mixture of silicones, from which dimethyldichlorosilane is removed by distillation.65 Subsequent hydrolysis gives the cyclic dimethylsiloxane.

Polymer Production

The polymerization process involves an equilibrium ring-opening reaction carried out in the bulk state, which can be catalyzed by acids or bases:



polydimethylsiloxane

Uses

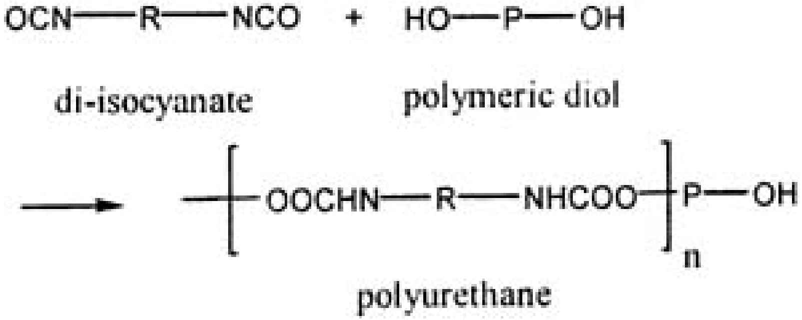
Silicone rubber offers a set of unique properties to the market, which cannot be obtained by other elastomers. The SiO backbone provides excellent thermal stability and, with no unsaturation in the backbone, outstanding ozone and oxidative stability. The very low glass transition temperature, combined with the absence of low-temperature crystallization, puts silicones among the materials of choice for low-temperature performance. The fluoro-substituted versions provide solvent, fuel, and oil resistance along with the above-mentioned stability advantages inherent with the silicone backbone.

The gum polymer has rather poor tensile properties when cured, but these properties can be greatly improved by the use of silica-reinforcing agents. These systems exhibit some of the greatest improvements in properties by filler addition; and because this improvement is significantly higher for silica than for other reinforcing agents, it is assumed that direct bonding occurs between the silica and the polymer. The silicone materials also may have very low surface energy, which accounts for their nonstick characteristics. Because of the inherent inertness of the materials, they have been widely used for medical purposes within the body. The largest use of silicone is in sealant and adhesive applications.

Several different methods have been developed to cure silicones. Free-radical cures are possible for those polymers containing vinyl groups. The largest-volume process, however, involves room-temperature vulcanizations, which can employ either a one-component or a two-component system. In the one-component system, a cross-linking agent such as methyltriacetoxysilane is used. With exposure to moisture, hydrolysis of the cross-linking agent leads to the silanol reactive cure site, so such materials must be compounded and stored free of moisture.66

18.16 Polyurethane Rubber

A wide range of materials is included in this class. The common feature is the use of chain extension reactions to provide products with acceptable commercial properties. The chain extension reaction effectively reduces the actual number of chain ends, thereby eliminating the generally poor properties observed when very low-molecular-weight polymers are cross-linked. The chain extension step involves the reaction of a difunctional polymeric polyol with difunctional organic isocyanates to give the polyurethane:



The most used polyols have a polyester or a polyether backbone. A wide variety of isocyanates are used, with toluene di-isocyanate, *m*-phenylene di-isocyanate, and hexamethylene di-isocyanate the most common.

Raw Materials

The largest-volume polyether used is obtained from propylene oxide polymerized under basic conditions. Polyester polyols are produced from a number of different materials involving diacids and diols to give the ester linkage. Aliphatic polyesters generally are used for elastomers to impart chain flexibility. The production of isocyanates is based on the reaction of phosgene with primary amines. Toluene di-isocyanate is the most frequently used di-isocyanate.

Uses

The urethane elastomers are complexsegmented or block polymers. Soft, noncrystalline blocks are provided by the polyether or aliphatic polyester long chains, whereas stiff, hard blocks are produced by the reaction of aromatic di-isocyanates with low-molecular-weight materials such as diols or diamines. The hard blocks can phase-separate to provide a physical rather than a chemical cross-link, similar to those to be discussed for the butadienestyrene thermoplastic elastomers, although the domain size is much smaller for the polyurethanes. Chemical cross-links can be introduced by using tri-functional materials or by adjusting the stoichiometry to allow additional reactions with the urethane or urea structures to give allophanate or biuret linkages.

Polyurethane rubbers can have high tensile strengths, excellent tear strength, and good abrasion and chemical resistance. The greatest disadvantage is the hydrolytic instability of the urethane linkage. A major use is in automotive bumpers and facias. These materials are made in a reaction injection molding process. Castable urethanes are used to produce solid rubber wheels and printing rolls. Millable urethanes can be processed on conventional rubber equipment. Unsaturation can be introduced to allow the use of conventional rubber cure systems. Cures also are possible by reaction of the active hydrogens with materials such as high molecular weight polyfunctional isocyanates that are nonvolatile at cure temperatures.

Thermoplastic polyurethane elastomers are processed by injection molding and other processes used for thermoplastics. Small gears, seals, and even automotive fender extensions can be produced by this means.

By selecting from the large number of possible reactions and stoichiometry, properties can be tailored to meet a very wide range of applications.6769

18.17 Modified Polyethylene Rubbers

Elastomeric polymers can be produced by the chlorination or chlorosulfonation of polyethylene. Both products start with polyethylene, either in solution or in aqueous suspension, which then is reacted to give the specified degree of substitution to obtain the desired properties. Sufficient substitution is necessary to disrupt the regularity of the polymer chain, changing it from the crystalline polyethylene plastic into amorphous elastomers.

Chlorinated Polyethylene

The chlorinated products contain around 40 percent chlorine. These materials must be stabilized with metal salts, like other chlorinated elastomers and plastics. Peroxide cross-linking generally is used. Being saturated, the materials have excellent weather and ozone resistance and can be used over a temperature range of -65300°F. The high chlorine content imparts oil resistance and relatively slow rates of burning. Typical applications where this combination of properties is required include hoses for chemical or oil resistance, tubing, and belting. In comparison with plasticized PVC, these materials have better low-temperature properties and do not suffer a loss of plasticizer because none is required.

Chlorosulfonated Polyethylene

Reaction with sulfur dioxide in addition to chlorine introduces cross-linking sites into the polymer chain. Sulfur contents in the range of 1.01.5 percent are used, with chlorine contents of 2540 percent. Curing is accomplished by using metallic oxides, sulfur- bearing organic compounds, and epoxy resins. These materials have outstanding ozone resistance and show little color change upon light exposure. Good resistance to oils, heat, oxidation, weather, and corrosive materials also is exhibited. Applications include pond and pit liners, coated fabrics, light-colored roofing membranes, wire and cable insulation, chemical hose, and belting.70,71

18.18 Thermoplastic Elastomers (TPE)

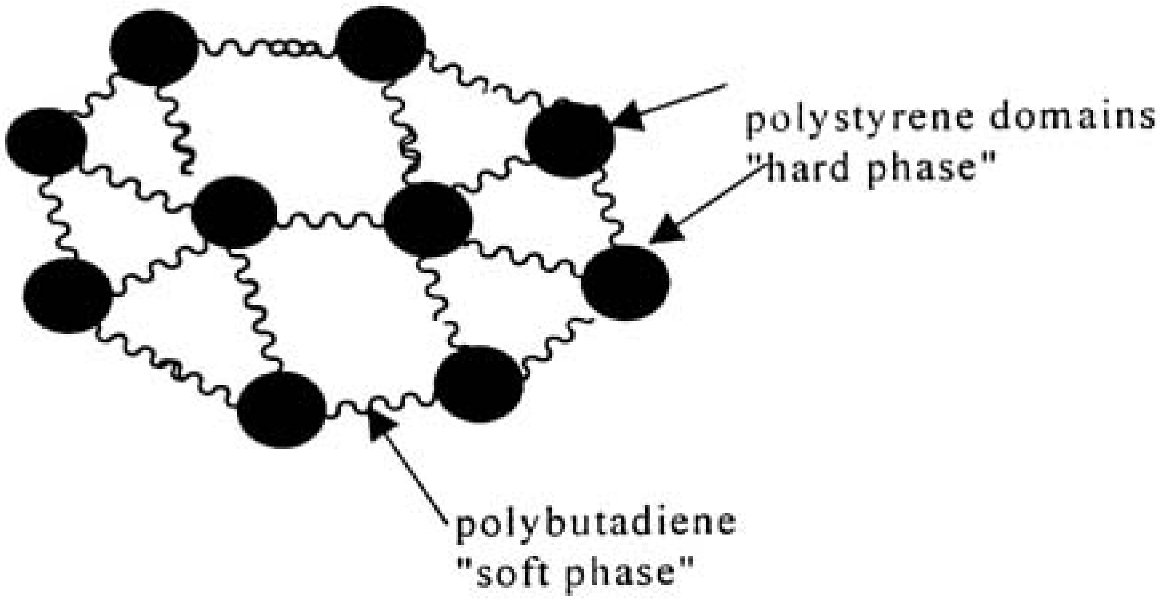
This class of elastomeric materials is called "thermoplastic" because they contain thermally reversible cross-links of various types. The types of crosslinks vary from phase-separated polystyrene domains, such as in styrenebutadienestyrene (SBS) elastomers to ionic cluster cross-links in the ionomers. The beauty of these noncovalent interactions to form cross-links is that when the material is heated, the cross-links are broken. This allows the polymer to flow and be processed, and also recycled. When cooled, the cross-links reform and the material becomes strong again. All TPEs are two-phase systems where there is a soft, rubbery "continuous phase" and a hard "dispersed phase" which does not flow at room temperature. The first two polymers to be discussed are "block copolymers" and "ionomers."

Block Copolymers

A very popular and useful TPE is made from blocks of styrene and butadiene monomers using anionic polymerization techniques, which was described in the solution SBR section above. They are made up of short chains of polystyrene (usually 800015,000 MW), followed by a much longer chain of polybutadiene (about 60,000 MW), and capped off by another short chain of polystyrene, hence the name SBS. Similar polymers are prepared using isoprene instead of butadiene (SIS). The differences between SBS and SIS will be discussed later in the subsection "Uses."

The linear polymers, as shown above, can be built up by the sequential addition of monomer or by coupling the living anionic chains using compounds like dichloro dimethylsilane. Hence, the base polymer would have styrene polymerized first, followed by butadiene, and then addition of the coupling agent. If a multifunctional coupling agent such as silicon tetrachloride is used, a radial block or "star-branched" SBS is formed.

The polystyrene is highly insoluble in the polybutadiene so the PS chains cluster together and phase-separate into domains. Since there is much more polybutadiene (PBD) than polystyrene, the PBD becomes the continuous phase containing dispersed particles of PS which act both as cross-links and reinforcing agents. Every PBD chain is tied to a PS chain on both ends so a very strong cross-linked network is formed. If the PBD is tied to only one PS, then a "diblock" polymer is formed which has very little strength. When this network is heated to above the glass transition of polystyrene (100°C) the PS domains break down and begin to flow, so the polymer can be processed by injection molding or extrusion. Upon cooling to below 100°C, the domains (cross-links) reform and the material becomes strong again. SBS elastomers can have tensile strength as high as conventional thermoset elastomers which may approach 4000 psi.



The butadiene blocks can be hydrogenated, as mentioned above with hydrogenated nitrile, to form SEBS polymers having better thermal stability and chemical resistance. The EB stands for ethylenebutylene, which are the structures formed after the butadiene segments have been hydrogenated.

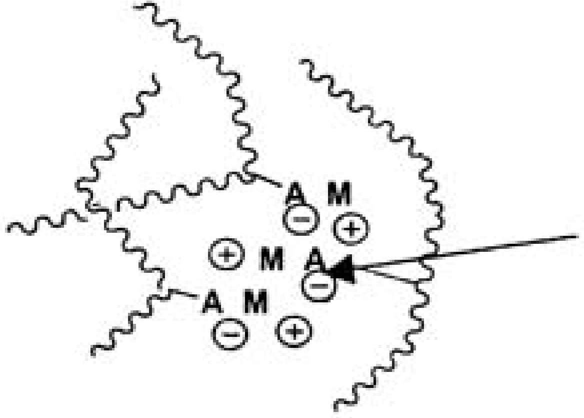
Other block copolymers which are useful are based on polyesters, polyurethanes, and ethylenepropylene. The first two have been discussed in other sections and the ethylenepropylene blocks will be discussed below in the subsection "Metallocene Polymers."

Uses

SBS copolymers are used in a wide variety of applications because of their clarity, toughness, and ease of processing. A major application is hot melt adhesives where they are compounded with hydrocarbon resins and oil.72 The SIS polymers are very popular in adhesives because the isoprene segments tend to undergo chain scission during aging instead of cross-linking, which is observed in butadiene polymers. This leads to better retention of adhesion after aging; however, the SIS polymers usually have poorer initial strength compared with SBS. Another high-volume use is in toughening of asphalt compounds for paving, crack sealants, and roofing. The SBS improves rutting and low-temperature performance in paving and crack resistance in roofing applications.73 SEBS would be the elastomer of choice in many adhesive and asphalt applications because of its superior aging properties; however, its high cost is prohibitive. SBS also is used widely to toughen polystyrene and high-impact polystyrene. This polyblending technique is used to toughen a number of plastics. Various other injection molding and extrusion applications include shoe soles and toys. SBS is limited in use because of its poor high-temperature performance.

Ionomers

Ionomers are copolymers in which a small portion of the repeat units have ionic pendant groups on usually a nonpolar backbone. The ionic groups tend to separate themselves into domains similar to the polystyrene segments in the SBS rubber because they are insoluble in the nonpolar polymer chains. Therefore, these ionic clusters serve as cross-links up to temperatures where they tend to disassociate. Most commercial grades of ionic elastomers are based on ethylene and propylene monomers. The polymer backbone usually contains sulfonyl or carboxylic acid groups and the metal counterion can be zinc, calcium, sodium, or lithium. The properties are highly dependent on the metal cation because they determine the temperature at which the ion clusters disassociate. The ionic cross-links usually impart outstanding tensile and tear strength properties because these cross-links are very good energy absorbers (hysteretic) due to their mobility. A process known as ion hopping provides one source for absorbing energy.



clusters of ionic groups forming a crosslink

M^+ = metal cation, A^- = anion

Uses

Solid golf balls are a good example of ionic cross-linked materials.74 The covers are usually lithium or sodium cross-linked plastics but the cores are also a type of ionomer. The cores are high *cis* polybutadiene which is compounded with zinc diacrylate (ZDA) and then peroxide cured. During the curing process, the ZDA grafts to the BR backbone creating a material with a very high cross-link density where the cross-links are clusters of zinc carboxylates with very high resilience. Other applications include heat-sealable food packaging, automotive trim, footwear, foamed sheets (mats), and interlayers for bulletproof glass.

Metallocene Elastomers

Metallocene catalysts are the latest innovations to make a big impact in the polymer industry. They have been used mostly to make new polyolefin plastics, such as very-high-molecular-weight, bullet-proof polyethylene, but they have also been used to make elastomers. The catalysts make very regular "stereospecific" polymers similar to the ZieglerNatta catalysts. They are based on various metals. Such as zirconium, complexed with cyclopentadienide anions. This type of compound is called a "zirconocene" and is used with organoaluminum to make highly regular polymers. The catalyst has the ability to flip back and forth from making *atactic* to *isotactic* polypropylene in the same polymerization. The alternating tacticity of the polymer breaks up the crystallinity of the chains and yields an elastomer. Metallocene catalysts are currently very expensive and cannot yet polymerize dienes such as butadiene, so they have only enjoyed limited commercial success in elastomers. However, this is one of the most intense fields of polymer research and many new product breakthroughs are expected in the near future.

RubberPlastic Alloys

A discussion on thermoplastic elastomers would not be complete without mentioning the elastomers produced by simply blending rubbers and plastics in an internal mixer using a process known as "dynamic vulcanization." The simplest of these elastomers is based on polypropylene and EPDM rubber. Products are made with various rubber contents (hardness values) by simply mixing the rubber and plastic in a Banbury or an extruder at high temperature while cross-linking the EPDM in situ with a phenolic curing resin or other curing agents such as sulfur or peroxides. The resulting blend processes like polypropylene but is actually an elastomer because of the cross-linked rubber phase which it contains. It can be reprocessed and recycled like other TPEs. The process is described in an excellent review article by S. Abdou-Sabet.75 Other alloys are based on nylon and NBR to obtain better thermal stability and solvent resistance. Constant improvements are being made in polymer alloys by using different combinations of plastics and rubbers and also new grafting and cross-linking chemistry to achieve properties more like conventional thermoset rubbers.

Uses

The applications for this type of elastomer have been limited because of the melting point of the plastic phase and poor compression set. For these reasons, alloys probably will never be successful as tire materials but they have many other automotive applications, including instrument panels, cowl vents, body panels, and bumpers. They also are used in appliances, lawn and garden equipment, and as grips on tools.

18.19 Plasticized Polyvinyl Chloride

Generally one thinks of polyvinyl chloride as a rigid plastic, which it is, with a glass transition temperature around 85°C. However, the addition of polar chemicals such as dioctyl phthalate can reduce the glass transition temperature below room temperature, producing a rubbery material.

Monomer Production

Vinyl chloride is produced primarily from ethylene, which is converted to ethylene dichloride either by chlorination or oxychlorination. The ethylene dichloride is cracked to give vinyl chloride and hydrogen chloride.

Production Process

Polyvinyl chloride is produced by the free-radical polymerization of vinyl chloride. Bulk, emulsion, solution, and suspension polymerization processes have been used.

The plasticized product can be produced by mixing the polymer and plasticizers at elevated temperatures, also by dry blending in which the plasticizer is absorbed into the resin and then heated. Solution blending is sometimes used, as well as the plastisol process, in which fine polyvinylchloride powder is dispersed in the plasticizer which is relatively stable until it is heated.

Properties and Use

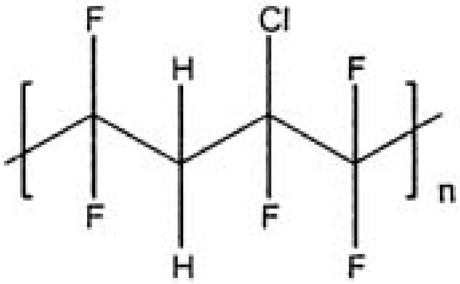
Plasticized polyvinyl chloride can be regarded as the first thermoplastic elastomer, as it is used in an uncross-linked form. Because of the lack of cross-linking, this material exhibits high rates of creep and stress relaxation. As with other thermoplastic elastomers, these disadvantages worsen as the temperature is increased. Although the polymer is saturated, it must be stabilized for use to prevent dehydrochlorination at processing temperatures. Because of the high chlorine content, polyvinyl chloride has excellent flame resistance as well as good electrical properties. Even at high plasticizer levels, these materials have marginal elasticity when compared with most other elastomers. Despite their shortcomings, they are used in many areas where they compete with other rubbers.

One of the larger uses is for wire and cable insulation. The flame resistance makes this the material of choice for residential wiring, extension cords, and so on. Inexpensive garden hose represents another large-volume use. Sports balls can be produced by rotational molding of plastisols. The low fabrication cost allows such products to dominate the lower-price-range market.76,77

18.20 Fluorocarbon Elastomers

The fluoroelastomers were developed as specialty materials for high heat applications and solvent resistance. By 1999 there was over 13,000 metric tons of fluoroelastomer capacity worldwide.78 The elastomers were by-products of Plunkett's 1938 discovery of polytetrafluoroethylene. Copolymers of olefins with vinylidene fluoride were found to be leathery, whereas elastomers were made if tetrafluoroethylene (TFE) or trifluoropropene were used. The first commercial product was produced as a copolymer of vinylidene fluoride and chlorotrifluoroethylene (CTFE) and was called Kel-F. The later, more stable polymers used hexafluoropropene (HFP) with TFE. They were first commercialized by DuPont under the trade name Viton and then a few other companies added their own fluoroelastomers. These elastomers are usually prepared by radical polymerization in emulsion using catalysts such as ammonium persulfate and sometimes with chain transfer agents such as carbon tetrachloride or halogen salts.79 Caution must be exercised when using these polymerizations because the fluoromonomers can be explosive. They are usually isolated by coagulation and normal rubber drying techniques and sold in the form of pellets, slabs, or rubber crumb.

As a result of being very unreactive, the fluoroelastomers cure very slowly and usually require a high-temperature post-cure. The curatives are designed to remove hydrogen fluoride to generate a cure site which can react with a diamine or bisphenol. They are also cured with organic peroxides. Most fluoroelastomers can be compounded using normal rubber processing equipment such as rubber mills and internal mixers. Processing aids such as dioctyl phthalate or waxes can be used to obtain smoother extrusions and better mold release.



copolymer of TFE with vinylidene fluoride (VF)

Uses

Fluoroelastomers have outstanding heat resistance when cured. Some vulcanizates have almost indefinite service life at temperatures up to 200°C. The perfluoro polymers, such as DuPont's Kalrez, have short-term useage at up to 316°C and extended service as high as 288°C. This polymer is extremely expensive and is only offered in the form of finished parts, usually O-rings, seals, or gaskets. The fluoroelastomers also have excellent solvent and ozone resistance making them ideal for automotive fuel hoses. Although many of the applications involve small seals, O-rings, gaskets, and hose, the single largest applications is flue-duct expansion joints. The polymer's resistance to high temperature and wet acidic flue gases are critical in this application.

18.21 References

Whitley, G. S., Davis, C. C., and Dunbrook, R. F. (Eds.), Synthetic Rubber, John Wiley & Sons, New York, 1954.
Blackely, D. C., Synthetic Rubber: Their Chemistry and Technology, Chapter 2, Applied Science Publishers, London, 1983.
Worldwide Rubber Statistics 2000, p. 6, International Inst. of Synthetic Rubber Producers, Inc., Houston, 2000.
Fox, T. G., Grateh, S., and Loshaek, S., RheologyTheory and Applications, Vol. 1, Academic Press, New York, 1956.
Billmeyer, F. W., Jr., Textbook of Polymer Science, John Wiley & Sons, New York, 1984.
Odian, G., Principles of Polymerization, pp. 1924, McGraw-Hill, New York, 1970.
Flory, P. J., Principles of Polymer Chemistry, Chapter 7, Cornell University Press, Ithaca, NY, 1953.
Rodriguez, F., Principles of Polymer Systems, pp. 138142, McGraw-Hill, New York, 1982.
Rudin, A., The Elements of Polymer Science and Engineering, p. 126, Academic Press, New York, 1982.
Rudin, A., The Elements of Polymer Science and Engineering, p. 402, Academic Press, New York, 1982.
Krause, S., J. Macromol. Sci., C7(2), 251 (1972).
Aklonis, J. J., Introduction to Polymer Viscoelasticity, John Wiley & Sons, New York, 1972.
Oberster, A. E., Bouton, T. C., and Valaitis, J. K., Die Angewandte Markomoleculare Chemie, 29/30, 291 (1973).
Flory, P. J., J. Chem. Phys., 17, 223 (1949).
Natta, G., in Polymer Chemistry of Synthetic Elastomers, Part 1, J. Kennedy and E. Tornqvist (Eds.), Chapter 7, John Wiley & Sons, New York, 1969.
Semegen, S. T., and Cheong, S. F., Vanderbilt Rubber Handbook, pp. 1841, R.T. Vanderbilt Company, Inc., 1978.
Webster, C. C., and Baulkwill, W. J., Chapters 1 and 11, Rubber, John Wiley & Sons, New York, 1989.
Brandrup, J., and Immergut, E., Polymer Handbook, Vol. III, p. 54, John Wiley & Sons, New York, 1975.
Treloar, L., The Physics of Rubber Elasticity, Clarendon Press, Oxford, 1949.
Poh, W. N., Elastomers, p. 12 (1989).
Synthetic Rubber: The Story of an Industry, International Inst. of Synthetic Rubber Producers, New York, 1973.
Stavelv, F. W., Ind. Eng. Chem., 48, 778 (1956), presented at Div. Rubber Chem., ACS, Philadelphia, PA, 1955.
Eng. Patent 827365, to GoodrichGulf Chem., Dec. 1954, priority data.
Worldwide Rubber Statistics 2000, International Inst. of Synthetic Rubber Producers, Inc., pp. 6263, Houston, 2000.
Worldwide Rubber Statistics 2000, International Inst. of Synthetic Rubber Producers, Inc., pp. 19, Houston, 2000.
Renninger, T. J., presentation to International Rubber Study Group, Ottawa, Sept. 1990.
Faith, W. L. et al., Industrial Chemicals, p. 731, John Wiley & Sons, New York, 1957.
Dunbrook, R. F., India Rubber World, 117, 203207 (1947).
Gardon, J. L., Rubber Chem. Technol., 43, 7494 (1970).
Storey, E. B., Rubber Chem. Technol., 34, 1402 (1961).
Kuntz, L., J. Poly. Sci., 54, 569 (1969).
Bouton, T., and Futamura, S., Rubber Age, 3, 33 (1974).
Mostert, S., and Van Amergongen, G., British Patent 1,136,189, 1968.
Day, G., and Moore, D., paper at 26th Annual Meeting, IISRP, May 1985.
Quirk, R. P., Rubber Chem. Technol., 69(3), 444 (1996).
Oshima, N., Salcacobore, M., and Tsutsumi, F., paper at 27th Annual Meeting, IISRP, May 1986.
Worldwide Rubber Statistics 2000, International Inst. of Synthetic Rubber Producers, Inc., p. 9, Houston, 2000.
Ponder, T., Hydrocarbon Processing, 55(10), 119121 (1976).
Womeldroph, D., Am. Petrol. Inst. (May 14, 1958).
Welch, M., Hydrocarbon Processing, 57(11), 131136 (1978).
Adams, H., Farhat, K., and Johnson, B., Ind. Eng. Chem. Prod. Dev., 5(2), 127 (1966).
Halasa, A., Schulz, D., Tate, D., and Mochel, V., in Advances in Organometallic Chemistry, F. Stone and R. West (Eds.), Vol. 18, Academic Press, New York, 1980.
Tate, D., and Bethea, T., Encyclopedia of Polymer Science and Engineering, Vol. 2., pp. 568572, John Wiley & Sons, New York, 1985.
Lauretti, F., and Gargani, L., 27th Annual Meeting, IISRP, June 1987.
DeChirico, A., Lamzani, P., Eaggi, E., and Bruzzone, M., Makromol. Che., 175, 2029 (1974).
Buckenell, C., Toughened Plastics, Applied Science, London, 1977.
Maruko, T., US Patent No. 6,071,201 (2000) Solid Golf Ball.
Natta, G., J. Poly. Sci., 51, 411 (1961).
Lukach, C., and Spurlin, H., Copolymerization, G. Ham (Ed.), Interscience, New York, 1964.
Crespi, G., and DiDrusco, G., Hydrocarbon Processing, 48, 103107 (Feb. 1969).
Vandenberg, E., and Repka, B., in Ref. 12, Chapter 11.
Scagliotti, F., Milani, F., and Galli, P., paper at 26th Annual Meeting, IISRP, May 1985.
Miles, D., and Briston, J., Poly Technol, p. 299, Chemical Publishing Company, New York, 1963.
Powers, K. W., US Patent No. 5,162,445 (1993) *Para-alkylstyrene/isoolefin copolymers*.
"Butyl Rubber Reaches 50 Year Mark," Elastomerics, 3031 (Mar. 1988).
Konrad, E., and Tschunkur, E., U.S. Patent 1,973,000 to 1. G. Farbenindustrie, 1934.
Worldwide Rubber Statistics 2000, International Inst. of Synthetic Rubber Producers, Inc., p. 9, Houston, 2000.
McKenzie, G., paper at 27th Annual Meeting, IISRP, May 1986.
Hashimoto, K., et al., paper at 26th Annual Meeting, IISRP, May 1985.
Buding, H., US Patent No. 4,581,417 (1986), *Production of Hydrogenated Nitrile Rubber*.
Whitley, G. S., Davis, C. C., and Dunbrook, R. F. (Eds.), Synthetic Rubber, John Wiley & Sons, New York, 1954, p. 770.
Miles, D., and Briston, J., Poly Technol, p. 305, Chemical Publishing Company, New York, 1963.
Blackely, D. C., Synthetic Rubber: Their Chemistry and Technology, Chapter 2, Applied Science Publishers, London, 1983, pp. 175194.
Brodrecht, L., Chemical Economics Handbook, SRI International, 1989.
Polmateer, K., Rubber Chemistry and Technology, 16(3), 470 (1988).
Semegen, S. T., and Cheong, S. F., Vanderbilt Rubber Handbook, pp. 216232, R.T. Vanderbilt Company, Inc., 1978.
Natta, G., in Polymer Chemistry of Synthetic Elastomers, Part 1, J. Kennedy and E. Tornqvist (Eds.), Chapter 8, John Wiley & Sons, New York, 1969.
Morton, M., and Fetters, L., Polymerization Processes, Chapter 9, John Wiley & Sons, New York, 1977.
Handbook of Elastomers, pp. 643659, Marcel Dekker, Inc., New York, 1988.
Broadrecht, L., Mulach, R., and Tauchiya, K., in Chemical Economics HandbookElastomers, SRI International, Menlo Park, CA, 1989.
Semegen, S. T., and Cheong S., Vanderbilt Rubber Handbook, pp. 1841, R.T. Vanderbilt Company, Inc. 1978.
Kraton Polymers for Adhesives & Sealants, Shell online literature, Shell Chemical Website.
Bull, A., and Vonk, W., Shell Chemicals Technical Manual TPE 8: 15 Report, 1988 (Shell Website).
Maruko, T., US Patent No. 6,071,201, (2000) Solid Golf Ball.
Abdou-Sabet, S., Rubber Chem. Technol., 69(3), 476 (1996).
Penn, W., PVC Technology, Applied Science Publishers, London, 1971.
Miles and Briston, Polymer Technology, p. 159, Chemical Publishing Company, New York, 1963.
Worldwide Rubber Statistics 2000, International Inst. of Synthetic Rubber Producers, Inc., p., Houston, 2000.
Morton, M., Rubber Technology, p. 410, Chapman & Hall, London 1995.

19.1 Introduction

Definition

Plastic (*adj.*) is defined by Webster as "capable of being molded or modeled (e.g., clay…capable of being deformed continuously and permanently in any direction without rupture." Plastic (*n.*) in modern industry covers high-molecular-weight organic compounds which can be formed into any desired shape and then solidified into a useful product which can withstand the mechanical stresses normally applied to it.

History

Commercial plastics began in 1868 when John and Isaiah Hyatt plasticized cellulose nitrate and molded it into billiard balls, and later into combs, brushes, and other useful articles. Commercial *synthetic* plastics began in 1908 when Leo Baekeland reacted phenol, formaldehyde, and wood flour and molded them into electrical insulators and a growing variety of other product specialties. During the next 30 years, a variety of plastics appeared as interesting specialties. During World War II the U.S. and German governments met the severe demands of advanced military technology by supporting vast research, development, plant construction, and manufacturing of a number of major polymers for plastics and rubber applications. With the end of the war in 1945, all of this technology and production capacity converted to civilian products, and commodity plastics began their tremendous growth, first polyvinyl chloride and polystyrene, then polyethylenes, polypropylene, polyesters, and polyurethanes, along with a constantly expanding range of more specialized polymers for more demanding applications.

Advantages of Plastics over Conventional Materials

We have had conventional structural materialsmetals, ceramics, glass, wood, leather, textiles, paperfor thousands of years, during which time we have been able to explore and exploit them thoroughly to full maturity. By comparison, plastics are so new that we are still learning to develop and use them. Their exponential growth is due to the fact that they offer many advantages over conventional materials.

Processability and Product Design

It is much easier to convert plastic materials into an almost unlimited range of products.

Modulus

Plastics cover an extremely broad range from extremely rigid to stiffly flexible to extremely soft and rubbery, sometimes even within a single chemical family.

Elasticity

Their recovery from deformation is superior to almost all conventional materials.

Impact Strength

Plastics are much less brittle than ceramics, glass, and paper, and some families can be made almost unbreakable.

Lubricity and Abrasion Resistance

Specific plastic materials offer outstanding self-lubricating performance and abrasion resistance.

Thermal Insulation

Plastics offer very good insulation against heat and cold, and can be foamed to increase their insulating qualities much further.

Flame-Retardance

Many plastics are less flammable than wood and paper, and most can be formulated to make them much more resistant to burning.

Electrical Insulation

Plastics are excellent electrical insulators. They can also be formulated to provide semi-conductivity, and high or low dielectric constant and loss.

Transparency/Opacity

Some plastics can approach or equal the transparency of glass. All can be formulated to a wide range of translucency to opacity.

Color and Appearance

Most conventional materials have a very limited color range, and they simply look like what they are. Plastics can be produced in an almost infinite range of colors, transparent or opaque, and surfaces to simulate all of the conventional materials or create totally new ones.

Chemical Resistance

Plastics are generally far superior to metals in corrosion resistance. Among the 100 families of commercial plastics, they offer a wide variety of chemical resistance, or solubility/reactivity, as desired for different types of products.

Water Resistance

It is far superior to wood and paper.

Permeability

Different families of plastics offer a wide range from highly impermeable barrier materials to membranes of different separation abilities.

Weathering

Some plastics are very resistant to weather. Others have moderate resistance. Still others can be designed for self-destruction to alleviate collection of solid waste.

Cost

Contrary to some popular belief, plastics are not "cheap." But their superior processability often makes the finished product less expensive.

This economic advantage, plus all their other advantages listed above, accounts for their success in replacing conventional materials, and in leading to new products which were not even possible before.

Markets for Plastics

The U.S. plastics market has grown past 100 billion pounds/year. The largest share (24%) goes to packaging, both rigid packages and film. Close behind is building and construction (20%), mainly piping, plywood, siding, insulation, and flooring. Smaller, more specialized amounts go into agriculture, aircraft, appliances, autos and trucks, electrical and electronics, furniture, glazing, housewares, luggage, marine, medical, office equipment, optical, tools, toys, and miscellaneous industrial and consumer products. Thus, taken as a whole, plastics is one of the largest and fastestgrowing industries in the United States.

Major Classes of Plastic Materials

The primary binary classification is the distinction between thermoplastics and thermosets. Thermoplastics are stable large molecules, typically molecular weights 104106, which soften on heating to permit melt processing, and solidify on cooling to give solid finished poducts; the process is reversible, so they are essentially recyclable within the limits of their thermal stability. Thermosetting plastics are reactive low-molecular-weight polymers, which may be melted or even poured, shaped into final products, and then reacted further into cross-linked molecules of essentially infinite molecular weight; the process is essentially irreversible, so they are difficult or impossible to recycle. Thermoplastic processing is simpler and more economical, so it accounts for about 85 percent of the plastics market. On the other hand, thermosetting plastics permit many special processes and offer outstanding final properties, which accounts for the health of their share of their more specialized markets.

Another important distinction is based on: (1) filled and reinforced plastics, and (2) foams. (1) When any family of polymers is combined with particulate inorganic fillers, this produces major increase in density, modulus, dimensional stability, heat transfer, dielectric constant, and opacity, and frequently a decrease in cost. When the fillers are reinforcing fibers, they can further produce a great increase in strength, impact resistance, and dimensional stability. Thus, these properties may depend more upon the use of fillers and fibers, than upon the choice of the particular polymer in which they are used.

(2) When a polymer is liquified, foamed, and solidified to trap the air spaces within it, air contributes so much to the final properties that it may be more important than the particular polymer in which it is dispersed. The most outstanding effects are flotation by closed-cell foams, softness in open-cell foams, impact cushioning, thermal and electrical insulation, and permeability in open-cell foams.

Following these introductory remarks, we turn to a study of the fundamental aspects of Polymer Chemistry (Part I), followed with a discussion of Commercial Plastic Materials (Part II), and conclude with Plastic Processing (Part III).

Part I. POLYMER CHEMISTRY

Materials are often classified as metals, ceramics, or polymers. Polymers differ from metals and ceramics, by their lower densities, thermal conductivities, and moduli. A vast array of products utilize plastic materials. For example, in applications requiring lighter weight, plastics offer an advantage over other choices as a result of their lower density. Polymeric materials are used in automotive, packaging, and consumer goods, just to name a few. The requirements for these diverse applications vary greatly, but through proper control, plastic materials can be synthesized to meet these varied service conditions.

19.2 Molecular Weight

A polymer is prepared by linking a low molecular weight species, called a monomer (such as ethylene), into an extremely long chain, called a polymer (such as polyethylene), much as one would string together a series of beads to make a necklace (see Fig. 19.1). As molecular weight increases, the properties of the material change. Looking at the alkane hydrocarbon series with the general structure H(CH2)nH, we can see the material change from a gas for values of n = 14, a liquid for n = 511, a high viscosity liquid from n = 1625, a crystalline solid for n = 2550, to a tough plastic solid for n = 10005000.1 The molecular weight affects both the mechanical and processing behavior of the polymer. In general, higher molecular weights result in improved mechanical properties, but face more difficulty in processing. Unlike low molecular weight species, polymeric materials do not possess one unique molecular weight, but rather a distribution of molecular weights as depicted in Fig. 19.2. Molecular weights for polymers are usually described by two different average molecular weights, the number average molecular weight, $\overline{M}_{\mathrm{n}}$, and the weight average molecular weight, $\overline{M}_{\mathrm{w}}$. These averages are calculated using the equations below:

$$\overline{M}_{\mathrm{n}} = \sum_{i=1}^{\infty} \frac{n_i M_i}{n_i}$$

$$\overline{M}_{\mathrm{w}} = \sum_{i=1}^{\infty} \frac{n_i M_i^2}{n_i M_i}$$

where ni is the number of moles of species i and Mi is the molecular weight of species i.

Fig. 19.1 Polymerization.

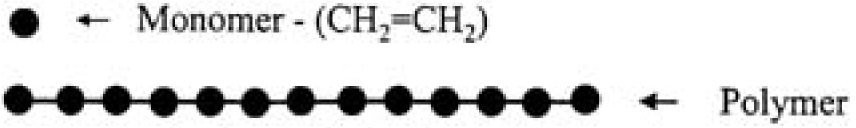
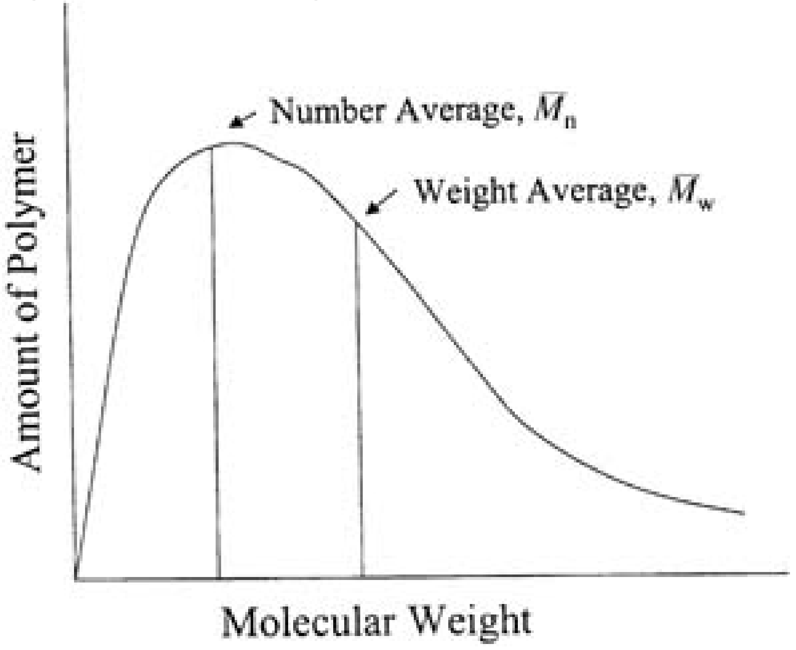


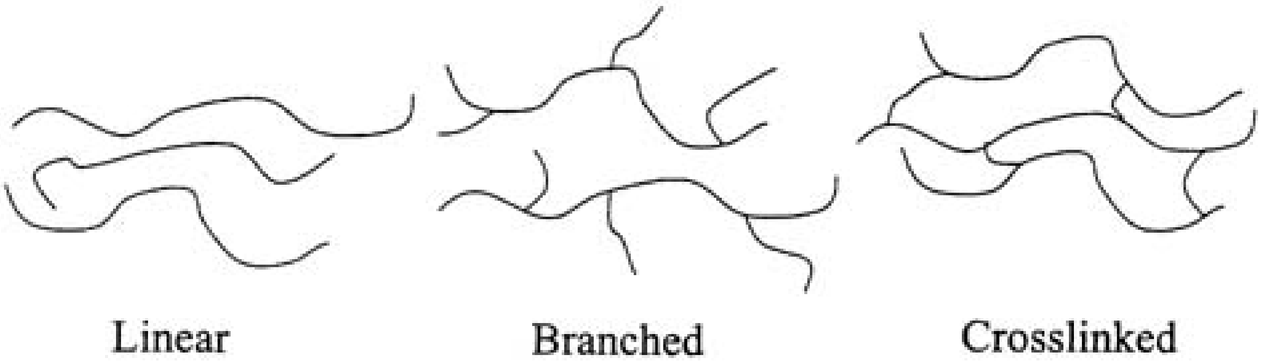
Fig. 19.2 Molecular weight distribution.



19.3 Chain Structure

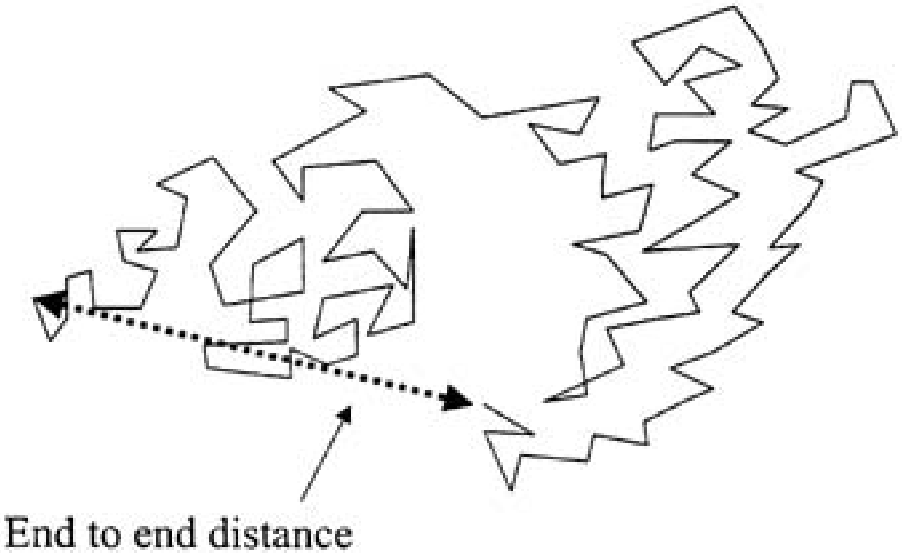
Polymerization can produce linear chains, but other structures can exist as well. As shown in Fig. 19.3 branched and crosslinked structures can be formed. Linear and branched structures can be shaped and reshaped simply by heating and are called *thermoplastics*. In the case of a crosslinked structure a three-dimensional network is formed that cannot be reshaped by heating. This type of structure is called a *thermoset*.

Fig. 19.3 Polymer structures.



Macromolecular conformations describe the positions of the atoms that occur due to rotation about the single bonds in the main chain.[2] Polymer chains in solution, melt, or amorphous state exist in what is termed a random coil. The chains may take up a number of different conformations, varying with time. Figure 19.4 shows one possible conformation for a single polymer chain. In order to describe the chain, polymer scientists utilize the root mean square end-to-end distance ($\langle r^2 \rangle^{1/2}$), which is the average over many conformations. This end-to-end distance is a function of the bond lengths, the number of bonds, and a characteristic ratio, *C*, for the specific polymer.

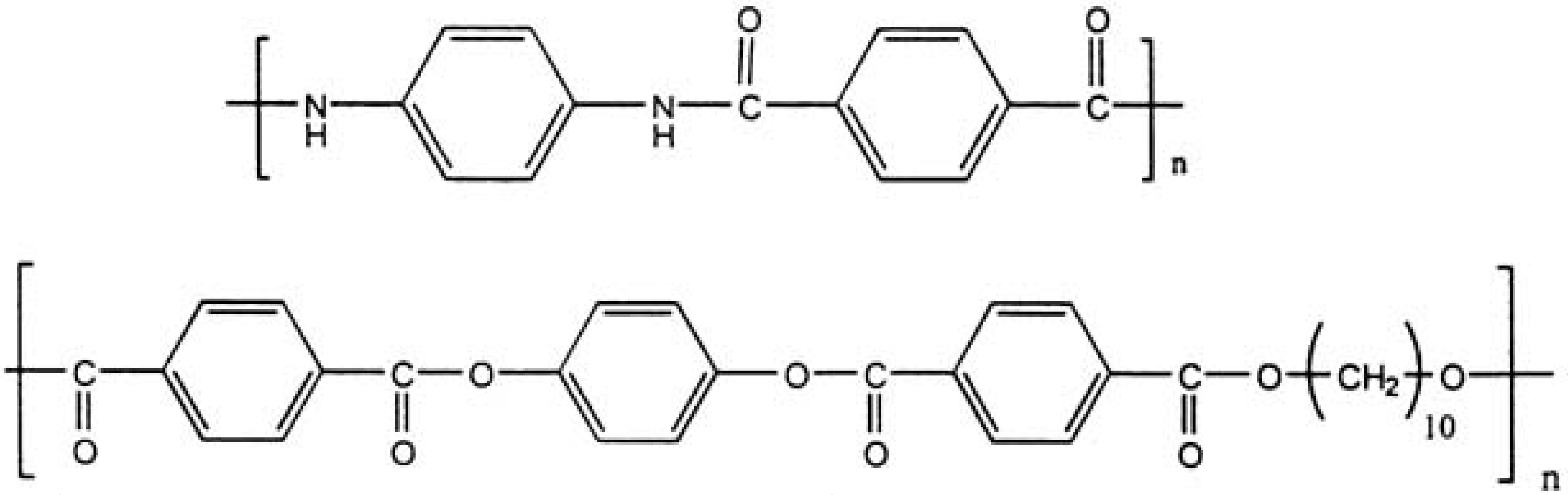
Fig. 19.4 Random coil chain.



19.4 Chemical Structure

The chemical characteristics of the starting low molecular weight species will help determine the properties of the final polymer. Along the chain axis primary bonds hold the atoms together and determine molecular properties such as flexibility and glass transition temperatures. Flexibility of the chain is governed by the ease of rotation about main chain bonds. The presence of methylene units or carbonoxygen single bonds act to increase the flexibility of the chain. Groups or interchain interactions (described below) that tend to restrict rotation will decrease the flexibility of the molecule. An extreme example of this principle is the rigid rod polymers. Figure 19.5 shows several examples of these types of structures. The aromatic groups along the backbone act to stiffen the polymer chain and restrict rotation, causing the polymer to remain straight, much like a log.3 Materials such as these may exhibit liquid crystalline behavior.

Fig. 19.5 Rigid rod-type polymers.

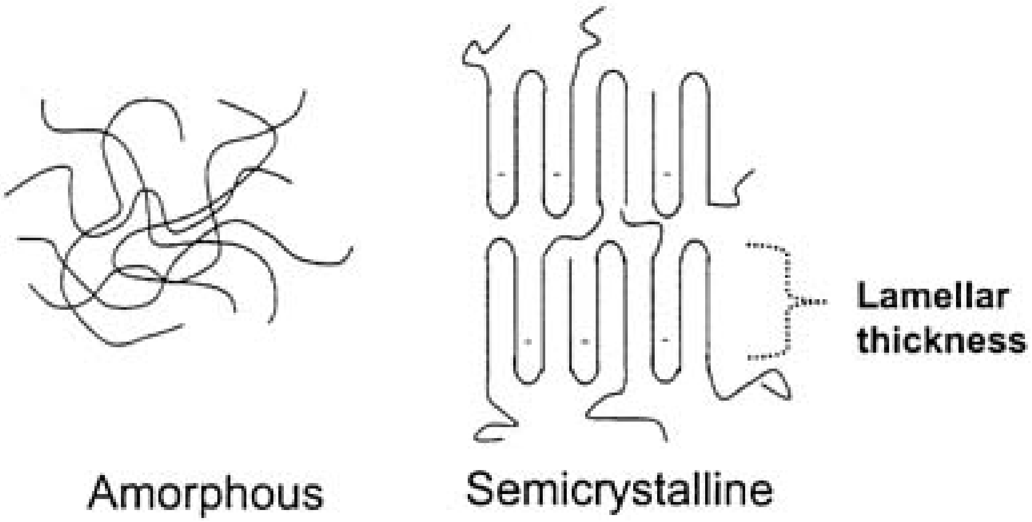


The forces holding the many individual chains together (interchain forces) are determined by secondary bonds, except in the case of thermosets where primary bonds hold the chains together. The type and strength of the secondary bonds (often termed van der Waals forces)4 will depend on the structure of the polymer. In the case of hydrocarbon polymers, such as polyethylene, the secondary bonds are dispersion forces. For polymers containing carbon and oxygen groups, such as the polyesters, the presence of the C=O bond results in a dipole due to the different electronegativities of the carbon and oxygen atoms. The presence of polarity in the polymer will act to increase the intermolecular forces. In the case of polyamides, hydrogen bonding between the polymer chains leads to high intermolecular forces. The strength of the intermolecular forces will affect the properties of the polymer such as viscosity, solubility, miscibility, surface tension, and melting point. As discussed above, interchain forces will also affect the flexibility of the chain if they restrict free rotation.

19.5 Morphology

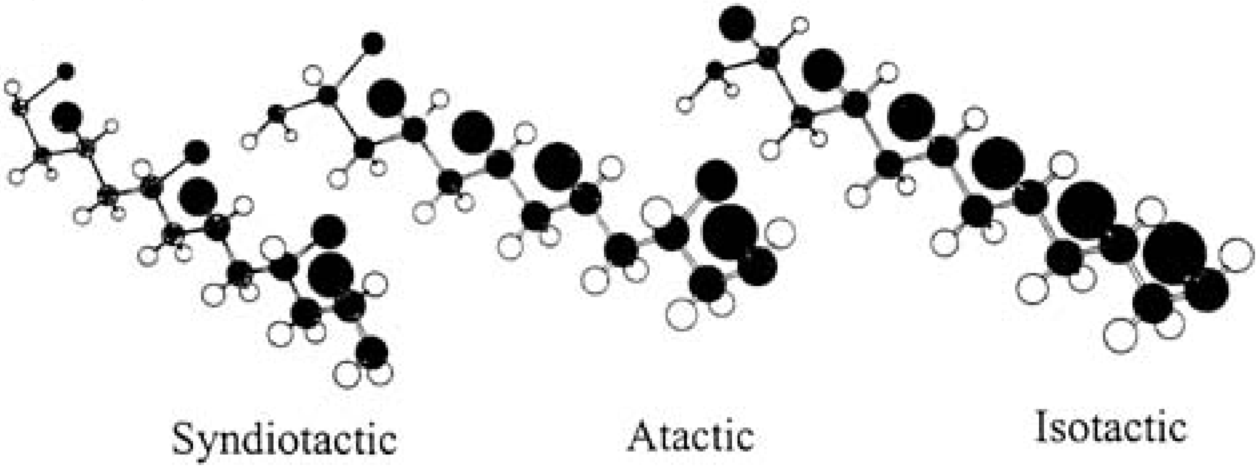
In its solid form a polymer can exhibit different morphologies depending on processing conditions and the structure of the polymer chain. Amorphous polymers show no order to the arrangement of the chains. The chains are entangled with each other, much like the strands of spaghetti on one's plate. An example of an amorphous polymer is polystyrene. If the polymer backbone has a regular, ordered microstructure, then the polymer can pack tightly into an ordered crystalline structure, although the material will generally be only semicrystalline. Examples of semicrystalline polymers are polyethylene and polypropylene. The amorphous and crystalline chain structures are depicted in Fig. 19.6.

Fig. 19.6 Amorphous and semicrystalline structures.



The exact make-up and architecture of the polymer backbone will determine the ability of the polymer to crystallize. Figure 19.7 shows the different types of microstructure that can be obtained for a vinyl polymer. Isotactic and syndiotactic structures are considered stereospecific polymers, and their highly regular backbone structure allows them to crystallize. The atactic form is irregular and would produce an amorphous material. This nature of the polymer microstructure can be controlled by using different synthetic methods. As will be discussed below, the ZieglerNatta catalysts are capable of controlling the microstructure to produce different types of stereospecific polymers.

Fig. 19.7 Syndiotactic, atactic, and isotactic structures.

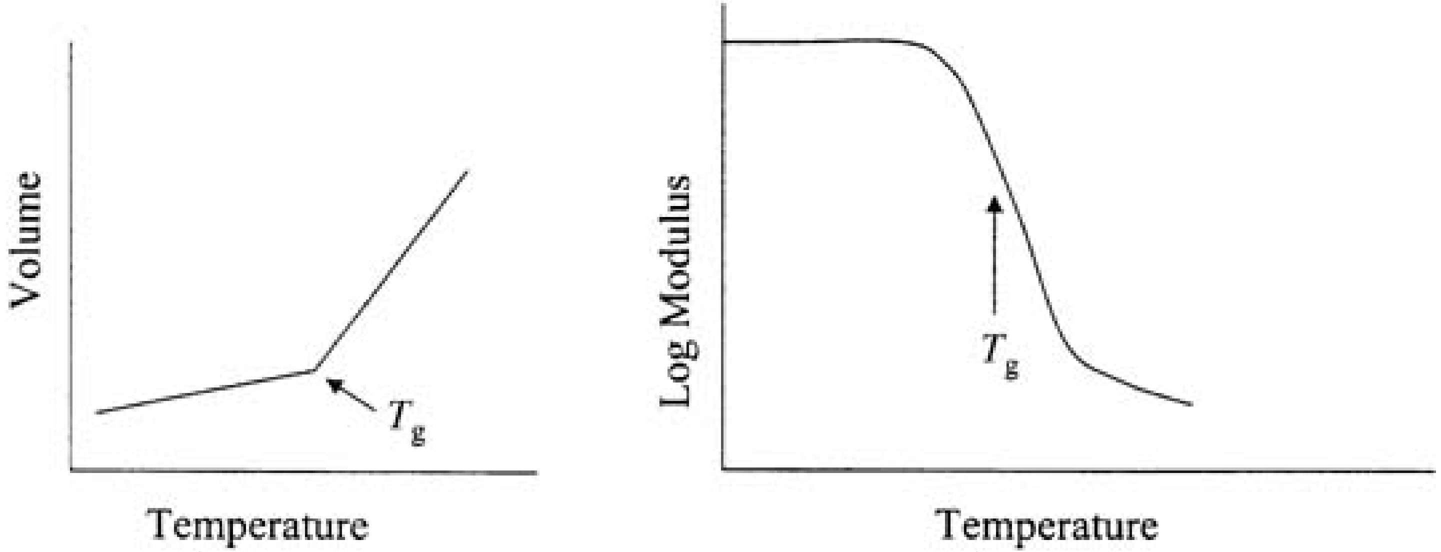


19.6 Transition Temperatures

Glass Transition Temperature (*T*g)

Chain flexibility is governed by molecular structure, but is also affected by temperature. As the temperature is reduced, amorphous polymers reach a temperature where large-scale (2050 chain atoms) segmental motion ceases.5 This temperature is called the glass transition temperature or *T*g. Volumetemperature plots can be used to indicate the *T*g. At the *T*g transition temperature a several decade change in the modulus occurs and the material changes from a rigid solid to a rubbery material. At still higher temperatures, provided the material is a thermoplastic, it becomes a liquid, which can flow and be processed. The behavior in the glass transition region is depicted in Fig. 19.8.

Fig. 19.8 Behavior at the glass transition temperature.



The location of the glass transition temperature will depend on the nature of the polymer. Generally, a plastic differs from a rubbery material due to the location of its glass transition temperature. A plastic has a *T*g above room temperature, while a rubber has a *T*g below room temperature. As previously mentioned the flexibility of the chain will affect the value of *T*g. Flexible groups will tend to lower the *T*g, while stiffening groups will act to increase it. Side groups can also affect the value. The effect of various factors on the value of *T*g is shown in Table 19.1. Aliphatic side groups will tend to have decreasing *T*g's as the length of the side chain increases; however, rigid side groups will tend to increase the *T*g. Increased polarity of the polymer will also increase the glass transition temperature as shown by the polymers polypropylene, polyvinyl chloride, and polyacrylonitrile. Symmetry about the backbone can act to decrease the *T*g as illustrated by the pairs, polypropylene and polyisobutylene, and polyvinyl chloride (PVC), and polyvinylidene chloride.

TABLE 19.1 Effect of Chain Structure on *T*g6

	Tg(°C)
Side groups	
Polypropylene	-10
Polystyrene	100
Polarity	
Polypropylene	-10
PVC	85
Polyacrylonitrile	101
Symmetry	
Polypropylene	-10
Polyisobutylene	-70
PVC	87
Polyvinylidene chloride	-19

Increases in number average molecular weight, $\overline{M}_n$, and cross-link density will both act to increase the *T*g. The addition of plasticizers, such as in the case of PVC, will decrease the value of the glass transition temperature. The glass transition temperature of the plasticized material can be estimated if the glass transition temperature of the two components (A and B) and their weight fractions are known.7

$$\frac{1}{T_g} = \frac{W_A}{T_{gA}} + \frac{W_B}{T_{gB}}$$

where *T*gA and *T*gB are the glass transition temperatures of components A and B, respectively, and *W*A and *W*B are the weight fractions.

One difficulty in obtaining values for the glass transition temperature is its dependence on measurement rate. When experiments are conducted at slow rates, the measured values will be lower than those measured at more rapid rates. Other difficulties include experimental problems and many definitions and interpretations on how to measure the values. Measurement of the temperature at which a step change in the volumetric thermal expansion coefficient occurs, when heating and cooling rates are 1°C/min, is perhaps one of the less ambiguous methods.8 The measured value of the glass transition temperature will increase approximately 3°C (volumetric measurements) to 7°C (maximum in tan δ from dynamic mechanical analysis) for a decade change in rate.

Crystallization and Melting Points (*T*M)

A number of polymers may exhibit the ability to crystallize. As previously mentioned, such polymers are semicrystalline, meaning they will have regions of amorphous and crystalline material dispersed throughout the part. Both the morphology and degree of crystallinity can be affected by the processing conditions used to manufacture the part. These changes can greatly affect the mechanical behavior of the material. The degree of crystallinity and the melting point are often measured using differential scanning calorimetry.

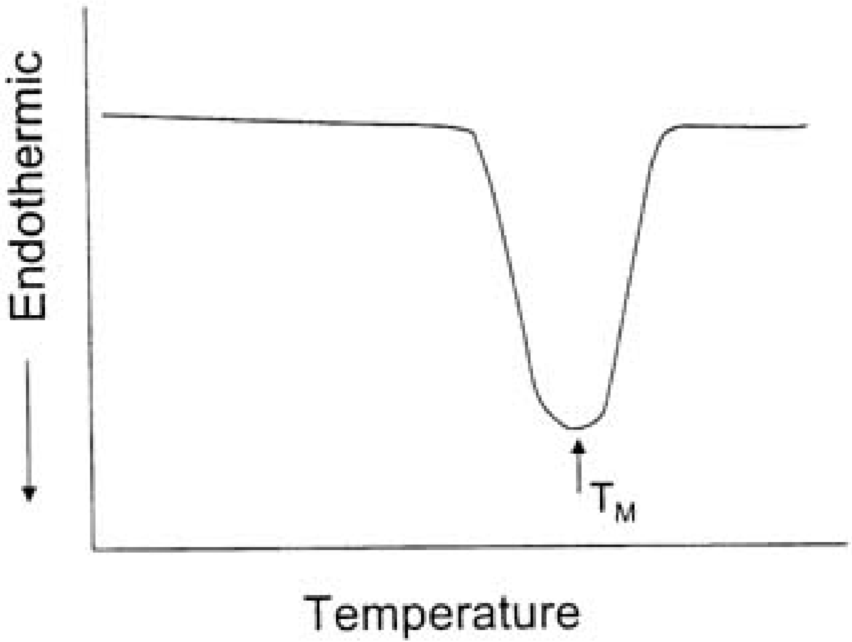
The general structure of the crystalline regions is rather complex. Polymer chains appear to fold back and forth into a lamellar structure as shown in Fig. 19.6. These lamellae form layers of ribbon like structures, tied together with amorphous regions.

Semicrystalline polymers do not exhibit a single sharp melting point, but rather melt over a range of temperatures. In addition, the melting point of the material will be higher the greater the lamellar thickness, which can be controlled by the crystallization temperature.9 Higher crystallization temperatures generally lead to greater lamellar thickness.

Polymer structure will affect the melting point of the polymer in a number of ways. Polymers containing polar groups will have higher melting points. Polymers with hydrogen bonding would be expected to have even higher melting points. For example, polyethylene has a melting point of 135°C, while nylon 6 has a melting point of 265°C. The presence of chain stiffening groups in the backbone will act to raise the melting point of the polymer; contrast the melting point of polyethylene (135°C) with poly *p*-xylene (400°C). In general, polymers with high interchain forces and rigid chains will tend to have the highest melting points.

In addition to the melting point, the degree of crystallinity will influence the behavior of the polymer. Differential scanning calorimetry (DSC) can be used to determine both the melting point and the percent crystallinity. Figure 19.9 shows a representative DSC curve for a melting point. The area under the melting peak is related to the percent crystallinity of the polymer sample. The percent crystallinity may have a dramatic effect on the mechanical behavior of the material. Above the *T*g of the material, the presence of crystalline regions will act to increase the rigidity of the polymer, resulting in an increasing modulus with increasing percent crystallinity.10

Fig. 19.9 Differential scanning calorimetrymelting point.



The kinetics of crystallization are a function of the crystallization temperature.11 As the crystallization temperature is decreased the rate of crystallization increases, up to a maximum. When the crystallization temperature is lowered beyond this point the rate begins to slow down as the molecular motion of the chains is reduced. At temperatures below the *T*g crystallization is suppressed.

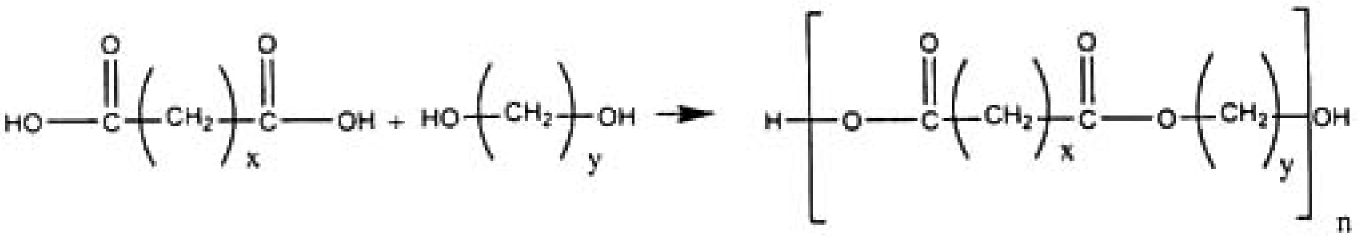
19.7 Polymerization

There are two primary polymerization approaches: step-reaction polymerization and chain-reaction polymerization.12

Step-Reaction Polymerization

In step-reaction polymerization (also referred to as step-reaction polymerization condensation polymerization) any two polyfunctional (typically bifunctional) monomers of the correct chemical species can react, often liberating a small molecule such as water. As the reaction continues, higher molecular weight species are produced as longer and longer polyfunctional groups react together. For example, two monomers can react to form a dimer, then react with another monomer to form a trimer. The reaction can be described as n-mer + m-mer → $(n + m)$mer, where n and m refer to the number of monomer units for each reactant. Monomer functionalities of the two will produce linear polymers. With functionalities greater than two, branching or crosslinking can occur. Molecular weight of the polymer builds up gradually with time and in order to produce high molecular weight polymers high conversions are usually required. Polymers synthesized by this method typically have atoms other than carbon in their backbone. Examples of polymers produced by step-growth polymerization include polyesters and polyamides. A typical step growth reaction is shown in Fig. 19.10.

Fig. 19.10 Step growth polymerization.

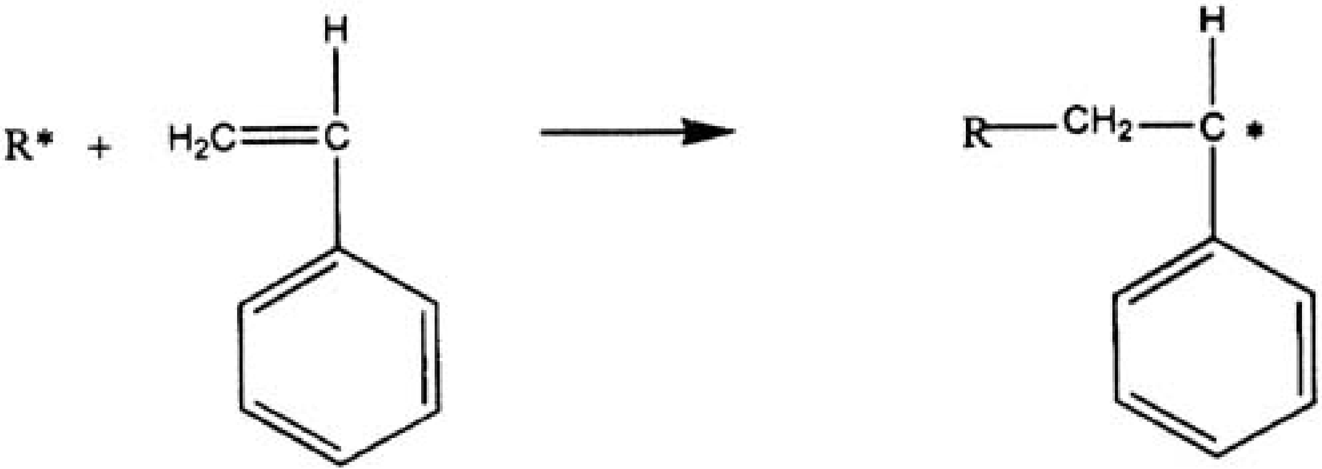


Molecular weight control can be accomplished by stopping the reaction; however, further heating may result in changes in the molecular weight. A stoichiometric imbalance (excess of one of the reactants over another) can also be used to control the molecular weight. Finally, a monofunctional reagent can be used to control the molecular weight.13

Chain-Reaction Polymerization

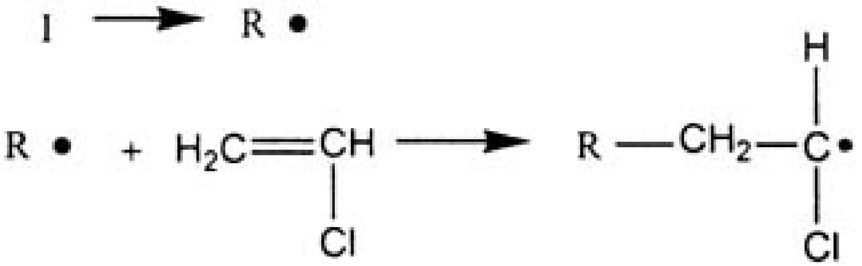
In contrast to the slow step-reaction polymerizations, chain-reaction polymerizations are fairly rapid.14 Chain-reaction polymerizations (often referred to as chain-reaction polymerization addition polymerizations) require the presence of an initiator for polymerization to occur. Initiation can occur by a free radical, an anionic, or a cationic species, which open the double bond of a vinyl monomer and the reaction proceeds as shown in Fig. 19.11 where * may be a radical, cationic, or anionic species. Chain-reaction polymers typically contain only carbon in their backbone and include such polymers as polystyrene and polyvinyl chloride. Once initiated, the chain will continue to add monomer until a termination reaction stops the growth. Termination generally occurs in radical type reactions, but in anionic and some cationic polymerizations termination reactions may not be present and the polymerization is termed a "living polymerization." This has important commercial implications in that a second monomer may be added to the living end to form block copolymers.

Fig. 19.11 Addition polymerization.



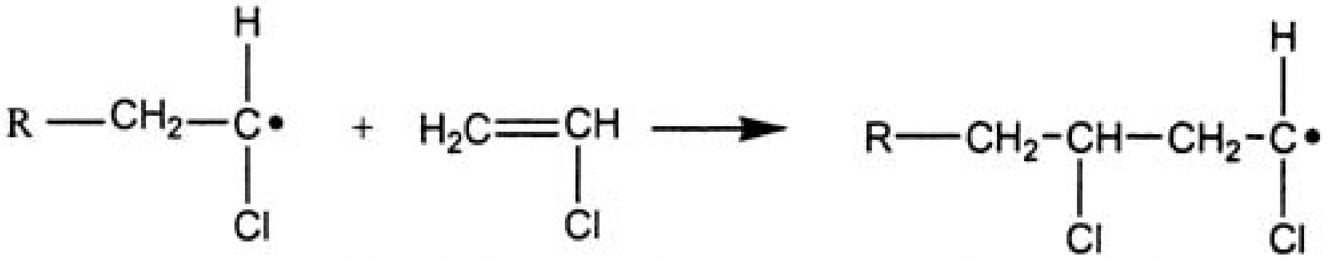
Radical Polymerization

Free radical polymerization consists of three steps, initiation, propagation, and termination or chain transfer. Initiation consists of two steps, decomposition of the initiator to form a radical species, followed by addition of the initiator to the first monomer unit.



Typical initiators include benzoyl peroxide or azobisisobutyronitrile, but other radical generating methods, such as high-energy radiation or photolytic decomposition of compounds, may also be used.12

The next step in the reaction is propagation.



Propagation continues until the radical is terminated. Termination occurs when two radical species meet and react either by coupling or by disproportionation as shown in Fig. 19.12.

Fig. 19.12 Termination mechanisms.

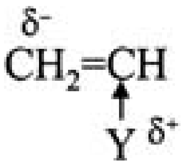
Coupling

Disproportionation

In radical polymerizations other reactions occur to prematurely stop chain growth. These reactions are termed chain transfer and when they occur they reduce the molecular weight of the chain. Chain transfer acts to transfer the radical from the growing end of the chain to another species. This may include solvent, initiator, polymer, or a deliberately added chain transfer agent. Propagation now begins from the new radical generated, while the original chain is terminated. When chain transfer to polymer occurs this produces branching along the polymer backbone. Low-density polyethylene is produced via free radical processes, with considerable transfer to polymer. Branching along the polyethylene backbone suppresses the degree of crystallinity, resulting in a lower density material.

Cationic Polymerization

This follows similar steps as with free radical polymerization. Initiator types are typically acids, such as H_2SO_4 and H_3PO_4, and Lewis acids, such as $AlCl_3$, BF_3, $TiCl_4$, and $SnCl_4$. Lewis acids generally require the presence of a proton source, such as trace amounts of water.15 Monomers exhibit considerable selectivity in their ability to be polymerized via ionic mechanisms. The nature of the substituent (Y) will influence the electron density of the double bond and thus its ability to polymerize.16 Substituents that are electron donating, such as alkoxy, phenyl, or alkyl, will increase the

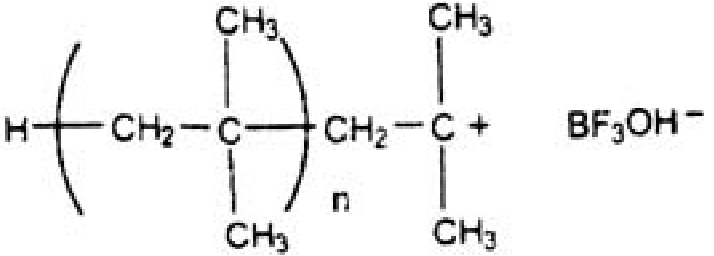


$$\overset{\delta^-}{\mathrm{CH_2}}=\underset{\overset{\uparrow}{\mathrm{Y}^{\,\delta^+}}}{\mathrm{CH}}$$

electron density of the double bond and allow for polymerization by cationic methods.

Typical monomers that may be polymerized by cationic methods include styrene, isobutylene, and vinyl ethers. Unlike radical polymerizations, solvent polarity can influence the rate of polymerization. This is due to the presence of the counterion (see Fig. 19.13). For example, more polar solvents can increase the degree of separation between the growing end and the counterion during the propagation step, increasing the rate of propagation.17 Chain transfer is the most common chain terminating reaction in cationic polymerization and can include transfer to monomer, solvent, and polymer. Termination by combination with the counterion can also occur in some systems. In some cases, cationic polymerization may be used to prepare stereoregular polymers. Although the exact mechanism is unclear, it is known that stereoregularity varies with initiator and solvent.18 Lower temperatures also tend to favor more stereoregular polymers.

Fig. 19.13 Active cationic propagating species.



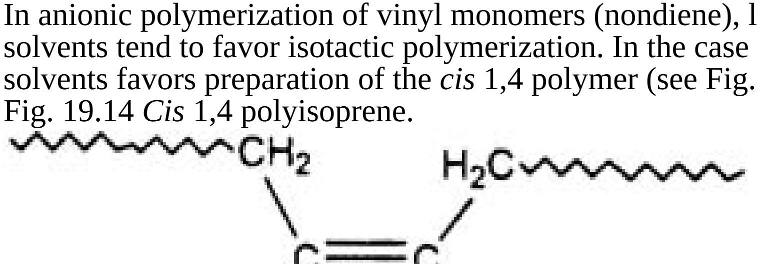
Anionic Polymerization

This is similar to cationic polymerizations, except that the propagating species is anionic. Initiator types are typically alkali metals or their compounds. In the case of anionic polymerizations, electron withdrawing substituents, such as cyano, nitro, carboxyl, vinyl, and phenyl, facilitate polymerization by anionic means. Termination in anionic polymerization is generally by chain transfer. If the system is purified so that chain transfer is suppressed, the propagating species may remain active resulting in what is termed "living polymerization".19 This allows for the preparation of block copolymers, where one monomer is polymerized, followed by addition of a second monomer to the living end.

As with cationic polymerization the propagation rate is highly dependent on the solvent. Solvents that reduce the association between the growing chain end and the counterion result in faster rates of propagation.20

In anionic polymerization of vinyl monomers (nondiene), low temperatures and polar solvents favor the preparation of syndiotactic polymers.21 Nonpolar solvents tend to favor isotactic polymerization. In the case of diene monomers such as butadiene and isoprene, the use of lithium based initiators in nonpolar solvents favors preparation of the *cis* 1,4 polymer (see Fig. 19.14).

Fig. 19.14 *Cis* 1,4 polyisoprene.



Coordination Polymerization

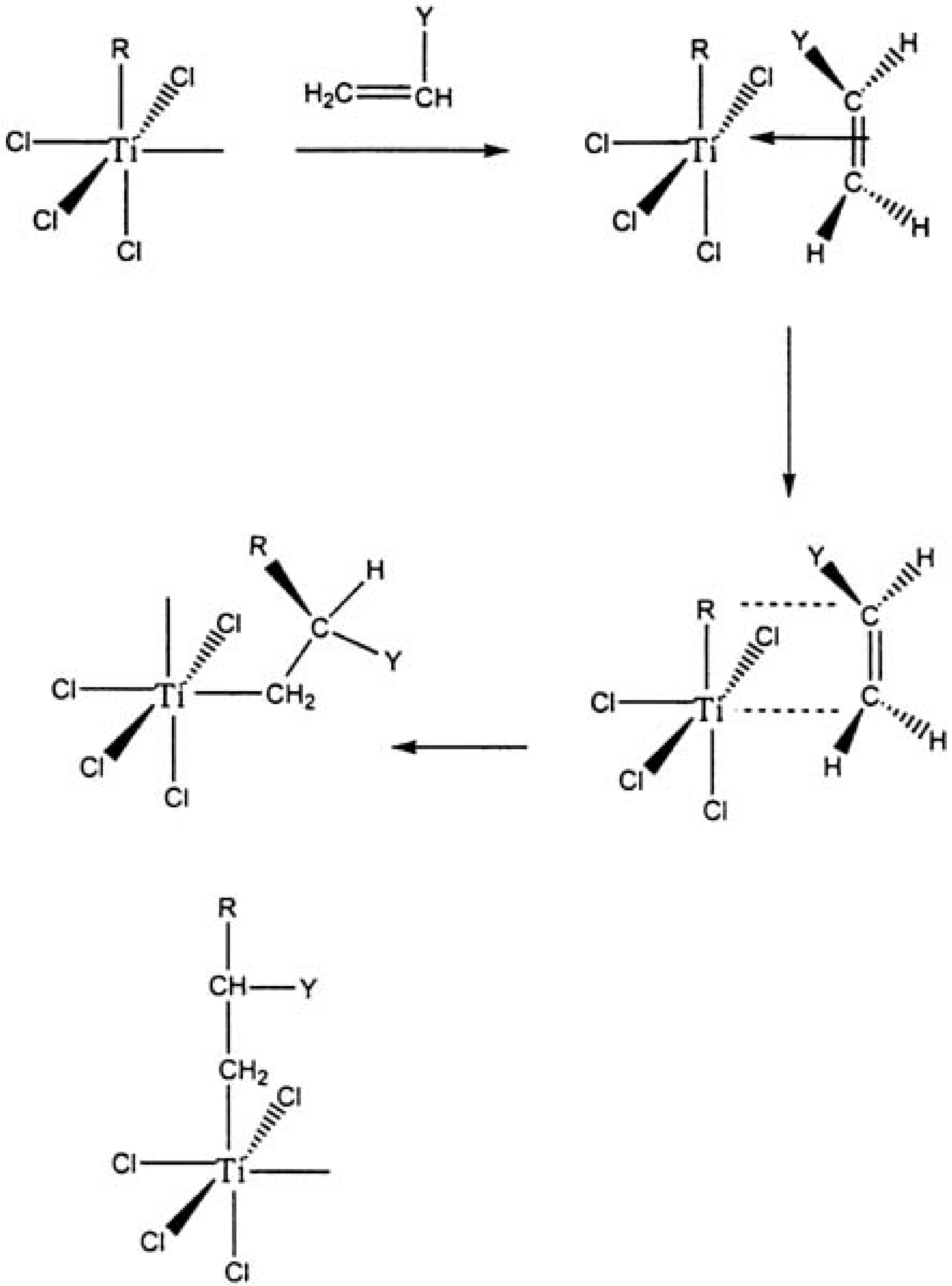
In the 1950s, Karl Ziegler discovered a way to polymerize ethylene in a linear structure to produce high-density polyethylene (HDPE) using transition metal compounds and organometallic compounds. Using similar catalysts, Giulio Natta polymerized alpha olefins, for example propene, with controlled stereoregularity. These catalyst systems are called ZieglerNatta catalysts, and are widely used for the synthesis of a number of commodity plastics, such as high-density polyethylene and polypropylene.22 ZieglerNatta catalysts may be either insoluble (heterogeneous) or soluble (homogeneous) systems.

ZieglerNatta Catalysts (Heterogeneous)

These systems consist of a combination of a transition metal compound from groups IV to VIII and an organometallic compound of a group IIII metal.23 The transition metal compound is called the catalyst and the organometallic compound the cocatalyst. Typically the catalyst is a halide or oxyhalide of titanium, chromium, vanadium, zirconium, or molybdenum. The cocatalyst is often an alkyl, aryl, or halide of aluminum, lithium, zinc, tin, cadmium, magnesium, or beryllium.24 One of the most important catalyst systems is the titanium trihalides or tetrahalides combined with a trialkylaluminum compound.

The catalyst system is prepared by mixing the two compounds in the solvent, usually at low temperatures. Polymerization occurs at specific sites on the catalyst surface. There are several proposed mechanisms for polymerization, but the important aspect of both is that the polymerization occurs in coordination with the catalyst. The pi bond of the monomer complexes with the transition metal and is then inserted in between the transition metal and the carbon of the coordinated polymer chain. One of the proposed mechanisms is shown in Fig. 19.15. Isotactic polymers are generally formed with the insoluble catalysts. Syndiotactic polypropylene has been formed with both heterogeneous and homogeneous catalysts.

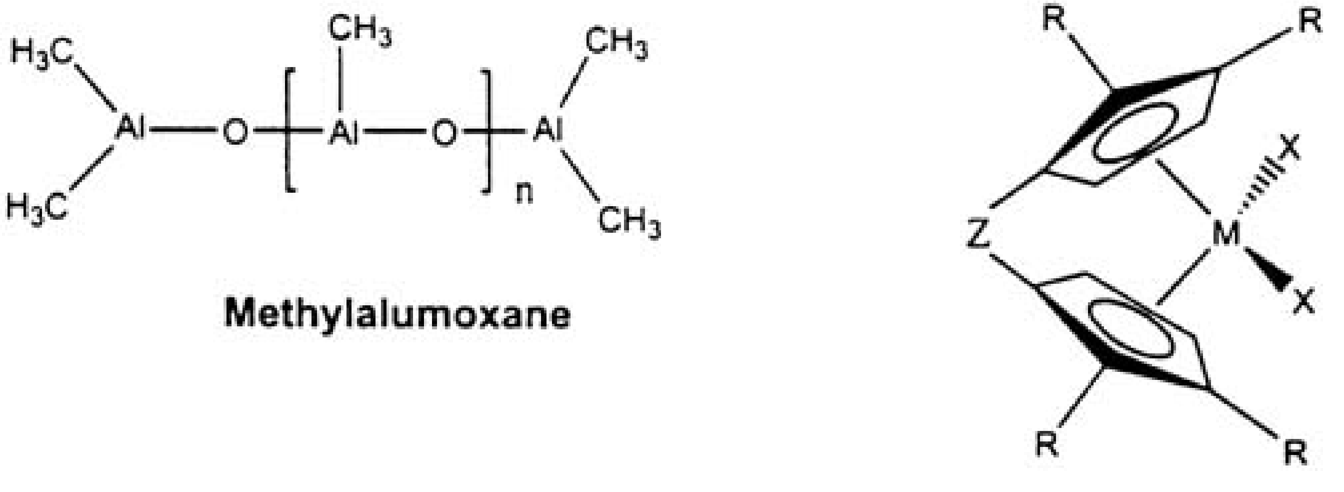
Fig. 19.15 ZieglerNatta polymerization (nonmetallic mechanism).



Metallocene Catalysts (Homogeneous ZieglerNatta)

Solid ZieglerNatta catalysts suffered from several problems, including the presence of multiple polymerization sites on the catalyst surface and catalyst residue in the final polymer, requiring a secondary purification step.25 The use of soluble catalysts offered an answer to some of these problems. These catalysts are composed of a metal atom (the active site), a cocatalyst, and a ligand system. Zirconium is the most commonly used metal although other metals such as Ti, Hf, Sc, and Th have been used. The most commonly used ligand is cyclopentadienyl. Methylalumoxane is typically the counterion. Figure 19.16 shows a proposed structure for methylalumoxane and a generalized metallocene structure.

Fig. 19.16 Metallocene catalysts.



Methylalumoxane

Metallocene

M = metal (Zr, Ti, or Hf) X = Cl or alkyl

Z = optional bridging group R = H or alkyl

The soluble catalysts can prepare polymers with very good stereospecificity and narrow molecular weight distributions, as a result of the uniformity of the active sites. In fact, these catalysts are often referred to as "single site catalysts".26 The polydispersity $(\overline{M}_w/\overline{M}_n)$, which is a measure of the molecular weight distribution, is about 22.5 for the soluble catalysts and 56 for the heterogeneous ZieglerNatta systems. By proper selection of the catalyst, syndiotactic, atactic and isotactic polypropylene and higher alpha-olefins can be synthesized. In addition, it has been possible to prepare polypropylene with alternating blocks of isotactic and atactic chains. The resulting material exhibits elastomeric properties.

19.8 Polymerization Methods

While polymerization by ionic methods is usually performed in solution, free radical polymerizations can be performed in solution, bulk, suspension, or emulsion.27 Each of these methods is described below.

Bulk Polymerization

This is one of the simplest methods of polymerization. It is often used in the polymerization of step-growth polymers.28 In these types of systems the viscosity remains low for a large portion of the reaction and heat transfer is easily controlled. Chain-growth polymers are more difficult to polymerize by this method due to the rapid and highly exothermic reactions. As the viscosity increases, thermal control becomes more difficult and may result in thermal runaway or localized hot spots. Commercial use of bulk polymerization for vinyl polymers is rather limited for those reasons. Bulk polymerization may be either homogeneous or heterogeneous, as in the case where the polymer is insoluble in the monomer.29

Solution Polymerization

Solution polymerization offers improved heat transfer over bulk polymerizations. Proper selection of the solvent is critical to avoid chain transfer reactions. Coupled with environmental concerns over organic solvents, the complete removal of solvents from the polymer also poses a potential problem. Recent work has been performed on the use of supercritical carbon dioxide as a solvent, which is easy to remove and poses less environmental concerns.30

Suspension Polymerization

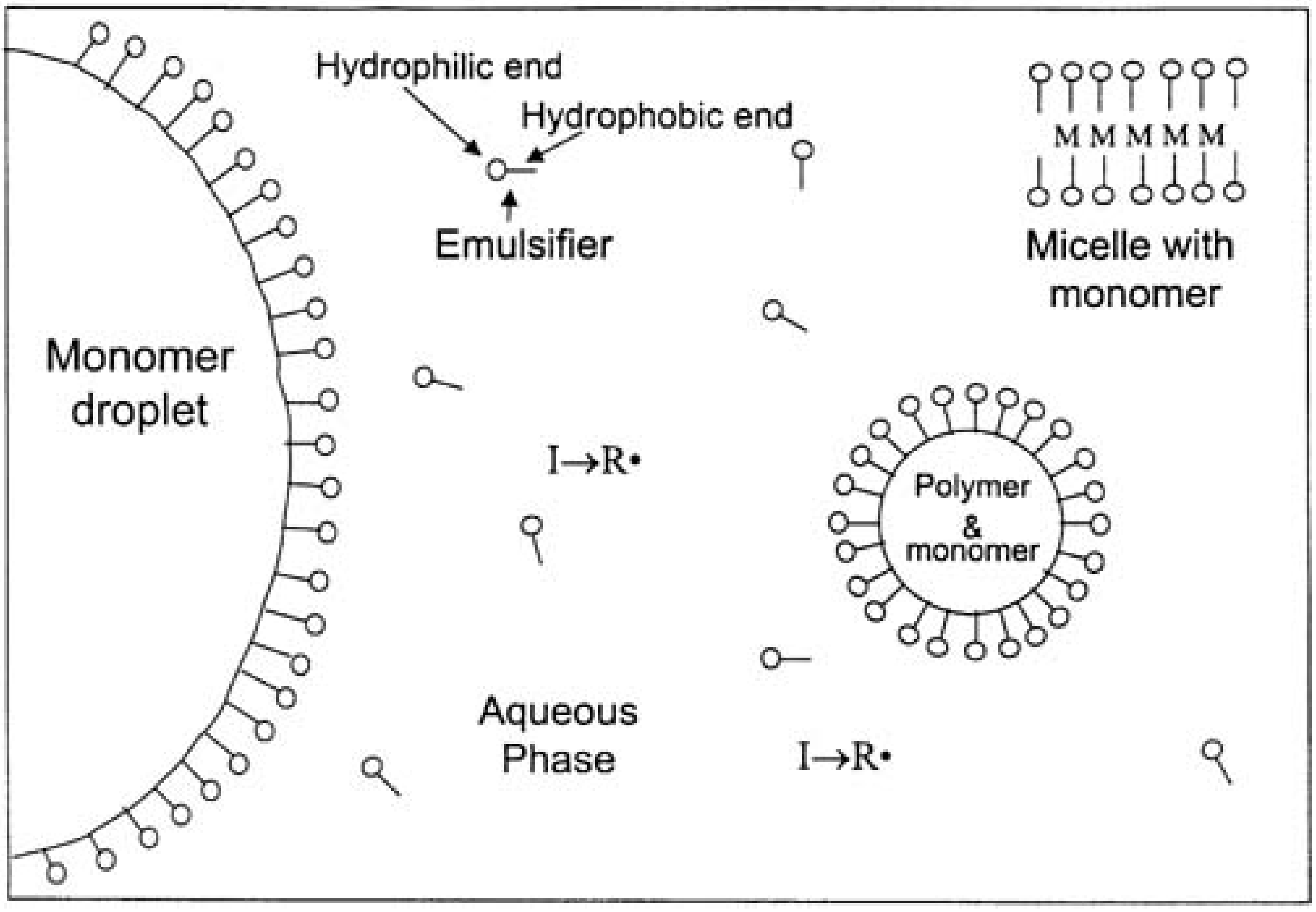
From an environmental standpoint, the use of water as a solvent is desirable. Unfortunately, many of the monomers of interest are insoluble in water, but suspension polymerization offers a way to utilize water. Suspension polymerization is performed by mechanically dispersing a monomer in an incompatible solvent, most often water. The system is heterogeneous and when polymerization is complete the polymer is collected as granular beads. This method is not suitable for tacky materials, such as elastomers, as the beads will tend to clump together.31

Monomer droplets are suspended in the water through the use of agitation and stabilizers, such as methyl cellulose, gelatin, polyvinyl alcohol, and sodium polyacrylate.32 Typical droplet sizes are 0.010.5 cm. A monomer soluble initiator is added to begin the polymerization. The kinetics of suspension polymerization are the same as for bulk polymerization, but suspension polymerization offers the advantage of good heat transfer. Polymers such as polystyrene, PVC, and polymethyl methacrylate are prepared by suspension polymerization.

Emulsion Polymerization

Superficially, emulsion polymerization resembles suspension polymerization, but there are a number of important differences. Water is used as the continuous phase and heat transfer is very good for both suspension and emulsion polymerization. In contrast to suspension polymerization, the polymer particles produced in emulsion polymerization are on the order of 0.1 µm in diameter.33 Another important difference is the presence of an emulsifying agent or soap. At the beginning of polymerization the soap molecules aggregate together in a group of about 50100 molecules to form what is called a micelle. Some of the monomer enters the micelles, but most of it is contained in monomer droplets. A water-soluble initiator is added, which migrates to the micelles as a result of their large surface to volume ratio and initiates polymerization primarily in the micelles. As polymerization continues, the micelles grow by addition of monomer from the water. The monomer droplets provide additional monomer to the aqueous phase as polymerization continues. Polymerization continues in the micelles until a second radical enters to terminate the reaction. As a result, very high molecular weight polymers may be synthesized.34 Figure 19.17 shows an overall view of the emulsion polymerization process. Emulsion polymerization is a widely used technique, especially useful for making synthetic rubber, latex paints, and adhesives.

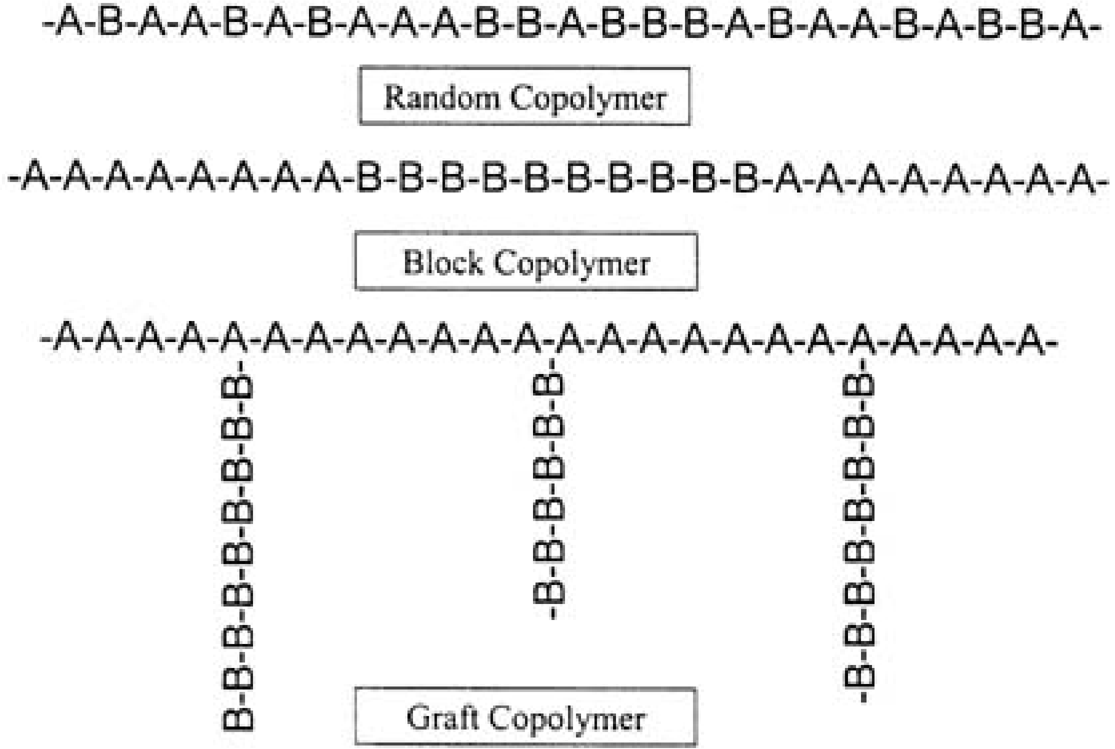
Fig. 19.17 Emulsion polymerization.



19.9 Copolymerization

In many cases the properties of a single type of polymer cannot meet the demands of a particular application. One approach to solving this problem is to combine two monomers into a single polymer through copolymerization. The properties of the resulting copolymer will then depend on the chemical nature of the monomers used and the microstructure of the chain. Monomers may be placed in the chain in a variety of ways, including random, block, and graft copolymers as depicted in Fig. 19.18.

Fig. 19.18 Copolymer structures.



Random Copolymerization

Copolymerization can occur through any of the chain reaction polymerization mechanisms described above; however, the reactivity of a given monomer toward the second monomer can vary. Thus, not all combinations of monomers may be copolymerized. Each active end will exhibit different reactivity toward each monomer, which can be expressed as reactivity ratios, *r*1 and *r*2.35 These reactivity ratios (*r*1 in this example) show the tendency of a given active end, for example M1*, to add its own monomer (M1) over the other monomer (M2). The copolymer composition at any instant can be determined by the composition of the feedstock and the reactivity ratios by

$$F_1 = \frac{r_1 f_1^2 + f_1 f_2}{r_1 f_1^2 + 2 f_1 f_2 + r_2 f_2^2}$$

where *F*1 is the mole fraction of monomer 1 in the copolymer, *f*1 and *f*2 are the mole fractions of monomer 1 and 2, respectively, in the feedstock, and *r*1 and *r*2 are the corresponding reactivity ratios. In general, the copolymer composition will not be the same as the feed composition.

The properties of random copolymers are often a weighted average of the two polymers. For example, the *T*g of a single-phase copolymer typically falls somewhere in between the *T*gs of the two homopolymers. This can be estimated using36

$$a_1 c_1 (T_g - T_{g1}) + a_2 c_2 (T_g - T_{g2}) = 0$$

where *T*g1 and *T*g2 are the glass transition temperatures for the pure homopolymers, *a*1 and *a*2 depend on the monomer type, and *c*1 and *c*2 are the weight fractions of monomers 1 and 2, respectively. For crystalline polymers, the degree of crystallinity and melting point decrease as the second monomer is added.37

Block and Graft Copolymers

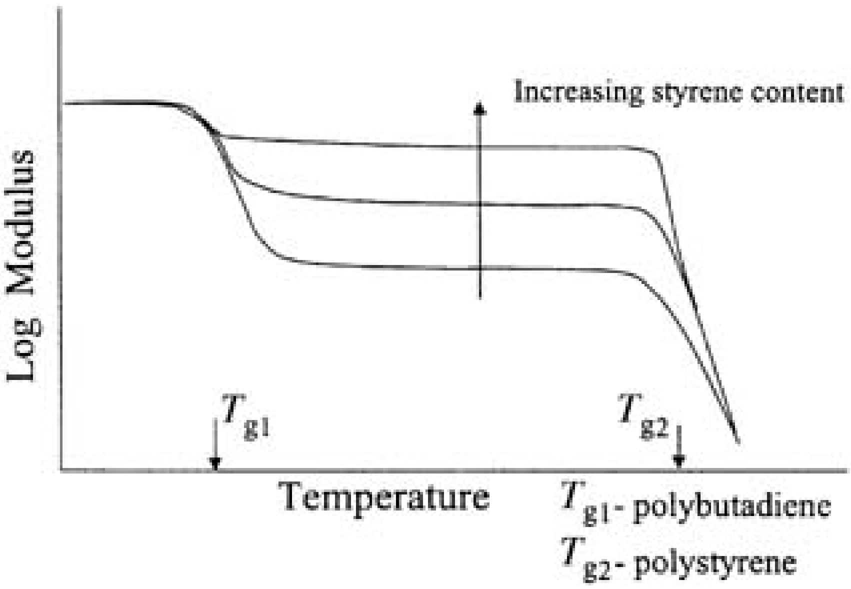
As mentioned above, the ability to have "living polymerizations" offered the potential to make block copolymers. In the preparation of a block copolymer the sequence of addition can be important to ensure that the second monomer is capable of adding to the living end. An example is the formation of a polystyrenepolymethyl methacrylate block copolymer.38 In this case polystyrene is polymerized first, followed by addition of the methyl methacrylate. The block copolymer could not be formed if methyl methacrylate were polymerized first, as styrene will not add to the methyl methacrylate living end. An important example of block copolymerization is the synthesis of ABA triblock copolymers from styrene and diene monomers to form thermoplastic elastomers. They may be prepared by sequential polymerization from one end or from the middle using a difunctional initiator. Block copolymers may also be prepared with polymers containing functionalized end groups that are linked together.

There are three basic methods to produce graft copolymers.39 In the first method a monomer is polymerized in the presence of a polymer and chain transfer provides the branching. Initiator, monomer and polymer are combined to prepare the copolymer. The initiator may function by polymerizing the monomer, which then reacts with the polymer to form the graft, or by forming a reactive site on the polymer, which then polymerizes the monomer onto the polymer. The second method of forming a graft copolymer is to polymerize a monomer with a polymer containing a reactive functional group (or sites that may be activated). Irradiation is commonly used to form active sites for graft copolymerization, but other methods may also be used. When irradiation is used the grafting mechanism is free radical in nature. In cases where the monomer is present when the polymer is irradiated, homopolymerization may also occur. If homopolymerization is a concern, then the polymer may be irradiated in the presence of air (or oxygen) to form hydroperoxide groups. The pretreated polymer is then mixed with monomer and heated to initiate the grafting reaction by decomposition of the peroxide groups.

The third method of preparing a graft copolymer is to combine two polymers with functional groups that can react together. An example is the reaction of oxazoline-substituted polystyrene with polymers containing functional groups such as alcohols, amines, and carboxylic acids.

Unlike the random copolymers, block and graft copolymers separate into two phases, with each phase exhibiting its own *T*g (or *T*M).40 The modulus-temperature behavior of a series of triblock copolymers prepared from styrene and butadiene (SBS) is shown in Fig. 19.19. With increasing styrene content the modulus in the plateau region (between the two glass transition temperatures) increases. Mixing of the two phases may shift the values of the two transition temperatures.

Fig. 19.19 Modulus-temperature behavior of a triblock copolymer.


Log Modulus
Increasing styrene content
T_{g1}
T_{g2}
Temperature
T_{g1}- polybutadiene
T_{g2}- polystyrene

19.10 Mechanical Properties

Many factors influence the mechanical behavior of polymers including polymer type, molecular weight, and test procedure. Modulus values may be obtained from a standard tensile test with a given rate of crosshead separation. In the initial linear region, the slope of a stressstrain curve will give the elastic or Young's modulus, *E*. Unlike many other materials, polymeric material behavior may be affected by factors such as test temperature and rates. This can be especially important to the designer when the product is used or tested at temperatures near the glass transition temperature, where dramatic changes in properties occur as depicted in Fig. 19.8. The time dependent behavior of these materials is discussed below.

Viscoelasticity

Polymer properties may exhibit considerable time dependent behavior, depending on the polymer type and test conditions. Increases in testing rate or decreases in temperature cause the material to appear more rigid, while an increase in temperature or decrease in rate will cause the material to appear softer. Viscoelastic response of a polymer to changes in testing rate or temperature is shown in Fig. 19.20. This time dependent behavior can also result in long-term effects such as stress-relaxation or creep.41 Creep and stress relaxation are shown in Fig. 19.21. A specimen held under a fixed load will continue to elongate with time, a process called creep. A spring and dashpot in parallel is often used as a simple model to predict the creep behavior (see Fig. 19.22). This model predicts the time dependent strain to be

$$\varepsilon(t) = \varepsilon_o e^{-t/\tau}$$

where τ is the characteristic relaxation time (η/k).

Fig. 19.20 Viscoelastic behavior of polymers.

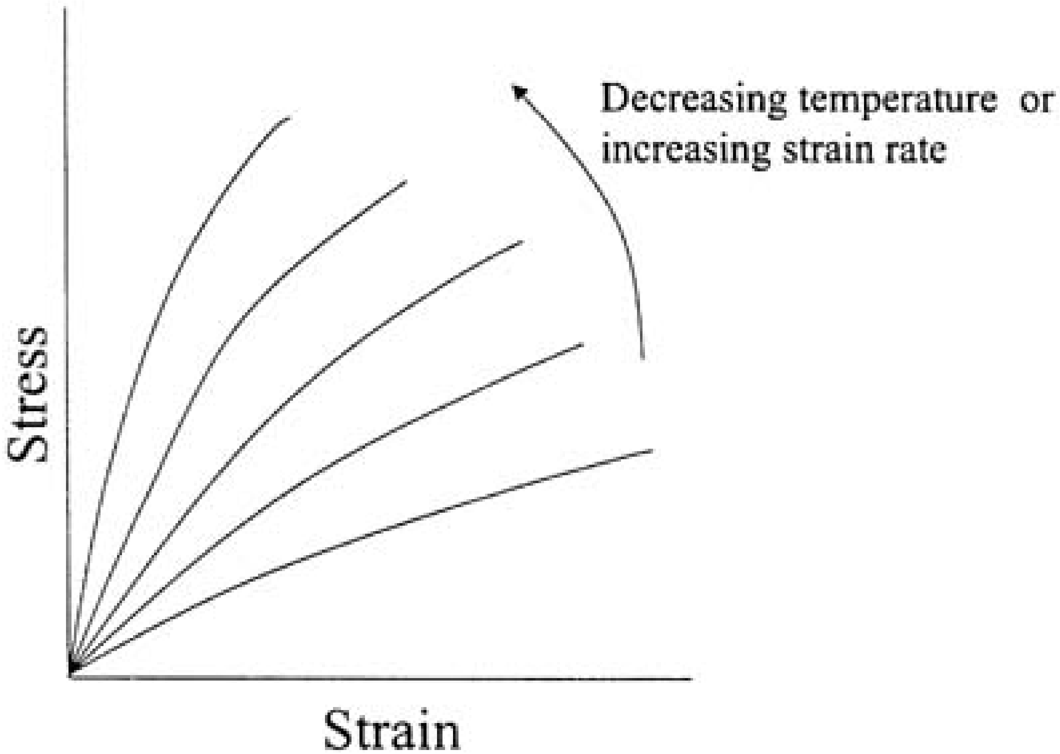


Fig. 19.21 Creep and stress relaxation.

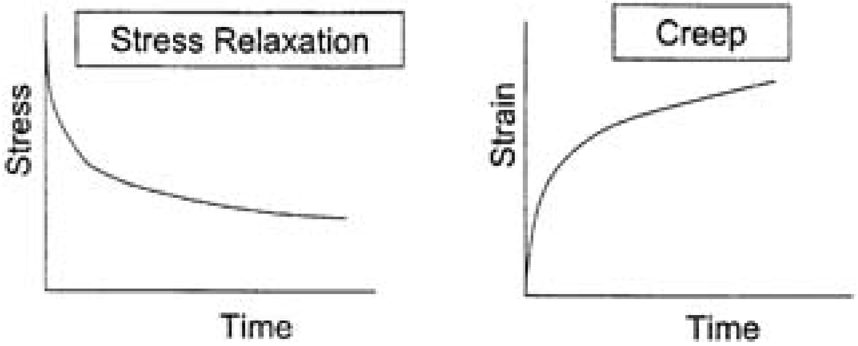
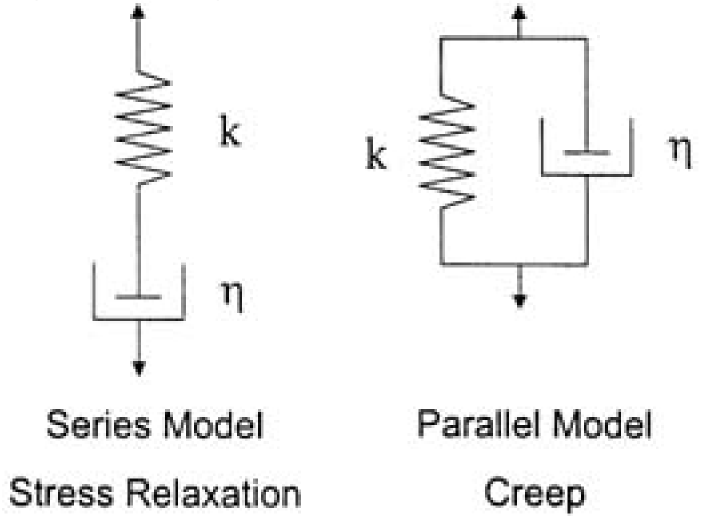


Fig. 19.22 Spring and dashpot models for stress relaxation and creep.



Stress relaxation is the decrease in load of a material held at a fixed displacement. Figure 19.22 shows the spring and dashpot in series that can be used to model the stress relaxation behavior. Using this model one obtains the following equation for the stress behavior.

$$\sigma(t) = \sigma_o e^{-t/\tau}$$

These models are useful for understanding the general concept of viscoelasticity, but are typically unable to accurately model the time dependent behavior. More accurate prediction can be obtained by using models with more elements.

If stress relaxation curves are obtained at a number of different temperatures, it is found that these curves can be superimposed by horizontal shifts to produce what is called a "master curve".42 This concept of timetemperature equivalence is very important to understanding and predicting polymer behavior. As an example, a polymer at very low temperatures will behave as if it were tested at higher temperatures at much higher testing rates. This principle can be applied to predict material behavior under testing rates or times that are not experimentally accessible, through the use of shift factors (*aT*) and the equation below

$$\ln a_T = \ln\left(\frac{t}{t_o}\right) = -\frac{17.44(T - T_g)}{51.6 + T - T_g}$$

where *Tg* is the glass transition temperature of the polymer. This has practical applications where one may be interested in the mechanical properties of a material at low temperatures under high rate (perhaps impact rate) conditions. It is also important to recognize that if material properties are measured at room temperature, they may not adequately reflect the material behavior at much lower temperatures or higher rates.

Failure Behavior

Part design requires the avoidance of failure without overdesign of the part, which leads to increased part weight and cost. Failure behavior depends on material type, service temperature, and rates. A tensile stressstrain test can be used to gather some information on material strength and behavior, for example, stress (or strain) at break, σB. Figure 19.23 shows some different types of failure behavior. Materials failing at rather low elongations (1% strain or less) are considered to have undergone brittle failure.43 General purpose polystyrene and acrylics are polymers that show this type of failure. Failure usually starts at a defect where stresses are concentrated. Once a crack is formed it will grow as a result of stress concentrations at the crack tip. Many amorphous polymers will also exhibit what are called "crazes." Crazing is a form of yielding and can enhance the toughness of a material. Although crazes appear to look like cracks, they are load bearing, with fibrils of material bridging the two surfaces as shown in Fig. 19.24.

Fig. 19.23 Types of stressstrain behavior.

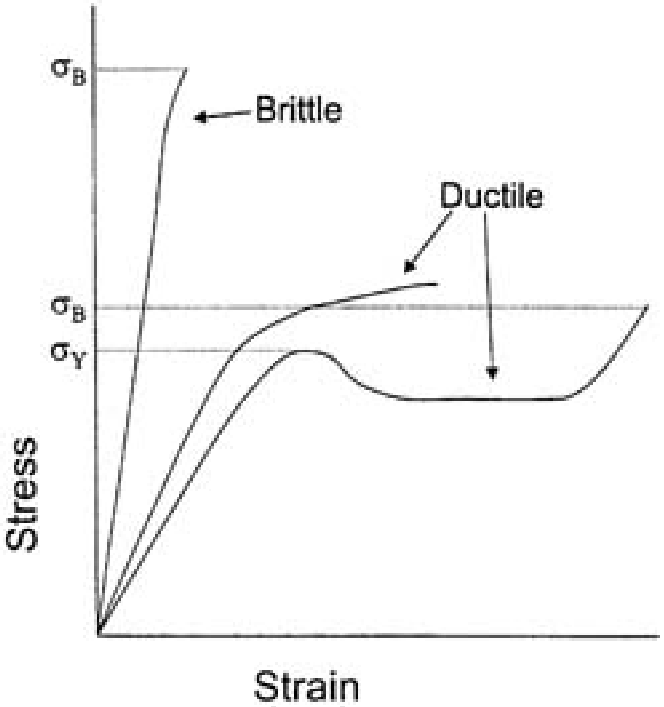
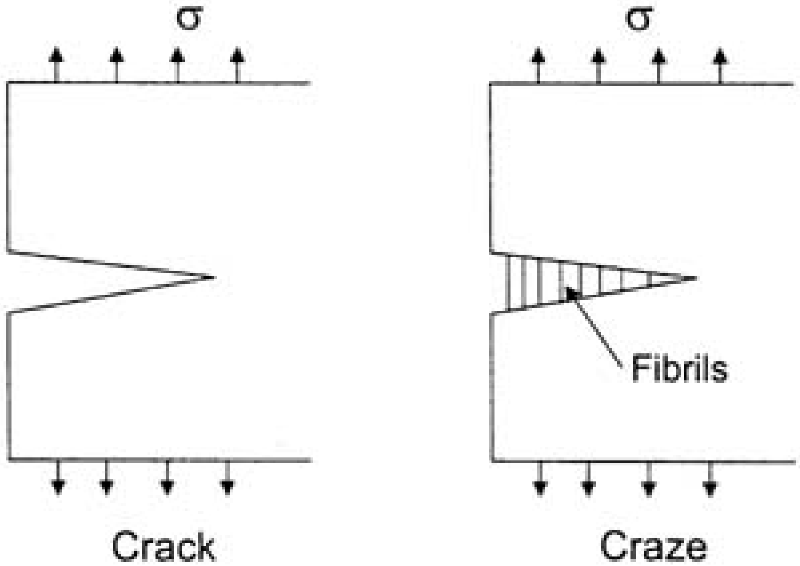
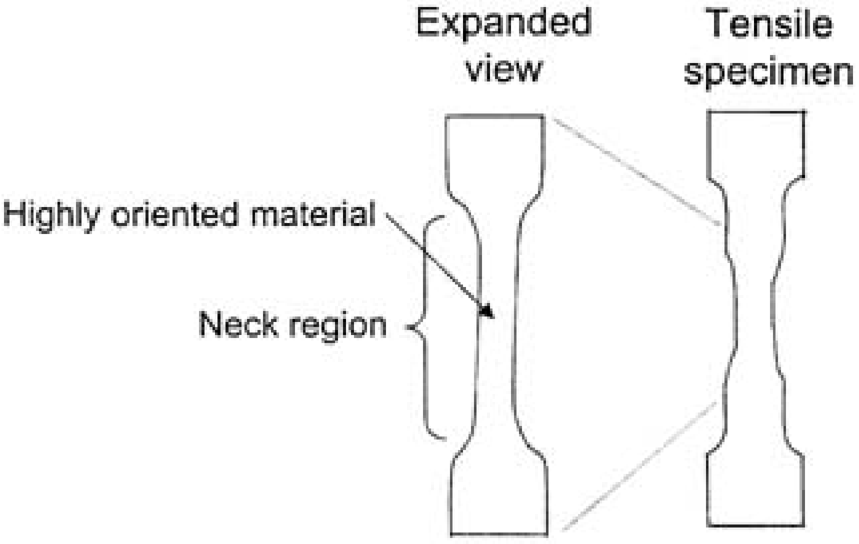


Fig. 19.24 Crazes and cracks.



Polymers also exhibit what is termed ductile failure by yielding of the polymer or slip of the molecular chains past one another. This is most often indicated by a maximum in the tensile stressstrain test or the yield point, σY (see Fig. 19.23). Above this point the material may exhibit lateral contraction upon further extension, called necking.44 In the necked regions, molecules are oriented in the direction of deformation, resulting in increased stiffness in the necked region. As a result of this localized stiffness increase, material adjacent to the neck is deformed and the neck region continues to grow. This process is known as cold-drawing (see Fig. 19.25), which may result in elongations of several hundred percent.

Fig. 19.25 Cold-drawing.



As might be expected, temperature will influence the behavior of the material. The effect of temperature on modulus has been discussed above. Very different behavior may be seen in a single polymer simply by changing the temperature.45 Beginning at temperatures well below the *T*g, the stressstrain behavior will show very low elongations, with no yield point. As the temperature is raised, the material will exhibit a yield point and elongation increases. The yield point is typically seen at temperatures near the glass transition temperature of the polymer. The higher the testing rate, the higher the temperature must be for yielding to occur.

Under repeated cyclic loading, a material may fail at stresses well below the single-cycle failure stress found in a typical tensile test, a process called fatigue.46 Fatigue behavior is usually described by plotting the maximum stress versus the number of cycles to failure. The exact conditions of the fatigue test can be varied and are specified by the service requirements. Type of loading (compression, tension, shear), load or displacement control, frequency, waveform (typically sinusoidal), and ratio of maximum to minimum stress may all be varied. Thermal effects and the presence or absence of cracks are other variables that may be included when evaluating the fatigue life of a material.

19.11 Classes of Families of Commercial Plastics

Commercial plastic materials may conveniently be divided into five major classes. Commodity thermoplastics are families which are produced in volumes of a billion pounds or more per year. Engineering thermoplastics are generally designed to withstand use at higher temperatures. Specialty thermoplastics are designed for a greater variety of performance and applications. Thermoplastic elastomers can be processed like thermoplastics, and give many of the properties of conventional thermoset vulcanized rubbers. And thermoset plastics are permanently cross-linked during processing, generally to give higher performance than would be possible with linear thermoplastics.

Commodity Thermoplastics

Six families of thermoplastics exceed a billion pounds per year in the United States: polyethylenes, polypropylene, PVC, polystyrene and its copolymers, polyethylene terephthalate and acrylonitrile-butadiene-styrene (ABS) (Table 19.2).

TABLE 19.2 Commodity Plastics in the United States and Canada, 2000

	Million Metric Tonsa
Polyethylenes	14.0
Polypropylene	6.5
PVC	6.1
Polystyrene and its copolymers	2.9
Polyethylene terephthalate	2.3
ABS	0.6

aTo convert MMT to pounds, multiply by 2200.

Polyethylenes

The major members of the polyethylene family are distinguished primarily by the amount of side-branching, which decreases regularity, crystallinity, density, melting point, modulus (rigidity), and strength (Table 19.3).

TABLE 19.3 Major Members of the Polyethylene Family: Typical Properties

Polyethylene	HDPE	LLDPE	LDPE
Density	0.96	0.93	0.92
Melting point, °C	130	124	108
Modulus, MPaa	1400	450	180
Tensile strength, MPaa	26	14	12

aTo convert MPa to psi, multiply by 145.

High-density polyethylene (HDPE)

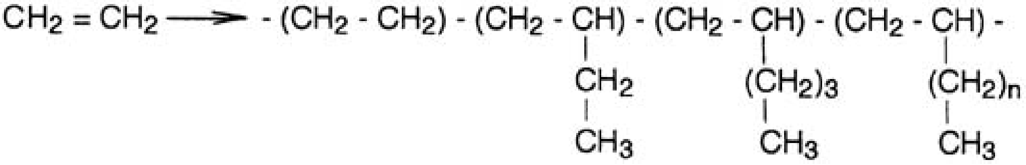
Polymerization of ethylene to polyethylenes is most often carried out at low temperature and pressure, using either the Ziegler aluminum triethyl plus titanium tetrachloride catalyst system, the Phillips chromic oxide plus silica plus alumina system, or more recently the newer metallocene single-site catalyst systems.

$$nCH_2{=}CH_2 \rightarrow —(CH_2—CH_2)_n—$$

This produces a regular, linear polymer which is 5090 percent crystalline, making it rigid and strong. Its largest use is blow-molded bottles for milk, water, juice, noncarbonated drinks, laundry detergent, antifreeze, and motor oil. Other large uses are large molded containers, pipe and conduit, and shopping bags. For detergent bottles, environmental stress-crack resistance is increased by increasing molecular weight, adding ethylene/propylene rubber, or copolymerization with a small amount (12%) of l-butene to reduce crystallinity. For auto gasoline tanks, impermeability may be increased by dispersing nylon in the polyethylene. For extreme abrasion resistance in materials-handling machinery, and in hip and knee bone replacement, Ultra-High Molecular Weight (UHMWPE: MW 36 × 106) is outstanding.

Low-density polyethylene (LDPE)

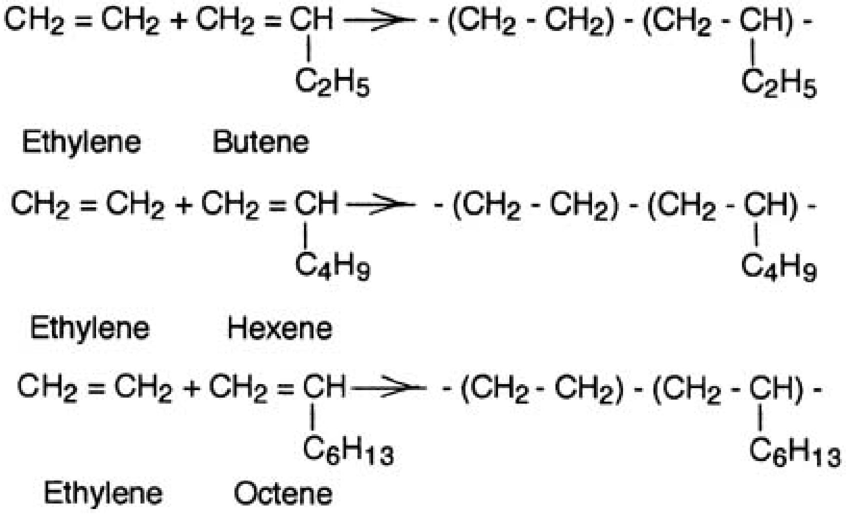
Polyethylene was originally made by polymerization at high temperature and pressure; during the polymerization, side-reactions produced several percent of short-chain branching (ethyl and butyl groups) which reduced regularity/crystallinity (e.g. to 3040%) and gave a stiffly flexible plastic material;



and chain-transfer reactions also produced some long-chain branching which gave pseudoplastic melt processability, important in the production of extruded blown film. Largest use is in packaging film; second largest is for coating and laminating on cardboard, paper, and foil.

Linear low-density polyethylene (LLDPE)

Ziegler and metallocene polyethylene can be modified by copolymerization with several percent (e.g. 10%) of butene, hexene, or octene,



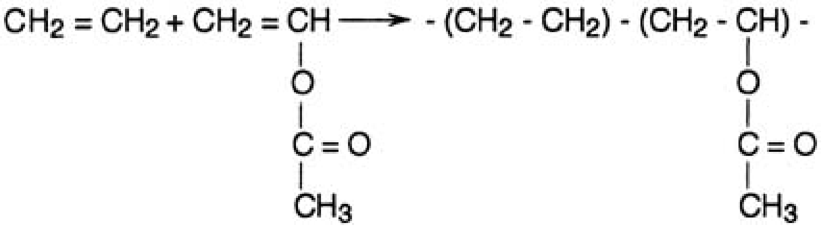
to reduce regularity/crystallinity and modulus, and thus produce low-density polyethylene at low temperature and pressure. The resulting LLDPE is stronger than the original LDPE, and therefore preferred for packaging film and trash-can liners. Lack of long-chain branching makes it difficult for extrusion-blown film, so processors add enough LDPE to improve processability.

Very low density (VLDPE) and ultra low density (ULDPE) polyethylenes

These are made by copolymerization with increasing amounts of comonomers, especially 1-octene, reducing regularity/crystallinity (density 0.910.86) down toward ethylene/propylene rubber. These are soft and flexible enough to compete with plasticized polyvinyl chloride and thermoplastic elastomers in some applications.

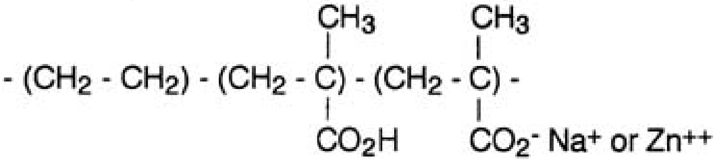
Ethylene/vinyl acetate copolymers (EVA)

LDPE is easily internally plasticized by copolymerization with increasing amounts (220%) of vinyl acetate comonomer.



This reduces regularity/crystallinity (melting points 10870°C), increasing softness/flexibility (modulus 13819 MPa), transparency, and polar adhesion, again competing with plasticized PVC and thermoplastic elastomers, particularly in packaging film, coatings, and adhesives.

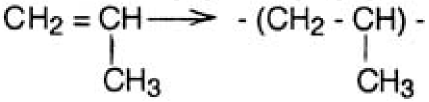
Ionomers are generally ethylene copolymers with 510 percent of methacrylic acid, half-neutralized by sodium or zinc.



This reduces regularity/crystallinity, increasing softness/flexibility (modulus 250 MPa) and transparency. The ionic groups increase strength (20 MPa) and polar adhesion, again competing with plasticized PVC and thermoplastic elastomers, particularly in packaging films and sporting goods.

Polypropylene

Polymerization of propylene to polypropylene (Table 19.4)



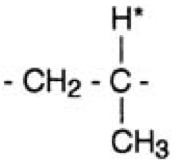
produces an asymmetric carbon atom, which can be either right-handed or left-handed. Peroxide initiation of free-radical polymerization produces a random ("atactic") mixture of right- and left-handed carbon atoms, which is too irregular to crystallize, so the resulting polymer is a soft gummy rubber. Ziegler triethyl aluminum plus titanium chloride catalysis of polymerization pre-orients each monomer unit before inserting it into the polymer chain, and produces all-right-handed or all-left-handed ("isotactic") polymer which is very regular and therefore able to crystallize (e.g. 60%). This is a high-melting (165°C), rigid, strong, low-cost commodity plastic, very competitive with more costly polystyrene and ABS. Another unique quality is its ability to form an "integral hinge," so that rigid packaging container plus lid can be molded as a single part.

TABLE 19.4 Typical Properties of Polypropylenes

Type	General Purpose	High-Impact	40% Talc	40% Glass Fiber
Modulus, MPa	1450	1100	3100	7000
Tensile strength, MPa	36		32	83
Ultimate elongation, %	500		10	2
Notched izod impact strength, J/cma	1.1	5.6	1.3	3.9
Heat deflection temperature, °C at 1.82 MPa	57	53	79	149

aTo change J/cm to FPI, multiply by 1.87.

It has two major weaknesses: (1) Methyl branching leaves tertiary hydrogen atoms,


$-CH_2-\overset{H^*}{\underset{CH_3}{C}}-$

which are sensitive to thermal oxidative aging. (2) Bulky methyl groups cause steric hindrance and stiffen the polymer molecule; while this increases modulus, it also causes embrittlement below room temperature. The polypropylene industry has successfully retarded aging by adding phenolic and aliphatic sulfide antioxidants; and it has reduced low-temperature embrittlement by adding or grafting ethylenepropylenediene (EPDM) rubber. A third common improvement is the addition of fillers to increase rigidity, making it even more competitive with polystyrene.

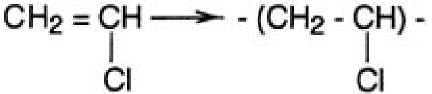
With these problems solved, polypropylene use has been growing rapidly. Extrusion and stretch orientation produces excellent synthetic fiber for carpeting and synthetic turf, and excellent film for packaging. Other large uses are rigid packaging (replacing polystyrene), automotive parts (replacing ABS), and a great variety of injection-molded consumer products.

Polyvinyl Chloride (Table 19.5) is the most versatile of the commercial thermoplastic polymers. It is used mainly for rigid and flexible plastics, for rubberlike products, for coatings on steel, cloth, and paper, and in smaller amounts for specialty fibers. It is processed mainly by extrusion and calendering, and in smaller amounts by injection, compression, and blow molding, thermoforming, rotomolding, plastisol/organosol technology, casting, pouring, and foaming. This extreme versatility comes primarily from the use of plasticizers, aided by a host of other types of additives.

TABLE 19.5 Typical Properties of Rigid PVC

Type	General Purpose	High-Impact	30% Glass Fiber	Chlorinated PVC
Modulus, MPa	3,300	2,900	8,300	2,800
Tensile strength, MPa	48	43	90	56
Ultimate elongation, %	145	130	2	35
Notched izod impact strength, J/cm	0.32	8.56	0.70	0.91
Heat deflection temperature, °C at 1.82 MPa	67	66	76	103

Vinyl chloride is polymerized primarily by peroxide-initiated polymerization in aqueous suspension,


$CH_2=\underset{Cl}{CH} \longrightarrow -(CH_2-\underset{Cl}{CH})-$

producing a fine porous powder which is easy to blend with compounding ingredients. A smaller amount is polymerized in emulsion and spray-dried for plastisols and organosols. And an even smaller amount is copolymerized with vinyl acetate in organic solution, to produce a uniform copolymer which precipitates at a constant composition and molecular weight.

Rigid PVC is polymerized at 5570°C and then compounded with organotin stabilizer, acrylic processing aid, lubricants, and optionally with rubbery impact modifier.

It has high modulus, strength, and resistance to creep, weathering, and chemicals; and compounding with rubbery impact modifier produces high resistance to brittle failure. It is mainly extruded into pipe, house-siding, and window and door frames, while smaller amounts are injection molded into pipe fittings and blow-molded into bottles for detergents and other chemicals.

Flexible PVC (Table 19.6) is made by polymerizing at 4055°C and then compounding with 2080 PHR (parts per hundred of resin) of dioctyl phthalate and/or other monomeric liquid plasticizers (e.g. dioctyl adipate for low-temperature flexibility, oligomeric polyesters for permanence, organic phosphates for flame-retardance), plus a synergistic stabilizer system usually composed of barium or calcium soap, zinc soap, epoxidized fatty ester, and organic phosphite.

TABLE 19.6 Typical Properties of Flexible PVC

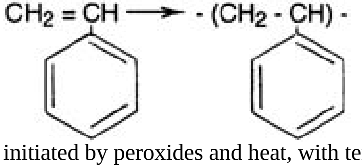
Plasticizer	Dioctyl Phthalate			Dioctyl Adipate		
Parts/hundred of resin	*30*	*50*	*75*	*30*	*50*	*70*
Shore a hardness	95	79	62	90	74	60
Tensile strength, MPa	31	21	14	27	20	14
Ultimate elongation, %	220	300	400	270	395	410
Flex temperature, °C	+20	0	-26	-15	-48	-75

This flexible-rubbery material is most often calendered into film and sheet for clothing, luggage, raincoats, upholstery, and flooring, or extruded into garden hose, wire and cable insulation, and medical tubing.

For plastisol/organosol technology, high-molecular-weight PVC powder is slurried in liquid plasticizer, and poured, dip- or spread-coated onto metal, fabric, or paper to produce dishwasher racks, upholstery, and wallpaper, respectively.

Polystyrene and its Copolymers

Polystyrene (Table 19.7) is made by continuous bulk polymerization,



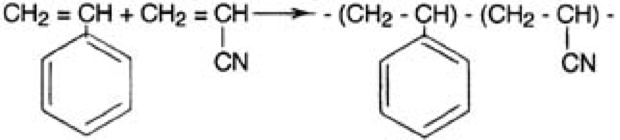
initiated by peroxides and heat, with temperature rising continuously to keep the system molten, and optionally adding solvent ("solution polymerization") to facilitate the process. It was the first commodity thermoplastic, combining easy melt processability, rigidity, and glass-like transparency for major applications in packaging, toys, and housewares. When swollen with 10 percent pentane and heated, it expands to rigid closed-cell foams with densities as low as 0.01 (expanded polystyrene or EPS), which are popular for packaging hot and cold foods, delicate instruments and appliances.

TABLE 19.7 Typical Properties of Polystyrenes

		Impact Styrene		
Polymer	Polystyrene	Moderate	Very High	Styrene/Acrylonitrile
Modulus, MPa	2900	1700	1100	3500
Tensile Strength, MPa	46	21	13	73
Notched izod impact strength, J/cm	0.17	1.28	3.70	0.26
Heat deflection temperature, °C at 1.82 MPa	83	81	81	100

The major weaknesses of polystyrene are brittleness, and softening in hot water. Brittleness is remedied by dissolving 210 percent of rubber in styrene monomer before polymerization, producing "high-impact styrene" (HIPS), in which 10-μm rubber particles improve impact strength by an order of magnitude, with some sacrifice of other mechanical properties and transparency; this accounts for more than half of the total "polystyrene" market.

Heat resistance is improved by copolymerization with 1530 percent of acrylonitrile, producing styrene/acrylonitrile (SAN).



Polarity and hydrogen-bonding stiffen the polymer molecule, improving mechanical properties and especially resistance to hot water. This improvement is useful in household products, autos, and appliances.

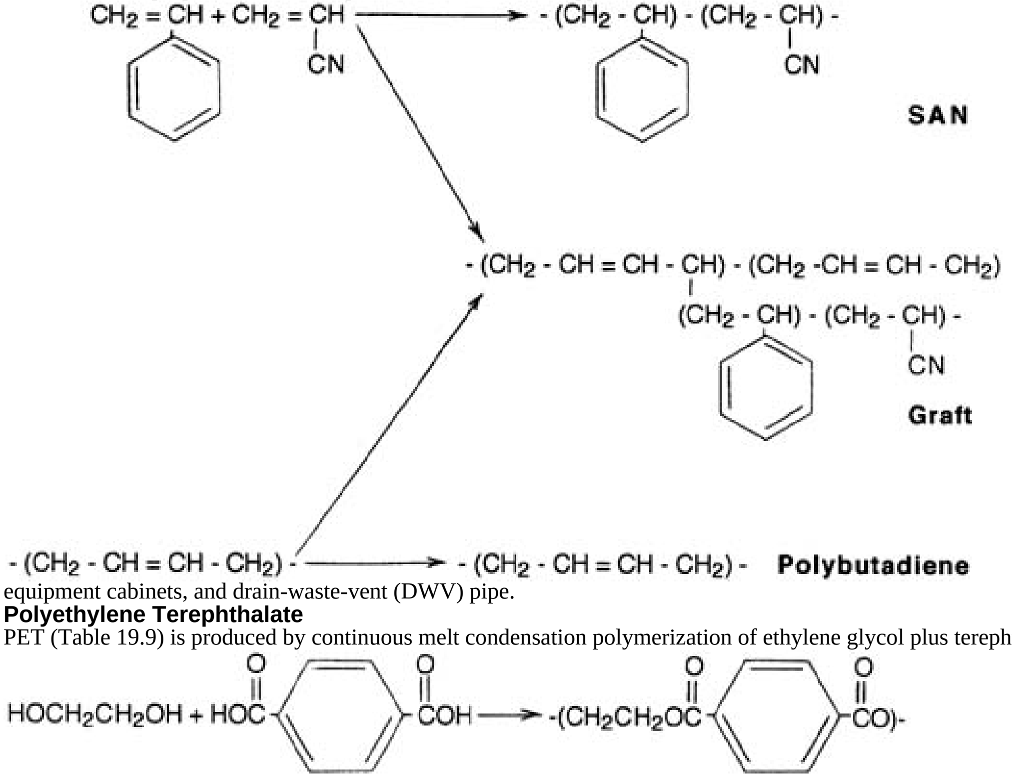
Acrylonitrile/Butadiene/Styrene (ABS)

The benefits of impact styrene and SAN are combined in ABS (Table 19.8). Typically, 1530 percent acrylonitrile and 4575 percent styrene are copolymerized in the presence of 530 percent of polybutadiene rubber, producing some graft terpolymer, and dispersing 1-μm rubber domains in an SAN matrix.

TABLE 19.8 Typical Properties of ABS

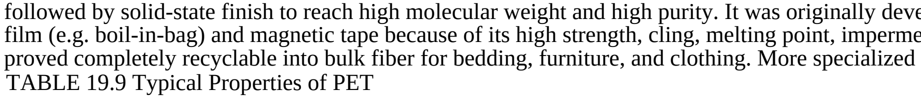
Type	Moderate Impact	Very High Impact
Modulus, MPa	2,500	2,000
Tensile strength, MPa	46	38
Notched izod impact strength, J/cm	3.5	9.0
Heat deflection temperature, °C at 1.82 MPa	94	88

This balance of properties has found particular usefulness in appliances, autos, electronic equipment cabinets, and drain-waste-vent (DWV) pipe.



Polyethylene Terephthalate

PET (Table 19.9) is produced by continuous melt condensation polymerization of ethylene glycol plus terephthalic acid,



followed by solid-state finish to reach high molecular weight and high purity. It was originally developed for synthetic fibers, replacing cotton as the leading textile fiber worldwide. Later it became popular in packaging film (e.g. boil-in-bag) and magnetic tape because of its high strength, cling, melting point, impermeability, and clarity. Finally, it became the basic material for blow-molding bottles for carbonated beverages, after which it proved completely recyclable into bulk fiber for bedding, furniture, and clothing. More specialized engineering grades for solid molding applications are generally reinforced by short glass fiber.

TABLE 19.9 Typical Properties of PET

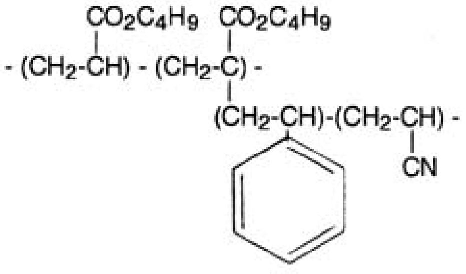
				Reinforced Molding	
Form	Fiber	Film	Bottles	15% Glass	55% Glass
Modulus, MPa		3600	3200	5900	18000
Tensile strength, MPa	550	200	120	110	200
Ultimate elongation, %	29	125		2	1
Notched izod impact strength, J/cm				0.64	1.07
Heat deflection temperature, °C at 1.82 MPa				210	229

In summary, commodity thermoplastics are manufactured readily at low cost, and offer a combination of processability, mechanical, thermal, optical, and chemical properties that are useful in a wide range of mass markets and products.

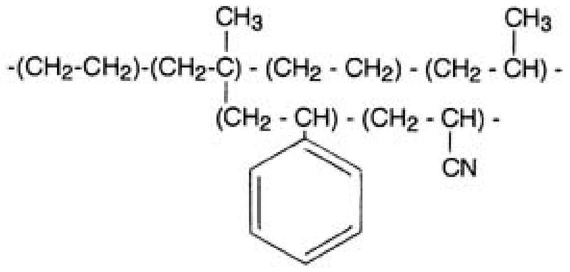
Engineering and Specialty Thermoplastics

Many present and future applications of thermoplastics make greater demands for higher properties, and especially combinations of properties, than are available from the commodity materials. To satisfy these demands, organic polymer chemists and chemical engineers have developed and commercialized over four dozen major types of polymers, offering many improved properties to meet these demands. They may be listed as follows, and then compared in their abilities to satisfy these requirements.

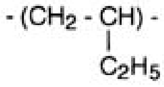
Acrylonitrile/Styrene/Acrylic Rubber (ASA)



Acrylonitrile/Styrene/Olefin Rubber (ASO)

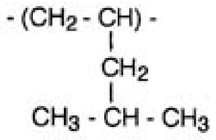


Poly(1-Butene)



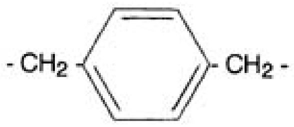
$-(CH_2-\underset{C_2H_5}{CH})-$

Poly(4-Methyl Pentene-1) (TPX)

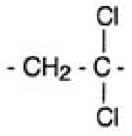


$-(CH_2-CH)-$ with side chain $CH_2-CH(CH_3)-CH_3$

Poly-p-Xylylene (Parylene)

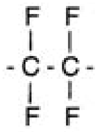


Poly(Vinylidene Chloride) (PVDC)



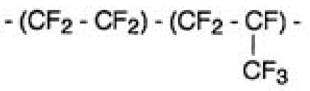
$-CH_2-\overset{Cl}{\underset{Cl}{C}}-$

Poly(Tetrafluoroethylene) (PTFE)



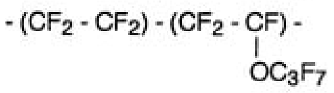
$-\overset{F}{\underset{F}{C}}-\overset{F}{\underset{F}{C}}-$

Fluorinated Ethylene/Propylene (FEP)



$-(CF_2-CF_2)-(CF_2-\underset{CF_3}{CF})-$

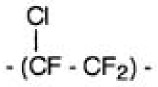
Perfluoroalkoxy Ethylene (PFA)



$-(CF_2-CF_2)-(CF_2-\underset{OC_3F_7}{CF})-$

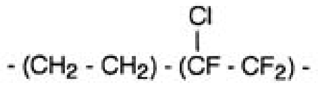
Ethylene/Tetrafluoroethylene (ETFE) $-(CH_2-CH_2)-(CF_2-CF_2)-$

Chlorotrifluoroethylene (CTFE)



$-(\overset{Cl}{CF}-CF_2)-$

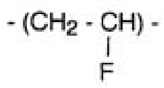
Ethylene/Chlorotrifluoroethylene (ECTFE)



$-(CH_2-CH_2)-(\overset{Cl}{CF}-CF_2)-$

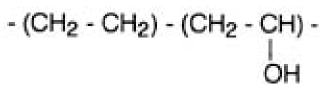
Poly(Vinylidene Fluoride) (PVDF) $-(CH_2-CF_2)-$

Poly(Vinyl Fluoride) (PVF)



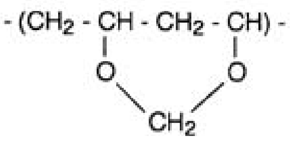
$-(CH_2-\underset{F}{CH})-$

Ethylene/Vinyl Alcohol (EVAL or EVOH)

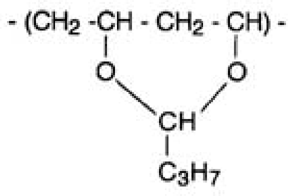


$-(CH_2-CH_2)-(CH_2-\underset{OH}{CH})-$

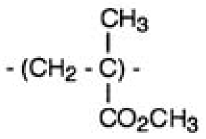
Poly(Vinyl Formal)



Poly(Vinyl Butyral)

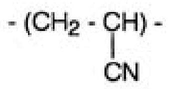


Poly(Methyl Methacrylate) (PMMA)



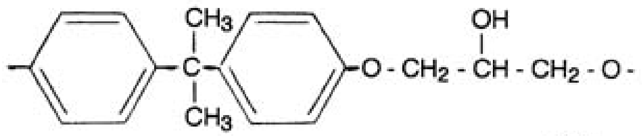
$$-(CH_2-\overset{CH_3}{\underset{CO_2CH_3}{C}})-$$

Poly(Acrylonitrile) (PAN)

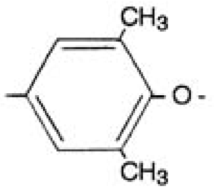


$$-(CH_2-\underset{CN}{CH})-$$

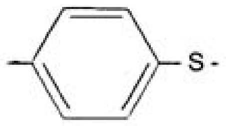
Phenoxy Resin



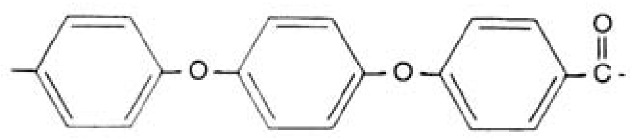
Poly(Phenylene Ether) (PPE or PPO)



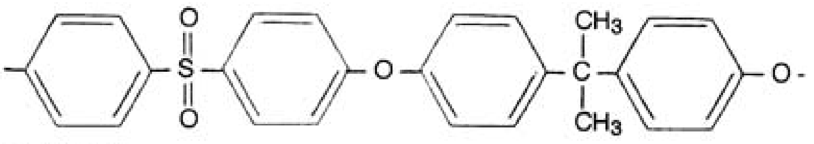
Poly(Phenylene Sulfide) (PPS)



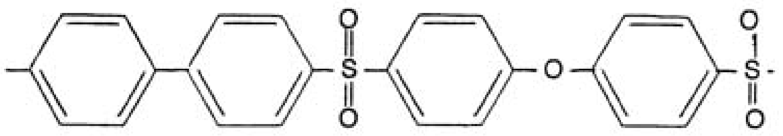
Poly(Ether Ether Ketone) (PEEK)



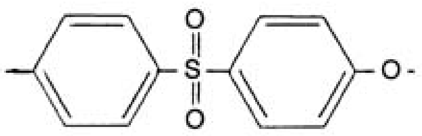
Polysulfone (Udel)



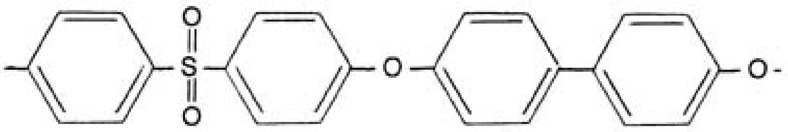
Polyaryl Sulfone



Polyether Sulfone



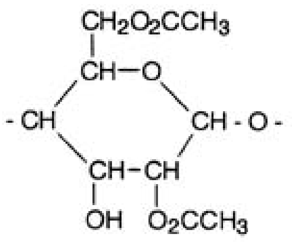
Polyphenyl Sulfone (Radel)



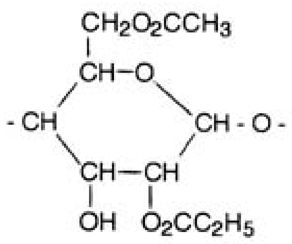
Poly(Oxymethylene) (Polyacetal)

$-(CH_2-O)-$

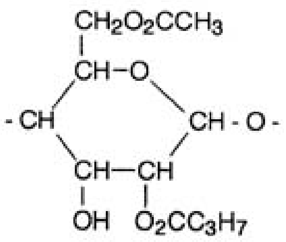
Cellulose Acetate (CA)



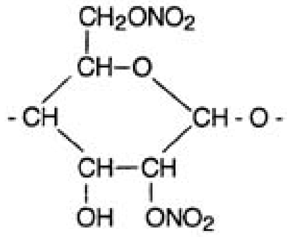
Cellulose Acetate/Propionate (CAP)



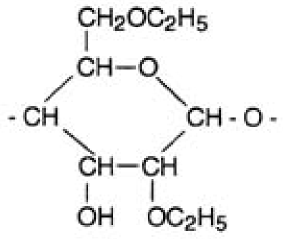
Cellulose Acetate/Butyrate (CAB)



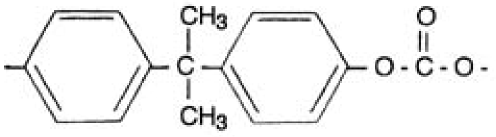
Cellulose Nitrate



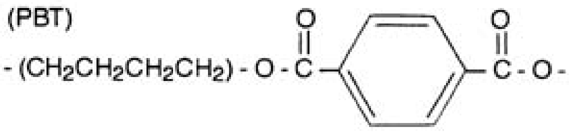
Ethyl Cellulose (EC)



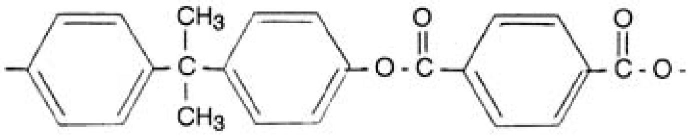
Polycarbonate (PC)



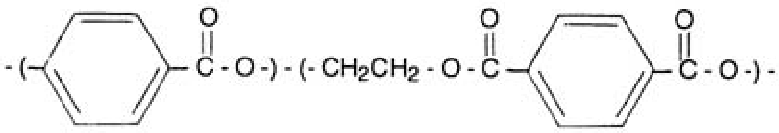
Poly(Butylene Terephthalate) (PBT)



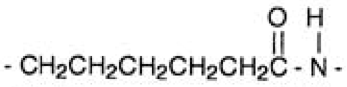
Polyarylate



Liquid Crystal Polyesters (LCP)

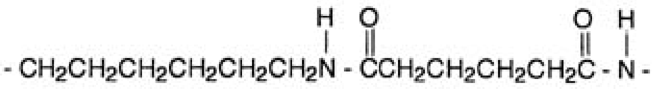


Nylon 6 (PA6)



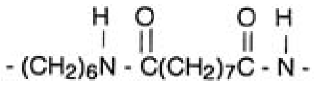
$$-CH_2CH_2CH_2CH_2CH_2\overset{\overset{O}{\|}}{C}-\overset{\overset{H}{|}}{N}-$$

Nylon 66 (PA66)



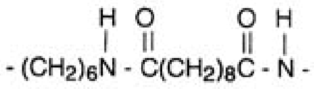
$$-CH_2CH_2CH_2CH_2CH_2CH_2\overset{\overset{H}{|}}{N}-\overset{\overset{O}{\|}}{C}CH_2CH_2CH_2CH_2\overset{\overset{O}{\|}}{C}-\overset{\overset{H}{|}}{N}-$$

Nylon 69 (PA69)



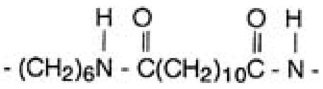
$$-(CH_2)_6\overset{\overset{H}{|}}{N}-\overset{\overset{O}{\|}}{C}(CH_2)_7\overset{\overset{O}{\|}}{C}-\overset{\overset{H}{|}}{N}-$$

Nylon 610 (PA610)



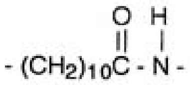
$$-(CH_2)_6\overset{\overset{H}{|}}{N}-\overset{\overset{O}{\|}}{C}(CH_2)_8\overset{\overset{O}{\|}}{C}-\overset{\overset{H}{|}}{N}-$$

Nylon 612 (PA612)



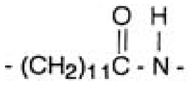
$$-(CH_2)_6\overset{\overset{H}{|}}{N}-\overset{\overset{O}{\|}}{C}(CH_2)_{10}\overset{\overset{O}{\|}}{C}-\overset{\overset{H}{|}}{N}-$$

Nylon 11 (PA11)



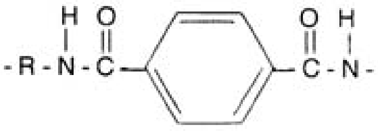
$$-(CH_2)_{10}\overset{\overset{O}{\|}}{C}-\overset{\overset{H}{|}}{N}-$$

Nylon 12 (PA12)

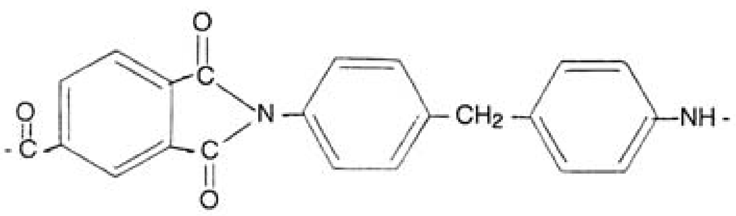


$$-(CH_2)_{11}\overset{\overset{O}{\|}}{C}-\overset{\overset{H}{|}}{N}-$$

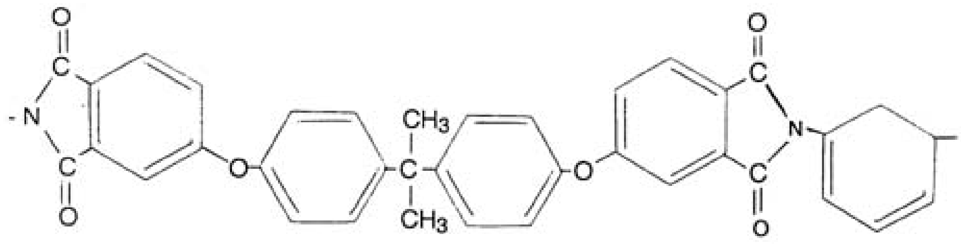
Poly(Phthalamide) (PPA)



Poly(Amide Imide) (PAI)



Poly(Ether Imide) (PEI)



The leading materials, in terms of market volume, are: nylon, polycarbonate, polybutylene terephthalate, polyphenylene ether, polyoxymethylene, and polyethylene terephthalate.

Perhaps even more important than their structures, most of these polymers are frequently reinforced by glass fibers or even carbon fibers, which contribute tremendously to their properties, and must be considered in any comparison of their practical performance. Reinforcing fibers generally raise modulus 24 fold and usually increase breaking strength somewhat. In crystalline plastics, they often raise maximum use temperatures dramatically.

Maximum Use Temperature

The most frequent requirement for higher engineering performance is retention of properties at higher temperatures. Whereas most commodity thermoplastics soften and distort in boiling water, engineering thermoplastics are most often characterized by their ability to stand much higher temperatures. This is measured most often by the short-term Heat Deflection Temperature (HDT) under a load of 1.82 MPa, less often but perhaps more practically by an estimated Continuous Service Temperature (CST) in long-term use (Table 19.10).

TABLE 19.10 Typical Temperature Resistance of Engineering Thermoplastics

Polymer	Reinforcing Fiber	HDT (°C)	CST (°C)
Liquid crystal polyester		347	355

Polyetheretherketone	30% Glass	316	250
Polyamideimide	Filled	280	80
Polyphthalamide	15% Glass	277	
Polyphenylene sulfide	30% Glass	265	220
Perfluoroalkoxy ethylene	20% Carbon	260	260
Nylon 66	33% Glass	250	130
Nylon 610	30% Glass	215	110
Polyaryl sulfone	30% Glass	213	
Nylon 612	33% Glass	210	110
Nylon 6	33% Glass	210	121
Polyetherimide		210	170
Polyethylene terephthalate	30% Glass	210	140
Ethylene/tetrafluoroethylene	25% Glass	210	177
Polyether sulfone		204	179
Polybutylene terephthalate	33% Glass	204	
Polyphenyl sulfone		190	
Polysulfone	10% Glass	183	160
Nylon 11	23% Glass	176	90
Fluorinated ethylene/propylene	20% Glass	176	204
Polyvinylidene fluoride	20% Carbon Fiber	173	121
Polyarylate		172	
Polyoxymethylene	30% Glass	160	105
Nylon 12	23% Glass	160	100
Polyphenylene ether		146	105
Ethylene/Chlorotrifluoroethylene	20% Glass	136	150
Polycarbonate		130	

Major applications are primarily in electrical and electronic products, auto parts, industrial products, and appliances.

Impact Strength

When plastics are compared with metals and wood, they often fail under high-speed impact. Flexible molecules such as nylons have some inherent impact strength. Of the rigid molecules, only polycarbonate and polyphenyl sulfone combine inherent rigidity and high impact strength, and we still do not understand the secret of their success. Most plastics can be reinforced with long fibers to increase their impact strength, but processing becomes much more difficult. Some plastics have been toughened by dispersing tiny rubber domains in them, and this technique is currently expanding with the development of compatibilization technology. One unique material is poly(1-butene), whose high creep- and puncture-resistance make it particularly desirable in pipe and tubing.

Transparency

Some applications of plastics require transparency. Amorphous plastics should be able to transmit light. Some factors which prevent transparency include unsaturation/light absorption, crystallinity, fillers and reinforcing fibers, and use of rubber particles to increase impact strength. The plastics most often used for their transparency are poly(4-methylpentene-1) (TPX), poly(methyl methacrylate) (almost equal to glass), cellulose acetate, propionate, and butyrate, polycarbonate, and polysulfones (slightly yellow). As a research challenge, it is quite possible that fillers and rubber particles could be modified to match the refractive index of the matrix polymers, and thus retain transparency while offering their reinforcing effects on mechanical properties.

Processability

Engineering performance generally requires rigid molecules to give maximum rigidity, strength, and high-temperature performance. Melt processability, on the other hand, generally requires flexible molecules to give a fluid melt. Several polymers which manage to combine easy melt processability, with high rigidity/strength/heat resistance, include nylons, polyoxymethylene (polyacetal), and liquid crystal polymers. The first two combine flexible molecules in the melt plus high crystallinity in the solid form. The molecular rigidity of liquid crystal polymers explains their high modulus and strength, but their easy melt processability remains something of a mystery.

Lubricity

Plastic gears and bearings are less polar than metals, and therefore are relatively self-lubricating, without the need for lubricating oil. Ultrahigh molecular weight polyethylene, fluoropolymers, polyoxymethylene (polyacetal), and nylons are the best. To improve other plastics even further, adding powdered polytetrafluoroethylene to them can produce an even greater increase in lubricity, and simultaneously also an increase in abrasion resistance.

Barrier Plastics

When plastics replace metals and glass in packaging, their permeability is often a limiting property. Barrier performance generally increases with density and crystallinity. The most promising barrier plastics include ethylene/vinyl alcohol, polyvinylidene chloride, polyacrylonitrile, and polyethylene naphthoate. These are used most efficiently by laminating them to commodity plastics such as polyethylene and polyethylene terephthalate.

Weather-Resistance

For long-lived outdoor products, most plastics can be stabilized somewhat by opaque UV reflectors or at least dissolved UV stabilizers. For inherent resistance to sunlight, rain, and other components of weather, some preferred plastics include acrylonitrile/styrene/acrylic rubber, acrylonitrile/styrene/ethylene-propylene rubber, polyvinyl chloride, fluoropolymers, and polymethyl methacrylate.

Chemical Resistance

Plastics are generally superior to metals in resistance to aqueous inorganic environments. Their resistance to organic solvents depends mainly on crystallinity and difference in polarity: nonpolar polymers are more resistant to polar organics, while polar polymers are more resistant to nonpolar organics.

Individual Specialties

There are a number of special plastics which are used for their individual special properties and applications. Poly(4-methylpentene-1) combines rigidity, impact resistance, heat resistance, transparency, and chemical resistance, making it a unique replacement for glass in chemical equipment. Poly-*p*-xylylene is an extreme specialty coating, applied by vapor deposition to produce thin uniform films for electrical insulation. Polyvinyl formal is used in specialty wire-coating. Polyvinyl butyral offers the high toughness, adhesion, and clarity which make it the critical component in safety glass. Phenoxy resin was one of the first engineering plastics, offering a combination of processability, rigidity, strength, toughness, transparency, adhesion, and chemical resistance; its present use is primarily in coatings and adhesives. Cellulose nitrate was the first commercial plastic; its present uses are primarily as high-quality coatings on wood furniture and leather goods. Ethyl cellulose is a very tough transparent adhesive material, mainly used in coating bowling pins and specialty papers. Many other commercial polymers are more important in non-plastic applications such as rubber, textiles, paper, coatings, and adhesives.

Thermoplastic Elastomers

Soft flexible rubbery behavior depends on long flexible polymer molecules in the form of random coils. Strength, heat and chemical resistance depend on attachment between the coils. Conventional rubber chemistry uses vulcanization, permanent thermoset primary covalent cross-links, usually by sulfur plus metal oxides, to hold the coils together; but this makes processing more difficult, and recycling very difficult. In the past 40 years, this technology has been supplemented by the development of thermoplastic elastomers. These are based mainly on block copolymers, in which long flexible blocks form the continuous rubbery matrix, and short glassy or crystalline blocks form the thermoplastic "cross-links," secondary attractions that give strength, heat, and chemical resistance. These are much more attractive to the plastics processing industry, offering easy thermoplastic processing and good recyclability. They now account for about 10 percent of the total rubber market. There are six families of commercial thermoplastic elastomers (Table 19.11).

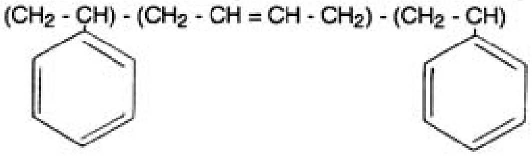
TABLE 19.11 Typical Property Ranges of Thermoplastic Elastomers

Polymer System	Shore Hardness	Tensile Strength (MPa)	Ultimate Elongation (%)	Low Temperature Limit (°C)	High Temperature Limit (°C)
Styrene-Diene	A23-D65	234	2001750	-75	100
TPOlefin	A67-D72	336	20900	-60	120
TPVulcanizate	A40-D60	319	300650	-63	135
Polyurethane	A60-D78	1447	250966	-50	120
Polyetherester	D35-D82	947	170600	-50	137
Polyetheramide	D25-D69	662	250760	-40	150

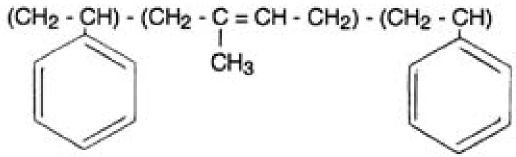
The primary variable in each of these families is the ratio of rubbery soft block to glassy or crystalline hard block, thus offering a wide range in balance of soft flexible properties vs. strength, heat, and chemical resistance.

Styrene-Diene ("styrenic") thermoplastic elastomers are block copolymers of styrene with butadiene (SBS) or isoprene (SIS) in about 30/70 monomer ratio.

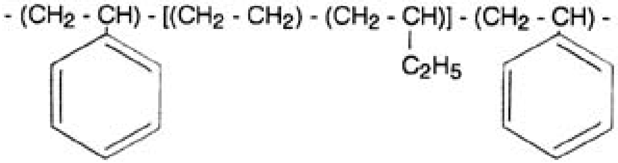
SBS



SIS

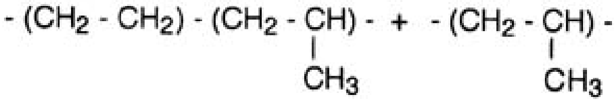


They have all the rubberiness of the butadiene rubber matrix, and the glassy polystyrene domains hold them together up to the softening point of polystyrene. Since their unsaturation is sensitive to oxygen and ozone aging, SBS is often saturated by hydrogenation (SEBS) to improve age-resistance.



They are the leading class of thermoplastic elastomers, 45 percent of the total market, used mainly in adhesives, shoe soles, wire and cable insulation, kitchen utensils, medical products, and auto parts.

Thermoplastic olefin thermoplastic elastomers are primarily blends, or block or graft copolymers, of ethylene/propylene rubber with polypropylene.

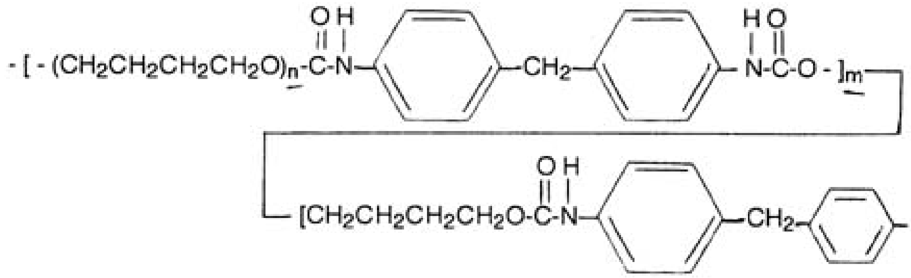


$$-(CH_2-CH_2)-(CH_2-\underset{CH_3}{\underset{|}{C}}H)- \;+\; -(CH_2-\underset{CH_3}{\underset{|}{C}}H)-$$

They have all the rubberiness of the ethylene/propylene (EPR) rubber matrix, and the crystalline polypropylene (PP) domains hold them together. As saturated elastomers, they have natural resistance to oxygen and ozone aging. They are the second largest class of thermoplastic elastomers, 25 percent of the total market, used mainly in mechanical rubber parts.

Thermoplastic vulcanizates are a surprising improvement over conventional thermoplastic olefins. Vulcanized ethylene/propylene/diene (EPDM) rubber particles are dispersed in a thermoplastic polypropylene (PP) matrix. The vulcanized EPDM has higher heat and chemical resistance than ordinary EPR, and the PP matrix provides thermoplastic processability. In more specialized grades, the vulcanized rubber may be nitrile rubber for greater oil resistance, or butyl rubber for impermeability. They are used where greater heat and/or chemical resistance are required, for example, in oil wells, mechanical goods, and building and construction.

Polyurethane rubber is mainly thermoset, but thermoplastic processability can be achieved by block copolymers of amorphous polyurethane rubber with strongly hydrogen-bonded crystalline polyurethane blocks.

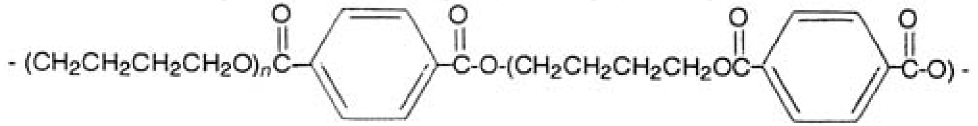


They combine the high strength, oil- and gas-resistance of polyurethanes with the advantage of thermoplastic processability. They are about 15 percent of the total thermoplastic elastomer market, used in auto parts, wire and cable, medical products, and fuel hose.

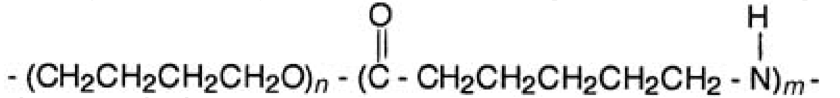
Polyetherester thermoplastic elastomers are typically block copolymers of polyoxybutylene rubber with polybutylene terephthalate crystalline domains.

A number of other rubber and crystalline blocks are mentioned in the literature, and may be offered by various manufacturers. For the most part they are balanced towards less rubbery block and more crystalline block, combining moderate flexibility with greater strength, heat, and chemical resistance. They are used mainly in automotive and other mechanical parts requiring this combination of properties.

Polyetheramide thermoplastic elastomers are typically block copolymers of polyether rubber with nylon crystalline domains.



A number of such combinations are mentioned in the literature, and may be offered by various manufacturers. They combine the soft flexible rubberiness of polyether or polyester elastomers with the high strength, heat-, oil-, and gas-resistance of nylons, and thus find use in auto parts, wire and cable, and sporting goods.



$$-(CH_2CH_2CH_2CH_2O)_n-(\overset{O}{\overset{\|}{C}}-CH_2CH_2CH_2CH_2CH_2-\overset{H}{\overset{|}{N}})_m-$$

Thermoset Plastics

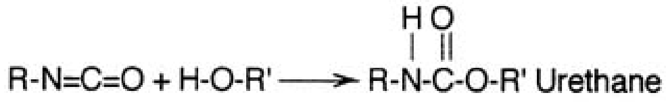
Whereas difunctional monomers produce linear thermoplastic polymers, monomers with higher functionality can react further during processing, cross-linking up to infinite molecular weight. Such thermosetting processing may be more difficult, but infinite cross-linking produces extreme increases of rigidity, creep-resistance, dimensional stability, heat-resistance, and chemical resistance, which are valuable in many demanding engineering applications. Thus thermoset plastics account for about 15 percent of the total plastics market. The major thermoset plastics families may be ranked in order of market size as shown in Table 19.12.

TABLE 19.12 Thermoset Plastics Market

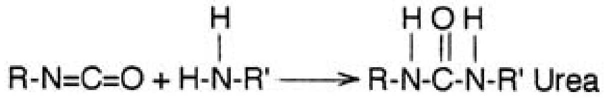
Family	Percent of Thermoset Market
Polyurethane	37
Phenolformaldehyde (Phenolic)	24
Ureaformaldehyde	15
Polyester	10
Epoxy	4
Melamineformaldehyde	2
All Others	8

Polyurethanes

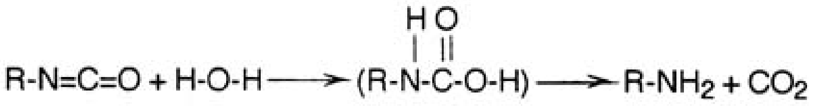
Most polyurethane chemistry may be simplified down to three basic reactions:



$$R-N{=}C{=}O + H-O-R' \longrightarrow R-\overset{H}{\overset{|}{N}}-\overset{O}{\overset{\|}{C}}-O-R' \quad \text{Urethane}$$

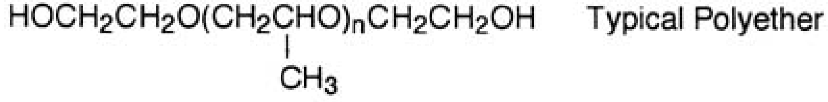


$$R-N{=}C{=}O + H-\overset{H}{\overset{|}{N}}-R' \longrightarrow R-\overset{H}{\overset{|}{N}}-\overset{O}{\overset{\|}{C}}-\overset{H}{\overset{|}{N}}-R' \quad \text{Urea}$$

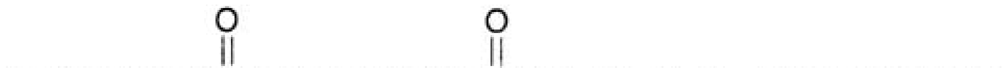


$$R-N{=}C{=}O + H-O-H \longrightarrow (R-\overset{H}{\overset{|}{N}}-\overset{O}{\overset{\|}{C}}-O-H) \longrightarrow R-NH_2 + CO_2$$

Polyols are usually aliphatic polyethers or polyesters.

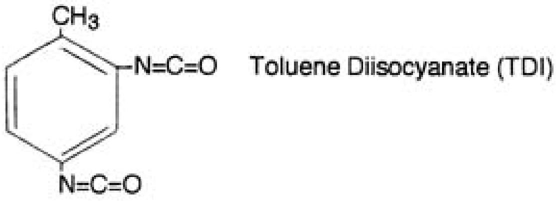


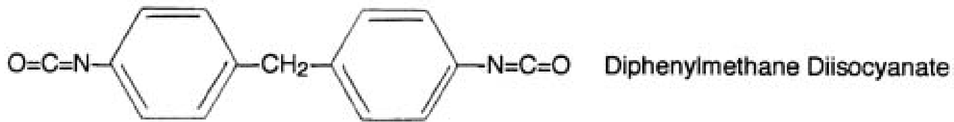
$$HOCH_2CH_2O(CH_2\underset{\underset{CH_3}{|}}{C}HO)_nCH_2CH_2OH \quad \text{Typical Polyether}$$



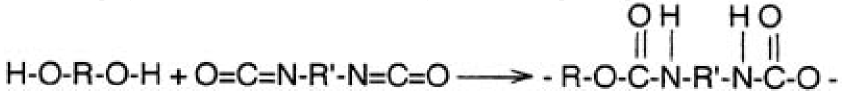
$$HOCH_2CH_2O(\overset{O}{\overset{\|}{C}}CH_2CH_2CH_2CH_2\overset{O}{\overset{\|}{C}}OCH_2CH_2O)_nH \quad \text{Typical Polyester}$$

Polyisocyanates are usually toluene diisocyanate or diphenylmethane diisocyanate.



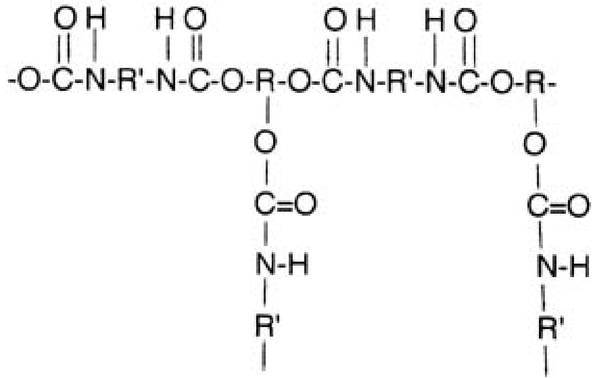


When the polyols are linear (difunctional) they form thermoplastic polyurethanes.



$$H-O-R-O-H + O{=}C{=}N-R'-N{=}C{=}O \longrightarrow -R-O-\overset{O}{\overset{\|}{C}}-\overset{H}{\overset{|}{N}}-R'-\overset{H}{\overset{|}{N}}-\overset{O}{\overset{\|}{C}}-O-$$

When the polyols are trifunctional or higher, they form thermoset polyurethanes.



In most processes, the reactive liquids are mixed and poured, and polymerized and cured rapidly to the final products (Table 19.13).

TABLE 19.13 Polyurethane Markets

Form	Percent of Polyurethane Market
Flexible foam	48
Rigid foam	28
Reaction injection molding (RIM)	6
Rubber, spandex, sealants, adhesives, coatings	17

Most polyurethane is foamed during the polymerization/cure reaction.

Flexible foam is made by mixing long trifunctional polyol with isocyanate to form the polyurethane, and adding a little excess isocyanate and water to the reaction to produce carbon dioxide which produces the foam. The largest use is in furniture, with smaller amounts in auto seating, mattresses, rug underlay, textiles, and packaging.

Rigid foam is made by mixing short polyfunctional polyol with di- or higher polyisocyanate, and foaming either with volatile liquid or with isocyanate and water. The largest use is in building insulation, with smaller amounts in refrigeration, industrial insulation, packaging, autos, and marine flotation.

Reaction injection molding (RIM) mixes polyol, polyamine, polyisocyanate, and strong catalyst, and injects the mixture rapidly into a mold, where it cures rapidly to form large parts very economically. It is used primarily for producing front ends and other large parts of autos.

PhenolFormaldehyde

Phenolic plastics were the first commercial synthetic plastics in 1908, and were the leading commodity plastic for 40 years, until the growth of vinyl and styrenic thermoplastics (Table 19.14). Now quite mature, they remain the second largest family of thermoset plastics.

TABLE 19.14 Phenolic Plastics Markets

	% of Total
Binder for plywood, composition board, paper laminate board	72
Binder for glass wool insulation	15
Molding powders	8
Foundry resins	3
Binder for abrasive products	2

While their largest use is as adhesives for outdoor plywood and glass wool insulation, they provide a group of compression-molding fiber-reinforced plastics which meet high engineering performance requirements (Table 19.15).

TABLE 19.15 Typical Properties of Molded Phenolic Plastics

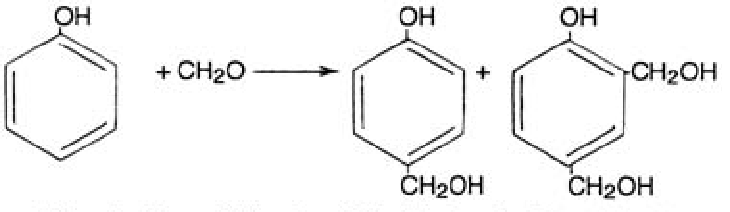
Grade	General-Purpose	Engineering
Fibrous filler	Wood flour	Glass
Modulus, MPa	9,000	18,000
Flexural strength, MPa	76	275
Notched izod impact strength, J/cm	0.20	5.3
Heat deflection temperature, °C at 1.82 MPa	168	240
Maximum continuous use temperature, °C	149	194

For performance under severe conditions, they compare very favorably with more expensive engineering thermoplastics as shown in Table 19.16.

TABLE 19.16 Phenolics as Engineering Plastics

Property	Phenolic	Engineering Thermoplastics
Price, in cents/cubic inch	2.78.0	9.518.4
Creep modulus, 100 hr/14 MPa, in MPa	28	5.510
Compressive creep, 14 MPa/50°C, in %	0.02	0.11.4
Heat resistance, in °C	315	121260

The chemistry of phenolic molding powders begins with the reaction of phenol with formaldehyde.



The high reactivity of the methylol groups makes it easy to polymerize and cure phenolic polymers,

$$
\begin{aligned}
\mathrm{P{—}CH_2OH + HOCH_2{—}P} &\rightarrow \mathrm{P{—}CH_2OCH_2{—}P}\\
&\rightarrow \mathrm{P{—}CH_2{—}P}\\
\mathrm{P{—}CH_2OH + H{—}P} &\rightarrow \mathrm{P{—}CH_2{—}P}
\end{aligned}
$$

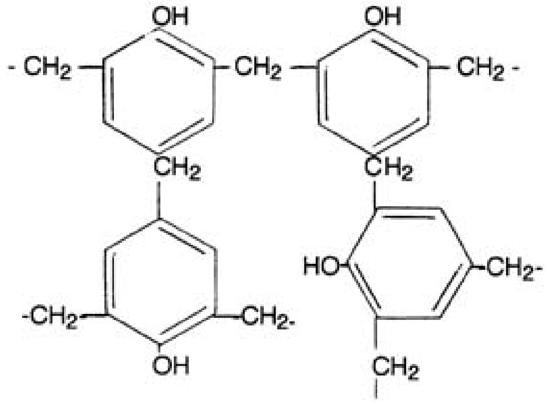
and also to copolymerize them with melamine and furan plastics,

$$\mathrm{P{—}CH_2OH + HOCH_2{—}M \rightarrow P{—}CH_2{—}M}$$

and with cellulosic reinforcements such as wood, cotton, and paper.

$$\mathrm{P{–}CH_2OH + HO{–}Cellulose} \longrightarrow \mathrm{P{–}CH_2{–}O{–}Cellulose}$$

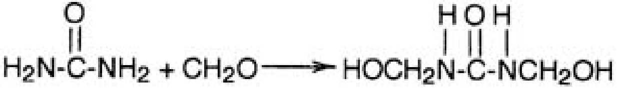
It is generally a three-stage process: (1) initial reaction to A-stage resin produces low-molecular-weight oligomers, which are still soluble, fusible, and reactive. (2) Melt compounding advances them to B-stage resins which are fairly fusible to doughy melts, and still reactive. (3) Molding them into finished products advances them to C-stage resins which are fully cross-linked to stable thermoset plastics of three-dimensional infinite molecular weight,



high modulus and strength, and very resistant to heat and chemical environments.

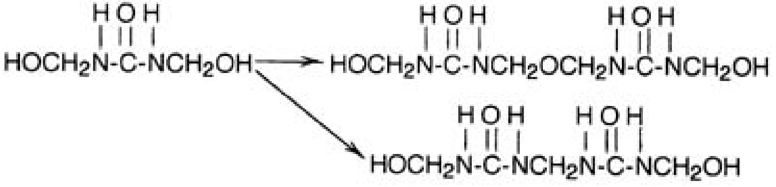
UreaFormaldehyde

These plastics became commercial about 1929. Urea and formaldehyde react very readily to form methylol compounds, mainly dimethylol urea.

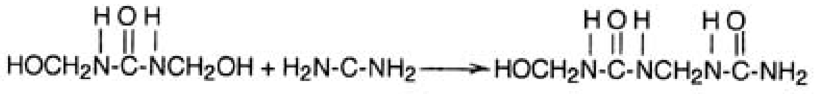


$$\mathrm{H_2N{-}C({=}O){-}NH_2 + CH_2O \longrightarrow HOCH_2N(H){-}C({=}O){-}N(H)CH_2OH}$$

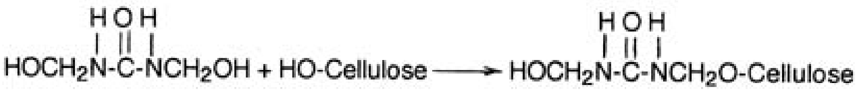
The methylol groups are very reactive, condensing with each other, with the NH groups in urea, and with the OH groups in cellulose.



$$\mathrm{HOCH_2N(H){-}C({=}O){-}N(H)CH_2OH \longrightarrow HOCH_2N(H){-}C({=}O){-}N(H)CH_2OCH_2N(H){-}C({=}O){-}N(H)CH_2OH}$$
$$\mathrm{\longrightarrow HOCH_2N(H){-}C({=}O){-}N(H)CH_2N(H){-}C({=}O){-}N(H)CH_2OH}$$

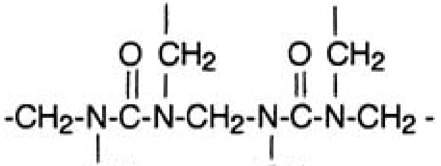


$$\mathrm{HOCH_2N(H){-}C({=}O){-}N(H)CH_2OH + H_2N{-}C({=}O){-}NH_2 \longrightarrow HOCH_2N(H){-}C({=}O){-}N(H)CH_2N(H){-}C({=}O){-}NH_2}$$



$$\mathrm{HOCH_2N(H){-}C({=}O){-}N(H)CH_2OH + HO{-}Cellulose \longrightarrow HOCH_2N(H){-}C({=}O){-}N(H)CH_2O{-}Cellulose}$$

Thermosetting cure produces highly cross-linked three-dimensional molecules of infinite size.



The largest use is for binding fibrous and granulated wood into indoor composition board. Smaller uses are for wet-strength paper and permanent-crease textiles. About 4 percent of ureaformaldehyde resin is combined with alpha-cellulose to make molding powders (Table 19.17). These find use mainly in electrical parts such as switches, wall plates and receptacles, circuit breakers, electric blankets, handles and knobs.

TABLE 19.17 Typical Properties of UreaFormaldehyde-Alpha-Cellulose Moldings

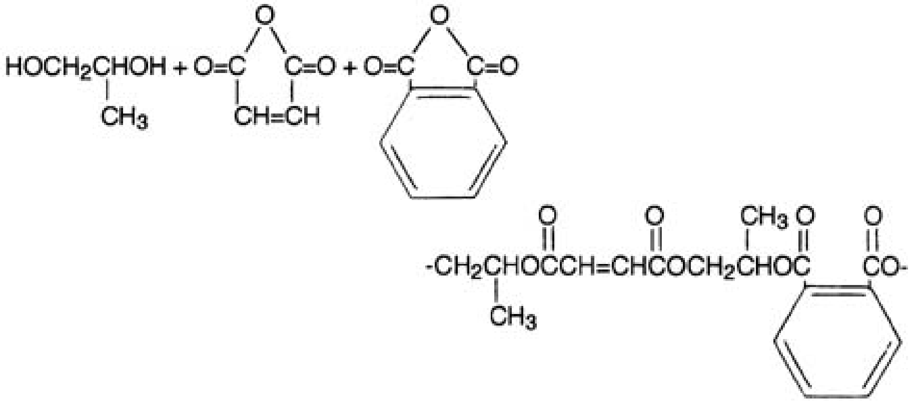
Flexural modulus, MPa	10,000

Flexural strength, MPa	100
Notched izod impact strength, J/cm	0.16
Heat deflection temperature, °C at 1.82 MPa	132

Unsaturated Polyesters

The chemistry of unsaturated polyesters was developed in the 1930s, and manufacture of glassfiber-reinforced polyesters began in the early 1940s. Their outstanding performance was recognized early in the typical consumer compliment: "That's not cheap plastic, that's high-performance fiberglass." They matured early, and form about 10 percent of the present thermosetting plastics market.

Their chemistry is fairly complex. The most common material is made from propylene glycol plus maleic anhydride plus phthalic anhydride.



When these are cooked together, maleic anhydride isomerizes to fumaric acid, and they condense to form low-molecular-weight propylene fumarate phthalate copolyester oligomers. These are mixed with styrene monomer, reinforced by glass fibers, usually extended with low-cost fillers, and cured by peroxide to form rigid strong products which are very resistant to impact and heat (Table 19.18).

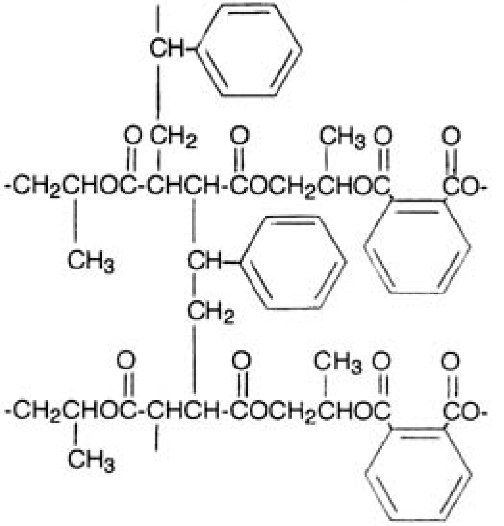


TABLE 19.18 Typical Properties of GlassFiber-Reinforced Thermoset Polyesters

Process	BMC	SMC	Woven Cloth	Filament Wound
Modulus, MPa	14,000	11,000	14,000	600024,000
Strength, MPa	121	159	414	283586
Notched izod impact strength, J/cm	4.0	7.8	9.4	
Heat deflection temperature, °C at 1.82 MPa	182	225	20+	

The mechanics of processing are carried out by a number of techniques. *Hand layup or sprayup* produces large shapes such as boat hulls, recreational vehicles, mobile homes, truck cabs, and tub-shower units. *Continuous panel processing* produces room dividers and skylights. *Compression molding* of sheet molding compound (SMC) and bulk molding compound (BMC) produces autobody parts, bathtubs, septic tanks, trays, tote boxes, and equipment housings. *Pultrusion* produces flagpoles, archery bows, and park benches. And *filament winding* is highly engineered to produce maximum strength in pipes and storage tanks.

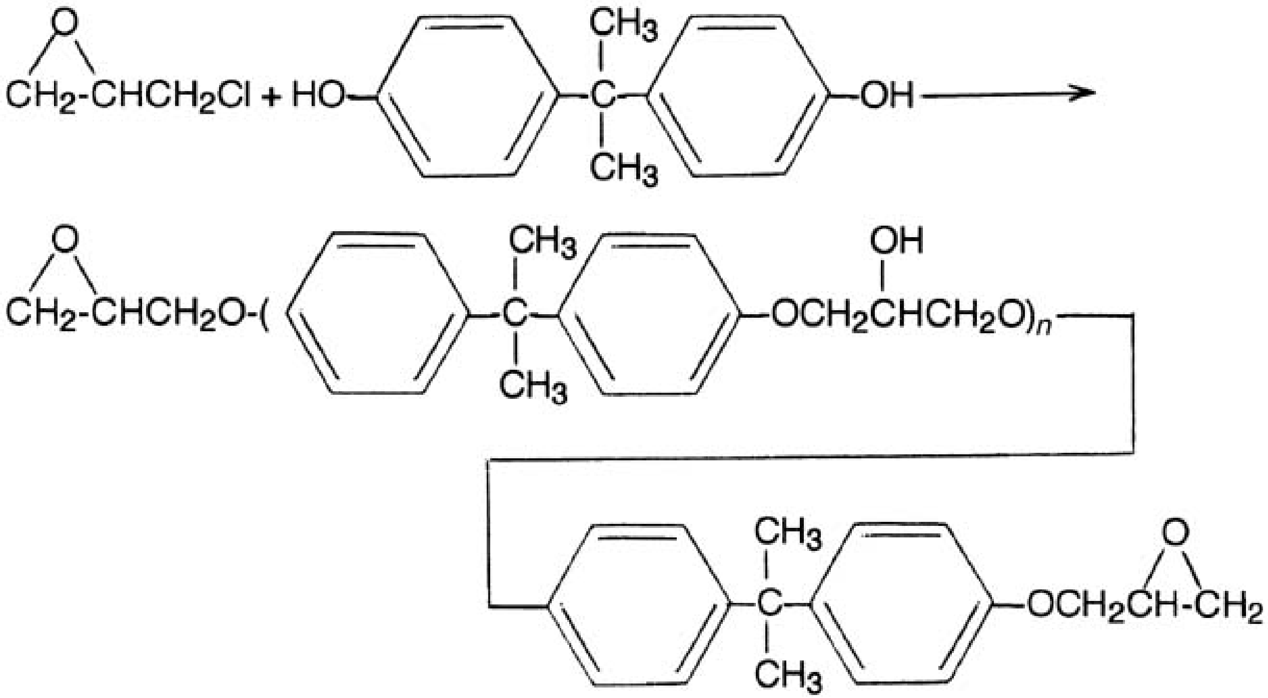
Epoxy Resins

These were developed in the 1940s and offered a unique combination of engineering performance which made them a popular family of thermoset plastics: fast low-temperature cure with low pressure and low shrinkage; high adhesion to polar surfaces; hardness, heat- and chemical-resistance (Table 19.19). Their largest use is in coatings for corrosion protection and electronic equipment. The second largest is in printed circuit boards. And three other important uses are adhesives, flooring, and high-performance fiber-reinforced plastics.

TABLE 19.19 Property Ranges for Cured Epoxy Resins

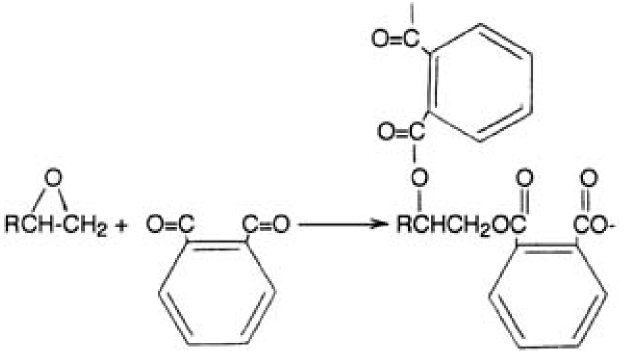
Flexural modulus, MPa	14,00034,000
Flexural strength, MPa	55655
Notched izod impact strength, J/cm	0.1621.0
Heat deflection temperature, °C at 1.82 MPa	93288

Epoxy chemistry is complex. Most epoxy resins are made by reaction of epichlorohydrin with bisphenol A.



While they are called "resins," they are really monomers to low-molecular-weight oligomers, liquids to soluble fusible solids, with high reactivity in the epoxy rings, and fair reactivity in the internal hydroxyl groups. They are cured most often by room-temperature reaction with polyamines or polyamide amines.

For higher-temperature cure and heat-resistance, they are cured most often by acid anhydrides.



There are also many more types of epoxy resins and curing agents for more specialized applications.

MelamineFormaldehyde

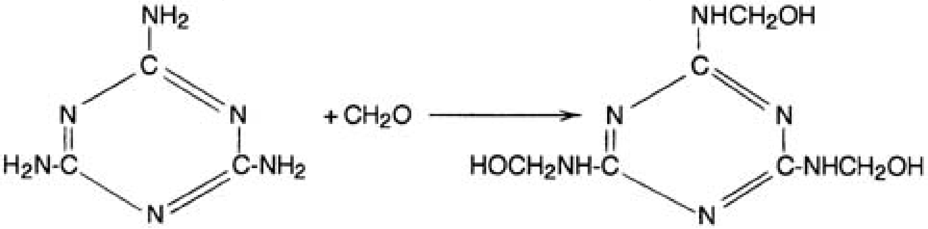
These resins became commercial in the 1930s. Their combination of high thermosetting reactivity, cured hardness, and resistance to heat, weather, and chemical environments made them particularly valuable for their good appearance and durability (Table 19.20).

TABLE 19.20 Typical Properties of Cellulose-Filled MelamineFormaldehyde

Flexural modulus, MPa	900011,000
Flexural strength, MPa	70124
Notched izod impact strength, J/cm	0.130.21
Heat deflection temperature, °C at 1.82 MPa	127143

Their largest use is in coatings, where they are used to cure acrylic automotive coatings and polyester appliance coatings. Their second largest use is in counter-tops, where they protect the decorative surfaces against abrasion, heat, and chemical attack. Their third use is in dinnerware, where their light weight, impact resistance, and attractive appearance are very competitive with China.

Melamine chemistry begins with the addition of 23 mols of formaldehyde to form methylol melamines.



These are very reactive with each other, with the remaining N-H bonds on melamine, with the hydroxyl groups in acrylic and polyester coatings, and with the hydroxyl groups in paper for countertops and in alpha-cellulose for molded dinnerware.

$$\mathrm{RNHCH_2OH + HOCH_2NHR'} \rightarrow$$

$$\mathrm{RNHCH_2OCH_2NHR'}\text{——}\mathrm{RNHCH_2NHR'}$$

$$\mathrm{RNHCH_2OH + R'NH_2} \rightarrow \mathrm{RNHCH_2NHR'}$$

$$\mathrm{RNHCH_2OH + HOR'} \rightarrow \mathrm{RNHCH_2R'}$$

The combination of resonance stabilization in the melamine heterocycle, and the high cross-linking between methylol melamines and with the other polymers, all produce the outstanding properties which make it a valuable specialty member of the thermoset plastics spectrum.

19.12 General Considerations

StructureProperty Relationships

When plastics engineers want to improve properties in an existing product, or when they want to select the optimum material for a new product, a routine search of existing tables of properties may sometimes be sufficient. For more professional judgment and problem-solving, however, and for planning development of new materials, they need to understand the basic relationships between polymer structure and practical properties. Here are some starting guidelines.

Molecular Weight

Low molecular weight gives lower melt viscosity for injection molding, and easier solution processing in general. High molecular weight is preferred for extrusion, and particularly for blow molding, thermoforming, stretch orientation, and thermoplastic foaming. In finished products, high molecular weight generally gives higher mechanical strength and chemical resistance.

Molecular Flexibility/Rigidity

Flexible molecules generally give lower melt viscosity for easier processing; softer, more flexible, more extensible products; and higher impact strength, friction, and acoustic absorption. Rigid molecules generally give higher rigidity, strength, creep resistance, heat deflection temperature, and impermeability; and lower coefficient of thermal expansion, and dielectric constant and loss.

Crystallinity

When polymers crystallize, their melting points are much higher and sharper than the softening points of amorphous plastics. This requires higher processing temperatures, but gives higher maximum use temperatures in the final products. Increasing crystallinity generally increases rigidity, strength, creep resistance, dimensional stability, impermeability, and chemical resistance; but decreases impact strength and transparency.

Orientation

Stretch orientation of extruded fibers and films greatly increases modulus, strength, transparency, and impermeability. It is also useful for producing shrink-packaging; conversely, a disadvantage is thermal dimensional instability. When orientation occurs accidentally in injection molding, calendering, thermoforming, and other processes, it generally produces undesirable anisotropy of final structure and properties.

Polarity and Hydrogen-Bonding

Whereas hydrocarbon polymers are nonpolar and have weak intermolecular attraction, introduction of negative atoms into the polymer moleculeoxygen, nitrogen, chlorineproduces permanent polarity, giving strong intermolecular attractions. Increasing polarity generally requires higher processing temperatures, and gives higher modulus, strength, creep resistance, heat deflection temperature, crystalline melting point, dielectric constant and loss, and gasoline and oil resistance.

When the polymer contains oxygen and especially nitrogen, polarity also produces hydrogen-bonding, which is an even stronger intermolecular attraction, and produces all the same effects to an even greater extent. Another effect of hydrogen-bonding is water absorption. In fabrics this produces greater comfort; but in plastics it decreases modulus, strength, and dimensional and chemical stability. These effects are most noticeable in nylons and cellulosics.

Cross-Linking

Thermoplastics are stable linear molecules which are softened by heat and soluble in solvents of similar polarity; this makes for easy processability. Primary covalent cross-linking in thermosets converts them into three-dimensional molecules of infinite size, with tremendous changes in properties: insolubility and infusibility; higher modulus, creep resistance, maximum use temperature, and chemical resistance; and lower extensibility, impact strength, thermal expansion, dielectric constant and loss, solvent swelling, and permeability. Cross-linking produces shrinkage strains which embrittle the polymer, so most thermoset plastics must be reinforced with fibers; the result is synergistic improvement of modulus, strength, impact resistance, and dimensional stability, producing enhanced engineering performance.

Additives

Polymers are rarely used in pure form. They are almost always improved by use of additives to enhance specific properties. The major classes of additives may be briefly summarized as follows.

Stabilizers

Organic polymers are not perfectly stable. Specific polymers and specific products require additives to improve their stability during processing and/or long-term use of finished products. *Antioxidants* are added to polyolefins and rubber-modified impact plastics to protect against atmospheric oxygen; these are primarily hindered phenols and polyphenols, sometimes synergized by aliphatic sulfides or organic phosphites, used in fractions of a percent up to several percent. *Thermal stabilizers* must be added to PVC to prevent dehydrochlorination, discoloration, and cross-linking during melt processing: organotin esters are strongest, used as a fraction of a percent for processing rigid PVC; barium/zinc soaps plus epoxidized fatty esters plus organic phosphites, total concentration several percent, form a synergistic stabilizer system for plasticized flexible PVC; and basic lead oxide compounds, several percent, are best for wire and cable insulation. *Ultraviolet light stabilizers* are needed in products for outdoor applications: cyclic hindered amines and *o*-hydroxy benzophenones and benzotriazoles are used at a fraction of a percent in clear products, while carbon black and especially aluminum flake are extremely effective in opaque products. *Biostabilizers* are used to protect natural polymers and monomeric additives against attack by microorganisms; these are chemicals which require a delicate balance between toxicity to microorganisms vs. safety for macroorganisms like ourselves.

Fillers are inorganic powders added in large amounts to increase modulus, dimensional stability, and opacity, and often to reduce cost. *Reinforcing fibers* are mostly glass, occasionally carbon or organic fibers, typically added in concentrations of 1040 percent, to increase modulus, strength, impact strength, creep resistance, and dimensional stability; long and continuous fibers give the greatest improvement in properties, while short chopped fibers ($\frac{1}{16}$2 in.) permit fairly conventional melt processing.

Coupling agents are chemical surface treatments applied to fibers, and sometimes to fillers, to strengthen the interface between inorganic fillers and fibers and organic polymer matrixes, to improve dispersion and stress transfer across the interface. Most common are organosilicon compounds of the type (RO) 23SiX21, where RO is typically methoxy or ethoxy to react with the silanol surface of glass fibers, and X is an organic group designed to react with a thermosetting polymer matrix, or at least to be attracted toward a thermoplastic polymer matrix.

Plasticizers are typically organic liquids of very low volatility, which are miscible with a polymer, and are added to it to improve processability and, in larger amounts (2080 parts per hundred of resin), to make it soft and flexible, or even rubbery and/or adhesive. The major portion of the plasticizer market (80%) is aliphatic and aromatic esters, which go to convert rigid PVC into flexible PVC. The remainder goes to improve the processability of cellulosics, and for a variety of specialized uses in other polymers.

Lubricants are a variety of proprietary additives, which are used either to improve melt flow, release from steel process equipment, or self-lubricity in final products such as gears and bearings.

Flame-retardants may be built into the polymer during polymerization or cure, or they may be physical additives to the finished composition. Organic phosphorus compounds are the most effective, typically requiring about 2 percent of phosphorus to prevent burning. Organobromine (1020%) and chlorine (1040%) are effective when used in larger amounts. Antimony oxide, and some other metal oxides, can synergize the action of bromine and chlorine, reducing total flame-retardant concentrations to 510 percent. And alumina trihydrate and magnesium hydroxide, which release water when heated, are becoming increasing popular when used in large enough amounts to be effective (4060%).

Colorants

One third of plastic materials are used in their natural color. The other two thirds are colored for esthetic and/or functional reasons, typically using about 1 percent of colorant. Inorganic minerals and synthetic colors give greater stability and opacity; while organic colors are available in greater variety, miscibility, and efficiency, and less likely to raise questions of toxicity. Thus use of inorganics is decreasing, while use of organics is increasing.

Organic peroxides are used to initiate free-radical polymerization of ethylene, butadiene, styrene, vinyl chloride, vinyl acetate, and methyl methacrylate. They are also used to cure unsaturated polyesters, occasionally to cross-link thermoplastics such as polyethylene and polyacrylates, and increasingly for grafting and compatibilization of polymer blends. A variety of organic peroxides offer useful reactivity over a temperature range from 0 to 130°C or more, for different polymers and different processes.

Polymer Blends

Blending of polymers with each other accounts for approximately 40 percent of the present plastics market, and the practice is growing continually, because it permits the development of improved properties without the cost of inventing new polymers. When polymers are fairly miscible, as in the polyethylenes, and in polyphenylene ether plus polystyrene, blending can be used to produce intermediate properties and balance of properties. Most polymer blends are not miscible, and separate into microphases; if these can be strongly bonded at the interface, it is often possible to produce synergistic improvement of properties, particularly the balance between rigidity and heat deflection temperature on the one hand, plus ductility and impact strength on the other; or between soft, flexible, rubbery properties on the one hand, plus the strength of "thermoplastic cross-linking" on the other. Some of the newer blends are used to produce barrier properties and other valuable improvements.

Critical Properties: Challenges to the Plastics Industry

It is easy to be positive and proud of the accomplishments of the plastics industry. On the other hand, there are a number of areas in which it is obvious that plastics are not yet perfect, areas where major breakthroughs could open major new markets and uses for plastics in the future. It may be stimulating to explore some of these here.

One-step conversion of monomer to finished product could reduce processing steps and costs dramatically. Epoxy cure and polyurethane RIM are examples of very fast reactions producing finished products. Monomer casting of acrylics and nylon 6 are commercial one-step processes. Polymer chemists have many more polymerization reactions which can rapidly convert monomers into high-molecular-weight or even thermoset polymers. What is needed are strong cooperative programs between polymerization chemists and plastics process engineers to develop these possibilities into commercial realities.

Continuous-fiber reinforcement gives plastic products which are not simply quantitatively, but often qualitatively, superior to most present commercial practice. Most plastic processing is limited to conventional melt flow of short-fiber reinforcements, which sacrifices much of the potential benefits of reinforcement. There are a few processes for incorporating continuous fiber reinforcementfilament-winding, pultrusion, swirl conformation of fibers in polymer sheets, and mixed fabrics of reinforcing fibers and plastic fibers. More vigorous development of such techniques could rapidly produce plastics products with far superior properties.

Modulus and creep-resistance of plastics are still inferior to metals, ceramics, and glass, which means there are areas where they cannot compete. We know that molecular rigidity, crystallinity, polarity, and reinforcing fibers can all go a long way toward closing the gap.

Abrasion-resistance of transparent plastics still cannot equal that of glass. We have ways of improving or coating the polymer, but customers keep telling the industry that there is still a long way to go.

Coefficient of thermal expansion of organic polymers is 12 orders of magnitude higher than metals, ceramics, and glass, which gives serious difficulties in product design, and especially in mating plastics parts with inorganic parts in an assembled product. Molecular rigidity, crystallinity, fillers, and especially fibrous reinforcement can go a long way toward bridging the gap, but there still is a gap to be overcome.

Thermal conductivity of plastics is very low, which makes them excellent insulators against heat and cold. On the other hand, there are times when high thermal conductivity is preferred, for example, in processing, cooking, and heating equipment. Inorganic fillers can help, in proportion to their volume concentration. Perhaps this approach can be carried further; or perhaps there are totally different mechanisms waiting to be discovered.

Heat resistance of organic polymers is far lower than that of metals, ceramics, and glass. There have been major improvements, based on aromatic and heterocyclic resonance, ladder structures, and other mechanisms, and we may see further improvement in the future. Perhaps more serious limitations are the high cost of synthesis and the difficulty of processing these polymers into the desired final products. This is an area where the polymer chemist could use more help from the plastics engineer.

Electrical conductivity of polymers is very low, making them very useful as insulation. On the other hand, there are products in which conductivity would be very desirable. Semi-conductivity is fairly easily achieved by adding semi-compatible hydrophilic organic compounds, and fairly high conductivity can be achieved by metallic fillers, especially fibers. Research is developing polymers which are inherently conductive due to conjugated unsaturation plus doping with inorganic electron donors. This is an area where research support and activity are making good progress at present.

Dielectric breakdown occurs when high voltage drop across an insulator causes some current to leak through, turning to heat, and ultimately decomposing the polymer and burning a conductive carbon track right through the insulator. Practically, some polymers are more resistant than others; but more theoretical understanding is needed in order to design polymer structures which will offer superior resistance to dielectric breakdown.

Outdoor weathering of plastic products has been the subject of both theoretical and practical study. Some plastics can last for many years, others for more limited times, and a few can actually be designed to self-destruct rapidly. Mechanisms involve ultraviolet light, atmospheric oxygen, water, transition metals, acid rain, wind-blown dust, and microbiological action. Polymer structure and additives respond to these mechanisms in various ways. Whereas there presently exist a fair theoretical understanding and practical control measures, there remains much to learn in order to achieve the ultimate goals of long-term weather stability and efficient control of solid waste.

Solvent-resistance of organic polymers varies with polarity, crystallinity, and cross-linking, so it is usually possible to solve solvent-attack problems by proper choice of polymer; but it is important to remember than they do not easily compete with metals, ceramics, and glass.

Barrier properties of organic polymers cannot equal metals and glass. Plastics offer so many other advantages that we often try to compromise or laminate to optimize overall balance of properties. It is known that molecular rigidity and crystallinity improve barrier performance, and from practical experience we can identify several very high-barrier polymers; but the details that make one much better than another are not understood so discovery of superior barriers is slow and uncertain.

Cost of engineering performance is a major factor retarding growth. Organic polymer chemists can easily design and synthesize polymer structures with higher and higher properties, but the cost of synthesis and difficulty of processing often inhibit their use for many years. Closer cooperation between chemists and engineers should be the optimum route to more efficient development.

Fire Performance

Wood burns. Most of the fires throughout history, causing death and destruction, have been caused by wood. It is rare that anyone will say "Ban wood." Plastics are also carbon compounds, and they also burn. When plastics are involved in a fire, even to the slightest degree, there is often an outcry, "Ban plastics." Thus the use of plastics in building and transportation is seriously restricted by this prejudice.

Plastics are not all equally flammable. Some burn as readily as fuels. Some do not burn spontaneously; but when exposed to a severe fire, they can be burned. And some require enriched or pure oxygen environment before they will burn. Most plastics can be made more resistant to burning, by incorporating flame-retardant elementsphosphorus, bromine, chlorine, antimony, even watereither in the polymer molecule or in physical additives.

Related problems must be considered in individual products. Bromine, chlorine, and antimony add to the smoke of a fire, while phosphorus and water do not, and some metal oxides can actually reduce it. Toxicity of combustion gases is a major concern; but the main problem is that oxidation of carbon compounds in an enclosed spaceindoorsproduces carbon monoxide, no matter whether the carbon compounds are wood or plastics. Other problems include the cost of flame-retardants, difficulties in processing, and loss of mechanical or thermal properties. The designer must balance all of these in each product, and choose the optimum solution to the problem.

Health and Environment

Whenever new chemistry is introduced into the environment, there arises the question of its effect on our health. On the positive side, plastics packaging of food prevents contamination and spoilage, and prolongs its useful life; and use of plastics in medicine has made major contributions to health and longevity. On the negative side, there have been several occasions where plastics chemistry has caused health problems; whenever these have been identified, they have been solved successfully and quickly.

This leaves a large grey area, in which people who do not understand chemistry may combine ignorance and fear of anything new, and try to roll back the material progress of modern science and industry. Unfounded fear of plastics has led to many popular and even political attempts to limit or ban their use. This has certainly had a retarding effect on our ability to develop their new uses to maximum advantage. Such problems have been seen throughout historysteel plows, balanced diets, immunization and medication, all have had to overcome popular fears before they could offer their benefits to mankind; so the industry should not be discouraged when plastics encounter similar difficulties.

Recycling

Modern science and industry have provided a growing supply of material products. When they reach the end of their useful life, they become solid waste, and disposing of it has become a growing problem. Worst of all is over-packaging to stimulate sales, so discarded packaging is the major contributor to this solid waste. Plastics are not the major component of solid waste; but because of their low density, bright colors, and relative weather-resistance, they are the most obvious component. It would be desirable to remove them from solid waste by recycling.

Since 85 percent of the plastics market is thermoplastic, it is theoretically recyclable. In industry, individual thermoplastic materials are almost always recycled immediately for purely economic reasons. Post-consumer waste, on the other hand, presents serious problems. Voluntary recycling of PET soda bottles has been most successful, because they are easily separated, cleaned, and converted into bulk fiber products. Voluntary recycling of HDPE milk and water bottles has been fairly successful, because they are easily separated and cleaned, but development of markets has been slower in coming. Most other plastics occur in smaller amounts and more diversity, so voluntary efforts have been very limited. The greatest problem is that, when several materials are combined in a single product, separation is difficult to impossible. Ultimately, the success of recycling will depend very much on greater cooperative efforts by government, industry, and consumers.

Recycling rarely reproduces virgin plastic materials. With repeated recycling, quality decreases, and potential uses decrease. At some point, the final step should be incineration. This produces useful energy, returns carbon dioxide and water to the ecocycle, and reduces final solid waste to an absolute minimum. Present incineration technology suffers from old inefficient equipment and non optimal operation. Sooner or later, we will have to make the effort to develop, build, and operate incinerators for maximum efficiency and minimal harm to the environment.

Part III. PLASTIC PROCESSING

19.13 Rheology

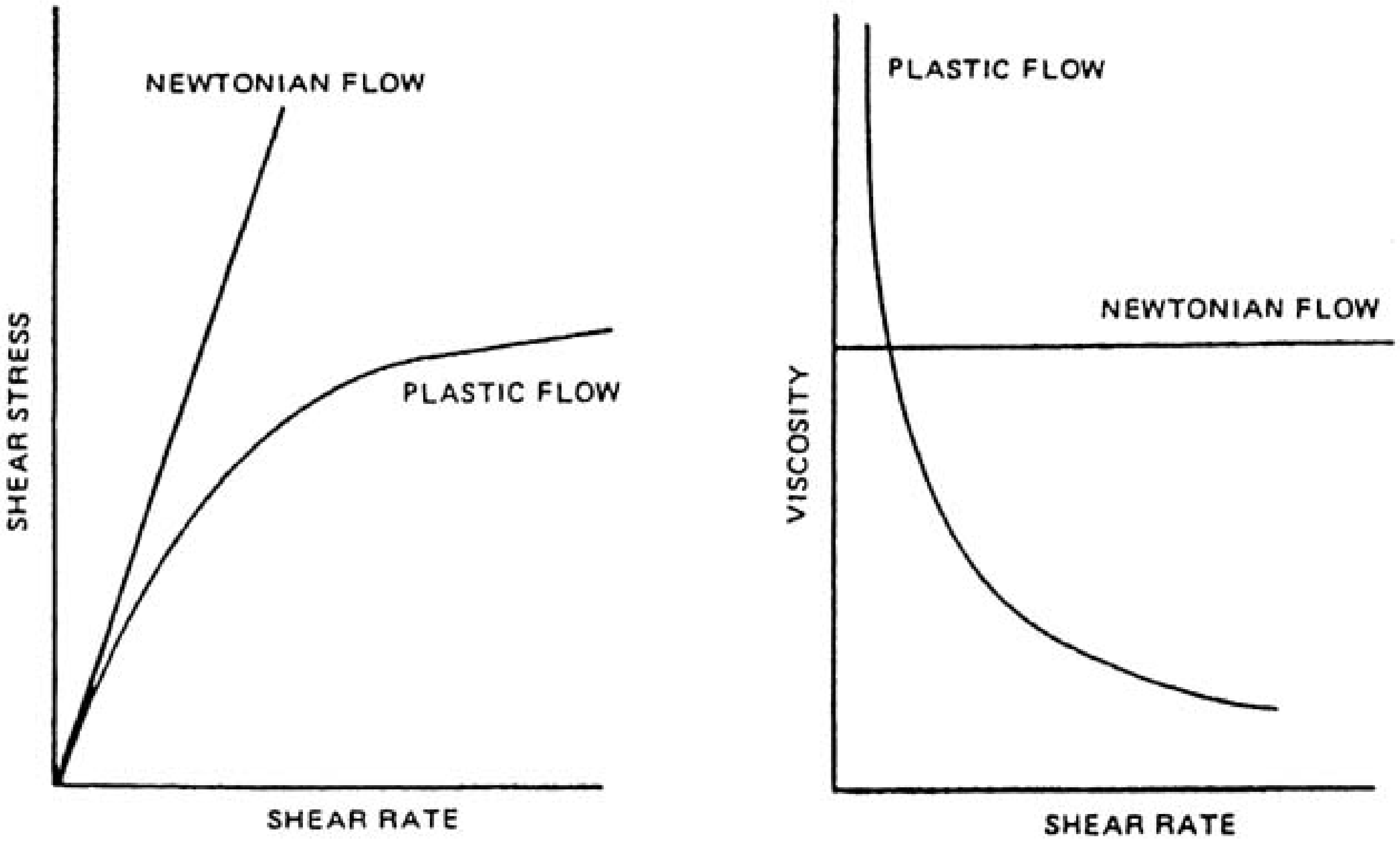
Fundamental Concepts

Plastic processing is primarily the flow and shaping of viscous liquids. The scientific study of this flow is called *Rheology*. Assuming laminar shear flow, viscosity is defined as the ratio of shear stress to shear rate.

$$\eta = \frac{\sigma}{\gamma}$$

If this ratio is constant, it indicates a simple Newtonian fluid. For most plastic materials, however, increasing shear rate disentangles polymer molecules and aligns them in the direction of flow, so increasing shear rate decreases resistance to flow (viscosity) (Fig. 19.26), and this non-Newtonian behavior is defined as *pseudoplastic*.

Fig. 19.26 Polymer rheology. (Berins, M. L. (Ed.), *SPI Plastics Handbook*, p. 56, 1991, Copyright © Kluwer Academic/Plenum Publisher, New York. Used by permission of the copyright owner.)



With increasing temperature, viscosity decreases in a manner approximately described by the Arrhenius equation,

$$\eta = Ke^{-E/RT}$$

and the resulting activation energy E correlates with the rigidity of the polymer molecule (Table 19.21). The terms are defined as follows: K is a constant characteristic of the polymer and the test method, e is the natural log base 2.718, E is activation energy, R is the gas constant, and T is absolute temperature in °K.

TABLE 19.21 Polymer Activation Energy

Polymer	Activation Energy, kJ/gm·mol
Silicone	16.7
Polyethylene	27.8
Polypropylene	39.6
Polyethylene terephthalate	79.2
Polystyrene	104.2
Polycarbonate	116.7
Poly-*a*-methyl styrene	133.3

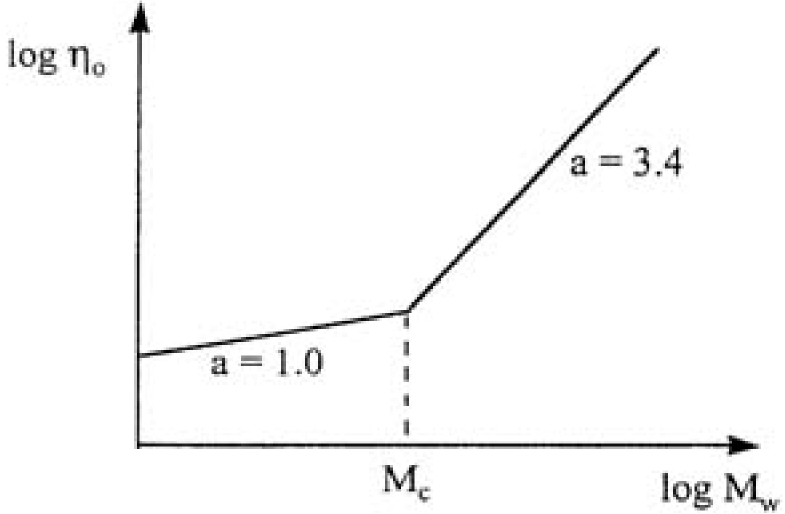
With increasing pressure, free volume between polymer molecules decreases, flow becomes more difficult, and viscosity increases.

A major factor in polymer viscosity is molecular weight, M,

$$\eta = KM^a$$

where the experimental exponent "a" represents the kinetics of disentangling polymer molecules from each other in the melt (Fig. 19.27) ($a = 1$). Beyond a critical molecular weight Mc (typically 500040,000), the difficulty of disentangling molecules multiplies viscosity exponentially ($a = 3.4$).

Fig. 19.27 Viscosity vs. molecular weight. (Harper, Charles A., *Modern Plastics Handbook*, p. 5.13, Copyright © 2000 by The McGraw-Hill Companies and used with permission of the copyright owner.)



Instrumental Measurement of Flow Properties

Capillary rheometers measure the effect of pressure on volumetric flow through a cylindrical capillary. They are popular in practical work because shear rate and flow geometry are similar to conditions in extrusion and injection molding. They cover a wide range of shear rates, and they give practical information on die swell, melt instability, and extrudate defects. Their main disadvantage is that they require a number of mathematical corrections to convert to true viscosity.

Cone and plate rheometers solve one problem, by providing constant shear rate. They can also be designed to measure torque, dynamic properties, normal stresses, and forces in other directions. A disadvantage is that they are limited to low shear rates.

Parallel plate viscometers are used for very high viscosities at low shear rates. Measurement of shear rate is difficult.

Coaxial (concentric) cylinder viscometers provide nearly constant shear rate. A disadvantage is that they are limited to liquids of low viscosity.

Extensional viscometers are useful to measure tensile viscosity in processes such as stretch orientation.

Dynamic or oscillatory rheometers measure viscous and elastic modulus in shear or tension. Energy dissipation produces a phase difference, so stress, strain, and phase angle can be used to characterize complex viscosity behavior.

Practical Effects of Flow Properties

Melt viscosity is the most critical practical property for the process engineer. When stiff molecules give high viscosity and slow flow rates, higher temperature and/or shear-sensitivity (pseudoplasticity) can often be used to decrease the melt viscosity and thus increase processability.

Extrusion defects are primarily due to melt elasticity. When shear rate occurs faster than polymer molecules can disentangle from each other, they simply stretch elastically and store potential energy. When they exit from the die, they release this energy and recover elastically. The resulting extrudate diameter is then greater than the die diameter. This is called *die swell*. The effect is aggravated by high molecular weight, high pressure, low temperature, high shear rate, and short L/D ratio. Two practical ways to compensaste for die swell are to: (1) reduce the die diameter, or (2) pull the extrudate away from the die at a velocity at least as great as it went through the die.

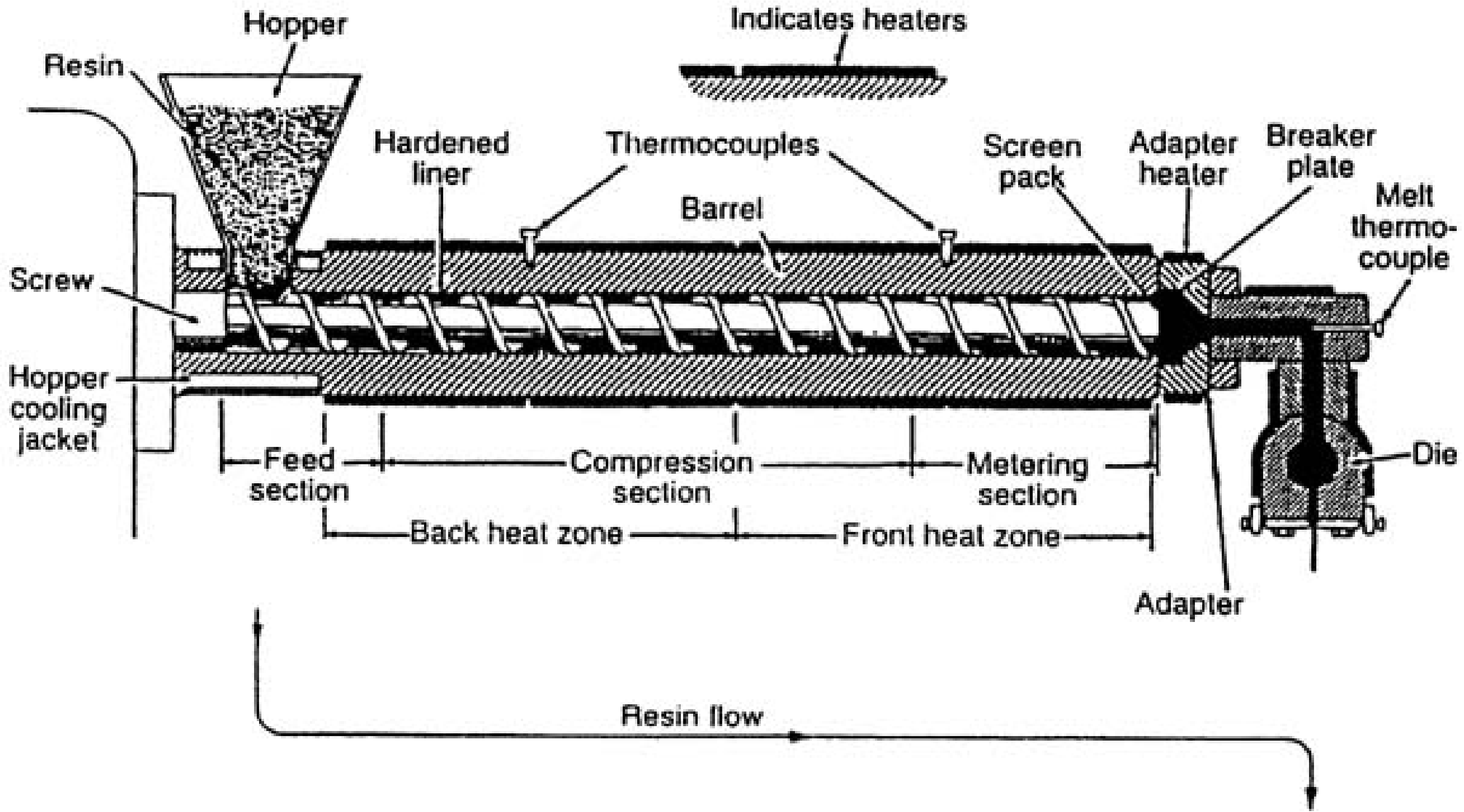
Melt fracture is the occurrence of distorted extrudate coming from the extruder. It is caused by flow disturbance at the point where flow cross-sectional area is rapidly reduced from the large diameter of the melt feed to the much smaller cross-sectional area of the die orifice. Here again it is aggravated by melt elasticity.

Land fracture is a fine surface roughness on the extruded product. It is caused by friction between the melt and the wall (land) of the die. It is solved by addition of lubricants to reduce melt/die adhesion.

19.14 Extrusion

Extrusion is the process of forming a material continuously through an opening. Most extruders do this by rotating a screw inside a stationary heated cylindrical barrel, to melt the polymer and pump the melt through a suitably shaped orifice (Fig. 19.28). This is used for direct manufacture of finished products such as film or pipe. It may also be used to feed a second process such as injection molding, blow molding, coating, laminating, or thermoforming. It is also important in compoundingblending polyers with additives to improve overall balance of propertiesand in recycling plastics out of solid waste.

Fig. 19.28 Extruder. (Harper, Charles A., *Modern Plastics Handbook*, p. 5.19, Copyright © 2000 by McGraw-Hill Companies and used with permission of the copyright owner.)



Basic Functions

The extruder generally performs six successive functions: solids conveying, melting ("plasticating"), melt conveying (pumping), mixing, devolatilization (venting), and forming. While a complete process may involve all six, some processes may omit one or more of these steps.

Solids conveying is carried out in two steps: the feed hopper and the back (entrance) portion of the screw. The feed hopper is an inverted cone or pyramid, in which solid pellets or powder flow downward from the force of gravity. If they flow poorly ("arcing" or "bridging"), the problem may be solved by installing a vibrator or a stirrer ("crammer/feeder") in the hopper, or by coating the particles with a small amount of a chemical that acts as an "external lubricant." Once the solid particles enter the back end of the screw, they are carried forward and compressed by the rotation of the screw and the friction between the solids and the barrel. Friction can be increased by roughening the surface of the barrel, and particularly by grooving the internal surface of the barrel.

Melting of the solid polymer is the result of heat from two sources: conduction from the heated barrel, and friction between the barrel and the polymer. Most of the heat comes from friction between the barrel, the viscous melt film, and the remaining solid polymer. This frictional heating is most efficient when the melt viscosity is high, and when the melt film is thin. Thinness of the melt film depends on designing and maintaining a very small clearance between the screw flights and the barrel surface.

Melt conveying is the forward motion of the molten polymer through the extruder, due to the pumping action of the rotating screw. This simple "drag flow" *M*d is proportional to melt density, down-channel velocity, and cross-sectional area of the screw channel. In most cases, however, there is also a pressure gradient as the melt moves downstream, either positive pressure approaching the exit die, or negative pressure following the solid/melting zones; so this "pressure flow" *M*p must be subtracted from the drag flow to calculate the net output of the extruder. Since pressure flow is inverse to melt viscosity, which is non-Newtonian and temperature sensitive, this complicates the calculation considerably.

Mixing in the extruder is important for homogenization of temperature and pressure, and especially for uniform blending with additives. In a simple single-screw extruder, melt flow is fairly linear and provides little mixing. A variety of ingenious modifications of screw design have been developed to build in mixing elements. Some improve distributive mixing of the liquid melt, to homogenize temperature and pressure fluctuations. Others build in higher shear (dispersive mixing), to break down particles of additives and blend them uniformly into the molten polymer. And still others combine the two types of action.

Twin-screw extruders permit much wider variation in design and performance. The screws may be non-intermeshing, just touching ("tangential"), partially or fully intermeshing. They may be co-rotating or counter-rotating. They are assembled ("programmed") of different sections ("elements"), designed for feed, melting, conveying, distributive melt mixing, dispersive shear for additives mixing, and sealing pressure or vacuum at the vent or the die. The barrels are also programmed of elements to provide functions such as feed ports, venting, and abrasion resistance. Their higher mixing efficiency makes them particularly useful in compounding with additives, processing polyvinyl chloride, reactive extrusion, and devolatilization.

Devolatilization can be used to remove up to 5 percent of volatile impurities from the plastic melt. The first melt conveying (metering) zone builds up melt pressure. Then channel depth is increased abruptly in the vent zone, the melt is decompressed, and volatiles escape through the vent. After this the melt enters a second metering zone, which builds up melt pressure again, and feeds it to the die.

Die forming forces the melt into the shape and dimensions desired in the final product. Temperature, pressure, viscosity, die design, flow rate, and flow patterns must all be optimized and controlled closely to make a suitable product. The sytem is so complex that it is managed more by experience and rule-of-thumb than by theoretical design. Typical guidelines include small approach angles, land length = 10 × land clearance, avoid abrupt changes in geometry, no dead spots, generous radii, and thin uniform wall sections.

Calibration is a technique for maintaining the shape and dimensions of the product from the time the melt exits from the die until it can be cooled enough to solidify and stabilize it. Depending on the size and shape of the product, calibrators can use water-cooled plates, internal mandrel, internal air pressure, external vacuum, or a post-extrusion die to change the shape of the molten extrudate before it solidifies.

Major Processes and Products

Blown film is produced from a single-screw extruder by extruding a tube, cooling it with external and/or internal air streams, stretching it in the machine direction by pulling it away from the die ("draw-down"), stretching it in the transverse direction (typically 24 × "blow-up" ratio) by internal air pressure up to 34 KPa (5 psi), flattening it by passing through nip (pinch) rolls, and winding it onto a cylindrical roll. Optional post-stretching operations may include flame or corona surface treatment for wettability/adhesion, sealing, slitting, and bag-making.

Flat film, sheet, and coating are produced from a single-screw extruder with a high L/D ratio (2733/1), which feeds the molten polymer through a flat die. The die opening is adjusted to control the thickness of the film, and the film is solidified in a cold water bath, or preferably, for transparency and gloss, over two or more water-cooled steel chill rolls. Compared to air-cooled blown film, water-cooled flat film generally has higher clarity. The take-off and wind-up line may include automatic feed-back thickness control, surface treatment, and/or slitting.

For extrusion coating, the substratepaper, plastic, or metal foilis preheated, and may be pretreated, before the extruder deposits a layer of molten polymer onto its surface. Low molecular weight and high temperature help the polymer to flow into a uniform adhesive coating. The laminated layers pass between pressure and chill rolls, and optionally through surface treatment, printing, and slitting before collecting on the final windup rolls.

Coextrusion produces multi-layer laminates in a single process step. Two or more extruders feed different molten polymers into a multi-manifold die which layers them directly, or into a modular feedblock which layers them before feeding them into the die. This is used primarily in the packaging field, to sandwich an impermeable barrier layer between two commodity outer-film layers, and often includes adhesive tie layers to bond the barrier layer to the outer layers.

Pipe, hose, and tubing are extruded through an annular (ring) die. L/D ratio is typically 24/1 or greater. Rigid vinyl is the leading material, and often requires a conical twin-screw extruder. The molten pipe is solidified by water-cooling. Pipe dimensions are controlled initially by the die, but then finally calibrated by pull-off rate, internal mandrel, vacuum, or compressed air. Flexible tubing is collected on a wind-up unit; while rigid pipe is hauled off by a caterpillar puller to a cut-off saw and stacker.

Profile extrusion of siding, window frames, gasketing, and other shapes is complicated by the effects of their asymmetry on heating, viscosity, cooling, and dimensional control. This generally requires modification of the die, and vacuum-driven calibration of the extruded product.

Wire coating extrudes plastic insulation around electrical wire and cable as it passes through a T-shaped crosshead die. The entire process line includes pay-off and capstan to feed and preheat the wire, extrusion-coating, water-cooling, spark testing, diameter and eccentricity controls, and take-up capstan and wind-up. The most common "pressure-coating" die applies the plastic coating inside the die; whereas, for larger wire and cable, the "tubing (tool) die" applies the coating as wire and plastic exit from the die, using internal vacuum to pull the molten coating onto the wire. A fine wire coating line can run at speeds up to 10,000 ft/min.

19.15 Injection Molding

Introduction

Extrusion and injection molding are the two leading methods of converting plastic materials into solid products. Extrusion produces continuous products with a fixed cross-section; injection molding produces discrete products with more complex shapes. Modern injection molding is a very fast, automated process for large-scale manufacture of complex products at minimum cost. An injection molding machine can be used for many different jobs and for many years, so the amortization per job can be very low. On the other hand, it requires a different mold to make each product, so the cost of the mold must be amortized over the life of its individual production run.

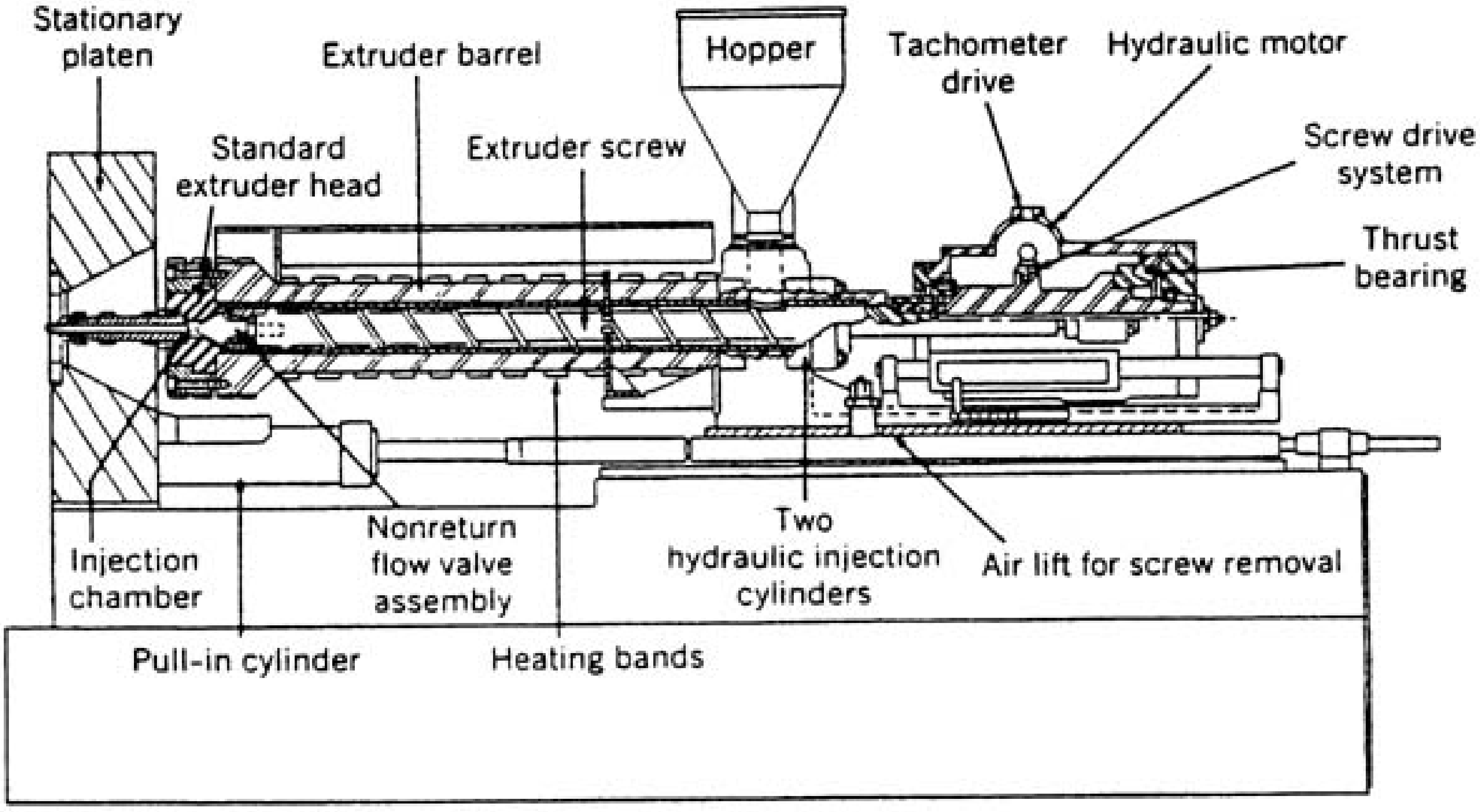
Injection Molding Cycle (Fig. 19.29)

The injection molding cycle involves a sequence of events:

1. closing the mold
2. melting the plastic material
3. injecting the melt into the mold
4. filling the mold cavity to form the product
5. cooling the molten product to solidify it in the mold
6. opening the mold and ejecting the solid product

The entire cycle is usually completed within a minute or less.

Fig. 19.29 Injection molder. (Rubin, Irvin I., *Handbook of Plastic Materials and Technology*, p. 1235, Copyright © 1990 by John Wiley & Sons, Inc.; used by permission of John Wiley & Sons, Inc.)



Closing the Mold

One half of the mold is attached firmly to the melting/feeding end of the injection molding machine. The other half slides horizontally to open or close the mold. The moving half is driven mechanically or hydraulically. When it closes the mold, it is held firmly by clamping action to withstand the injection pressure.

Melting the Plastic Material

The plastic pellets are fed through a hopper into an extruder screw, typically with an L/D ratio of 20/1. The first half of the screw compresses the pellets and squeezes out air. The third quarter of the screw melts the pellets, 70 percent by friction, 30 percent by conduction from heater bands on the extruder barrel. This is called "preplastication." The fourth quarter of the screw pumps the melt forward to the front of the extruder, where it goes through a one-way valve to prevent backflow. As the molten plastic accumulates at the front of the screw, this pushes the screw back to form a reservoir of molten material.

Injecting the Melt into the Mold

When the reservoir contains the desired (preset) volume of molten material, hydraulic force pushes the screw forward rapidly, injecting the melt into the mold cavity. For this reason it is called a "reciprocating screw." In order to fill the mold completely, and make up for shrinkage during cooling with or without crystallization, the screw and the one-way valve maintain high "boost" ("hold") pressure during this part of the cycle. An important auxiliary function is "venting," tiny openings to allow air to escape from the mold, so that the molten material can fill the mold completely and flawlessly.

Filling the Mold Cavity to Form the Product

The mold cavity is designed and machined to form the shape of the finished product. This is itself a complete art and science, based partly on experience, and increasingly on computerized engineering principles. Some major considerations are fast uniform flow, avoidance of degradation, minimization of orientation/anisotropy, fast cooling/solidification, shrinkage and dimensional tolerances, and of course final properties of the product.

Cooling the Molten Product to Solidify it in the Mold

The molten product must be solidified before it can be removed from the mold. This is accomplished by flowing cold water through channels machined into the mold. Computerized design of the cooling channels, and refrigeration of the cooling water to speed the cooling process, are major considerations here.

Since heat transfer through organic polymers is slow, the design favors thin-wall products, generally under $\frac{1}{4}$ in. Polymers which can crystallize rapidly offer promise of faster, more economical molding cycles. Polymers which crystallize slowly, and amorphous polymers which stiffen gradually, often require longer molding cycles.

Opening the Mold and Ejecting the Solid Product

When the plastic product has cooled sufficiently to be solid and retain its shape, the moving half of the mold is opened automatically. With luck, some products will eject from the mold spontaneously; this depends on design of the mold and the product, low polarity of the plastics versus the high polarity of the metal mold, and the use of external lubricants to help release the product from the mold. In most cases, however, the moving half of the mold is fitted with "knockout pins." When the mold is opened, the knockout pins automatically project into the mold cavity, and press gently on the product to force it away from the mold surface. Then, as the mold closes for its next cycle, the knockout pins automatically retract again.

Variations and Details

Drying

Although most plastics are quite resistant to water, the powder or pellets may still absorb or adsorb small amounts of moisture during storage and handling. In the heat of molding, this moisture evaporates, causing microscopic voids ("blushing") or macroscopic voids (bubbles), or even hydrolysis to lower molecular weight and less desirable properties. Thus many or most plastics require pre-drying before molding. They may be dried on shallow trays in a circulating-air oven, or by passing hot dry air through the hopper as they are fed into the extruder. Or they may be dried by opening a vent midway down the extruder screw to allow the steam to escape.

"Two-Stage" Injection Molding Machine

Instead of using a reciprocating screw to melt the plastic and to push the melt into the mold, the extruder screw may feed the molten plastic into a separate reservoir, and then a separate plunger can feed the melt into the mold.

Electric Drive

This may replace the hydraulic drive in some injection molding machines. There has been very active discussion of this procedure in recent years.

Multicavity Molds

While the mold can be machined with a single cavity to produce a single product, in most cases it is more efficient to machine multiple cavities into the mold and fill them all with a single shot of molten material. This requires a system of runners (tunnels) to distribute the melt to all the cavities, which in turn requires much more sophisticated engineering design to balance them all equally.

Hot Runners

When the molten plastic is pumped into the water-cooled mold, the cooling system solidifies both the plastic product in the mold cavities and also the plastic material in the runners. Later the solid runners must be separated, reground, and reused. This is an extra burden on the process. An alternative is to avoid cooling the runners, and actually keep them hot, so that the molten polymer in them remains ready for the next shot into the mold.

Thermosetting Plastics

While injection molding is used primarily for thermoplastic polymers, it is so efficient and economical that processors working with thermosetting plastics may suffer from higher processing costs. This has led some of them to adapt the injection molding process to thermosets. This may involve some modification of the reactivity of the thermosetting plastics and/or the injection molding machinery, gentle heating to melt the reactive materials without curing them prematurely, and use of warm or hot molds to finish the curing process quickly and thus speed the molding cycle.

Instant Inplant Recycling

Injection molding produces a certain amount of scrap: runners, gates, flash, and rejects of imperfect products. In theory, thermoplastic polymers should be perfectly recyclable. In practice, inplant scrap can be kept clean, reground, and recycled. While 100 percent recycle is theoretically conceivable, most processors simply determine how high a percent of recycle they can blend with their next batch of virgin material, without any harm to their process or product. In most cases they reuse all of their scrap internally, without any contribution to solid waste. This is simply both good economics and good environmental practice.

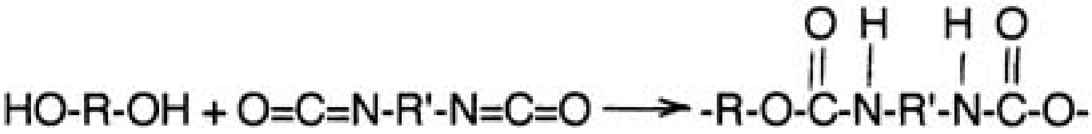
19.16 Reaction Injection Molding

Overview

Reaction injection molding (RIM) is a fast, low-pressure, low-temperature, low-cost process for one-step conversion of reactive liquids into large finished solid plastic products. Liquid polyol and liquid diisocyanate are mixed by impingement, pumped instantly to fill a large mold cavity, and polymerize/cure rapidly to form a thermoset polyurethane product. The cured polymer may be a stiffly flexible product such as automotive bumper covers, front ends, and trim; or a rigid foamed product such as furniture and housings (cabinets) for computers, business machines, TV, and radio.

Ingredients

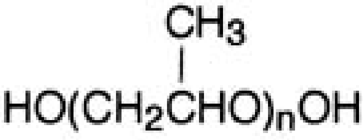
The basic reaction to form polyurethanes is



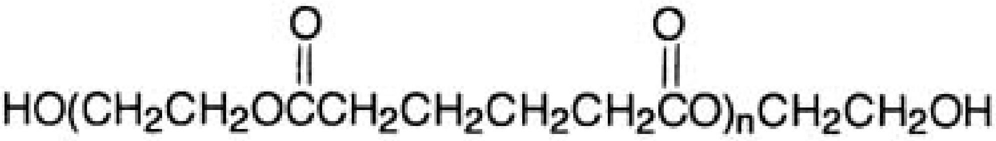
$$\text{HO-R-OH} + \text{O=C=N-R'-N=C=O} \longrightarrow \text{-R-O-}\overset{\overset{\displaystyle O}{\|}}{C}\text{-}\overset{\overset{\displaystyle H}{|}}{N}\text{-R'-}\overset{\overset{\displaystyle H}{|}}{N}\text{-}\overset{\overset{\displaystyle O}{\|}}{C}\text{-O-}$$

Polyol Diisocyanate Polyurethane

The *polyol* may be a polyether or a polyester:



Polyether



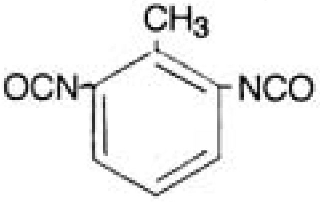
Polyester

Long polyols (high *n*) give flexible polyurethanes; short polyols give stiff, rigid, and/or crystalline polyurethanes. Branched polyols, with three or more OH groups, give cross-linked thermoset polyurethanes; short multi-branched polyols give rigid products. Replacing part or all of the polyol by a *polyamine*

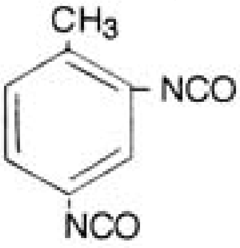
$H_2N—R—NH_2$

gives faster reaction and therefore shorter cure cycles; and also contributes increased hydrogen-bonding, which produces higher rigidity and strength.

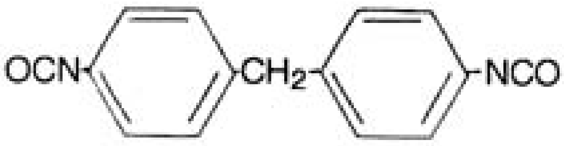
The *diisocyanate* is generally toluene diisocyanate (TDI) or methylene diphenyl isocyanate (MDI) or oligomers based on them.



2,6-TDI



2,4-TDI



MDI

Foaming agents are either volatile liquids or a trace of water. For flexible products, only a trace of foaming is needed to optimize the product. For rigid products, a higher degree of foaming is used to produce light-weight products.

Other ingredients include catalysts to control/balance the polymerization/cure reactions, surfactants for foam uniformity, fillers for stiffening, pigments for coloration, flame-retardants where needed, and especially short glass reinforcing fibers to increase strength and dimensional stability in Reinforced RIM (RRIM).

Equipment and Process

Raw material holding tanks are warmed to 3038°C (85100°F) to keep them at optimum viscosity. The raw materials are measured and pumped by piston pumps, to deliver exactly equivalent amounts of the two reactants, bring them up to 1421 MPa (20003000 psi), and shoot them into the mixing head. Here they are mixed by high-speed impingement. The mixed liquid system is very reactive, so it must be pumped into the mold cavity to fill it as rapidly as possible. Molding itself is a low-pressure process, rarely more than 0.7 MPa (100 psi), so the mold can be rather light construction; but since molds and products are generally large, the total force needed to clamp the mold is still considerable. The reaction is exothermic, and the temperature rise could over-cure the product, so mold cooling is required. The entire process is rapid, and can be accelerated further by use of polyamines and catalysts to speed the polymerization/cure reactions. It has proved particularly attractive for mass production of large polyurethane parts.

Other Polymers

Once RIM had been developed for polyurethane molding, the industry began asking whether it could be applied to other fast polymerization/cure reactions. It was quite readily applied to nylon 6 monomer casting, but the cost of the caprolactam monomer appeared non-competitive. It was also suggested for epoxy cure and possibly other fast reactions, but none of these have yet been developed commercially.

19.17 Structural Foam

Definition

When a solid plastic is expanded to moderately lower its density, it is called a "structural foam." Bone and wood are natural structural foams, which benefit from this moderate degree of expansion. Most plastic materials can similarly be expanded during melt processing, and also gain many benefits as a result.

General Description

A blowing agent is added to the plastic material, to compress gas into the melt during processing. It can be compressed nitrogen gas, which is inexpensive but difficult to disperse; or it can be a "chemical blowing agent," usually an organic azo compound RN=NR, which gives much more uniform dispersion in the melt, and reacts to produce compressed gases in the hot melt.

In extrusion, when the hot melt exits from the die, the compressed gas expands and foams the extrudate. In injection molding, the hot melt is injected into the mold, but the quantity is reduced ("short shot") so that there is not enough melt to fill the mold; the compressed gas expands, and tie melt foams and fills the mold. Either way, the foamed melt cools and solidifies, producing a product with a solid skin and a somewhat expanded internal foam structure. It looks like a solid product, but it is lighter in weight and offers a number of advantages.

Degree of Expansion

Plastics with solid densities of 0.91.2 are generally expanded 1325 percent, giving foamed densities of 0.750.9. At this modest degree of expansion, they look and perform like solid products.

Benefits

Use of 1325 percent less material is an obvious economic advantage. Low pressure in the mold (1.43.4 MPa, 200500 psi) permits construction of a lighter mold and operation at lower pressure, both of which provide further economic advantage. A low degree of foaming is enough to expand the molding against the walls of the mold, eliminating any accidental "sink marks." Higher degree of foaming increases rigidity 34 × over a solid part of equal weight. Low-pressure molding also eliminates molded-in stresses, which would weaken the product, and could also cause post-molding warpage. This in turn permits design and production of more complex parts in a single molding step, eliminating need for later machining and assembling.

Problems in Structural Foaming

Foamed products tend to be thicker, and contain gas bubbles. Both of these factors reduce heat transfer, taking longer to cool and solidify the product, and thus lengthening the manufacturing cycle.

Tiny foam bubbles in the skin of the product tend to form swirl patterns which are unsightly in consumer products. Several modified injection molding processes offer promise for reducing or eliminating this problem: (1) The *Farrel/USM Process* injects the melt into an expandable mold at high pressure, then expands the mold to lower the pressure and permit foaming. (2) The *Coinjection Process* uses two reciprocating screws to feed the mold; the first feeds feeds a non-foamable short shot which will form the solid skin, and then the second feeds a foamable second short shot which expands, pushing the skin ahead of it and filling the interior with foam. (3) The *Gas Counterpressure Process* seals the mold, compresses nitrogen gas into it, then injects a short shot of melt containing blowing agent, and then uses gradual controlled venting to release the nitrogen gas and allow the foamable melt to expand, forming a solid skin and a foamed core.

19.18 Low-Density Foams

When plastics are foamed to low densities, containing more air than polymer, they acquire unique new properties and applications. Major uses are in crash padding and thermal insulation. Closed-cell foams are outstanding for flotation, rigidity, and insulation; while open-cell foams are outstanding for softness, resilience, and comfort.

While foam production and properties may all belong to one unified theoretical basis, commercial practice is different for each of the major polymer families; so it is best to review them individually.

Polyurethane

The largest quantity of low-density foam is polyurethane, made by mixing liquid polyols with liquid diisocyanates, pouring the mixed liquid and allowing it to foam, polymerize, and cure rapidly into its finished form. The auxiliary ingredientsfoaming agent, tertiary amine and organotin catalysts, surfactant, and optionally flame-retardantare mixed with the polyol before it is mixed with the diisocyanate. The two liquid streams are metered by piston or gear pumps, and fed at high pressure into an impingement chamber where they mix by turbulence. This mixture is very reactive, so it is poured rapidly into the desired form or location, and allowed to polymerize, foam, and cure into the finished product. The largest amount is open-cell soft flexible foam; a smaller amount is closed-cell rigid foam for thermal insulation.

Flexible foam is made from long-chain diols with a small amount of triols for cross-linking to give strength and resilience. Foaming is produced by reaction between measured amounts of isocyanate and water to liberate carbon dioxide. Molded products are made by pouring the reactive liquid mixture into a mold cavity at 50°C. They foam and cure in 210 min, after which they are temporarily crushed or vacuum-shocked to open the cell walls and insure softness. Then they are allowed to condition for several hours to finish the cure. The major products are auto seating and headrests, and furniture cushions.

Large slabstock is made by pouring the reactive liquid mixture into a moving paper form, up to 8 × 5 × 50 feet long, which takes up to 5 min. Polymerization, foam rise, and cure are exothermic, and the interior of the slab can reach 140170°C. The rough surface is trimmed off, and the slab allowed to condition 1224 hr to finish the cure. The finished slab is sliced to the desired thickness for furniture, mattresses, public transportation seating, textile backing (sportswear), carpet backing, and packaging.

Rigid foam is made from short-chain polyols with a typical average functionality of four hydroxyls per molecule, to give high cross-linking for rigidity. Foaming is produced by volatile organic solvents, which are boiled by the exothermic polymerization/cure reaction. Sheet, slab, and block are made by pouring the reactive liquid mixture onto a moving paper form, or directly onto facing panels for laminated structures. Pour-in-Place technique is used to fill the wall cavities of refrigerators and freezers, holding them at about 40°C to control the reaction. Spray Coating insulation is applied to buildings, tanks, and pipes in the field, using a highly-catalyzed mixture that will foam and gel in less than 10 sec, so that it will stay in place without running down; the coating can be built up to 2 in. thick in a single pass, more by repeated passes.

Rigid polyurethane foam is used primarily as thermal insulation for buildings, trucks, rail cars, shipping containers, tanks, pipelines, cold-storage warehouses, and frozen food display cases.

Polystyrene

Polystyrene is foamed by swelling with pentane, heating to soften the polymer and vaporize the pentane, and allowing it to expand at atmospheric pressure. Extrusion thus produces foamed sheet and board stock, which are used mainly as thermal insulation in commercial refrigerators and freezers, and also in food packaging, roof and wall insulation, and pipe insulation.

Moldable beads are produced by suspension polymerization, swelling with pentane, warming to soften and "pre-expand" the beads, pouring them into a mold, and steaming to expand them fully, soften them, and fuse them together into a finished product. This produces drinking cups, molded packaging, board stock, and display and novelty products.

Loose fill for packaging is produced by extruding and chopping polystyrene into various shapes, swelling with pentane, and steaming to expand them into "peanuts."

Polyvinyl Chloride

Flexible PVC foams are generally laminated to layers of cloth and/or solid vinyl, and used for leatherlike clothing and luggage, upholstery in autos and furniture, and resilient flooring. They are made by polymerizing vinyl chloride in emulsion, spray-drying to a fine powder, mixing into liquid plasticizer at room temperature to form a viscous paste, compounding with barium/zinc stabilizer/catalyst and azodicarbonamide foaming agent, roller-coating onto a moving belt of textile or paper or solid vinyl film, oven-heating to dissolve the resin in the plasticizer and activate the foaming agent, and cooling to room temperature to gel the plasticized vinyl and stabilize the flexible foam. An alternative for large-scale production is to compound general-purpose PVC with plasticizer, stabilizer/catalyst, and foaming agent, calender and/or laminate sheet below the activation temperature of the azodicarbonamide, and then pass through a hotter oven to foam the sheet.

Polyethylene

Polyethylene foam sheet is used mainly in package cushioning, and also in roof insulation; and extruded profile is also used as construction sealants and pipe insulation. It is made mainly by tandem extrusion. The first extruder melts the polyethylene and blends in liquid volatile organic foaming agent. The second extruder cools it to the optimum foaming temperature, and pumps it through a die into room-temperature air, where it expands into foamed sheet.

For higher performance, cross-linked foamed sheet can be produced by radiation or chemical cross-linking. For radiation cross-linking, polyethylene is melted in an extruder and mixed with powdered chemical foaming agent, extruded into sheet, cross-linked by electron beam radiation, and foamed in an oven. For chemical cross-linking, polyethylene is melted in an extruder and mixed with peroxide cross-linking agent and chemical foaming agent, extruded into sheet, and passed through a two-stage oven for cross-linking and foaming.

Molded foam for package cushioning, flotation devices, and sports equipment is made from foamed pellets. Polyethylene pellets are foamed with volatile organic liquid and cross-linked by peroxide or electron beam. The foamed pellets are poured into a mold, the mold is closed, and the pellets are softened and expanded further to fuse into a molded foam product.

Blow Molding

Blow molding uses compressed air to blow and expand a hot plastic tube ("rubbery melt") in a female mold cavity, until the plastic conforms to the walls of the mold. This has developed into a major way to produce plastic bottles, and also more specialized hollow shapes such as fuel tanks, seat backs, tricycles, surfboards, and so on. The leading material is HDPE for milk, water, and household chemicals. Second largest is PET for carbonated beverages. A number of other thermoplastics are blow molded in smaller amounts for more specialized uses.

There are two ways to produce the plastic tube ("parison"): injection molding and extrusion. For injection blow molding, a test-tube shape is first injection molded, then transferred into a bottle mold and blown. For extrusion blow molding, the extruder produces the tube continuously, and a rotating or alternating series of bottle molds take turns clamping around the tube and blowing it. Injection blow molding is used for bottles up to 500 ml in size, while extrusion blow molding is used for bottles 250 ml and larger.

Stretch blow molding holds the parison above its glass transition temperature (Tg) and stretch-orients it to increase modulus, strength, impact resistance, transparency, and impermeability. This is most important for PET, and is also used for PVC, polypropylene, and polyacrylonitrile.

Multi-layer blow molding uses a parison containing two or more polymers in concentric layers, and produces a multi-layer laminated bottle or fuel tank. In general this can combine the best properties of each layer. Sandwich structure food packaging, with a mid-layer of ethylene/vinyl alcohol copolymer (EVOH), gains impermeability to oxygen. Sandwich structure fuel tanks of HDPE/nylon/HDPE gain impermeability to gasoline.

To speed the cooling portion of the blow molding cycle, chilled air can help. For more extreme action, liquid carbon dioxide or nitrogen can give even faster cooling.

Surface fluorination has been claimed to make polyethylene less permeable to gasoline and nonpolar solvents in general. Fluorine gas can be used to blow the container, which treats only the inside surface. Or the finished container can be fluorinated both inside and outside in a single post-treatment. This produces a fluorinated layer 2040 nm thick.

19.19 Thermoforming

Thermoplastic sheet can be softened by warming, placed in or on a mold, pressed gently but swiftly to conform against the walls of the mold, cooled to solidify, and trimmed to separate the product from the surrounding unused sheet.

Sheet can be produced by extrusion or calendering. For use in thermoforming, the sheet must be extremely uniform. For improved properties, the sheet may be biaxially stretch-oriented up to 300 percent. For large automatic production runs, it is fed continuously off a roll; for short runs, large, and/or specialty jobs, it may be cut into individual sheets and fed one at a time.

The sheet must be preheated to make it soft and pliable ("rubbery melt"). For fastest processing and best final uniformity and properties, it should be as hot as possible, without losing melt strength or beginning thermal degradation. There are three methods of preheating: (1) Convection oven is slow but very uniform. (2) Conduction heating is done with electrically-heated Teflon-coated aluminum plates. (3) Radiant heating is most efficient, especially when the infrared wavelengths are chosen to match the maximum absorption frequencies of the plastic material.

There are three ways to press the warm sheet gently and swiftly against the surface of the mold: (1) Vacuum forming is the most popular, pulling the soft sheet against the mold surface, with atmospheric pressure applying the driving force. The vacuum must be at least 25 in. Hg to give the fastest possible cycle. (2) Pressure forming (compressed air), at pressures of 140860 kPa (20125 psi), is faster and gives better final properties. (3) Mechanical force (tensile, flexural, or compressive) can also drive the warm sheet to conform to the mold surface. In some cases, two or all three of these methods may be combined for optimum performance.

The best molds are made of aluminum with water-cooling channels cut into them. Tooling costs are low, and heat transfer is high. The formed plastic sheet must be held against the mold surface until it is cooled to the solid state. This is often judged by the heat deflection temperature at 455 kPa (66 psi).

Whether the process uses continuous sheet or individual sheets, the product is formed from the center of the sheet, and the edges must be trimmed off to give the final product. Die cutting knives are probably most common, but a variety of other mechanical and thermal techniques are also in use. The trim may be 1070 percent of the original sheet. It is reground, blended up to 50 percent with virgin resin, and recycled directly into the same process. Considering the variability of recycle and blending, this requires careful control to maintain virgin quality.

Overall, thermoforming is a very useful method of test low-cost production for a great variety of plastic products, from packaging to building to automotive parts.

19.20 Rotational Molding ("Rotomolding")

Hollow products can be made by placing powdered plastic in a closed mold, tumbling and heating it until the plastic has coated the walls of the mold, cooling it to solidify the product, and opening the mold to remove the product.

Polyethylene is by far the most popular material. Others used occasionally include vinyl plastisol (liquid rather than powder), nylon, polypropylene, and polycarbonate. The powder is usually ground to 35 mesh.

The most popular machine is a carousel design with 34 arms operating independently, providing separate stations for loading, heating, cooling, and unloading. Molds are generally made of aluminum, two-piece plus clamps.

Slow rotation gives the most uniform products. The pattern of biaxial rotation must be determined by trial and error. Heating is best done in a gas-fired oven with a fan for circulation.

Cooling is a two-stage process, first a fan and then a cold water spray or mist.

Advantages are low capital investment, design freedom, strain-free products, and little or no post-molding secondary operations. For initial costing, multiply the cost of the raw material by 5.

Limitations: blow molding is faster and uses a wider variety of plastics.

Typical products are tanks (2085,000 L, 522,500 gal), containers for packaging and materials handling, portable outhouses, battery cases, light globes, vacuum cleaners, garbage cans, surfboards, toys, traffic barricades, display cases, and ducting.

19.21 Powder Coating

A solid plastic may be ground to a solid powder and then used to apply a plastic coating to a metal product, either for decorative reasons (color) or for functional performance such as insulation or corrosion resistance. Plastics most often used in this way include nylon, vinyl, acrylic, polyethylene, polypropylene, and epoxy. Coating without solvents is beneficial both economically and environmentally, and 100 percent utilization of material eliminates waste. After the powder is bonded to the metal surface, it is often reheated to flow into a more uniform coating and, in the case of thermosetting resins, to complete the cure reaction.

There are three techniques for applying the powder coating to the surface of the metal product: fluid bed, electrostatic fluid bed, and electrostatic spray.

Fluid Bed Coating

The equipment is simply a horizontal box with a finely-porous shelf near the bottom. Powdered plastic, ground to 20200 µm, is poured above the shelf. Compressed air is fed in below the shelf, percolating up through the pores, and percolating the powder so that it rises and flows much like a liquid. Sometimes the box is also vibrated to produce greater uniformity.

The metal product to be coated is preheated to a temperature which will melt the powder. Then the product is dipped into the fluid bed. The powder particles melt and flow onto the metal surface. Coatings up to 2.5 mm (0.1 in.) thick can be applied in a single dip.

Typical products are electric motors, electronics, transformers, valves, pumps, refinery equipment, and appliances.

Electrostatic Fluid Bed

In a modification of the fluid bed technique, the product to be coated, instead of being preheated, is passed over the bed, and the powder is attracted to it by a static charge. Then it is passed through a heating oven to fuse it into a finished coating. This process can be used for either discrete or continuous products.

Electrostatic Spray

Conventional electrostatic spray coating pumps a liquid coating formulation through a spray gun, which puts an electrostatic charge on the liquid, and sprays charged droplets toward a grounded metal product. The droplets are attracted to the product, where they discharge and adhere to the metal surface.

Plastic powders can be applied in a similar way. The solid plastic is ground to a 30200 µm powder, fluidized, and conveyed by compressed air through a spray gun which uses high voltage to apply an electrostatic charge to the particles, and then sprays them at the grounded metal product. Spraying is normally done in a hood, to protect the worker and to collect and recycle the excess powder. The charged powder is attracted to the metal product, attaching a coating 5075 µm thick on a cold product, or up to 250 µm on a preheated product. A post-fusion step melts and flows the powder into a uniform coating and, in the case of thermosetting resins, also completes the cure reaction.

19.22 Calendering

Basically, a viscous bank of excess thermoplastic or rubber is pressure-formed between a pair of parallel co-rotating rolls to form a thin "film" or a thicker "sheet." Most calenders consist of four rolls in L- or Z-shape, plus additional feed rolls and post-calender laminating, patterning, and/or wind-up rolls.

PVC is the most commonly calendered plastic, especially in plasticized flexible form. Other materials commonly calendered are rubber, ABS, polyurethane, and thermoplastic elastomers.

The plastic material delivered to the calender must be molten, homogeneous in composition and temperature, and at optimum viscosity for the calendering operation. For rigid vinyl, optimum temperature may be 180190°C; for flexible vinyl, 1020°C lower.

Nip pressures between the rolls are typically 1601050 kN/m (9006000 lb/in.) of roll face.

A surface pattern can be applied by the calender roll or by a post-embossing operation. Feeding hot sheet to a cold embossing roll works well. The pattern is measured by a profilometer.

Calendering can also be used to apply a plastic surface to a substrate web such as metal, cloth, or paper.

Capital investment for a calender line may be about $5 million. Operating cost may be about $500 per hr. Calendering is a very high-speed method of producing plastic film and sheet, but high-speed economy must be balanced against quality of the finished product.

19.23 Vinyl Plastisol Processing

Plastisol is a specialized technique for making flexible vinyl products. It is also sometimes called "paste" or "dispersion" technology. Basically, powdered PVC is stirred into liquid plasticizer at room temperature to form a viscous liquid or spreadable paste. This can be poured or spread into the shape of the final product, heated to fuse, and cooled to gel into the final solid product. Simple pouring or gentle spreading require no heavy equipment, and do not damage delicate substrates such as paper and cloth. Consequently, this accounts for about 10 percent of all vinyl production.

About 90 percent of PVC is made by suspension polymerization and processed as viscous melt at high pressure in heavy-duty equipment. Some 10 percent of PVC is made by emulsion polymerization and spray-drying, for use in plastisol processing. The resin is stirred into liquid plasticizer at room temperature, along with stabilizer and other optional additives, to form a viscous liquid or easily spreadable paste. This is poured or spread into the form of the desired final product. It is then heated, passing through several phases. At first the resin particles absorb the liquid plasticizer, swell, and form a gel. On further heating, the resin melts and dissolves in the hot plasticizer, forming a hot solution. On cooling, the resin crystallizes somewhat, turning the hot solution into a hot gel. On cooling to room temperature, the gel solidifies to a firm rubbery plastic.

If the plastisol is too viscous for pouring at room temperature, it may be thinned with volatile solvent; on heating, the solvent evaporates, leaving a firmer final product. This is called "organosol" technology. Alternatively, a polymerizable plasticizer may be added to thin the plastisol; on heating, it polymerizes to a solid ingredient, giving a firmer final product. This is sometimes called a "rigisol."

If molds are required to shape the plastisol into a finished product, they are generally light-weight low-cost molds of aluminum, electroformed nickel, or ceramic. Oven heating is slow but common. The major handicap is slow heat transfer, so overall production cycles may take 420 min. Several different methods are described below for converting plastisols into finished products.

Dip Coating

A product is preheated and dipped into the plastisol. The plastisol gels onto the surface of the product. It is withdrawn from the liquid, allowed to drain off excess liquid, and then placed in a 191204°C oven to heat it in a few seconds to about 177°C and fuse the gel to a homogeneous solution. This is then cooled in air or water to solidify it to the final rubbery coating. Typical products are tool handles, kitchen implements, and electrical insulation.

Dip Molding

Instead of a heated product, a heated male mold is dip-coated in the same way. In this case, the final flexible vinyl product is stripped off of the mold. A typical product is medical gloves.

Slush Molding

This is the reverse of dip molding. A female mold is used to give any desired surface finish on the product. The mold is preheated, plastisol is pour into it, and gels onto the surface of the mold. The excess liquid plastisol is poured out, and the internally- coated mold is heated to gel, then cooled to solidify the rubbery product, and the product is stripped out of the mold. Typical products are arm rests, head rests, road safety cones, anatomical models, dolls, toys, and auto parts.

HotMelt Molding

Hot molten plastisol is injected into a cooled mold to make products like fishing baits and novelties.

Rotational Casting

Like rotomolding, the plastisol is poured into a cold mold, which is rotated and heated to gel the plastisol onto the walls of the mold. It is then heated to fuse, cooled to solidify, opened, and the product removed. Typical products are volley balls, basketballs, dolls, and auto parts.

Open Molding

Plastisol is poured into an open mold, heated to gel and fuse, cooled to solidify, and stripped out from the mold. Typical products are auto air filters, oil filters, truck flaps, and place mats.

Closed Molding

Plastisol is filled into a closed mold, heated to gel and fuse, then cooled to solidify. A typical product is switch mats for automatic door openers.

Spray Coating

Liquid plastisol can be spray coated up to 1.25 mm (50 mils) in a single pass. The plastisol is formulated to be liquid enough to spray, then firm enough to hold on a vertical surface without running down. It is heated to gel and fuse, then cooled to form the final rubbery coating. A typical product is tank linings.

Continuous Coating

Doctor blade or roll coating applies plastisol continuously to a moving web of metal or cloth, or two successive coatings apply a solid vinyl skin and a foamable vinyl core to a fabric backing. Typical products are house siding, conveyor belting, and resilient flooring.

Silk-Screen Inks

Plastisol coatings can be applied a few mils thick on cloth to produce T-shirts and athletic uniforms.

Organosol Coatings

The low viscosity of organosol formulations permits coatings under 10 mils thick, particularly for chemical resistance.

19.24 Liquid Casting Processes

Reactive liquids may be mixed and poured to form a solid plastic product. They may be two co-reactive monomers or prepolymers, or one monomer/prepolymer plus a catalyst or curing agent. They are primarily thermosetting plastics and elastomers. Rigid plastics are primarily epoxy, polyester, and acrylic. Elastomers are generally room-temperature-vulcanizing (RTV) polymers: polysulfide, polyurethane, and silicone. They are used mostly to provide electrical insulation and mechanical and environmental protection for delicate electrical/electronic products such as resistors, coils, solenoids, capacitors, transformers, printed circuit boards, opti-electronics, and light-emitting diodes. There are a number of such liquid casting processes, generally carried out at room temperature with or without a later heating step for complete cure.

Casting

The reactive liquid mixture is poured into a mold, and allowed to polymerize and cure. This is used to make solid tires for industrial vehicles, and to make decorative items such as simulated wood trim, furniture, picture frames, and lamp bases.

Potting

Delicate electronic devices are placed inside an empty shell, and the space is filled by pouring the reactive liquid mixture into it and curing it.

Encapsulation/Embedment

After potting is completed, the outer shell is removed, and the solid plastic is the container.

Conformal Coating

Using a thixotropic reactive liquid, the product is dipped into the liquid and removed with a coating of the liquid on it. This is then heat-cured to solidify the coating and make it permanent.

Impregnation

A porous product can be dipped into a fluid reactive liquid, which soaks into the pores and fills them completely. The liquid is then cured to leave the product completely filled with solid polymer in the pores.

Syntactic Foam

Hollow glass, ceramic, or plastic spheres are dispersed in the reactive liquid system before it is cast. When the liquid is polymerized and cured, the hollow spheres make it a unicellular foam. The air "bubbles" in the cells make it low-density, low dielectric constant and loss, and very resistant to compressive forces such as hydrostatic head in deep-sea equipment.

Aside from simply casting at atmospheric pressure, vacuum is often used to remove air bubbles and volatiles which would give an imperfect casting, or to help promote porous impregnation. Small specialty runs can be done manually, with no capital investment but high labor cost; larger production runs can be automated with only modest capital investment. Since many of these liquid systems may be volatile and unpleasant to work with, worker protection often includes ventilation and/or protective clothing for eyes, skin, and lungs.

19.25 Compression Molding and Transfer Molding

Thermosetting plastics and rubber are heated to soften ("plasticate") them, and then pressed at 1428 MPa (20004000 psi) for about a minute to cross-link ("cure") them. Then the mold is opened and the finished product is removed.

Outstanding Properties

Compared to engineering thermoplastics, thermosets can offer equal or superior properties at lower cost. Major advantages include rigidity, creep resistance, dimensional stability, impact strength, heat resistance, and chemical resistance.

Typical Applications

Common compressionmolded thermoset products include electrical equipment, appliance handles and knobs, dinnerware, distributor caps, under-the-hood parts in general, automatic transmissions, brakes, and pumps.

Compression Molding

This press has two horizontal platens facing each other. The upper one is generally stationary; the lower one moves vertically, driven by mechanical, hydraulic, or pneumatic power. The upper mold half is fastened to the upper platen, the lower mold half to the lower platen. Most molds are electrically heated. In semi-automatic operation, the operator puts the granular molding powder in the lower half of the mold, and presses a button which closes the press, holds it till cure is complete, and opens the press again; then the operator removes the molding from the mold, and repeats the cycle. In fully automatic molding, the entire cycle is automatic and does not require an operator.

Transfer Molding

This preheats the molding powder in the upper part of the mold. Then a plunger pushes the fluid material down into the (lower) mold cavity and holds it there until cure is complete. This gives more fluid flow than compression molding, and avoids damage to delicate molds or molded parts.

Additional Considerations

Compared to injection molding of thermoplastics, compression molding is less capital intensive, more labor-intensive, and takes a longer molding cycle. Transfer molding can equal injection molding cycle times. Compression and transfer molding do less damage to reinforcing fibers, so they can give more impact-resistant products.

The molding cycle can be shortened by preforming and preheating. The granular molding powder is cold-pressed into a pill of the desired weight, and then dielectric preheating for 1020 sec brings it to molding temperature before it is inserted into the mold. This gives better flow and faster cure.

"Venting" and "breathing" must often be included in the molding cycle. Venting allows trapped air to escape through small grooves as the molding powder is compressed and flows into a solid part. Breathing opens the mold about 3 mm ($\frac{1}{8}$ in.) for a second or two, early in the molding cycle, to allow escape of water and other volatiles from the condensation cure reaction. This permits production of bubble-free parts.

Ejector pins help to separate the finished molding from the mold cavity. These are driven mechanically during the mold-opening stage of the cycle.

Post-cure ovens may be used to help complete the cure cycle, especially to control gradual cooling of the molded part.

Deflashing is often required to remove excess trim ("flash") from the edges of the molded product. This may be done by tumbling in a rotating drum, or by an air-blast grit.

19.26 Reinforced Plastics Processing

Adding short-fiber reinforcements to thermoplastics can produce major increases in modulus, strength, and dimensional stability; it makes melt processing more difficult, but does not change it qualitatively. Adding fibrous reinforcement to thermosetting plastics produces outstanding modulus, strength, impact strength, dimensional stability, and heat and chemical resistance; but it requires entirely new methods of processing to convert them into finished products. The polymers most often used are unsaturated polyesters and epoxy resins; for ultra-high performance, polyimides and other specialty resins are also used in small amounts. The reinforcing fibers are primarily glass; for ultra-high performance, aramide, carbon/graphite, and metallic fibers are also used in small amounts.

The polymers are generally polymerized to low-molecular-weight "A-stage" resins, which are still liquids or fusible solids, and potentially very reactive. In pure form they can be stored for months to years. The liquids are easily mixed with reinforcing fibers and other ingredients, and shaped into the form of the ultimate products. Addition of peroxide initiators to polyesters, or of amine or anhydride curing agents to epoxies, liberates their high reactivity to permit fast cure cycles to completely thermoset products. Some processes use heat and/or pressure to complete the shaping and cure cycle, while others are carried out simply at ambient conditions. These variables offer a wide variety of processes which can be used to manufacture reinforced thermoset plastic products.

Matched Die Molding Processes

By analogy with compression molding, thermosetting resins plus fibrous reinforcement can be pressed between matching mold halves, with heat and moderate pressure, and cured directly into finished products. Typical cure cycle is about a minute. Half of all reinforced thermoset products are made in this way. Major applications are in the automotive, electrical/electronics, appliance, and business machine industries.

Bulk Molding Compound (BMC)

Liquid resin, $\frac{1}{4}-\frac{1}{2}$ in. long glass fiber, and simple filler (typically calcium carbonate or clay) are mixed in a dough mixer or cold extruder. Group II metal oxide is added to gel the liquid resin (Table 19.22). This doughy mix ("premix") is placed in a hot mold, pressed to flow and fill the mold cavity, and held until cured.

TABLE 19.22 Typical Mix for a BMC

Resin	22%
Fiber	25%
Filler	53%

Sheet Molding Compound (SMC)

An SMC machine pours liquid resin and 1-in. glass fiber onto a moving belt, passes through calender rolls to make a good sheet, and then through an oven to begin cross-linking and thus gel the resin. The sheet is placed in a hot mold and pressed to shape and cure it. While the short fiber in BMC is more moldable, the longer fiber in SMC gives greater strength and impact resistance.

Cold Press Molding

Occasionally the mix may be placed in a cold mold and pressed at about 345 kPa (50 psi). The cure cycle is considerably longer.

Preform

A metal screen is made in the shape of the final product. Glass fiber is chopped 2 in. long and sprayed uniformly all over the shaped screen, using vacuum on the back side of the screen to assist the process. A small amount of binder, typically 5 percent of polymer in latex form, is sprayed onto the fiber to hold its shape. It is then removed from the screen, placed in the mold, saturated with an equal weight of liquid resin, and the mold is pressed at 1380 kPa (200 psi) and heated until cured, typically 315 min. This early process has been largely replaced by SMC.

Prepreg

Fabric is impregnated with 25 percent of liquid resin and laid in the mold. To insure isotropic properties, or to maximize properties in a specific direction, successive layers of impregnated fabric are carefully oriented in different directions. The mold is closed, pressed, and heated till cured. Products made from such impregnated fabrics have much higher strength than simple random fiber reinforcements.

Resin Transfer Molding (RTM)

Reinforcing fibers are distributed uniformly in the mold and the mold is closed. Liquid resin is injected into the mold until the excess comes out of the vents. The mold is pressed and heated, similarly to preform molding, until cure is complete.

Open Molding

Instead of applying pressure in a closed mold, the mix of liquid resin and reinforcing fiber may be laid into an open mold, and optionally pressed gently at room temperature until cured. To accelerate low-temperature cure, more active catalyst systems are added. Alternatively, the assembly may be UV or oven cured. This requires less capital investment but more skilled labor, so it is useful for prototype and small production runs. It permits unlimited size, so the largest reinforced thermoset products are made in this way, for example, large tanks and whole boat hulls.

Hand Lay-Up (Contact Molding)

A layer of liquid resin is applied to the surface of the mold. A layer of glass fiber mat (low strength) or fabric (high strength) is hand laid over it. Liquid resin is poured over it, and brushed or rolled (squeegeed) into it. The process is repeated to build up the desired thickness of the product. The assembly is allowed to stand until cured.

Spray-Up

Instead of hand lay-up, continuous glass roving and liquid resin are fed into a gun, which chops the glass fiber, mixes it with the resin, and sprays it into the mold. This can be automated for lower labor cost and greater uniformity. Spraying is often followed by hand-rolling to expel air and densify the assembly. Then it is allowed to stand until cured. Products are similar to those from simple hand lay-up.

Vacuum-Bag Molding

After hand lay-up, the assembly is covered with an air-tight film, typically polyvinyl alcohol, occasionally nylon or other material. Then a vacuum is pulled on the underside of the film, to let atmospheric pressure squeeze out air and excess resin. Use of a hand-held paddle may help. The assembly is allowed to stand until cured. Products are void-free, and quality is better than simple hand lay-up.

Pressure-Bag Molding

This is similar to vacuum-bag molding, but 345 kPa (50 psi) air pressure is applied to the outside of the cover film. A rubber bag may be used to facilitate the process.

Autoclave Molding

This is similar to pressure-bag molding, but it is carried out in an autoclave to apply the pressure to the outside of the cover film.

Rubber Plug Molding

Liquid silicone rubber is cast into a female mold cavity and cured. The plug is removed, and layers of mat or fabric are built onto the plug. This assembly is then inserted into the mold cavity, pressed, and heated until cured.

Special Processes

Filament-Winding

This requires a mandrel to shape the desired finished product. Continuous filament or woven tape is fed through a liquid resin bath to impregnate it, and then wound onto the mandrel in a calculated pattern to optimize the final properties (Table 19.23). The assembly is oven-cured. A collapsible mandrel can then be removed from the plastic product; or the mandrel can be left as a part of the finished product. These are the strongest plastic products ever made. Typical products are pipes, tanks, and pressure bottles. Other suggested products include rocket motor cases, railroad hopper cars, automotive springs, drive shafts, ship hulls, housing modules, helicopter rotor blades, and helicopter tail sections.

TABLE 19.23 Typical Properties of Filament-Wound Plastics

Glass fiber	5085%
Tensile modulus	3448 GPa (57 × 106 psi)
Tensile strength	5501700 MPa (80250 × 103 psi)
Flexural strength	7001400 Mpa (100200 × 103 psi)

Pultrusion

Filaments, woven tapes, or fabrics are fed continuously through a liquid resin bath, through a shaping die, and through a curing oven, then cut to any desired length. This makes continuous products of any desired cross-section. Typical products include fishing rods, flag poles, tool handles, ladder rails, tubing, and other electrical, corrosion-resistant, construction, and transportation applications.

Continuous Laminating

Reinforcing fiber and liquid resin are deposited on a moving belt, densified between squeeze rolls, passed through a curing oven, and cut to length. This process is used for mass production of glazing, panelling, and roofing.

Centrifugal Casting

Reinforcing fibers are distributed inside a circular mold. The mold is rotated, and liquid resin is distributed inside it to impregnate the fibers. Rotation of the assembly is continued inside an oven until it is cured. This process is used for making pipes, tanks, and hoops.

Foam Reservoir Molding

Flexible open-cell polyurethane foam is impregnated with liquid resin, faced with glass fiber mat, and gently compression molded. This squeezes the liquid resin into the glass fiber surface mat. Heat curing produces a sandwich structure of low density, high flexural and impact strength.

19.27 References for Part I

Sperling, L. H., Introduction to Polymer Science, 3rd ed., p. 2, John Wiley & Sons, New York, 2001.
Rudin, A., The Elements of Polymer Science and Engineering, 2nd ed., pp. 132141, Academic Press, San Diego, CA, 1999.
Rudin, A., The Elements of Polymer Science and Engineering, 2nd ed., p. 150, Academic Press, San Diego, CA, 1999.
Billmeyer, F. W., Textbook of Polymer Science, 2nd ed., p. 16, John Wiley & Sons, New York, 1971.
Stevens, M. P., Polymer Chemistry, 3rd ed., pp. 7074, Oxford University Press, New York, 1999.
Nielsen, L. E., and Landel, R. F., Mechanical Properties of Polymers and Composites, pp. 1823, Marcel Dekker, New York, 1994.
Nielsen, L. E., and Landel, R. F., Mechanical Properties of Polymers and Composites, p. 21, Marcel Dekker, New York, 1994.
Nielsen, L. E., and Landel, R. F., Mechanical Properties of Polymers and Composites, pp. 1718, Marcel Dekker, New York, 1994.
Painter, P. C., and Coleman, M. M., Fundamentals of Polymer Science, 2nd ed., pp. 284290, Technomic Publishing Company, Lancaster, PA, 1997.
Nielsen, L. E., and Landel, R. F., Mechanical Properties of Polymers and Composites, pp. 5051, Marcel Dekker, New York, 1994.
Sperling, L. H., Introduction to Polymer Science, 3rd ed., pp. 230233, John Wiley & Sons, New York, 2001.
Billmeyer, F. W., Textbook of Polymer Science, 2nd ed., John Wiley & Sons, New York, 1971.
Billmeyer, F. W., Textbook of Polymer Science, 2nd ed., pp. 270271, John Wiley & Sons, New York, 1971.
Cauaher, C. E., Jr., Polymer Chemistry, 4th ed., p. 265, Marcel Dekker, New York, 1996.
Steyens, M. P., Polymer Chemistry, 3rd ed., p. 206, Oxford University Press, New York, 1999.
Odian, G., Principles of Polymerization, 2nd ed., p. 182, John Wiley and Sons, New York, 1981.
Steyens, M. P., Polymer Chemistry, 3rd ed., pp. 208209, Oxford University Press, New York, 1999.
Stevens, M. P., Polymer Chemistry, 3rd ed., p. 213, Oxford University Press, New York, 1999.
Billmeyer, F. W., Textbook of Polymer Science, 2nd ed., pp. 317318, John Wiley & Sons, New York, 1971.
Stevens, M. P., Polymer Chemistry, 3rd ed., p. 222, Oxford University Press, New York, 1999.
Stevens, M. P., Polymer Chemistry, 3rd ed., pp. 223225, Oxford University Press, New York, 1999.
Stevens, M. P., Polymer Chemistry, 3rd ed., pp. 234235, Oxford University Press, New York, 1999.
Cauaher, C. E., Jr., Polymer Chemistry, 4th ed., p. 277, Marcel Dekker, New York, 1996.
Stevens, M. P., Polymer Chemistry, 3rd ed., pp. 237245, Oxford University Press, New York, 1999.
Carraher, C. E., Jr., Polymer Chemistry, 4th ed., p. 282, Marcel Dekker, New York, 1996.
Stevens, M. P., Polymer Chemistry, 3rd ed., pp. 246249, Oxford University Press, New York, 1999.
Stevens, M. P., Polymer Chemistry, 3rd ed., p. 173, Oxford University Press, New York, 1999.
Billmeyer, F. W., Textbook of Polymer Science, 2nd ed., pp. 355357, John Wiley & Sons, New York, 1971.
Stevens, M. P., Polymer Chemistry, 3rd ed., p. 174, Oxford University Press, New York, 1999.
Shaffer, K. A., and DeSimone, J. M., Trends Polym. Sci., 3, 146 (1995).
Stevens, M. P., Polymer Chemistry, 3rd ed., p. 174, Oxford University Press, New York, 1999.
Billmeyer, F. W., Textbook of Polymer Science, 2nd ed., pp. 358359, John Wiley & Sons, New York, 1971.
Billmeyer, F. W., Textbook of Polymer Science, 2nd ed., pp. 359361, John Wiley & Sons, New York, 1971.
Stevens, M. P., Polymer Chemistry 3rd ed., p. 175, Oxford University Press, New York, 1999.
Stevens, M. P., Polymer Chemistry, 3rd ed., pp. 194195, Oxford University Press, New York, 1999.
Billmeyer, F. W., Textbook of Polymer Science, 2nd ed., p. 231, John Wiley & Sons, New York, 1971.
Billmeyer, F. W., Textbook of Polymer Science, 2nd ed., p. 227, John Wiley & Sons, New York, 1971.
Stevens, M. P., Polymer Chemistry, 3rd ed., p. 227, Oxford University Press, New York, 1999.
Stevens, M. P., Polymer Chemistry, 3rd ed., pp. 273275, Oxford University Press, New York, 1999.
Sperling, L. H., Introduction to Polymer Science, 3rd ed., pp. 342343, John Wiley & Sons, New York, 2001.
Birley, A. W., Haworth, B., and Batchelor, J., Physics of Plastics, Carl Hanser Verlag, Munich, 1992.
Williams, M. L., Landel, R. F., and Ferry, J. D., J. Am. Chem. Soc., 77, 3701 (1955).
Powell, P. C., Engineering with Polymers, Chapman and Hall, London, 1983.
Birley, A. W., Haworth, B., and Batchelor, J., Physics of Plastics, pp. 283284, Carl Hanser Verlag, Munich, 1992.
Nielsen, L. E., and Landel, R. F., Mechanical Properties of Polymers and Composites, p. 253, Marcel Dekker, New York, 1994.
Nielsen, L. E., and Landel, R. F., Mechanical Properties of Polymers and Composites, pp. 342352, Marcel Dekker, New York, 1994.

19.28 References for Part II

Billmeyer, F. W. Jr., Textbook of Polymer Science, John Wiley & Sons, New York, 1984.
Brydson, J. A., Plastics Materials, Butterworth Scientific, London, 1982.
Craver, C. D., and Carraher, Jr., C. E., Applied Polymer Science, 21st Century, Elsevier, Amsterdam, 2000.
Deanin, R. D., Polymer Structure, Properties, and Applications, Cahners, Boston, 1972.
Goodman, S. H., Handbook of Thermoset Plastics, Noyes Publications, Westwood, NJ, 1998.
International Plastics Selector, Plastics Digest: Thermoplastics and Thermosets, D. A. T. A., Englewood, CO, 1995.
Harper, C. A., Modern Plastics Handbook, McGraw-Hill, New York, 2000.
Lubin, G., Handbook of Composites, Van Nostrand Reinhold, New York, 1982.
Modern Plastics Magazine, Modern Plastics Encyclopedia Handbook, McGraw-Hill, New York, 1994.
Rubin, I. I., Handbook of Plastic Materials and Technology, John Wiley & Sons, New York, 1990.
Sears, J. K., and Darby, J. R., The Technology of Plasticizers, John Wiley & Sons, New York, 1982.

19.29 References for Part III

Harper, C. A., Modern Plastics Handbook, McGraw-Hill, New York, 2000.
Klempner, D., and Frisch, K. C., Handbook of Polymeric Foams and Foam Technology, Hanser, Munich, 1991.
Rubin, I. I., Handbook of Plastic Materials and Technology, John Wiley & Sons, New York, 1990.

20.1 Introduction

Adhesives have been used successfully in a variety of applications for centuries. Today, adhesives are more important than ever in our daily lives, and their usefulness is increasing rapidly. In the past few decades there have been significant advances in materials and in bonding technology. People now routinely trust their fortunes and their lives to adhesively bonded structures and rarely think about it, but the subject of adhesives and adhesive bonding is of great importance to researchers.

The scientific literature on adhesives and adhesion is growing at a rapid pace. There are many texts devoted to adhesion and adhesives.111 New patents on adhesive compositions and processes are granted almost daily, and there are scientific journals devoted to the science and technology of adhesives and their use. A significant portion of the publications dealing with adhesives are concerned with epoxy, urethane, and acrylic structural adhesives, as they are used in a wide variety of commercially important applications. This chapter reviews some of the chemistry of these adhesive types.

Adhesion

Materials are generally defined as adhesives by what they do. Almost any organic polymer and even many inorganic materials can function as adhesives in some situations. However, whether they are organic polymers or inorganic, or perhaps even solders or brazing alloys, all adhesives share common traits in performing their functions:

1. An adhesive, by surface attachment only, transfers and distributes mechanical loads among the components of an assembly.
2. At some time in the course of the bond formation, the adhesive must be liquid or behave as a liquid in order to wet the adherends.
3. An adhesive carries some continuous, low, and often variable load throughout its life.
4. An adhesive must work with the other components of the assembly to provide a durable product that is resistant to degradation by elements of the environment in which it will be used.

We all seem to have some built-in operational definition of "adhesion." If we are given a bonded assembly to test or evaluate, we usually bend, twist, and stretch it to try to make it fail. We typically equate adhesion with resistance to failure: the greater the resistance, the better the adhesion, and, by implication, the better the adhesive. This is not a bad working definition of adhesion and serves adequately in most cases. However, there are instances in which one wants assemblies to come apart fairly easily, such as protective masking or packaging. In these cases one might regard too much resistance to destruction as a mark against the adhesive.

The expectations of the user are extremely important in determining whether an adhesive is "good" or "bad." Adhesives are really judged on the ability of the whole assembly to meet the user's expectations, which will, in turn, depend on the way the assembly is loaded and tested and on what and where are the weakest points of the assembly. Adhesion is not an intrinsic property of any polymer but is rather a property of the whole assembly. In a broad sense, adhesion is really a number we measure or assign to the response of the assembly to some destructive deformation. Thus, adhesion is defined by the test, and the result of the test (the number we assign) is determined by the weak points or flaws in the structure. The acceptability of the bonded assembly is determined by our expectations of how the adhesive will respond to the destructive forces.

Curing

The chemistry of a structural adhesive is designed to do at least two important things. First, the adhesive must at some time pass through a fluid state in order to wet the adherends. Second, the adhesive in its final state in the bond line must be a solid, high molecular weight polymer that is able to carry and transfer mechanical forces. The chemistry must make possible some manufacturing process for the assembly that allows for the liquid state, and there must be some mechanism for passing from the liquid to the solid, load-bearing state. The process of going from the liquid to the solid state is usually termed "curing."

There are three general ways in which adhesives are cured. The adhesive can be applied in a molten state and allowed to cool and solidify in the bond line. In a second method, the adhesive can be applied as a solution or dispersion and the carrier liquid allowed to evaporate, leaving behind the high molecular weight polymer. In the third method the adhesive consists of a low viscosity fluid containing reactive groups that undergo polymerization in the bond line to build the molecular weight sufficiently to carry a load.

The third method is the one most commonly used with epoxy, urethane, and acrylic structural adhesives. The chemistry, methods and mechanisms for accomplishing this polymerization will be covered in more detail in the following sections.

Adhesion Mechanisms

Once the liquid adhesive has been applied to the adherends and intimate contact and wetting have been established, the liquid mass is cured by polymerizing it to a solid, high molecular weight, load-bearing state. It now transfers load among the components of the assembly. But we are interested here in what actually accounts for the adhesion between the polymer and the adherend surface, the ability to transfer load: Three mechanisms have been postulated to explain the adhesion of one material to another: (1) mechanical interlocking, (2) electrostatic attraction, and (3) the formation of chemical bonds across the interface. It is probable that all three mechanisms play some role in any bonding process although often one contributes much more than the others.

Mechanical interlocking probably plays a part in the adhesion of ice to glass. Silver amalgam dental fillings are held in place largely by mechanical interlocking. Although there probably are some exceptions, mechanical interlocking usually is not a major factor in bond formation with structural adhesives.

The other extreme in bonding is the formation of direct chemical links across the interface. These bonds would be expected to be quite strong and durable, but they are difficult to form. It is necessary that there be mutually reactive chemical groups tightly bound on the adherend surface and in the adhesive, and there is evidence that such bonds can be formed under controlled conditions. Silane coupling agents are one example of using specific reactive groups to promote the formation of direct chemical bonds.

By far the dominant adhesion mechanism, particularly in the absence of tightly bound reactive groups, is the electrostatic attraction of the polar groups of the adhesive to polar groups of the adherends and to each other. These are mainly transient dispersion forces (London forces) and forces arising from the interaction of permanent dipoles. These forces provide much of the attraction between the adhesive and the adherend and also provide a significant portion of the cohesive strength of the adhesive polymer. Dispersion forces arise from the interaction of trasient dipoles present in all atoms and molecules, whether permanently charged or not. Dispersion forces act over a very short range and decrease with the sixth or seventh power of distance. The attractive forces between permanent dipoles can be somewhat stronger but still decrease with the sixth power of distance. With such a great dependence on distance, neither force is significant beyond about 0.5 nm. The need to obtain intimate contact between the adhesive and the adherend surface is obvious.

In two articles on the cohesive and adhesive strengths of polymers, H. F. Mark12,13 derives some estimates of what adhesive bond strengths might be achievable with covalent bonds or polar forces across an interface. He concludes that the bonds actually achieved in real life are only a small fraction of what he estimates for the situation in which covalent chemical bonds are the main contributors to adhesion. He further proposes that even if there are a significant number of covalent chemical bonds across the interface, the failing strength of the bond still will depend on the strength of the polar bonds. The polar bonds will fail at a lower strength than the covalent bonds, and the applied load then will be concentrated on the covalent bonds. The measurable mechanical strength of a partially covalent adhesive bond still will be dominated by the polar forces. The implication is that although increasing the proportion of covalent bonds across the interface can lead to enhanced durability, the ultimate load-carrying capacity probably will not be significantly affected.

Surfaces

Adhesives must function solely through surface attachment. Therefore, the nature of the condition of the adherend surface are crucial to the formation of strong and durable bonds. By "surface" we usually mean that region of a material which interacts with its surroundings. There is some region of a bonded assembly where the adhesive and the adherend interact, but only rarely is this a sharp boundary. Usually it is a very diffuse, somewhat ill-defined region of interaction that has become known as an interphase rather than an interface. For example, the interaction with a freshly cleaved single crystal of zinc might occur over only a few atomic layers or a few nanometers. Rough or porous surfaces present more surface area than smooth ones of the same dimensions, and the adhesive might reach a depth of several hundred nanometers on a porous adherend such as wood or paper.

One very important aspect of surfaces is that they rarely have the same chemical composition as the bulk material and often seem to be entirely unrelated to the bulk. The surface usually consists of several regions having no clear boundaries. A metal alloy might have a well-defined bulk composition, but at the surface there probably will be a region that is still metallic but is of different chemical composition because of alloying elements or impurities that have segregated at the surface. On top of this region there probably will be a layer of oxides and hydroxides formed by reaction with the atmosphere. There also will be many other contaminants such as nitrogen, sulfur, and halogen compounds formed by interaction with the pollutants in the atmosphere. Finally there will be several layers of adsorbed water. The surface of a metal also might be contaminated with rolling oils, cutting lubricants, drawing compounds, or corrosion inhibitors. Mechanical working of the metal might even mix these contaminants with the other surface materials to create something like a nonhomogeneous "frosting" on the surface.

Engineering plastics display some of the same surface phenomena as metals, in that the surface is very different from the bulk. The manufacturing process often introduces anisotropy so that the mechanical properties of the material are different in different directions. In addition, it is common to find that components of the plastic have accumulated at the surface. Low molecular weight polymers or oligomers, plasticizers, pigments, mold release agents, shrink control agents, and other processing aids as well as adsorbed contaminants often are present.

The nature of the surface of an engineering plastic can change rapidly in response to its surroundings. The bulk of the material might be in the glassy state, but because of the concentration of low molecular weight material and contaminants, the surface region can be quite mobile. Exposure of the surface to a polar environment, such as by wiping with a polar solvent, can cause polar groups in the plastic surface to preferentially orient themselves outward.

Exposure to a nonpolar medium can bring out the nonpolar nature of the surface.

The adhesive layer itself can be thought of as one example of an engineering plastic. There is no reason to expect that the adhesive should be homogeneous throughout. Especially because the adhesive must go through a liquid, highly mobile state and must intimately interact with the adherend surface, it is quite likely that the adhesive in the interphase region is quite different from the bulk adhesive. Low molecular weight materials can be drawn into the surface of a porous adherend, leaving higher-weight polymer and fillers behind. The polar or nonpolar nature of the adherend influences the orientation and morphology of adhesive polymers in the interphase. Compounds at the adherend surface can catalyze or inhibit polymerization. Solvents in the adhesive can swell the adherend or dissolve portions of the adherend surface.

The interphase region is complex, and its composition is usually unknown. Primers or surface treatments often are used to improve control of the interphase and provide increased adhesion, durability, and resistance to aggressive environments. The chemistry of primers and surface treatments is as varied as the chemistry of the adhesives.

Any bonded construction consists of at least two adherends and one adhesive and contains at least two interphase regions. It is important to remember that the performance of the construction, its durability, its mechanical properties, and its response to tests and challenges, are all properties of the entire assembly. The successful use of adhesives depends on taking account of all parts of the construction and the process. Whereas the adhesive is just one part of the assembly, its chemistry plays an important role in the bonding process.

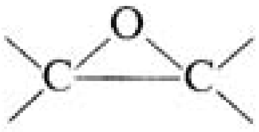
The following sections discuss the chemistry of some major classes of adhesives.

20.2 Epoxy Structural Adhesives

Introduction

Epoxy adhesives command a large portion of the structural adhesives market. Most people probably are familiar with epoxy structural adhesives, as these are the typical two-part adhesives found in hardware stores and supermarkets. Epoxy adhesives owe their popularity with both the general public and industry to their ease of use, their relative safety, and their compatibility with many adherends. The various chemical reactions involving the epoxy ring provide a fertile field for the development of a wide range of properties.14 Even a half century after the first epoxy patents were issued,15 new patents on epoxy adhesive technology continue to appear every month.

Epoxy adhesives get their name from the 1,2-epoxy, epoxide, or oxirane ring, which provides the reactive sites on the polymers. This three-member ring consists of two carbon atoms joined to an oxygen atom. The strained geometry contributes to the reactivity with nearly all nucleophilic compounds, not only curatives but also adherend surfaces.



Epoxy ring

Typical epoxy resins used to formulate epoxy adhesives have at least two epoxy rings, usually at the ends of a relatively short-chain prepolymer. The epoxy groups then are reacted with other epoxy groups in a chain-growth polymerization or with another curative in a step-growth polymerization to produce a polymer network, which can be either thermo-plastic or thermoset. The polymer linkages created by reaction of the epoxy ring are polar and provide adhesion to a variety of polar surfaces. With the proper backbone polymers and curing agents, cured epoxy adhesives can be very tough and resistant to chemical degradation. Bonded assemblies can have a high degree of durability and environmental resistance. Figure 20.1 shows the hand application of an epoxy adhesive to an automotive sheet molding compound (SMC) hood assembly.

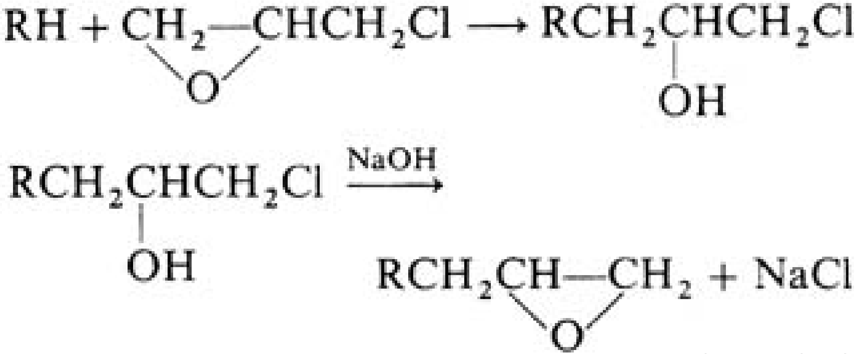
Fig. 20.1 Hand application of a two-part epoxy adhesive to a SMC (sheet molding compound) automotive hood assembly. (*Courtesy Lord Corp., Erie, PA*)

20.3 Commercial Synthesis of Epoxy Resins

Epichlorohydrin, or 1-chloro-2,3-epoxypropane, is a key raw material in the manufacture of epoxy resins. Epichlorohydrin is made by chlorination of propylene, treatment with hypochlorous acid, and dehydrohalogenation of the intermediate to give the mono-epoxide.

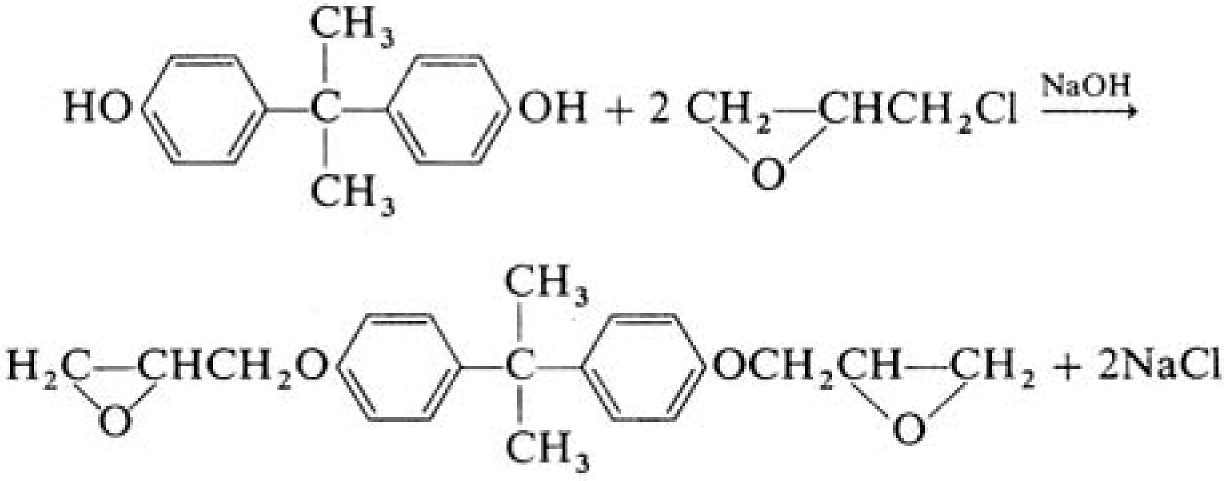
Epichloroydrin

Most epoxy resins used in adhesives are made by reacting a compound containing an active hydrogen atom with epichlorohydrin. This intermediate then can be dehydrohalogenated to give an epoxy functional material. Usually the starting material will contain at least two reactive hydrogen atoms to give a product having at least two epoxy groups per molecule.



Epoxides can also be prepared by direct chlorination of an olefin followed by dehydrohalogenation with base, or by the direct oxidation of an olefin by a peracid such as peracetic acid prepared in situ from the acid and hydrogen peroxide.

The most commercially important epoxy resins for adhesives are the diglycidyl ether of bisphenol A (DGEBA) and its higher homologues. Bisphenol A is made by condensing phenol and acetone with an acid catalyst. The bisphenol A then is reacted with epichlorohydrin, with dehydrohalogenation with sodium hydroxide following to give the product.

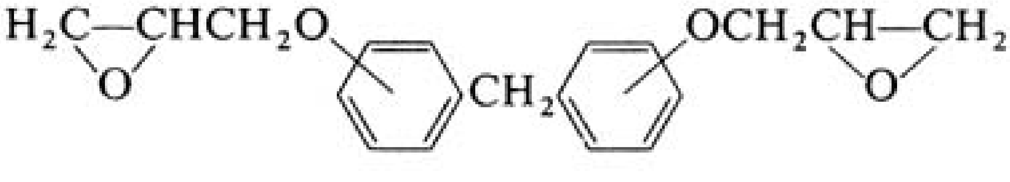


Diglycidyl ether of bisphenol A

The NaOH is present only in catalytic amounts until the addition of the epichlorohydrin is complete; then it is increased to a stoichiometric amount for the dehydrohalogenation. If this reaction is carried out with a large excess of epichlorohydrin, the principal product is the simplest diglycidyl ether. The pure DGEBA has a molecular weight of 340 and an epoxide equivalent weight of 170, and it melts at about 42°C. Because there are some side reactions, typical commercial resins have molecular weights of about 370 and epoxide equivalent weights in the range of about 180190, and will have some small percentage of material that is less than difunctional. The diglycidyl ether of bisphenol A is sold by several companies, such as Celanese (Epi-Rez 510), Reichhold (Epotuf 37-140), Shell (Epon 828), CIBA-GEIGY (Araldite GY 6010), and Dow Chemical (D.E.R. 331).

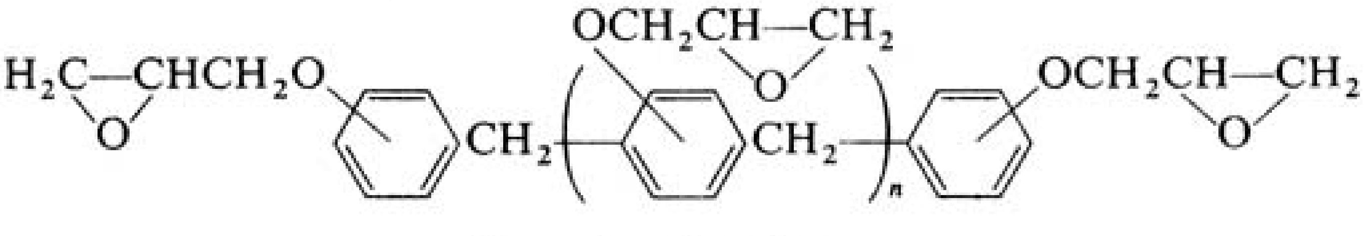
If the excess of epichlorohydrin is decreased and the reaction run with nearer the stoichiometric amounts of reactants, the molecular weight of the product increases. The higher molecular weight products are homopolymers of DGEBA, having pendant secondary hydroxyl groups. They can be produced commercially either through the direct reaction of bisphenol A, NaOH, and epichlorohydrin or by chain extension of the DGEBA with bisphenol A. The value of *n* can be as high as 30 for some resins. The higher-weight polymers are generally not practical in adhesives because the number of epoxy groups available for cross-linking becomes small, and the higher glass transition temperatures and lower reactivities make it impractical to drive curing reactions to completion.

Another important class of aromatic epoxy resins is the epoxy novolac resins. Novolacs are polymers made by condensing phenol or cresol with formaldehyde in the presence of an acid catalyst. The result is a chain of phenol or cresol groups held together by methylene bridges in a mixture of *ortho* and *para* isomers. The hydroxyl groups then can be reacted with epichlorohydrin and dehydrohalogenated with base to produce an epoxy functional novolac. The simplest, containing only two rings, is the diglycidyl ether of bisphenol F. It is a liquid resin having an epoxy equivalent weight of about 165.



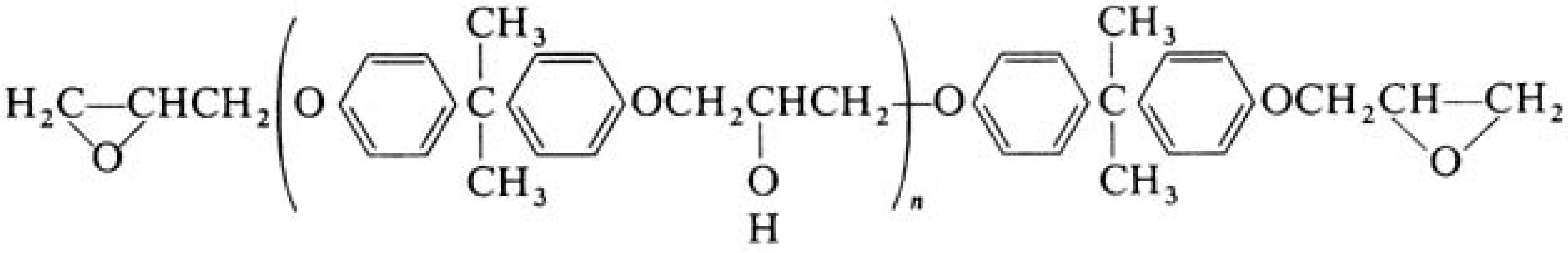
Diglycidyl ether of bisphenol F

Unlike that of the higher molecular weight bisphenol A resins, the epoxy functionality of the epoxy novolac resins remains relatively constant with increasing molecular weight because all of the aromatic rings contain hydroxyl groups that can be epoxidized. Higher molecular weight epoxy novolac resins can produce adhesives with higher cross-link densities.



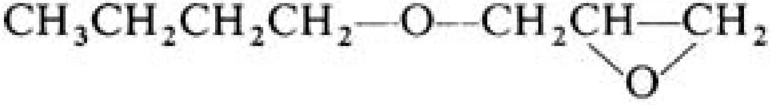
Epoxy phenol novolac

Value of *n* for epoxy novolac resins used in adhesives are usually in the range of about 0.23.5.



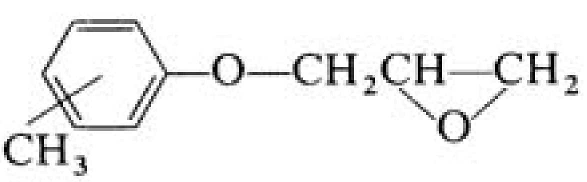
Bisphenol A resin

Low-viscosity monofunctional reactive diluents also are important components of many formulated epoxy structural adhesives. They are used to lower the viscosity of the composition without introducing low molecular weight unreactive species that might migrate out of the adhesive after it is cured. These compounds are made in much the same way that the polyfunctional epoxy resins are made. Some of the common reactive diluents are butyl glycidyl ether, cresyl glycidyl ether, and alkyl glycidyl ethers, among others.

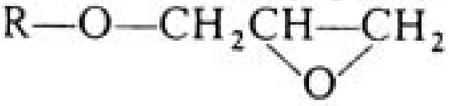


$CH_3CH_2CH_2CH_2{-}O{-}CH_2\underbrace{CH{-}CH_2}_{O}$

Butyl glycidyl ether



Cresyl glycidyl ether



$R{-}O{-}CH_2\underbrace{CH{-}CH_2}_{O}$

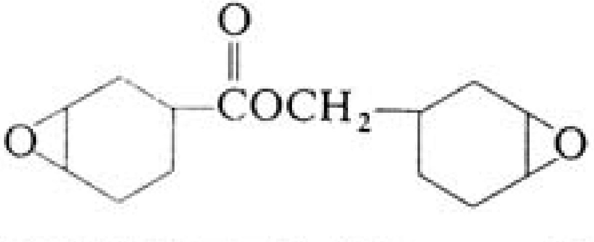
Alkyl glycidyl ether

Butyl glycidyl ether has the lowest viscosity and gives the greatest viscosity reduction at the least concentration. However, it has a higher vapor pressure than either cresyl glycidyl ether or the higher alkyl glycidyl ethers, and is more likely to cause problems with skin sensitivity and toxicity. For safety reasons the higher molecular weight, lower vapor pressure materials find more use even though they might be less efficient than those with a higher vapor pressure.

The diglycidyl ether of 1,4 butanediol is a commercially available (CIBA-GEIGY's RD-2) low-viscosity difunctional epoxy that is a somewhat less efficient diluent than the monoepoxides but does offer the advantage of having two reactive sites.

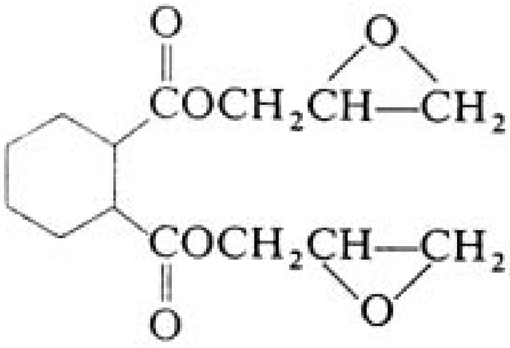
Epoxidized oils such as soybean oil or linseed oil, which are made by oxidizing the internal unsaturation in the fatty acid chain, find much use as plasticizers, particularly in vinyl resins. An internal epoxy ring is much less reactive than a terminal epoxy ring and usually does not take part in typical epoxy curing reactions under mild conditions. These materials can be used as plasticizers in epoxy adhesives but cannot really be considered reactive diluents.

Aliphatic epoxy compounds find some limited use in epoxy structural adhesives. They generally are low-viscosity liquids, so they can reduce viscosity with less impact on cross-link density than the monofunctional diluents have. The cycloaliphatic compounds such as 3′,4′-epoxycyclohexylmethyl 3,4-epoxycyclohexane carboxylate (CY-179 from CIBA-GEIGY) are made by direct epoxidation of the corresponding cyclohexane with peracetic acid. They produce high *T*g polymers too brittle for use as the sole epoxy compound in adhesives.



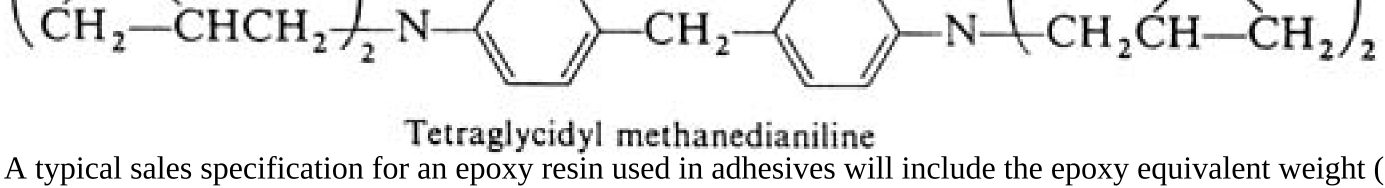
3′,4′-Epoxycyclohexylmethyl 3,4-epoxycyclohexane carboxylate

Tougher polymers of somewhat lower *T*g are obtained from glycidyl esters such as the diglycidyl ester of hexahydrophthalic acid (CY-184 from CIBA-GEIGY). The glycidyl esters are prepared by reacting the corresponding cycloaliphatic carboxylic acid with epichlorohydrin and dehydrohalogenating with NaOH.



Diglycidyl ester of hexahydrophthalic acid

There are also a variety of other specialty epoxy resins used in limited quantity for applications demanding some specific performance property. Tri- and tetra-functional materials such as triglycidyl-*p*-amino phenol, triglycidyl isocyanurate, or *N*,*N*,*N*′,*N*′-tetraglycidyl-4,4′-diaminophenyl methane (tetraglycidyl methanedianiline) have been used in adhesives requiring high heat resistance and good chemical resistance.



Tetraglycidyl methanedianiline

A typical sales specification for an epoxy resin used in adhesives will include the epoxy equivalent weight (EEW), also sometimes called the weight per epoxy (WPE), viscosity, and density at some specified temperature and the average functionality or number of epoxy groups per molecule. Sometimes a specification on total chlorine is included, which gives some indication of bound chlorine not removed by the NaOH in the dehydrohalogenation process. This bound chlorine is the result of epichlorohydrin side reactions during the epoxy synthesis. A high chlorine content indicates that there will be a large number of molecules of low functionality because each chlorine remaining represents one epoxy ring not formed in the dehydrohalogenation step.

Epoxy Cure Chemistry

Epoxy structural adhesives rely on the chemical reactions of the epoxy group with other reactants to pass from a liquid, wetting state to a solid, load-bearing state. There are a number of ways in which this is done, but all fall under one, or some combination, of three general schemes: step-growth polymerizations through reaction with curing agent, chain-growth polymerization initiated by Lewis acids, or chain-growth polymerizations initiated by Lewis bases.* Often the cure times of the slower step-growth curing adhesives are shortened by including Lewis acid or Lewis base catalysts.

*Lewis acids and Lewis bases are discussed below in the section on "Chain-Growth Polymerizations."

It would be tempting to consider the step-growth and chain-growth polymerization reactions as if they were independent and one could have the choice of either in any particular situation. The truth is that there are aspects of both types of polymerization in the cure of almost every epoxy structural adhesive. Such multiple-cure reactions often make it difficult to calculate the stoichiometry of an epoxy adhesive formulation. One type might predominate, depending on the formulation and cure conditions, but the effects of the other could not be completely discounted. The significance of this statement can be seen by looking at the two generalized reactions.

A typical step-growth reaction is one in which a single epoxy ring reacts with the active hydrogen of the curing agent. The general reaction is:

Model epoxy reaction

In order to form a high molecular weight polymer, both the epoxy-containing material and the curative must be at least difunctional. If both behave as difunctional materials, the resulting polymer is linear, and then it is necessary to drive the reaction nearly to completion to obtain a high-enough molecular weight to be useful. Neither of the reactants can contain a significant amount of monofunctional material. Monofunctional impurities or reactive diluents will act as chain terminators and limit the ultimate molecular weight obtainable.

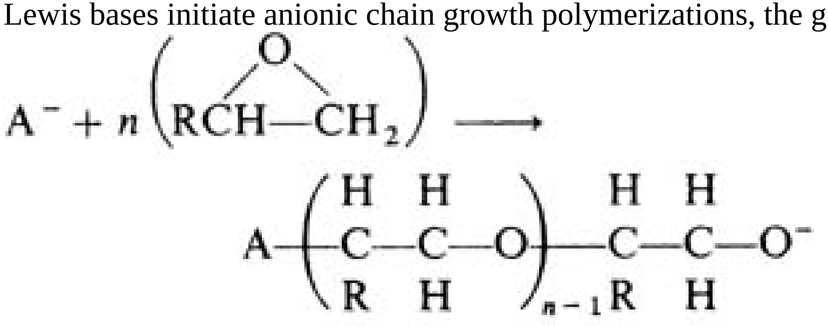
Chain-growth polymerizations occur through the reaction of epoxy rings with the active site on a growing chain and not with each other or with a second curing agent. After being initiated by a Lewis acid or a Lewis base, the growing chain will continue to consume epoxy groups and can reach a high molecular weight very rapidly. Because the epoxy groups are reacting with the growing chain, even a monofunctional epoxy compound can be polymerized. If A* is the initiator and M the monomer unit, the general reaction is:

$$A^* + nM \rightarrow A - M_{n-1}M^*$$

Chain-growth reaction

where the asterisk indicates the active site of the growing chain.

Lewis bases initiate anionic chain growth polymerizations, the generalized reaction being that of a propagating alkoxide anion.



Anionic chain growth

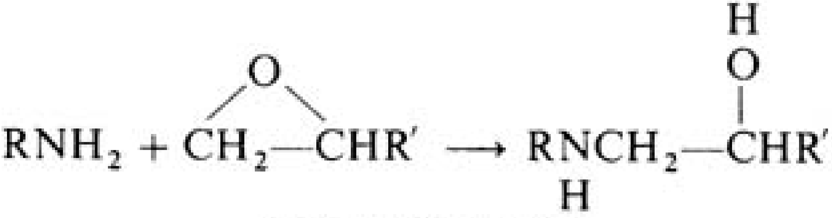
Lewis acids initiate cationic chain-growth polymerizations. There are several possible chain propagation reactions, and the mechanism of cationic chain growth is still open to some debate. However, the propagating species is a cation formed by reaction of the Lewis acid initiator with an epoxy ring.

In chain-growth polymerizations, epoxy reactants containing more than one epoxy ring per molecule can form tightly cross-linked, three-dimensional networks, as the epoxy groups on one molecule can easily react with more than one growing chain.

Step-Growth Polymerization

Only a relative few of the dozens of active hydrogen compounds that undergo reactions with the epoxy ring find widespread use in epoxy structural adhesives. The most common are amines, amides, acid anhydrides, phenols, thiols, and carboxylic acids.

Primary and secondary amines react with epoxy groups to form secondary or tertiary amine linkages. The reactions proceed at a reasonable rate at room temperature with the glycidyl ethers and glycidyl esters. The resulting polymer depends on the structure of the reactants and the degree of cure, and thus a variety of final adhesive properties can be had. Because the reaction is catalyzed by hydrogen ions, weakly acidic compounds such as phenol often are used to accelerate the reaction. The reaction of the secondary amine with the epoxy group produces a tertiary amine, which can in turn be the Lewis base that initiates an anionic chain growth polymerization of the remaining epoxy groups, depending on cure conditions.



Epoxy amine reaction

The simple linear aliphatic diamines, $H_2N(CH_2)_nNH_2$, can be used as curatives in adhesives. For small values of *n*, the short distance between the amine group can hinder the reaction of the second amine and slow the cure process. Also, the resulting products tend to be brittle for values of *n* less than about 6. More flexible, tougher products can be obtained by using liquid diamines or polyamines having more flexible backbones. For example, Jeffamine T.403 (Texaco Chemical) is a low-viscosity liquid (70 mPa · sec) having a molecular weight of about 440 and an amine equivalent weight of about 145. It is a poly (propylene oxide) triamine made from the polyether initiated by trimethylol propane, and is promoted as a flexible cross-linker for epoxy systems.

Various polyamines also are useful as curatives for epoxy adhesives. An example of a simple polyamine is diethylenetriamine (DETA):

$$H_2NC_2H_4NHC_2H_4NH_2$$

Diethylenetriamine

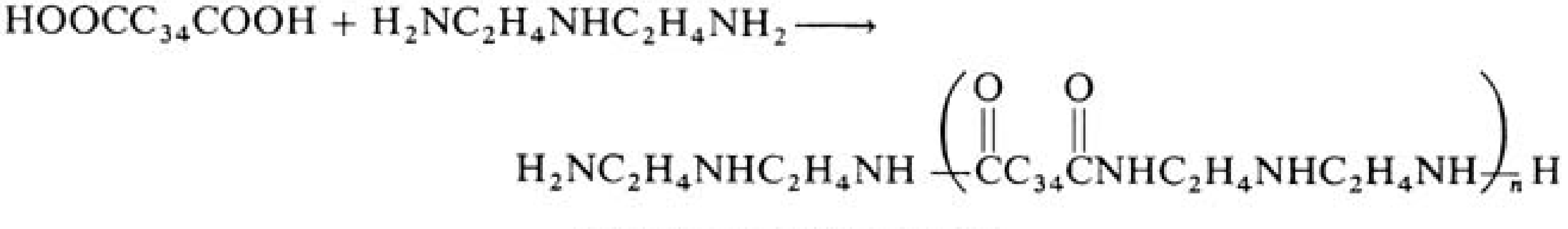
Its higher-weight homologues also are quite useful and are made by adding (CH_2CH_2NH) groups, leading to a homologous series of the form:

$$H_2N\!-\!\!\left(C_2H_4NH\!-\!\right)_n C_2H_4NH_2$$

where *n* is two, the material is triethylenetetraamine (TETA), and for *n* = 3 it is tetraethylenepentamine (TEPA), and so on. These three are the most important members of this series for adhesive applications, and are used more often than the simple aliphatic diamines. They often are used in combination with some other curative. Calculating stoichiometry can be difficult with these polyamines. All the active hydrogens might not be available because of steric factors introduced by the first reactions, or because once the first one or two active hydrogens are reacted with epoxy rings, the molecule is anchored into the chain and cannot readily diffuse to other epoxy groups. Therefore, an excess of amine is often added to the adhesive formulation. (Too much excess amine can lead to soft adhesives because all the epoxy rings are used up before the molecular weight becomes very high.)

Other series of polyamines can be made too. For example, some or all of the ethyl groups can be replaced by propyl groups, or the compound might be modified by reaction with an excess of a monoepoxide to give a hydroxy functional amine. Such modifications are made to improve adhesive properties and sometimes to lower toxicity of the curing agent or make it easier to handle.

Among the most important amine functional epoxy curing agents are the polyamide resins. These are made from dimerized unsaturated fatty acids by reaction of the dimer acid with a polyamine such as diethylenetriamine.



$$HOOCC_{34}COOH + H_2NC_2H_4NHC_2H_4NH_2 \longrightarrow H_2NC_2H_4NHC_2H_4NH\left(\overset{O}{\overset{\|}{C}}C_{34}\overset{O}{\overset{\|}{C}}NHC_2H_4NHC_2H_4NH\right)_n H$$

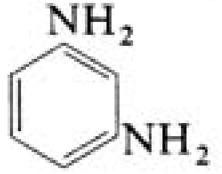
Polyamide Polyamine synthesis

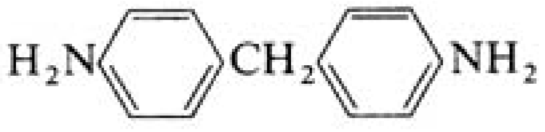
These polyamides are available from several suppliers worldwide and are among the most common curatives in the general purpose, "do-it-yourself" two-package epoxy adhesives. They have a distinctive odor somewhat like popcorn and are easily recognized in adhesive formulations. Chemically they are polyamides, but the activity with epoxy resins is primarily through the amine groups and not the amide hydrogens. The polyamide backbone does contribute to the overall good mechanical properties of the polyamide amine cured adhesives.

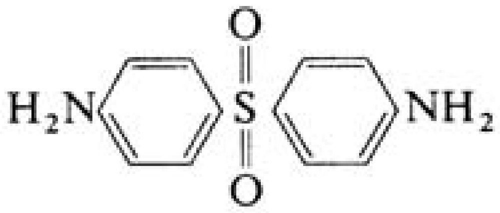
The polyamide amines are very high-viscosity liquids, some having viscosities over 50,000 mPa · sec. Typical amine equivalent weights are 100150. The polyamide amines react with bisphenol A epoxy resins at room temperature although the adhesives usually require several hours to reach sufficient molecular weight to carry a load. Cure times can be shortened to a few minutes at about 150°C. Tertiary amine catalysts such as dimethylaminomethylphenol also can be used to shorten cure time.

Epoxy adhesives cured with polyamide amines are flexible, tough, durable adhesives useful on a wide variety of adherends. They probably have contributed heavily to making the words "epoxy" and "adhesive" equivalent for many people.

A variety of aromatic polyfunctional amines also are used in curing epoxy adhesives. They generally are slower to react than the aliphatic amines and require a heat cure to be practical. They do provide generally better high-temperature properties than the aliphatic amines. The most commonly used aromatic amine curatives are *m*-phenylenediamine (MPDA), methylenedianiline (MDA), and diaminodiphenyl sulfone (DDS).


m-Phenylenediamine


Methylenedianiline


Diaminodiphenyl sulfone

All of the curatives described so far are used to make two-package adhesives in which the curing agent is packaged separately from the epoxy resin. Once they are mixed, they have a limited pot life, usually less than a few hours. It is possible to make one-package epoxy adhesives that can have very long shelf lives at room temperature but cure rapidly when heated. One amine curative widely used to make single-package heat-cured epoxy adhesives is dicyandiamide (cyanoguanadine), commonly known as dicy:

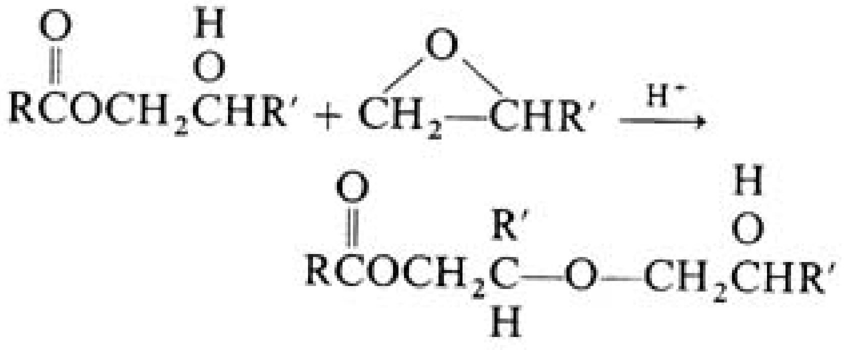

$$H_2N\overset{NH}{\overset{\|}{C}}NHC{\equiv}N$$

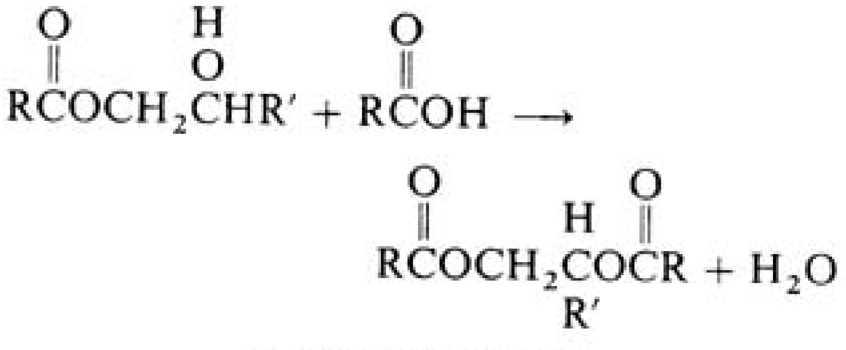
Dicyandiamide

It is made by dimerizing cyanamide in basic aqueous solution, and is a colorless solid melting at 208°C. Dicyandiamide is soluble in polar solvents, but at room temperature is insoluble in bisphenol A epoxy resins. It can be made into a very fine powder and milled into epoxy resins to form stable dispersions. Because the dicy is insoluble in the epoxy, the only possible reaction sites are at the particle surfaces. Although some reaction certainly occurs over a short time, the adhesives easily can have a useful shelf life of six months. On heating to about 150°C, the dicyandiamide becomes soluble in the epoxy resin, and the adhesive polymerizes rapidly.

Carboxylic acids can be used to cure epoxy adhesives or otherwise modify epoxy adhesives. The reactions can be complex. If no hydroxyl groups are present initially, the first reaction will be that of the active hydrogen with the epoxy ring to form an ester. This will produce a hydroxyl group on the backbone and allows for competing reactions. The organic acid can catalyze the etherification reaction with the hydroxyl group or undergo a condensation esterification reaction directly with the hydroxyl group.


Epoxy acid reaction


Etherification reaction

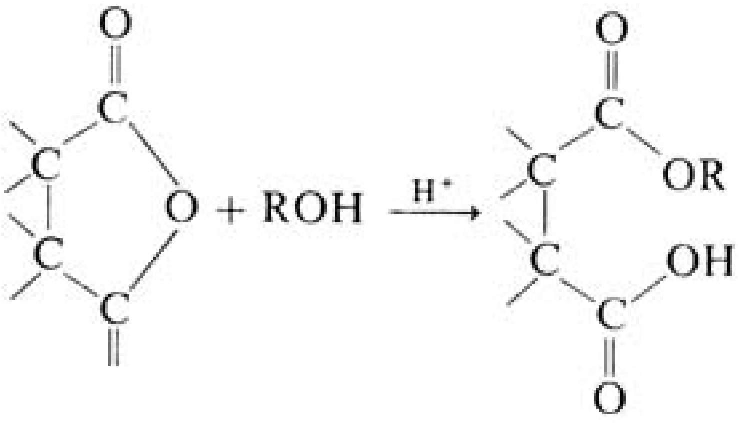

$$R\overset{O}{\overset{\|}{C}}OCH_2\overset{OH}{C}HR' + R\overset{O}{\overset{\|}{C}}OH \longrightarrow R\overset{O}{\overset{\|}{C}}OCH_2\underset{R'}{\overset{H}{C}}O\overset{O}{\overset{\|}{C}}R + H_2O$$

Condensation reaction

These reactions usually are slow at room temperature, and the adhesives must be cured with heat.

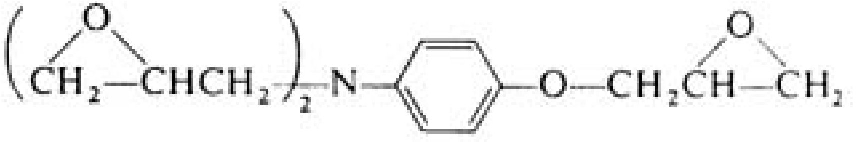
Acid anhydrides also can be used to cure epoxy adhesives although they usually are used only where good service at high temperatures is required. Most of the anhydride-cured epoxy adhesives are cured at high temperature. Because most of the anhydrides are relatively small molecules, the products tend to be tightly cross-linked and can be somewhat brittle.

The first step in the anhydride cure is ring opening of the anhydride by an active hydrogen, perhaps from water or hydroxyl groups already present on the epoxy resin. The resulting acid then reacts as a typical organic acid.

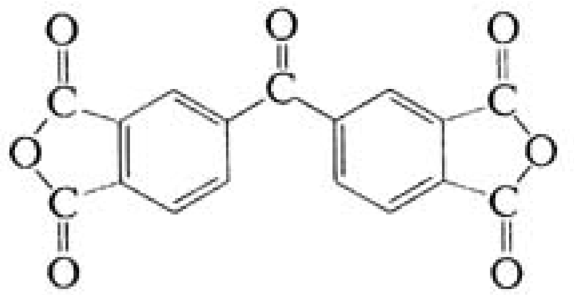


Anhydride ring opening by hydroxyl

An interesting anhydride-cured composition was reported,16 where 3,3′,4,4′benzophenone tetracarboxylic dianhydride (BDTA) was found to form a liquid solution with triglycidyl *p*-aminophenol (TGpAP) that cured at room temperature to give a product with good high-temperature properties.



Triglycidyl *p*-aminophenol



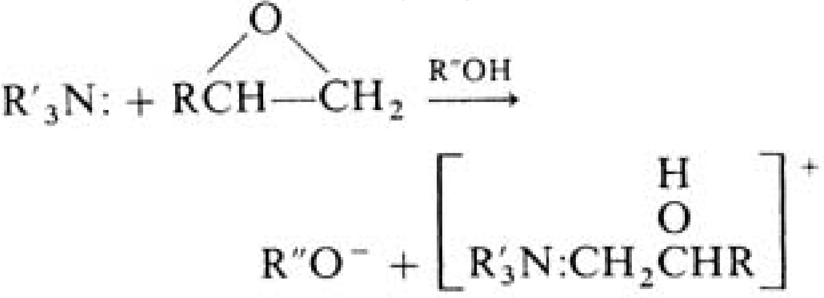
3,3′,4,4′-Benzophenone tetracarboxylic dianhydride

Another group of popular curing agents for epoxy resins is the thiol-terminated polysulfides. Thiols react with glycidyl ethers quite rapidly, particularly when a catalyst such as tris-(dimethylaminomethyl) phenol is present. The reaction is similar to that found with hydroxyl groups and produces a polythioether product. These are the familiar "five-minute" epoxy adhesives and have the characteristic odor of thiols.

Chain-Growth Polymerizations

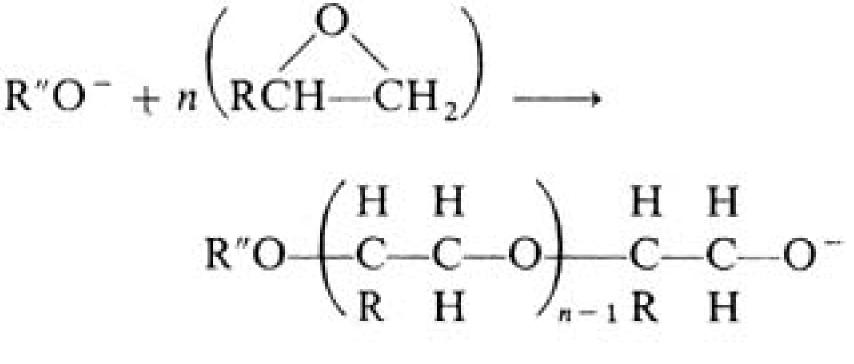
Chain-growth polymerizations are very important to many commercially successful epoxy structural adhesives. They can be extremely rapid and contribute to the fast cure times needed for high productivity in many manufacturing operations.

A Lewis base is a compound that contains an unshared pair of electrons capable of undergoing chemical reactions. Tertiary amines are examples of Lewis bases, and often are used in epoxy curing agents. In an anionic epoxy polymerization the propagating species is the alkoxide anion generated by the reaction of the Lewis base with an epoxy ring.



Formation of anionic chain initiator

This anion can continue to react with epoxy rings, adding them to the chain until the anion is destroyed in some side reaction.

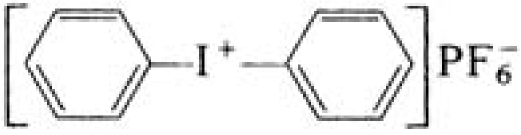


Anionic chain propagation

The product is a polyether, which can be tightly cross-linked when polyfunctional epoxides are used, as each epoxy ring can become part of a different chain.

Lewis acids, compounds with empty orbitals capable of accepting electron pairs, initiate cationic polymerization of epoxy resins. In this case the propagating species is a positive ion. The most commonly used Lewis acids are the boron trihalides, particularly BF_3 and BCl_3. They usually are used in the form of salts because both are gases at room temperature and are so reactive with epoxy resins that they can be difficult to control. Lewis base salts with the boron trihalides have much lower reactivity at room temperature but can react quickly on relatively mild heating, depending on the particular salt. Boron trifluoride readily forms salts with ethers, alcohols, and amines, and several of these salts are commercially available. A boron trifluoride ethanolamine salt can be included in the curative portion of a two-package epoxy adhesive. When the two packages are mixed at the time of use, the mixture can have a pot life of hours at room temperature but polymerize in minutes at temperatures of 100150°C.

One very interesting new application of Lewis acids in curing epoxy adhesives has appeared within the last 15 years. The Lewis acid initiator for the cationic polymerization is formed by the heat or ultraviolet light-induced decomposition of Lewis acid:Lewis base salts. Several patents by Crivello and coworkers1721 describe compounds containing aromatic onium salts such as iodonium or sulfonium, in which the cation is stabilized by the aromatic rings.



Diphenyl iodonium salt

The counterion generally is a large stable anion such a $[SbF_6]^-$, $[AsF_6]^-$, $[PF_6]^-$, or, $[BF_4]^-$. The more stable anions are less likely to terminate the growing cationic chains than are typical anions. The salts can be dissociated with heat to release the cation, which appears capable of initiating cationic polymerization of many materials in addition to epoxy rings. If the proper dye sensitizers are added, the cation can be liberated by ultraviolet light to initiate the polymerization. The cations persist for quite some time after the light source is removed. Acids of the form $H^+SbF_5X^-$, where X is a halogen, also have been used to catalyze epoxy reactions.22

Summary

Epoxy structural adhesives have proved to be versatile and reliable compositions. Their widespread use and acceptance is in part due to the varied chemistry of the epoxy ring and the skill of scientists and adhesive formulators in developing high-quality compositions that can produce reliable, reproducible structural joints even when applied by relatively unskilled users. New compositions and chemical reactions continue to be disclosed, and it is certain that the knowledge of epoxy chemistry will continue to grow. Many of the new discoveries will find their way into new epoxy structural adhesives.

20.4 Urethane Structural Adhesives

Introduction

The term "urethane adhesive" as it is generally used encompasses a lot of chemistry that is not necessarily urethane chemistry. "Urethane" is the common name for the compound ethyl carbamate. In common usage, "urethane adhesive" generally means an adhesive polymer derived from isocyanate chemistry and reactions of isocyanates with active hydrogen compounds. However, isocyanate reactions do not always lead to urethane linkages, and there are ways of arriving at urethane linkages without involving isocyanates. In this section the common approach will be taken; that is, an adhesive that uses reactions of the isocyanate group to bring about polymerization in the bond line will be considered a urethane adhesive.

The study of industrial applications of isocyanate chemistry and polymers derived from isocyanates received much attention in Europe, particularly Germany, in the 1930s and during World War II. Patents on aspects of urethane chemistry appeared as early as 1937.23 The effort going into understanding isocyanate chemistry and commercializing urethane products continues.24,25

The isocyanate group consists of a linear arrangement of nitrogen, carbon, and oxygen atoms, N=C=O. Several possible electronic configurations can be drawn, most of which involve a positive charge on the central carbon atom. This partial positive charge on the carbon atom accounts for much of the reactivity of the isocyanate group with nucleophilic groups. Not only does the isocyanate group react with a variety of potential curing agents, but also it is very reactive with many of the adherend surfaces on which urethane adhesives are used.

The same high reactivity that makes the isocyanate a desirable reactant for structural adhesives also renders the unreacted isocyanate more acutely toxic than, for instance, epoxy adhesives. Only a few isocyanates are safe enough and easy enough to handle that they find widespread use in urethane structural adhesives.

Isocyanate Preparation

Organic isocyanates are the major building blocks of urethane structural adhesives. They can be synthesized by a variety of routes, but most of the commercially available isocyanate compounds used in adhesives are made by the reaction of a precursor primary amine or amine salt with phosgene, followed by dehydrohalogenation. The reaction with phosgene usually is carried out at a relatively low temperature, less than 60°C, and then the temperature is raised to 100200°C to remove the HCl.

$$\underset{\text{amine}}{RNH_2} + \underset{\text{phosgene}}{COCl_2} \longrightarrow$$

$$\underset{\text{intermediate}}{[RNHCOCl + HCl]} \xrightarrow{\text{heat}} \underset{\text{isocyanate}}{RNCO + 2HCl}$$

Isocyanate synthesis by phosgenation

There are many possible side reactions, and the yield depends on the reaction conditions, which usually are specific to the starting materials and desired product. Much work has gone into determining the reaction conditions for manufacturing as clean a product as possible.

The amount of isocyanate in a commercial isocyanate or isocyanate containing formulation usually will be specified as weight percent isocyanate (as NCO) or as an amine equivalent weight, which is the weight of material containing sufficient isocyanate to react with one mole of amine hydrogen.

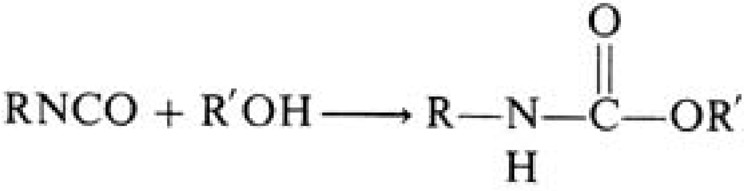
Isocyanate Reactions

Organic isocyanates can undergo a large number of reactions with active hydrogen compounds. One test commonly used to determine the presence of active hydrogen atoms and the number of active hydrogens per molecule is the TschugaeffZerewitinoff analysis or, more commonly, the Zerewitinoff test. An excess of a Grignard reagent, methyl magnesium iodide, is added to the sample to be tested, and the amount of methane evolved in measured.

$$CH_3MgI + RH \rightarrow CH_4 + RMgI$$

Zerewitinoff reaction

Urethane adhesives take their name from the product of the most common step-growth polymerization reaction used to generate the adhesive polymers. Isocyanates react with hydroxyl groups to create urethane (or carbamate) linkages:



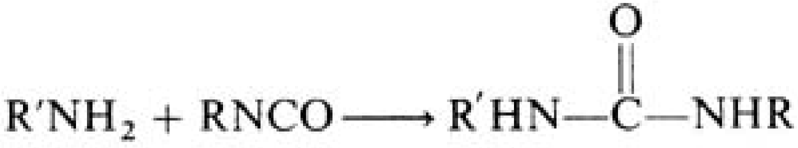
$$RNCO + R'OH \longrightarrow R{-}\underset{H}{N}{-}\overset{O}{\overset{\|}{C}}{-}OR'$$

Urethane formation

If both the isocyanate and the hydroxylcontaining material are difunctional, if the mixture is made up to have one isocyanate per hydroxyl, if there are no side reactions, and if the reaction can be driven to completion, a single linear thermoplastic polymer should result. If one or more of the reactants is more than difunctional, it is possible to create an infinite three-dimensional network.

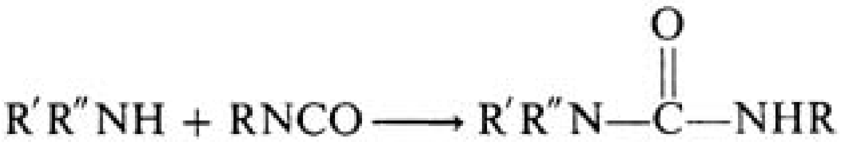
In general, primary hydroxyl groups are faster to react than secondary hydroxyls, which are in turn faster than tertiary hydroxyls, absent catalysts, and stearic factors. When reacted with hydroxyl-containing compounds, aliphatic isocyanates tend to be more sluggish than their aromatic counterparts. Urethane linkages made with tertiary hydroxyls tend to be less stable and at high temperature can dissociate into an olefin and an amine with loss of carbon dioxide. The isocyanatehydroxyl reaction is catalyzed by metal compounds, particularly tin compounds such as dibutyl tin dilaurate, and by various amines.

Isocyanates will react with amines to produce substituted ureas, primary amines being faster than secondary amines.



$$R'NH_2 + RNCO \longrightarrow R'HN{-}\overset{O}{\overset{\|}{C}}{-}NHR$$

Urea formation from 1° amine



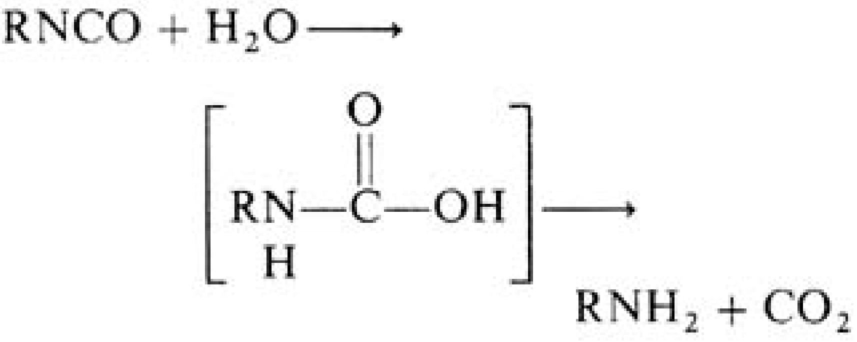
$$R'R''NH + RNCO \longrightarrow R'R''N{-}\overset{O}{\overset{\|}{C}}{-}NHR$$

Urea formation from 2° amine

The ureas are more rigid linkages than the urethane structure, but they also are generally more resistant to heat and chemical degradation. The reaction of isocyanates with amines is generally so rapid that it is nearly impossible to control the reaction well enough to make it useful in formulating urethane adhesives. What usually happens is that the reaction takes place faster than it is possible to mix the adhesive. If it is necessary to include urea linkages in the final production, that can be done by making a prepolymer which has the urea linkages in it already or by taking advantage of the slower reaction of isocyanates with water.

Small (much less than stoichiometric) amounts of diamines sometimes are added to the hydroxyl portion of the adhesive to provide a rapid but limited molecular weight increase as soon as the components are mixed.26,27 In this way a two-package adhesive can be made that will flow easily before mixing but will not readily flow after mixing. Such adhesives can be applied to vertical surfaces or overhead, and will remain in place until the bonds are closed and cured.

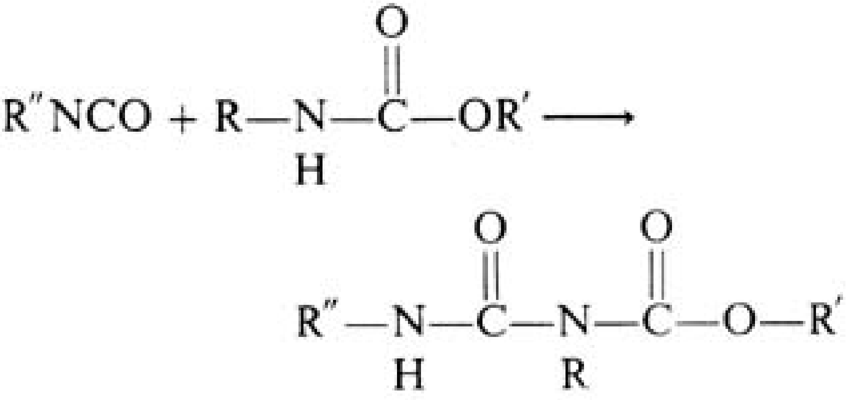
The reactions of amines and isocyanates are important in adhesives because of the possible reactions of isocyanates with water. Because isocyanates react readily with water, raw materials used in formulating adhesives must be dry, and the compositions must be protected from moisture, including atmospheric humidity, during storage. The first reaction with water is the formation of a carbamic acid, which rapidly loses carbon dioxide to form a primary amine.



Water NCO reaction

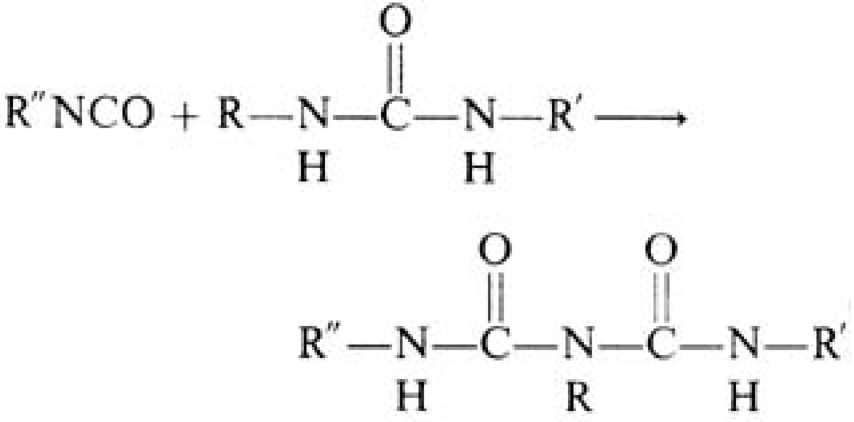
The primary amine then can react with another isocyanate to produce a urea. Thus one mole of water consumes at least two moles of isocyanate, builds molecular weight, and liberates carbon dioxide in the process. If this happens in a closed container such as a drum of adhesive, the result can be explosive, particularly as the reaction mass rises in temperature because of the exothermic reactions. On the other hand, these reactions can be useful in a bond line because under the proper conditions desirable urea linkages can be introduced into the curing adhesive through the reactions of small amounts of water normally present on the adherend surfaces.

The hydrogen atom attached to the nitrogen atom of the urethane group is active enough that it can react with another isocyanate group to produce an allophanate. This is an additional cross-linking mechanism for urethane polymers and can disturb the stoichiometry of the system by consuming an additional isocyanate group for each allophanate formed. Elevated temperatures usually are needed to produce allophanates in uncatalyzed systems; allophanation reactions can be catalyzed by tertiary amines.



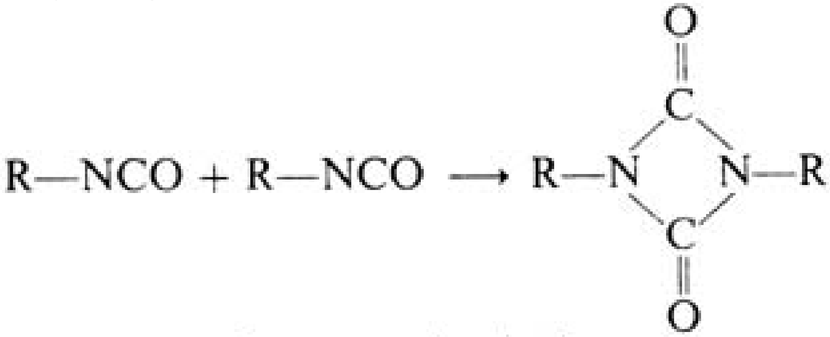
Allophanate formation

In a similar fashion, a urea hydrogen atom can react with an additional isocyanate group to produce a biuret.



Biuret formation

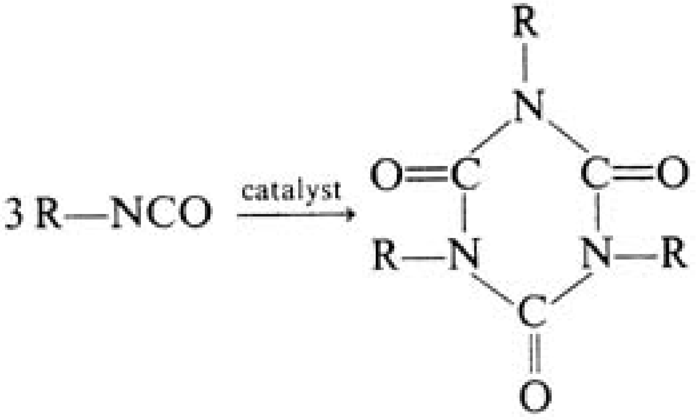
Isocyanate groups also will react with themselves to form a variety of compounds. Two isocyanate groups can react to form a dimer or uretidinedione.



Isocyanate dimerization

These dimers can be dissociated to regenerate the original isocyanates with heat, and some of the dimers, such as the dimer of toluene diisocyanate, have become commercially important.

Three isocyanate groups can react to form a trimer or substituted isocyanurate ring. This reaction can be catalyzed by phosphines or strong bases such as socium acetate or sodium formate. The isocyanurate ring is thermally stable, had good chemical resistance, and can enhance the resistance of a urethane adhesive to aggressive environments.

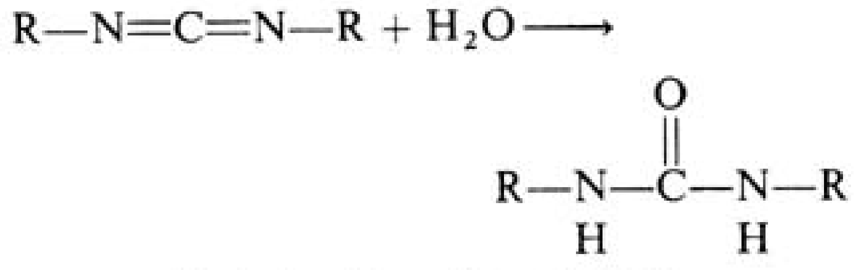


Isocyanurate formation

Isocyanates also can react with each other to produce carbodiimides with the loss of carbon dioxide. This reaction requires high temperatures unless catalyzed by specific phosphorus compounds. Formation of carbodiimides normally is not an important cross-linking mechanism in polyurethane adhesives. However, carbodiimides are sold by Union Carbide (Ucarlnk™) and have been recommended as water scavengers and stabilizers for polyurethanes. The carbodiimide can react with water to give a urea, which still can react with additional isocyanate to produce a biuret.

$$R{-}NCO + R{-}NCO \xrightarrow[\text{or heat}]{\text{catalyst}} R{-}N{=}C{=}N{-}R + CO_2$$

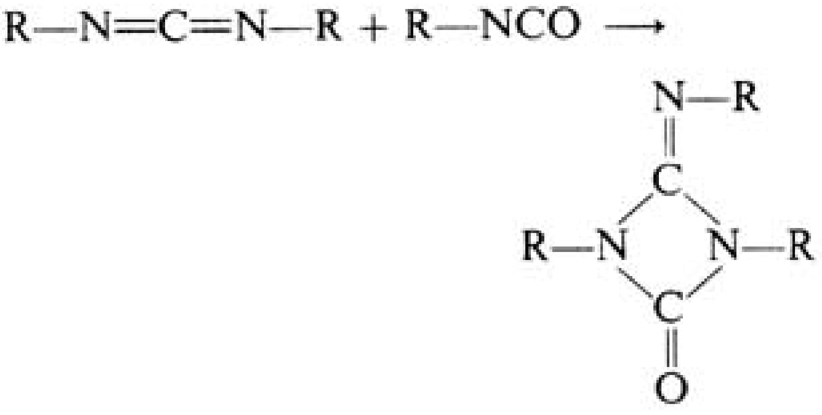
Carbodiimide formation



$$R{-}N{=}C{=}N{-}R + H_2O \longrightarrow R{-}NH{-}C({=}O){-}NH{-}R$$

Carbodiimide reaction with water

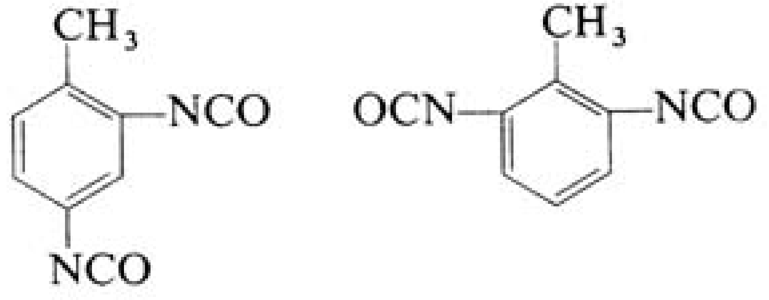
Carbodiimides can react with additional isocyanate groups to form uretone imines, which sometimes are used to modify polyisocyanates used in urethane structural adhesives.



Uretone imine formation

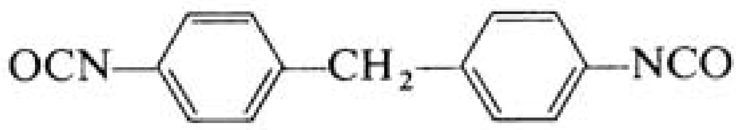
Important Isocyanates

Dozens of isocyanate functional compounds have been synthesized, but only a few find much use in urethane structural adhesives. The choices are largely dictated by a combination of performance, price, and safety considerations. Most of the materials used in adhesives are derived from the aromatic isocyanates, toluene diisocyanate (TDI) and diphenylmethane diisocyanate (MDI).



2,4-Toluene diisocyanate 2,6-Toluene diisocyanate

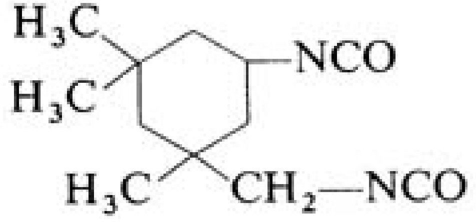
(TDI)



Diphenylmethane diisocyanate

4,4-Methylene bis(phenyl isocyanate) (MDI)

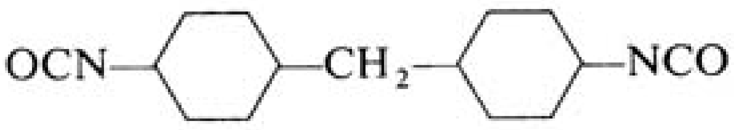
Where color and stability of the adhesive is important, and cure speed or cost is less important, aliphatic isocyanates are derived from isophorone diisocyanate (IPDI), hexamethylene diisocyanate (HDI), or 4,4′-dicyclo-hexylmethane diisocyanate (H12MDI).



Isophorone diisocyanate (IPDI)

$$OCN{-}CH_2(CH_2)_4CH_2{-}NCO$$

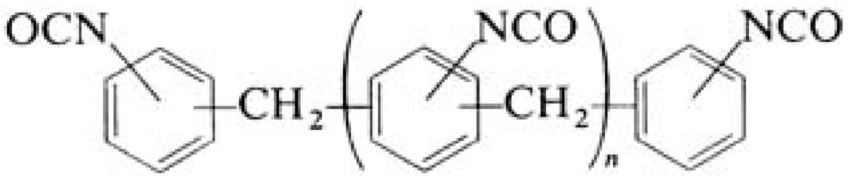
Hexamethylene diisocyanate (HDI)



4,4′-Dicyclohexylmethane diisocyanate (H_{12}MDI)

The common commercial TDI is an 80/20 mixture of the 2,4- and 2,6-isomers. The pure 2,6-isomer is available also and is sometimes called TDS. To make toluene diisocyanate, toluene first is nitrated to produce a mixture of the 2,4- and 2,6-dinitro isomers. The dinitrate is reduced to the diamine and reacted with phosgene, which is followed by dehydrohalogenation to give the diisocyanate. Because of its relatively high vapor pressure and toxicity, adhesives rarely contain toluene diisocyanate monomer. Typically, excess TDI is reacted with another material such as trimethylol propane or a polyester diol or polyether diol to produce a higher-molecular-weight isocyanate functional compound, which is safer and easier to handle than the free TDI. Still, there will be some free TDI present in the adduct, and much work has gone into finding synthesis schemes to minimize the free TDI in prepolymers and adducts.28

Diphenyl methane diisocyanate or 4,4′-methylene bis (phenyl isocyanate), commonly called MDI, can be considered the first member of a series of polyisocyanates of the general form:



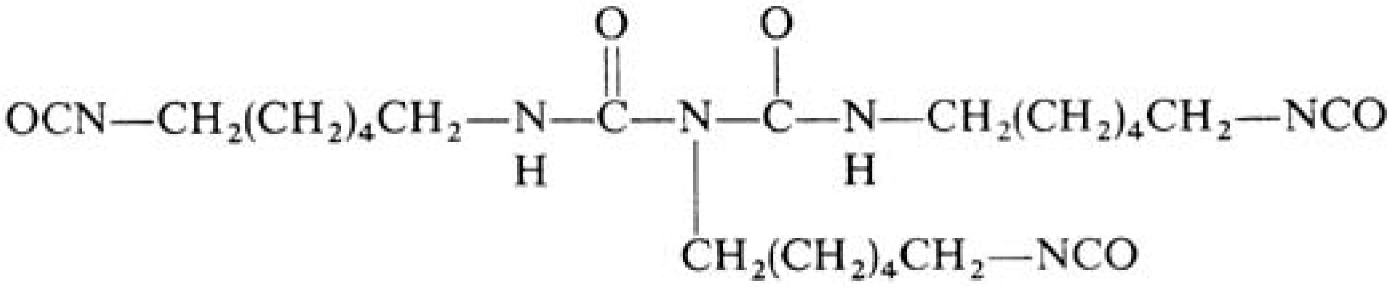
Aromatic polyisocyanates

MDI is the most important member of the series although materials with *n* of 1 or more also are available. The precurser amine is made by condensation of aniline hydrochloride with formaldehyde, followed by reaction with phosgene and dehydrohalogenation. When the aniline is present in excess, the diamine and consequently the diisocyanate are produced in greatest yield, with nearly all of this being the 4,4′-isomer.

MDI is a solid at room temperature with a melting point of about 38°C, which usually is stored and shipped in the molten state for convenience. However, on standing in the liquid state, the MDI slowly dimerizes, and the liquid MDI becomes saturated with dimer at about 1 percent dimer by weight. The dimer then begins to precipitate, and the isocyanate content of the liquid falls because isocyanate is consumed in the dimer formation and removed in the precipitation.

In general, the aliphatic isocyanates have better resistance to degradation by ultraviolet light than do the aromatic isocyanates. In spite of their higher cost and lower reactivity, the aliphatic isocyanates do find some use in adhesives where the adhesive is visible and color stability is important in the final assembly.

Hexamethylene diisocyanate (HDI) can be used as an aliphatic cross-linker in urethane adhesives but presents significant health risks because of its toxicity and high vapor pressure. HDI is more commonly used in the form of its biuret, which is much safer and easier to handle than HDI.



Biuret of hexamethylene diisocyanate

Isophorone diisocyanate (IPDI) is another low-viscosity aliphatic diisocyanate that is useful in formulating light-stable polyurethane adhesives, and is somewhat lower in cost than hexamethylene diisocyanate. IPDI has a low vapor pressure at room temperature but, like most low-weight isocyanates, still can present a health hazard. Higher-molecular-weight adducts of IPDI, such as its trimer, the isocyanurate of IPDI, are available. The trimer is a high-melting solid (100115°C), and one loses the advantage of the low viscosity liquid in using it but gains in safety and easy handling.

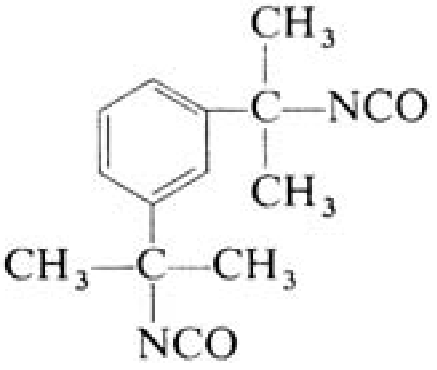
IPDI is a cycloaliphatic diisocyanate, but the two isocyanate groups are not equivalent. One is attached directly to the aliphatic ring, and the other is separated from the ring by a methylene group. Because their environments are different, the reactivities of the two groups are different; and the reactivity can depend on the choice of catalyst. For instance, Ono and coworkers29 have shown that when IPDI is reacted with a primary hydroxyl group, the primary NCO is most reactive when the catalyst is 1,4-diazabicyclo [2.2.2] octane. The secondary NCO attached to the ring is the most reactive when dibutyl tin dilaurate is the catalyst.

The saturated analog of MDI, 4,4′-dicyclohexyl methane diisocyanate, has found limited use as an aiphatic isocyanate in adhesives. This material is known by a variety of names including Desmodur W™ (Mobay), hydrogenated MDI (or HMDI or H12 MDI), reduced MDI (RMDI), and saturated MDI (SMDI). It is a low viscosity liquid with a fairly high vapor pressure; so it too must be handled with care. In adhesive compositions, the diisocyanate usually is used to make an isocyanate functional prepolymer by reacting excess diisocyanate with a hydroxyl or amine functional polymer such as a polyester diol.

Because it contains, two saturated sixmember rings, 4,4′-dicyclohexyl methane diisocyanate can exist in three isomeric forms, in which the orientations of the NCO groups with respect to the rings and each other are different. The *transtrans* isomer is a solid at room temperature so that over time this isomer can precipitate from the remaining liquid isomers. Samples of 4,4′-dicyclohexyl methane diisocyanate that have been stored at cool temperatures for a while often have a solid layer of *transtrans* precipitate in the bottom.

Two examples of a new family of diisocyanates have been offered by American Cyanamid, the *meta* and *para* isomers of tetramethyl xylene diisocyanate (TMXDI).30

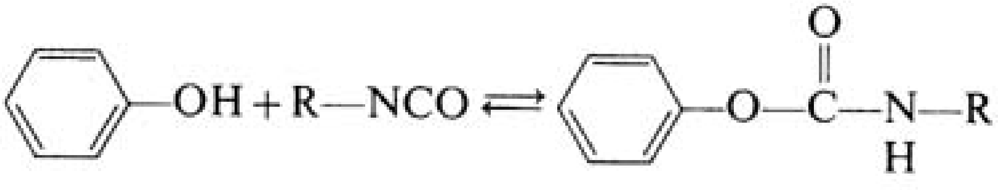
The *para* isomer of tetramethyl xylene diisocyanate is a solid melting at 72°C and so has received less attention in adhesives than has the liquid *meta* isomer (m.p. -10°C). These isocyanates are interesting because they contain an aromatic ring, but the NCO groups themselves are aliphatic isocyanates and have reaction characteristics typical of aliphatic diisocyanates. They are even somewhat more sluggish in reacting than the more standard aliphatic isocyanates, making the reactions easier to control. Compounds such as dimethyl tin dilaurate, lead octoate, or tetrabutyl diacetyl distannoxane have been shown to be effective catalysts for the isocyanatehydroxyl reaction. The manufacturers claim that although they are still toxic, the *p*-TMXDI and *m*-TMXDI appear to be less toxic than many other isocyanates.



m-Tetramethyl xylene diisocyanate (*m*-TMXDI)

Blocked Isocyanates

Blocked isocyanates are compounds formed by the reaction of an organic isocyanate with an active hydrogen compound where the reaction is reversible with heat. The blocked isocyanate can be used in formulating adhesives or other reactive compositions even in the presence of materials that normally would react rapidly with the isocyanate. The isocyanate can be regenerated by heating the mass to a temperature at which the blocked isocyanate begins to dissociate at an acceptable rate. Phenol is one example of a blocking agent used with isocyanates. There are many other blocking agents in use, including lactams, oximes, and malonates. Even isocyanate dimers such as TDI dimer could be considered blocked isocyanates because they will dissociate with heat to regenerate the isocyanates.



Phenol blocking reaction

There are two major limiting factors to the use of blocked isocyanates in urethane structural adhesives. First, the adherends must be sufficiently heat-resistant to withstand the temperatures needed to cause rapid dissociation of the blocked isocyanate. Second, and perhaps more important, there usually must be some mechanism for removing the blocking agent from the bond line. Otherwise the blocking agent is present to continue competition with the intended curing agent for reaction with the isocyanate. After the adhesive is cured, blocking agent remains trapped in the bond line and could contribute to poor mechanical properties in the bond or poor resistance to harsh environments. The need to remove the blocking agent makes the blocked isocyanates particularly attractive in applications such as laminating where a large exposed area of adhesive is cured before the bond is closed.

Summary

Urethane structural adhesives have proved quite successful in bonding applications that take advantage of their toughness, particularly when the adhesive must function at low temperatures. The reactivity of the isocyanate group may require some extra care in handling and storing the adhesives but the same reactivity provides many opportunities for the polymer chemist and adhesive formulator. The principal reaction used in curing urethane structural adhesives is the reaction of the isocyanate group with primary and secondary hydroxyl groups. Both the isocyanate groups and the hydroxyl groups can be carried on a wide variety of low molecular, weight oligomers such as polyesters, polyethers, polybutadienes, polyurethanes, and polymer blends so that many of the desired properties of the final adhesive can be built into the polymer before the adhesive is cured.

20.5 Acrylic Structural Adhesives

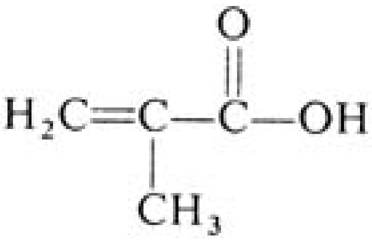
Introduction

Acrylic structural adhesives are relatively new products compared to epoxy or urethane adhesives. They first appeared in Europe in the mid-1960s and were commercialized in the United States a few years later. The first offerings were rather brittle products of use in limited applications where toughness and flexibility were not critical. However, because these new acrylic structural adhesives polymerized in the bond line through a free radical chain growth polymerization mechanism, they did have the big advantage of curing rapidly at room temperature. This polymerization mechanism offered a variety of potential advantages in manufacturing. Bonded structures could be assembled relatively rapidly without the need of curing ovens. Assemblies did not need to be stored for long periods while room-temperature cures were completed. The adhesives could be made at very low viscosities for easy handling and dispensing. In the past two decades there has been much creative work done with the chemistry of acrylic adhesives so that now they are available in many forms with a wide range of properties. Figure 20.2 shows the spray application of an acrylic structural adhesive to fiber-reinforced plastic panels that are assembled into farm silos in Japan.

Fig. 20.2 Spray application of an acrylic adhesive to fiber-reinforced plastic panels, which are assembled into silos in Japan. (*Courtesy Lord Corp., Erie, PA*)

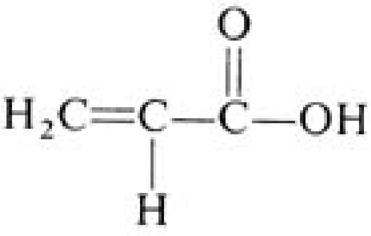
Acrylic Monomers

All acrylic structural adhesives consist basically of a solution or a mixture of polymers and unsaturated, low molecular weight, free-radical-polymerizable monomers with other materials added as needed for the particular intended use. In addition, some precursor of the polymerization initiator will be present. The most common monomers used are methacrylic acid (2-methyl propenoic acid) and its esters or, less commonly, acrylic acid (propenoic acid) and its esters.



$H_2C{=}C(CH_3){-}C({=}O){-}OH$

Methacrylic acid

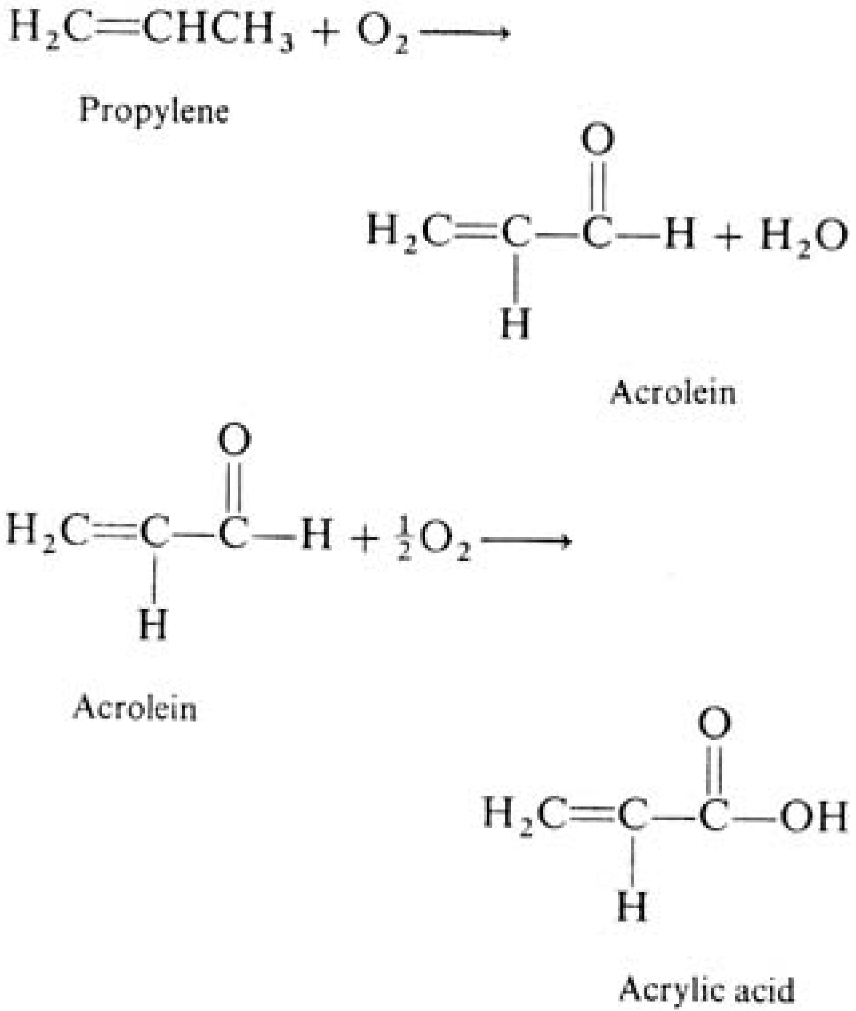


$H_2C{=}C(H){-}C({=}O){-}OH$

Acrylic acid

Several processes for making acrylic acid have been developed.31 Most manufacturers now use one of two processes, the oxidation of propylene or the Reppe process, which involves the oxidation of acetylene.

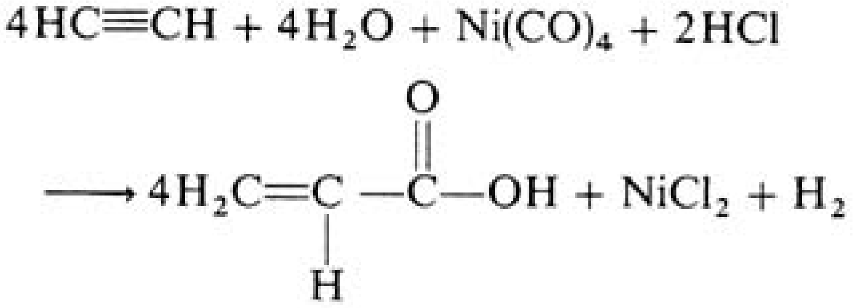
The oxidation of propylene is a two-step process with specific catalysts used in each step. In the first, the propylene is converted to acrolein, which then is further oxidized to acrylic acid in the second step. The acrylic acid then can be esterified with the appropriate alcohols.



$$H_2C{=}CHCH_3 + O_2 \longrightarrow H_2C{=}CH{-}CHO + H_2O$$

$$H_2C{=}CH{-}CHO + \tfrac{1}{2}O_2 \longrightarrow H_2C{=}CH{-}COOH$$

In the Reppe pocess, acetylene is reacted with carbon monoxide, nickel carbonyl, and water to produce acrylic acid. If the appropriate alcohol is used in place of the water, the acrylate ester can be produced directly. Acrylates made by the Reppe process tend to contain more propionates than do those made by oxidation of propylene.

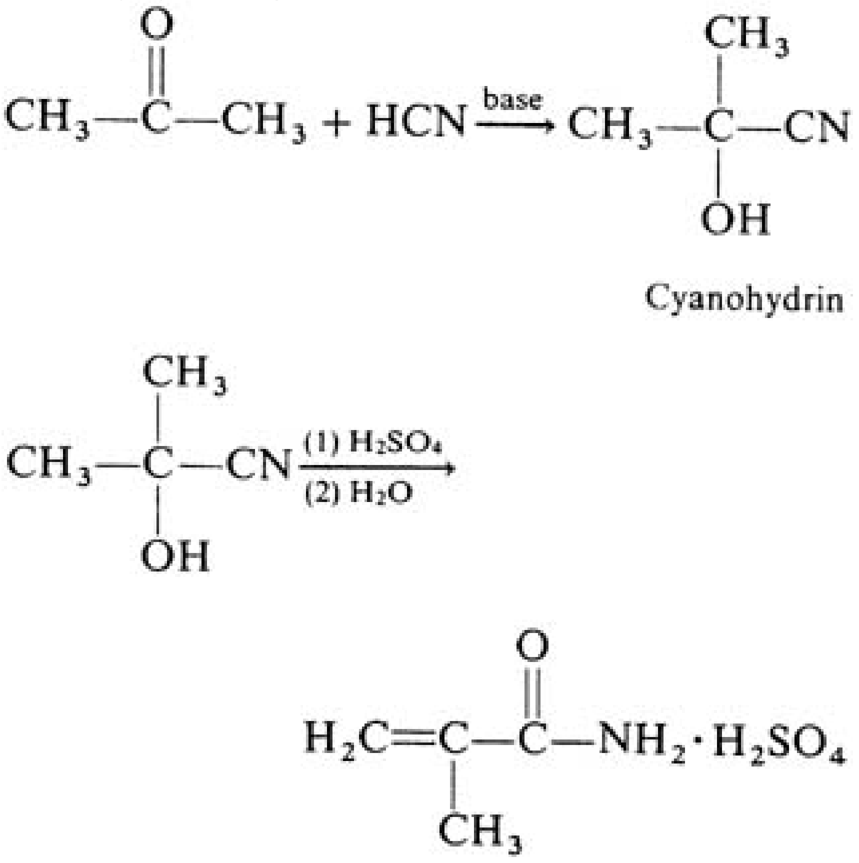


$$4HC{\equiv}CH + 4H_2O + Ni(CO)_4 + 2HCl \longrightarrow 4H_2C{=}CH{-}COOH + NiCl_2 + H_2$$

Reppe process for making acrylic acid

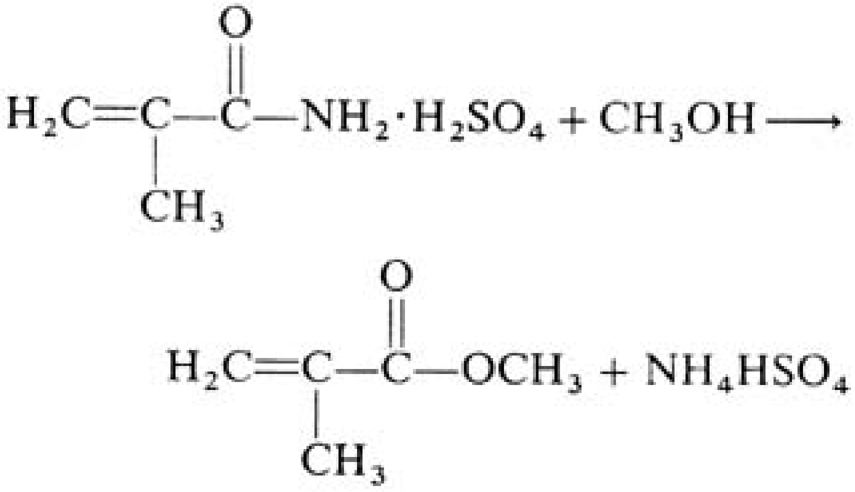
If carbon monoxide is supplied as a separate feed in the presence of small amounts of nickel carbonyl, the reaction also will proceed with the nickel functioning as a catalyst.

Methacrylic acid can also be produced by a variety of processes including oxidation of ethylene, propylene, and isobutylene.32 The most common commercial process for making methacrylic acid is the acetone cyanohydrin (ACN) process. The feedstocks for this process are acetone, hydrogen cyanide, and sulfuric acid. The acetone and HCN are reacted under alkaline conditions to produce the cyanohydrin. Reaction of the cyanohydrin with sulfuric acid ultimately produces methacrylamide sulfate, hydrolysis of which produces methacrylic acid. Methyl methacrylate can be made either by esterifying the acid or directly by reacting the amide sulfate with methanol.



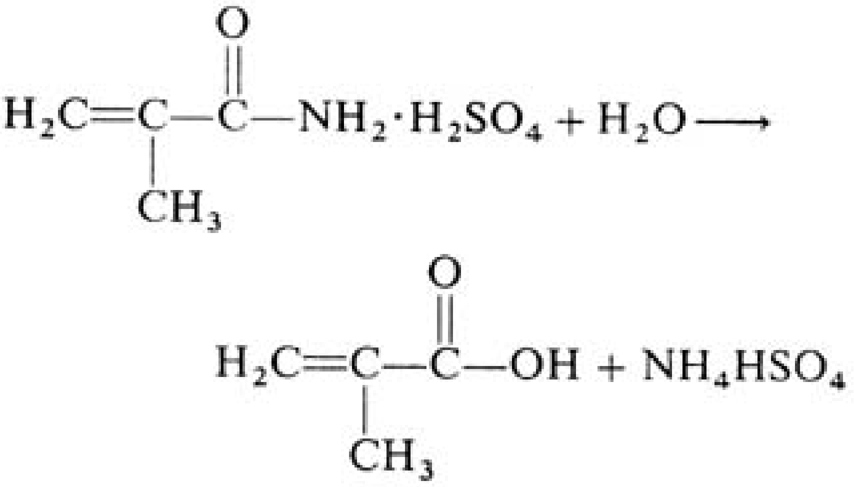
$$CH_3{-}CO{-}CH_3 + HCN \xrightarrow{\text{base}} (CH_3)_2C(OH){-}CN$$

$$(CH_3)_2C(OH){-}CN \xrightarrow[(2)\ H_2O]{(1)\ H_2SO_4} H_2C{=}C(CH_3){-}CO{-}NH_2 \cdot H_2SO_4$$



$$H_2C{=}C(CH_3){-}C({=}O){-}NH_2\cdot H_2SO_4 + CH_3OH \longrightarrow H_2C{=}C(CH_3){-}C({=}O){-}OCH_3 + NH_4HSO_4$$

Methyl methacrylate



$$H_2C{=}C(CH_3){-}C({=}O){-}NH_2\cdot H_2SO_4 + H_2O \longrightarrow H_2C{=}C(CH_3){-}C({=}O){-}OH + NH_4HSO_4$$

Methacrylic acid

The choice of monomers that are useful in acrylic structural adhesives is rather limited. Cost always is an important factor, and because acrylic structural adhesives consume only a very small portion of the world's output of acyrlic monomer, the formulator usually must rely on acrylic monomers which are made in large quantity for other uses. In addition, the monomers must polymerize readily at room temperature. If a mixture of monomers is to be used, the monomers must copolymerize easily. Finally, the monomers must be good solvents or dispersants for the polymers used in formulating the adhesive. The acrylic monomers finding most use in acrylic structural adhesives are methacrylic acid and its esters, particularly methyl methacrylate. These two usually make up the bulk of the monomer present.

Curing

The curing reaction of an acrylic structural adhesive is the chain-growth polymerization of the acrylic monomer. The monomer units are not reactive with each other but react only with a growing chain having an active site on one end. In order to begin a chain, one must generate an initiator in the monomer solution. In the case of acrylic structural adhesives, this initiator nearly always is a free radical, a species having an unpaired, reactive electron. After chains have been initiated, there are three general types of reaction that can occur: propagation, chain transfer, and chain termination. If "*" represents the active site, "A*" represents the initiator, and M is a monomer unit, then the four reaction types are:

$$\text{Initiation: } A^* + \rightarrow AM^*$$

$$\text{Propagation: } AM^* + n\mathrm{M} \rightarrow AM_nM^*$$

$$\text{Chain transfer: } AMnM^* + M \rightarrow AM_nM + M^*$$

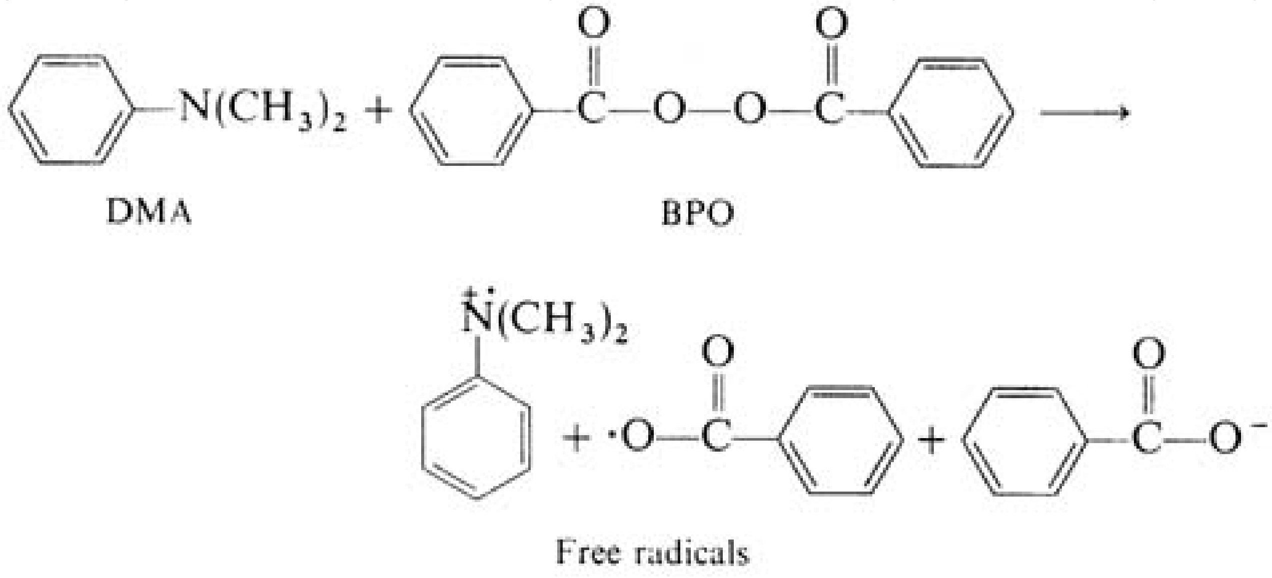
$$\text{Termination: } RM^* + R'M^* \rightarrow RMMR' \text{ (combination)}$$

$$RM^* + R'M^*$$

$$\rightarrow RMC{=}C + RMCC \text{ (disproportionation)}$$

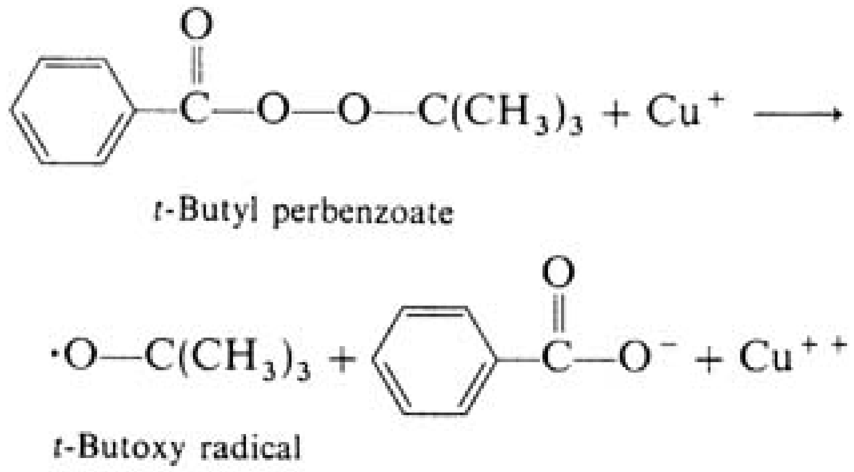
In addition, a growing chain might be terminated by reaction with a variety of unknown impurities invariably present in any mixture.

The initiator usually is created in a redox reaction that follows a free radical pathway. Common reactions involve the reduction of an organic peroxide by some reducing agent such as an amine or an ion capable of undergoing a one-electron transfer reaction. One example of an efficient free radical initiator generating reaction is the reduction of diacyl peroxides such as benzoyl peroxide (BPO) by tertiary aromatic amines such as *N*, *N*-dimethyl aniline (DMA).



Benzoyl peroxide reaction with N,N-dimethyl aniline

Peroxyesters can be reduced by metal ions such as copper or iron to produce free radicals that will initiate chain growth polymerization.



Reduction of a peroxyester

Free radical initiators also can be generated by thermal decomposition of certain compounds such as organic peroxides or azo compounds. Ultraviolet light or electron beams can initiate the polymerization of acrylic monomers.

The chain-growth polymerization, once it has been initiated, proceeds in two fairly distinct stages. The important point is that the monomers are not reactive with each other but only with the growing chains. Over its lifetime one propagating free radical could produce a chain many monomer units long. This chain is one single polymeric molecule dissolved in the remaining monomeric liquid. As free radicals are produced and propagate, more polymer "solute" chains are produced at the expense of the monomer "solvent." Eventually there will remain too little monomer to dissolve all the polymer being formed.

The point at which there is too little monomer to be an effective solvent for the acrylic polymer forming in the polymerization reaction (and also any oligomeric materials added to the formulation) can occur very early in the polymerization reaction. It is possible for the polymer to precipitate as a totally separate single phase, leaving the liquid monomer behind, but this is undesirable. Usually enough polymer forms that the polymer chains become entangled, and a gel saturated with monomer is formed. This is significant from two aspects. Once the gel has formed, the reacting mass in the bond line should not be disturbed until the cure is much further advanced because the composition may no longer be fluid enough to re-form the bond. Second, the rules of the polymerization process change with the onset of the Trommsdorf effect (or "gel" effect), and the remaining polymerization becomes much more rapid.

Prior to the gel point, the growing polymer chains with the reactive terminal sites are mobile and have easy access to monomer and to each other. Both chain growth and termination are significant reactions. After the gel point, the bimolecular termination reactions become much less significant because the chains no longer are freely mobile, and the probability of two radicals coming close enough to interact greatly decreases. The low molecular weight monomer still is able to diffuse to the reactive sites, so chain growth continues. The average life of the active radical site on the growing chain becomes much longer after the gel than it was before so that the overall polymerization proceeds at a much faster rate after the gel.

After the free radical initiators are generated in the adhesive, by mixing the redox couple, ultraviolet light, heat, or other means, the adhesive remains liquid for some period of time until the gel point is reached. Then the molecular weight, strength, and toughness of the adhesive build rapidly.

One difficulty sometimes encountered with free radical polymerizations in acrylic adhesives is the inhibition of the reaction by oxygen. Oxygen contains two unpaired electrons and readily attaches to the propagating free radical on a growing chain. The resulting oxy radical is much less reactive than the original carbon radical, and the chain growth effectively stops. Even if it proceeds, the chain backbone will contain a less stable peroxide linkage. The result is that the surface of acrylic adhesives cured in the presence of oxygen often will be of low molecular weight and will be tacky even though the bulk of the adhesive is fully cured.

Formulation

Although the free radical chain growth polymerization brings processing advantages, it also brings limitations. The product of the polymerization of the monofunctional acrylic monomers is a linear, thermoplastic polymer. In the case, of a copolymer that is largely methyl methacrylate, the glass trasition temperature of the polymer will be well above room temperature if the polymer has been driven to a high-enough molecular weight. The resulting adhesive is brittle, and although its shear strength might be quite high, its ability to withstand peeling forces or impacts will be low. One could use monomer mixtures yielding copolymers having lower glass transition temperatures and expect that the adhesive might be somewhat tougher. However, this parameter is difficult to control because in curing the acrylic adhesive in the bond line, one really is doing a bulk polymerization under uncontrolled conditions and trying to produce a specific polymer with a specific set of material and performance properties. The approach taken by most adhesive manufacturers or formulators is to use "prepolymers," high molecular weight oligomers that can be made under controlled conditions and then dissolved or dispersed in the acrylic monomer. By forming key elements of the polymer chains under controlled conditions, many of the desired properties of the final product can be built into the adhesive when it is formulated and before the material is cured.

The simplest approach to improving toughness is to dissolve or disperse a nonreactive, rubbery polymer in the monomer mixture that is to be polymerized. Examples of such materials are nitrile rubbers, polyethers, and acrylic rubbers. The choice will depend on cost, desired properties, and the solubility of the polymer in the monomer. As the monomer polymerizes, it will lose its solvating ability for the dissolved polymer. Eventually, when enough of the monomer has been consumed, it no longer can hold the polymer in solution, and the polymer precipitates. The polymer then can segregate as a totally separate phase or as distinct domains heterogeneously dispersed throughout the acrylic polymer matrix, or it can be homogeneously trapped in the acrylic matrix, depending on the rates of reaction, the compatibility with the acrylic polymer, and the relative mobility of the dissolved polymer.

One way of influencing the way in which the added oligomer is distributed in the final cured produce is to provide reactive sites on the oligomer so that it can be incorporated into the acrylic matrix as it forms. For instance, if the added oligomer contains terminal, active, acrylic, or methacrylic unsaturation, it can be easily incorporated into the growing acrylic polymer chains as the adhesive cures. Then it is unable to precipitate as a separate phase and must remain more or less uniformly distributed throughout the matrix. On the other hand, incompatible polymers can be added to deliberately encourage the formation of reinforcing domains. If the oligomer has two or more functional groups capable of participating in the polymerization, there is at least the possibility of forming cross-links in the polyacrylate matrix.

A technology developed at Du Pont[32,33] combines the use of reactive sites on the oligomers with the initiation reaction. The resulting family of acrylic structural adhesives has become popularly known as "second generation acrylics." They consist essentially of solutions of chlorosulfonated polyethylene (Du Pont Hypalon™) in acrylic or methacrylic monomers. The sulfonic acid groups present on the polymer will react with amine aldehyde condensation products to generate free radical initiators, at least some of which are claimed to be on the oligomer backbone. The claim is that the adhesive polymer chains then grow by graft polymerization from various sites on the rubber oligomer. Peroxides and metal ions can be added to accelerate the initiation and polymerization. One of the significant advantages of this type of acrylic structural adhesive is the ability to bond oily or dirty metals with only minimal surface preparation, although some surface preparation usually is necessary to improve bond durability.

Other acrylic structural adhesives can be used on poorly prepared metal surfaces. Inclusion of reactive monomers having phosphate groups can lead to significant improvements in primary bonds to many unprepared metals.[34]

Another innovative approach to initiating the polymerization of acrylic structural adhesives is illustrated by a family of adhesives that uses the adherend surface itself as part of the chemical reaction that generates the free radical initiator.[35,36] The compositions contain an acidic material, a sulfonyl halide group, and a transition metal compound. Upon contact with a metal surface such as iron, zinc, copper, cadmium, and their alloys, the components react with the metal surface through a free radical pathway, producing effective initiators for the polymerization of the acrylic monomers. Later extensions of this concept provide two-part adhesives employing the same initiation chemistry,[37] in which a metal powder is included in one portion of the adhesive.

Summary

Acrylic structural adhesives have the advantage of easy handling and processing and rapid cure at room temperature. The ability of some compositions to adhere to unprepared metal surfaces can allow substantial cost savings in manufacturing processes. The brittleness of the polymers made from the monomers often can be overcome by formulating to include tough, reactive oligomers in the liquid adhesive. The variety of possible initiators and mechanisms for generating them has brought many innovative compositions to market.

< previous page page_235 next page >

20.6 Hybrid Adhesives

Adhesives usually are classified as epoxy, urethane, or acrylic, based upon the chemistry used to bring about the polymerization of the liquid adhesive. However, the distinction among the various types of adhesives is not always clear, and many hybrid adhesives have been developed and marketed. The objective usually is to take advantage of the desirable mechanical properties or chemical resistance of one polymer while retaining the processing attributes of a different cure system.38

It is common in acrylic structural adhesives to use oligomers that have a desirable backbone and are terminated with freeradical-polymerizable bonds. A variety of isocyanate-terminated polyurethanes can be adapted to use in acrylic structural adhesives by reacting the terminal isocyanates with a hydroxy functional acrylic monomer such as 2-hydroxyethyl methacrylate.34,39,40

Sometimes the system is formulated so that multiple cure mechanisms are possible and can occur sequentially or simultaneously. Compositions that rely on both epoxy and urethane chemistry are examples.27,41,42 These are compositions containing the diglycidyl ether of bisphenol A, an isocyanate or isocyanate-terminated prepolymer, amines or other reactants for either epoxy or isocyanates, and catalysts.

< previous page page_235 next page >

20.7 Evolution

The fundamental chemistry of the structural adhesives described here can change very little. Vinyl and acrylic monomers polymerize by chain growth polymerization initiated by free radicals or ions. Isocyanate and epoxy compounds react with compounds containing active hydrogen in step growth polymerizations. Epoxy containing compounds can undergo chain growth polymerization initiated by certain Lewis acids and Lewis bases. These reactions will remain the most important ones for the polymerization of these raw materials.

Most of the raw materials used in large quantities in structural adhesives are used because they are widely available, relatively safe and inexpensive. Quite often they are made in large quantities for uses other than adhesives and the adhesive manufacturers have taken advantage of supply and price. Few new basic raw materials are being developed specifically for the adhesive industry although the traditional raw materials are being combined in new ways to enhance desired adhesive properties. New initiators, adhesion promoters, primers, and specialty chemicals are being developed for use in small quantities to provide wider application latitude and improved performance.

The interest in adhesives and adhesive joining continues to grow. A search of the U.S. Patent and Trademark Office patent files from 1996 to 2002 produces nearly 1500 patents associated with acrylic, urethane, and epoxy adhesive compositions. The overwhelming majority of these refer to uses of adhesive compositions to produce commercially useful structures and products. Many refer to novel application methods. A significant number do refer to chemical innovations in adhesive compositions to provide improved or specialized products.

There are many examples of the evolution of epoxy structural adhesives. The development of heat activated epoxy adhesives for use with induction heating apparatus43 can lead to increased assembly line productivity. Epoxy adhesives suitable for use on oily metal44 can reduce the number of manufacturing operations. Adhesion promoters such as dithiooxamides can be included in epoxy formulations to improve adhesion and durability.45 Epoxy resins have been modified with phosphorous46 to introduce flame retardance. Siloxanes have been used to modify both epoxy adhesives and adherends in order to improve adhesion and durability.47 Epoxy adhesives are being packaged in novel ways such as forming the two parts of the adhesive into sheets and interleaving them to produce a room temperature stable, heat curable construction.48

Urethane adhesives have also received much attention. New processes are being developed for the production of isocyanate containing raw materials49 and methods of stabilizing them.50 Formulators continue to produce new mixtures and intermediates from standard materials.51,52 New blocking agents for isocyanates are being developed to facilitate the formulation of stable urethane adhesives compositions rapidly curable when heated.53,54

Innovations in acrylic adhesive have led to new initiator reactions which provide rapid initiation and cure while at the same time allowing for improved shelf life.55 Some of the inherent brittleness of acrylic adhesives is being overcome by the inclusion of new toughening agents and better control of the molecular weight of mixtures of toughening agents.56

Some of the more interesting and innovative work has occurred in areas combining aspects of more than one chemistry type. For instance, moisture curable thermoplastic adhesives have received much attention. Hot melt adhesives have been developed which contain active, moisture curable isocyanate groups. The compositions provide rapid processing on assembly lines since a reasonable bond is formed as soon as the thermoplastic adhesive cools from the melt. However, bond strength and performance improve with time as the composition is slowly crosslinked to a thermoset by reaction of the isocyanates with atmospheric moisture.57,58

In combining urethane and acrylic technology, adhesives cured by ultraviolet light have been made from mixtures of acrylic functional urethane prepolymers mixed with free radically polymerizable monomers and photo initiators.59,60

Acrylic adhesives have been modified by including cyanate esters and organometallic compounds to provide an additional crosslinking mechanism and thereby improved resistance to heat.61

A very interesting initiator system has recently been patented which seems to allow for a great deal of flexibility in adhesive chemistry and formulation.6264 The initiator contains organoborane polyoxyalkylenepolyamine complexes. It can be used in concert with other reactants such as an organic acid to initiate polymerization of acrylic adhesive compositions.65 An interesting facet of the initiator is that it can be used in compositions which also contain materials reactive with amines, such as isocyanate groups or, presumably, epoxy groups in such a way as to provide simul-taneous polymerizations involving differing adhesive chemistries.

The evolution of structural adhesives will certainly continue. Each increment in strength, durability, processing speed and ease, safety, reliability and reproducibility opens new commercial markets, not only to displace older joining methods but also to allow for the manufacture of new structures not possible without adhesives.

20.8 Conclusion

Epoxy, urethane, and acrylic structural adhesives have been commercially successful because each can be used, under the proper conditions, to make reliable, durable, and useful adhesively joined assemblies. The adhesives are classified according to the polymerization reactions used to bring the liquid adhesive to a high-molecular-weight load- bearing state. The reactive sites on the uncured adhesive and the overall polarity of the cured adhesive are important in the adhesion of the polymer to any specific material. The polymerization reactions determine the processing requirements and are important to the mechanical properties and environmental resistance of the cured adhesive. Through the use of specific reactive oligomers it is possible to build certain molecular structures into the composition that will remain in the cured adhesive to provide desired mechanical or chemical properties.

The successful use of adhesive joining in producing any assembly depends on viewing the assembly as a whole from the very beginning of its design. The key is remembering that the adhesive is only one component of the assembly. Adhesion is a property of the whole assembly.

20.9 References

Patrick, R. L. (Ed.), Treatise on Adhesion and Adhesives, Vol. 1, Theory, Marcel Dekker, New York, 1967.
Patrick, R. L. (Ed.), Treatise on Adhesion and Adhesives, Vol. 2, *Materials*, Marcel Dekker, New York, 1969.
Kaelble, D. H., Physical Chemistry of Adhesion, Wiley-Interscience, New York, 1971.
Patrick, R. L. (Ed.), Treatise on Adhesion and Adhesives, Vol. 3, Marcel Dekker, New York, 1973.
Skeist, L. (Ed.), Handbook of Adhesives, 2nd ed., Van Nostrand Reinhold, New York, 1977.
Wake, W. C., Adhesion and the Formulation of Adhesives, 2nd ed., Applied Science, New York, 1982.
Lee, L.-H., Adhesive Chemistry, Developments and Trends, Plenum Press, New York, 1984.
Panek, J. R., and Cook, J. P., Construction Sealants and Adhesives, 2nd ed., Wiley-Interscience, New York, 1984.
Adams, R. D., and Wake, W. C., Structural Adhesive Joints in Engineering, Elsevier, New York, 1984.
Hartshorn, S. R. (Ed.), Structural Adhesives Chemistry and Technology, Plenum Press, New York, 1986.
Kinlock, A. J., Adhesion and Adhesives Science and Technology, Chapman and Hall, London, 1987.
Mark, H. F., Future Improvements in the Cohesive and Adhesive Strength of Polymers-Part I, Adhes. Age, 22(7), 3540 (1979).
Mark, H. F., Future Improvements in the Cohesive and Adhesive Strength of Polymers-Part II, Adhes. Age, 22(9), 4550 (1979).
Lee, H., and Neville, K., Handbook of Epoxy Resins, McGraw-Hill, New York, 1967.
Castan, P., Process of Preparing Synthetic Resins, U.S. Patent No. 2,324,483, 1943.
Graham, J. A., and O'Connor, J. E., "Epoxy with Low-Temperature Cure and High-Temperature Properties Developed," Adhes. Age, 21(7), 2023 (1978).
Crivello, J. V., Heat Curable Compositions, U.S. Patent No. 4,173,551, 1979.
Crivello, J. V., Heat Curable Cationically Polymerizable Compositions and Method of Curing same with Onium Salts and Reducing Agents, U.S. Patent No. 4,216,288, 1980.
Crivello, J. V., UV Curable Compositions and Substrates Treated Therewith, U.S. Patent No. 4,319,974, 1982.
Crivello, J. V., and Ashby, B. A., Methods of Adhesive Bonding using Visible Light Cured Epoxies, U.S. Patent No. 4,356,050, 1982.
Crivello, J. V., and Lee, J. L., Photocurable Compositions, U.S. Patent No. 4,442,197, 1984.
Tarbutton, K. S., and Robins, J., Acid Catalyzed, Toughened Epoxy Adhesives, U.S. Patent No. 4,846,905, 1989.
Bayer, O., German Patent No. 728.981, 1937.
Saunders, J. H., and Frisch, K. C., Polyurethanes, Chemistry and Technology, Vols. 1 and 2, Interscience, New York, 1962, 1964.
Oertel, G. (Ed.), Polyurethane Handbook, Hanser, Munich, 1985.
Fabris, H. J., Maxey, E. M., and Uelzmann, H., Urethane Adhesive having Improved Sag Resistance, U.S. Patent No. 3,714,127, 1973.
Goel, A. B., Sag Resistant Urethane Adhesives with Improved Antifoaming Property, U.S. Patent No. 4,728,710, 1988.
Baueriedel, H., Adhesives based on Polyurethane Prepolymers having a Low Residual Monomer Content, U.S. Patent No. 4,623,709, 1986.
Ono, H.-K., Jones, F. N., and Pappas, S. P., "Relative Reactivity of Isocyanate Groups of Isophorone Diisocyanate. Unexpected High Reactivity of the Secondary Isocyanate Group," J. Polym. Sci., Polym. Lett. Ed., 23, 509515 (1985).
Arendt, V. D., Logan, R. E., and Saxon, R., "*m*- and *p-TMXDI*: Two New Isocyanates for the Polyurethane Industry," J. Cell. Plast., 376383 (1982, November/December).
Nemec, J. W., and Bauer, W. Jr., "Acrylic Acid and Derivatives," in Encyclopedia of Chemical Technology, Vol. 1, Martin Grayson (Ed.), pp. 330354, Wiley-Interscience, New York, 1978.
Nemec, J. W., and Kirch, L. S., "Methacrylic Acid and Derivatives," in Encyclopedia of Chemical Technology, Vol. 15, Martin Grayson (Ed.), pp. 346376, Wiley-Interscience, New York, 1978.
Briggs, P. C., Bonding Method Employing Novel Methacrylate based Adhesive Compositions: Chlorinated or Chlorosulfonated Polyethylene as Adhesion Promoter, U.S. Patent No. 4,773,957, 1988.
Zalucha, D. J., Sexsmith, F. H., Hornaman, E. C., and Dawdy, T. H., Structural Adhesive Formulations, U.S. Patent No. 4,223,115, 1981.
Daico, D. J., Structural Adhesive Formulations, U.S. Patent No. 4,703,089, 1987.
Damico, D. J., and Bennett, R. M., Structural Adhesive Formulations and Bonding Method Employing Same, U.S. Patent No. 4,855,001, 1989.
Damico, D. J., Mushrush, K. W., and Bennett, R. M., Two-Part Adhesive for Bonding Plastics and Metals, U.S. Patent No. 4,857,131, 1989.
Pohl, E., and Osterholz, F. D., U.S. Patent No. 4,645,816, 1987, February 24.
Brownstein, A. M., U.S. Patent No. 3,428,614, 1969.
Su, W.-F. A., U.S. Patent No. 4,618,632, 1986, October 21.
Hawkins, J. M., Epoxy Resin Adhesive Compositions containing an Isocyanate Terminated Polyurethane Prepolymer and a Chain Extender, U.S. Patent No. 3,636,133, 1972.
Trieves, R., and Pratley, K. G. M., U.S. Patent No. 4,623,702, 1986, November.
Jorissen, S. A., Ferguson, G. A., and Imirowicz, K., Epoxy Compound Blend with Di(aminoalkyl) Ether of Diethylene Glycol. U.S. Patent No. 5,548,026, 1996.
Baldwin, J. M., and Robins, J., Epoxy Adhesive Composition Comprising a Calcium Salt and Mannich Base, U.S. Patent No. 5,629,380, 1997.
Markevka, V. C., Griggs, A. L., and Tarbutton, K. S., Epoxy Adhesives with Dithiooxamide Adhesion Promoters, U.S. Patent No. 5,712,039, 1998.
Harold, S., and Schmitz, H.-P., Phosphorus-modified Epoxy Resins Comprising Epoxy Resins and Phosphorus-containing Compounds, U.S. Patent No. 5,830,973, 1998.
Mowrere, N. R., Kane, J. F., and Hull, C. G., Siloxane-modified Adhesive/Adherend Systems, U.S. Patent No. 5,942,073, 1999.
Rijsdijk, H., Overbergh, N., DeBlick, G., Miles, G., and Kennan, A., Curable Adhesive System, U.S. Patent No. 5,952,071, 1999.
Okawa, T., Process for Producing Isocyanate Compound, U.S. Patent No. 5,166,414, 1992.
Nagata, T., Yamashita, H., Kusumoto, M., and Okazaki, K., Stabilizing Method of Isocyanate Compounds and Isocyanate Compositions Stabilized thereby, U.S. Patent No. 5,302,749, 1994.
Ohashi, Y., Matsuda, H., Nishi, E., Nishida, T., Moisture Curing Urethane Adhesive Composition, U.S. Patent No. 5,698,656, 1997.
Fukatsu, S., and Hattori, Y., Moisture Curable Polymer Composition and Process for Production thereof, U.S. Patent No. 5,767,197, 1998.
Schoener, T. E., and Housenick, J. B., Polyurethane Reaction System having a Blocked Catalyst Combination, U.S. Patent No. 6,348,121, 2002.
Ambrose, R., Retsch, W. R., Jr., and Chasser, A., Blocked Isocyanate-based Compounds and Compositions Containing the Same, U.S. Patent No. 6,288,199, 2001.
Edelman, R., and Catena, W., Rapid Curing Structural Acrylic Adhesive, U.S. Patent No. 5,865,936, 1999.
Huang, J., Righettini, R. F., and Dennis, F. G., Adhesive Formulations, U.S. Patent No. 6,225,408, 2001.
Anderson, G. J., and Zimmel, J. M., Therrmally Stable Hot Melt Moisture-cure Polyurethane Adhesive Composition, U.S. Patent No. 5,939,499, 1999.
McInnis, E. L., Santosusso, T. M., and Quay, J. R., Hot Melt Adhesives Comprising Low Free Monomer, Low Oligomer Isocyanate Prepolymers, U.S. Patent No. 6,280,561, 2001.
Usifer, D. A., and Broderick, I. C., Urethane Adhesive Compositions, U.S. Patent No. 5,426,166, 1995.
Usifer, D. A., and Broderick, I. C., Urethane Adhesive Compositions, U.S. Patent No. 5,484,864, 1996.
McCormick, F. B., Drath, D. J., Gorodisher, I., Kropp, M. A., Palazzotto, M. C., and Sahyun, M. R. V., Energy-curable Cyanate/Ethylenically Unsaturated Compositions, U.S. Patent No. 6,029,219, 2000.
Pocius, A. V., Organoborane Polyoxyalkylenepolyamine Complexes and Adhesive Compositions Made Therewith, U.S. Patent No. 5,681,910, 1997.
Pocius, A. V., Organoborane Polyamine Complexes and Adhesive Compositions Made Therewith, U.S. Patent No. 5,795,657, 1998.
Pocius, A. V., Organoborane Polyamine Complex Initiator Systems and Polymerizable Compositions Made Therewith, U.S. Patent No. 6,093,778, 2000.
Zahrov, J. V., and Krasnov, J. N., Polymerizable Compositions made with Polymerization Initiator Systems based on Organoborane Amine Complexes, U.S. Patent No. 5,690,780, 1997.

21.1 Textile Background

The first conversion of naturally occurring fibers into threads strong enough to be looped into snares, knitted to form nets, or woven into fabrics is lost in prehistory. Unlike stone weapons, such threads, cords, and fabricsbeing organic in naturehave in most part disappeared, although in some dry caves traces remain. There is ample evidence to indicate that spindles used to assist in the twisting of fibers together had been developed long before the dawn of recorded history. In that spinning process, fibers such as wool were drawn out of a loose mass, perhaps held in a distaff, and made parallel by human fingers. (A maidservant so spins in Giotto's *The Annunciation to Anne*, ca. A.D. 1306, Arena Chapel, Padua, Italy.1) A rod (spindle), hooked to the lengthening thread, was rotated so that the fibers while so held were twisted together to form additional thread. The finished length then was wound by hand around the spindle, which, in becoming the core on which the finished product was accumulated, served the dual role of twisting and storing, and, in so doing, established a principle still in use today. (Even now, a "spindle" is 14,400 yards of coarse linen thread.) Thus, the formation of any threadlike structure became known as spinning, and it followed that a spider spins a web, a silkworm spins a cocoon, and manufactured fibers are spun by extrusion, although norotation is involved.

It is not surprising that words from this ancient craft still carry specialized meanings within the textile industry and have entered everyday parlance, quite often with very different meanings. Explanations are in order for some of the words used in the following pages. For example, as already indicated, "spinning" describes either the twisting of a bundle of essentially parallel short pieces of wool, cotton, or precut manufactured fibers into thread or the extrusion of continuous long lengths of manufactured fibers. In the former case, the short lengths are known as "staple" fibers, and the resulting product is a "spun yarn," whereas the long lengths are called "continuous filament yarn," or merely "filament yarn." Neither is called a "thread," for in the textile industry that term is reserved for sewing thread and rubber or metallic threads. Although to the layperson "yarn" connotes a material used in hand knitting, the term will be used in the textile sense hereinafter.

Before manufactured fibers are discussed, it is necessary to define some terms.* The "denier" of a fiber or a yarn defines its linear density, that is, the mass in grams of a 9000 m length of the material at standard conditions of 70°F and 65 percent relative humidity. Although denier is actually a measure of linear density, in the textile industry, the word connotes the size of the filament or yarn. Fibers usually range from 1 to 15 denier, yarns from 15 to 1650. Single fibers, usually 15 denier or larger, used singly, are termed "monofils." The cross-sectional area of fibers of identical deniers will be inversely related to their densities, which range from 0.92 g/cc for polypropylene to 2.54 g/cc for glass. The approximate densities of some of the other commonly used fibers are given in Table 21.1. Because by definition denier is measured at standard conditions, it describes the amount of "bone-dry" material plus the moisture regain, which ranges from zero for glass and polypropylene to 14 percent for rayon. It should be mentioned that some years ago scientific organizations throughout the world accepted the word "tex," this being the mass (g) of 1 km of the material, as a more useful term than denier. "Tex" is an accepted adjunct to the SI, or International System of Units, but it has received only limited acceptance in commerce, whereas the SI units are being employed increasingly in scientific organizations. Furthermore, the sizes of cotton, wool, and worsted yarns, and yarns containing manufactured fibers but produced by the traditional cotton, wool, or worsted systems, still are expressed in the inverse-count system that has been used for centuries.

TABLE 21.1 Typical (Average) Values of Tensile and Physical Properties of Some Textile Fibers

Fiber	Breaking Stress (cN/tex)	Strain to Fail (%)	Moisture Regain (%)	Density (g/cc)
Natural				
1. Cotton	1. 40	1. 7	1. 8	1. 1.52
2. Flax	2. 54	2. 3	2. 12	2. 1.52
3. Silk	3. 38	3. 23	3. 11	3. 1.33
4. Wool	4. 14	4. 40	4. 14	4. 1.30
Regenerated				
1. Acetate	1. 13	1. 25	1. 6	1. 1.32
2. Rayon	2. 25	2. 20	2. 14	2. 1.51
Synthetic				
1. Acrylic	1. 26	1. 25	1. 1.5	1. 1.18
2. Modacrylic	2. 28	2. 32	2. 1.5	2. 1.32
3. Nylon	3. 50	3. 25	3. 4	3. 1.14
4. Polyester	4. 50	4. 15	4. 0.4	4. 1.38
5. Polypropylene	5. 53	5. 17	5. 0	5. 0.92

*Each year the ASTM publishes in its *Book of Standards*, the most recent and accepted definition and test methods used in the textile and fiber industries.

The "breaking tenacity" or more commonly, "tenacity," is the breaking strength of a fiber or a yarn expressed in force per unit denier, that is, in grams per denier, calculated from the denier of the original unstretched specimen. "Breaking length" expresses the theore-tical length of yarn that would break under its own weight, and is used mostly in Europe. "Elongation" means "breaking elongation" and is expressed in units of increase in length to break calculated as a percentage of the original specimen length.

Typical forceelongation curves of some manufactured and natural staple fibers and textile-type manufactured filaments are shown in Figs 21.1 and 21.2. Table 21.1 gives the values of some of the physical and tensile properties of textile fibers.

Fig. 21.1 Forceelongation curves of natural and manufactured staple fibers at standard conditions of 70°F and 65 percent humidity.

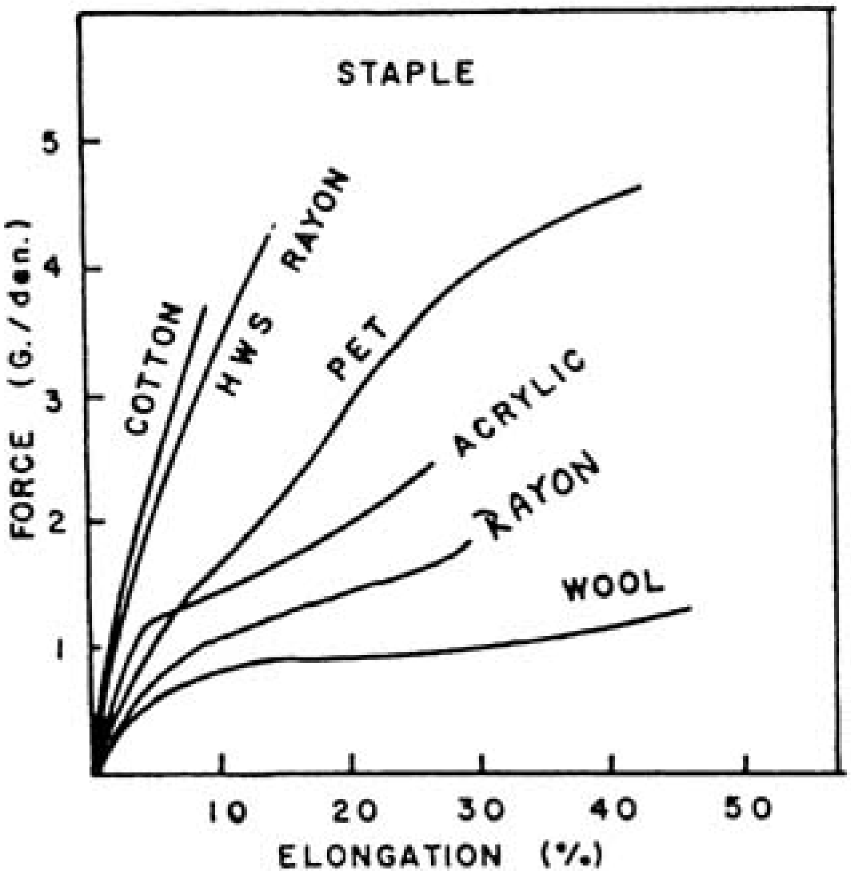
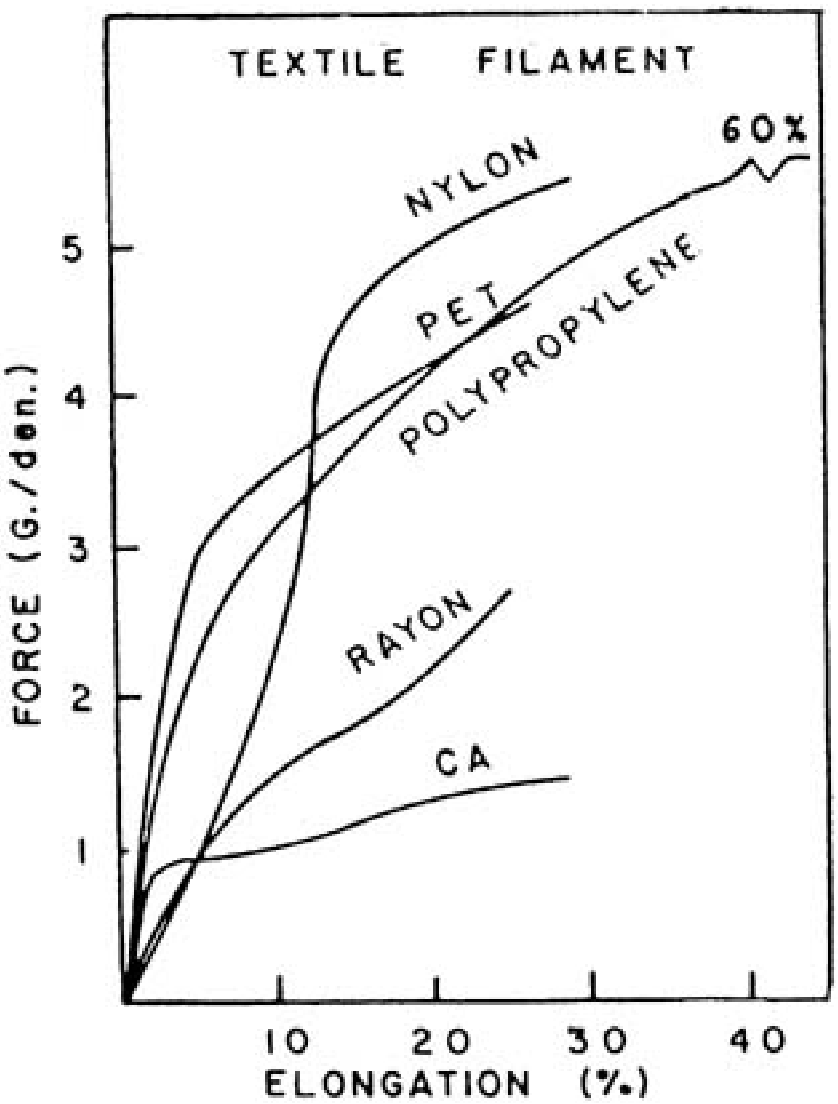


Fig. 21.2 Forceelongation curves of manufactured textile continuous filament yarns at standard conditions of 70°F and 65 percent relative humidity.



< previous page

next page >

21.2 History

Early humans, over time, became aware of the presence and usefulness of fibrous materials available from such sources as seeds, leaves, stems, animal coverings, and cocoons. They learned to spin, weave, knit, felt, or braid these fibers to protect their bodies and improve their lifestyle. A few hundred years ago, it has been suggested, someone took a clue from a busy worm and thought that it should be possible for humans to make a silk-like fiber that would be of commercial value. Curiosity combined with simple experiments strengthened that premise and much later led to the development of viscose rayon, only partially a manufactured fiber, in the 19th century. Further curiosity, war-time need, and superior commitment by modern researchers led to the synthesis of a totally synthetic fiber in the 1930s. The knowledge gained and the spark thus ignited resulted in the development of other fibers to such an extent that manufactured fibers now dominate the market in the industrialized nations in all major categories of apparel, home furnishing, and industrial end uses.

The story of the development of manufactured fibers is of great historical interest, beginning in 1664 when Robert Hooke, an Englishman, suggested that it should be possible to make a fiber much like silk that could be of value in the market place. Andemars, a Swiss chemist, received the first patent for making silk-like fiber in 1855. He drew fibers by dipping a needle and pulling it out from a solution of cellulose nitrate containing some rubber.

The credit for using a spinnerette and forcing a solution through it for producing a fiber, however, goes to the English scientist Sir Joseph W. Swan, in the early 1880s. The first person to put the idea into commercial practice was the French chemist Count Hilaire de Chardonnet, who built the first plant to commercially produce a fiber based on regenera-ted cellulose, called "artificial silk," at Besancon, France, in 1891. During the last years of the 19th century and the beginning of the 20th century, progress was so rapid that the production of this fiber increased from several thousand pounds in 1891 to over two million in 1910. Commercial production of the fiber in the United States began in 1910 with the opening of the first plant. In 1924, the industry gave the "artificial silk" fiber a new name, rayon.

By that year, 1910, the brothers Camille and Henry Dreyfus had discovered a practical method for producing cellulose acetate polymer and were making plastic film and toilet articles in Basel, Switzerland. During World War I, they built a plant in England to produce acetate dope for painting airplane wings to render them air-impervious. The success of the product led the U.S. government to invite the Dreyfus brothers to build a plant in the United States, which started commercial production in 1924.

The successful manufacture of these two fibers, although based on fibrous materials available in nature, marked the beginning of the development of manufactured fibers in the 1930s. This effort, initiated by a technological breakthrough, was marked by the work of W. H. Carothers, aimed at learning how and why certain molecules joined to form large molecules, or polymers.2 Fibers were described as being composed of high molecular weight linear polymers; and the first one to be manufactured, nylon 66, was synthesized and produced on a commercial scale in 1939. It was quickly followed by nylon 6, the second most widely used nylon, and modacrylic (1949), olefin (1949), acrylic (1950), polyester (1953), and triacetate (1954). Glass had joined this group of large-production items earlier in 1936.

Several other fibrous materials have been produced, but they are regarded as fibers with special performance characteristics, used either in limited textile or specialized industrial applications. Some of these worth noting are Spandex (1959), Aramid (1961), polybenzimidazole (PBI) (1983), and Sulfar (1983).

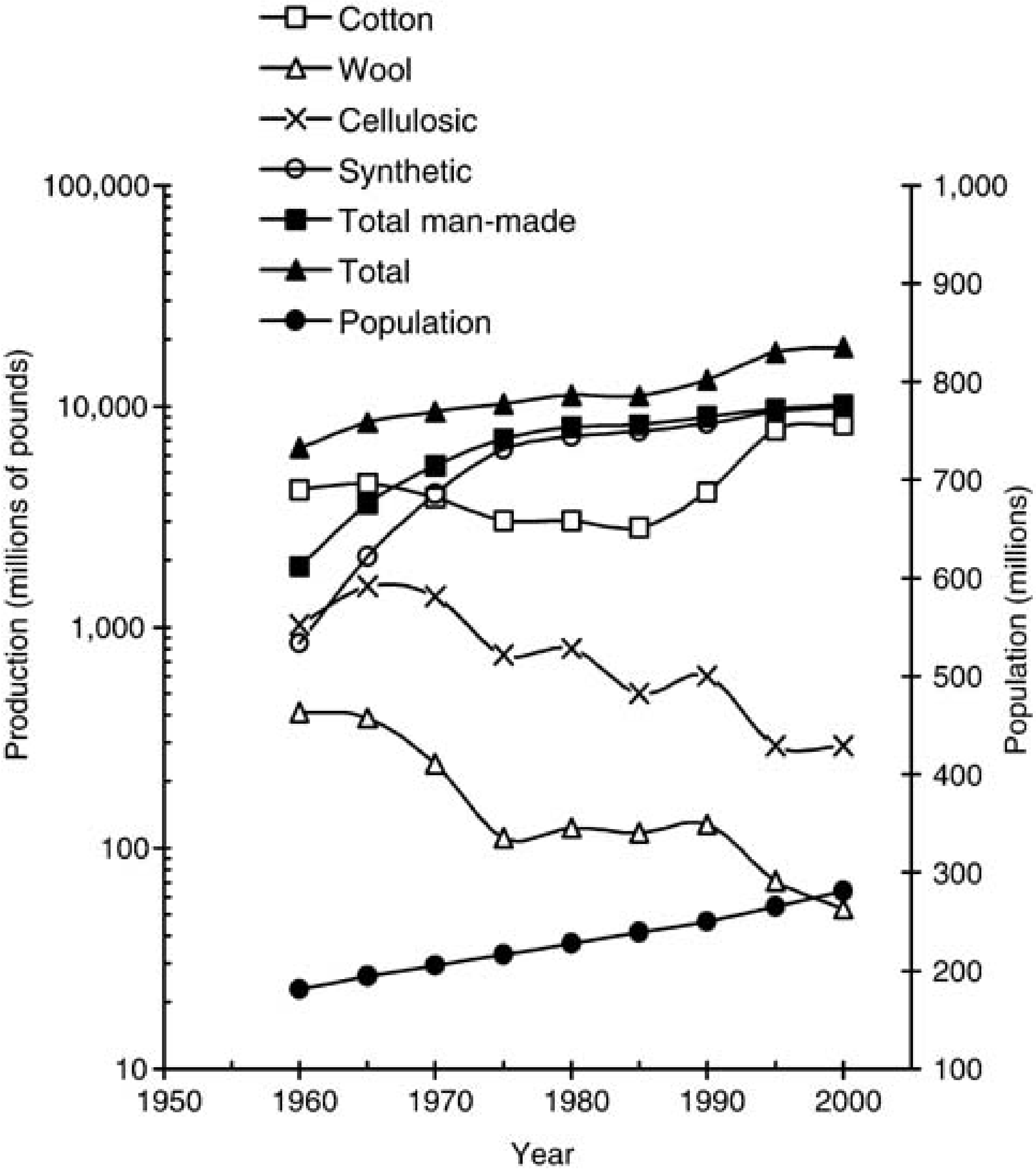
Thus, the period from the 1930s to the 1960s can be considered as a time of discoveries and innovations for manufactured fibers, when the majority of the basic fibers were developed. The years since then may be thought of as a period of modification of performance characteristics. The basic generic materials have been manipulated both chemically and physically to produce a wide variety of different fibers, tailored to secure the desired characteristics for specific end-use products. Thus, fibers can be extruded in different shapes and sizes for special purposes. They can be modified to offer greater comfort, flame resistance, or static-free behavior in apparel; they can offer soil-release and other desirable characteristics for carpets; they can be developed with unique surface characteristics, easier dyeability, or better blending qualities. The industry has begun to discover many possibilities for modifying the behavior of a given fibrous material. It has learned how to produce new fibers with greater strength, greater thermal resistance, or other special qualities.

Essentially, then, no new, large-volume, highly profitable fibers have been developed since the mid-1950s. Instead, the existing ones have become commodities with all the economic impact thereby implied. No major chemical engineering processes have been added, although the previously described ones have been modified to allow for spinning of liquid crystalline polymers or the formation of gel spun fibers. Research activity has been reduced and centered essentially on modifications of fiber size, shape, and properties, and many variants now are successfully marketed. Production volumes have increased enormously for nylon, polyester, and polyolefin.

21.3 Volume of Production

Figure 21.3 compares population growth with the production of manufactured fibers and the mill consumption of natural fibers in the United States. Per capita consumption of all fibers, starting at a level in the 1920s of about 30 lb, rose to approximately 40 lb following World War II and reached a level at or about 50 lb in the 1970s. It topped 60 lb in 1973, dropped below 45 lb in 1982, and in 1991 was at 55 lb. But clearly overshadowing the increases resulting from population growth and a higher standard of living are the volumes produced of, first, the cellulosic manufactured fibers and, second, the noncellulosic or completely synthetic fibers. The consumption of manufactured fibers increased from about 2 billion lb in 1960 to nearly 9 billion lb in 1990. During the 1980s, the increase was about 11 percent, made up of a decrease in cellulosics of about 200 million lb (25%) and an increase in synthetics of about 1.2 billion lb (17%). Of the 13.2 billion lb of fibers used in the United States in 1990, 4.6 percent was the manufactured cellulosics, 63.3 percent was the synthetics (67.9% manufactured), 31.1 percent was cotton, and less than 1 percent was wool. A significant increase was noted in the use of cotton during the 1980s, from 3.0 billion lb in 1980 to 4.1 billion lb in 1990, believed to be in large measure due to advances in chemical finishes that made caring for the natural fibers easier than it was with the older technology. To some extent, this increase also may have occurred because people were willing to accept a more wrinkled look in order to gain the comfort of hydrophilic fibers. Still, the general figures represent quite a reversal in the consumption of fiber types seen three or more decades ago. For example, in 1960, of the 6.5 billion lb of fibers used, manufactured fibers accounted for only 29 percent of the total, with cotton 65 percent and wool 6 percent.

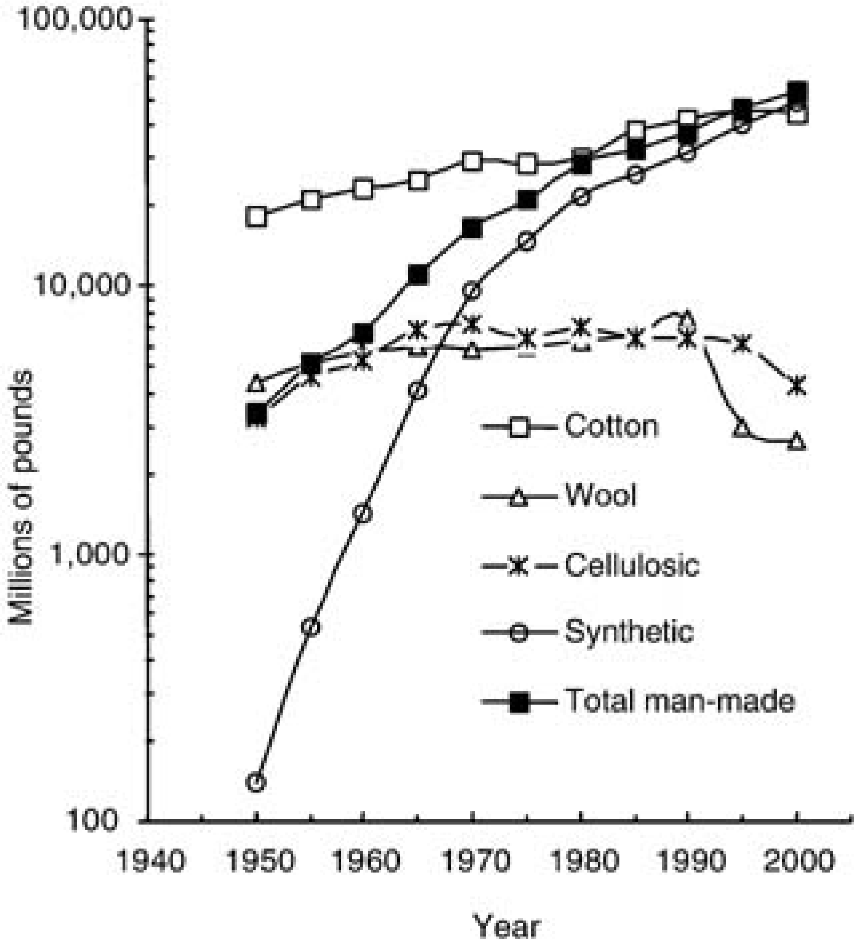
Fig. 21.3 Comparative growth of population and fiber consumption in the United States, 19602000.



Major applications of fibers lie in apparel, home furnishing, and industrial products. In each of these, manufactured fibers have made large inroads, and currently their usage dominates. As an illustration, consider the changes that have taken place in the use of the materials required in the manufacture of tire cords. Originally made from cotton, rayon took a commanding position during World War II. But as late as 1951, cotton comprised about 40 percent of the total output of tire cords of approximately half a billion lb, and nylon was at a negligible level of 4 million lb. By 1960, however, cotton had all but disappeared; nylon represented about 37 percent of the total (on a weight basis), even though only about 0.8 lb of nylon is needed to replace 1.0 lb of rayon. Whereas rayon for several years had dominated the so-called original-equipment tire market and nylon had held a corresponding position for replacement tires, more recently, glass and polyester have made heavy inroads into bothespecially in belted constructions. The situation continued to change in favor of noncellulosic manufactured fiber usage in tires, so that by 1972, rayon was down to 14 percent, nylon up to 42 percent, polyester up to 32 percent, glass up to 7 percent, and steel at 5 percent, all on a weight basis. By the late 1970s, tire markets were dominated wholly by manufactured fibers with polyester holding over 90 percent of the passenger car original-equipment market and nylon commanding over 90 percent of the truck original-equipment market. This division of markets is a direct result of the performance characteristics of the two fibers. Polyester-containing tires are free of "flat spotting" or cold-morning thump, and so are preferred in passenger cars for their smooth ride. On the other hand, nylon-containing tires are tougher and more durable, and so are the choice for trucks and off-road vehicles.

The production of manufactured fibers throughout the world has developed in a manner that rather parallels the situation in the United States, as may be seen in Fig. 21.4. There are some expected differences, and obviously the data for world usage are strongly influenced by the large components attributable to the United States, which currently accounts for about 23 percent of the manufactured fiber and about 15 percent of the total fiber consumption. The output of the world cellulosics has leveled off, but expansion of the noncellulosics has continued unabated. The use, or at least the recorded use, of the natural fibers, cotton and wool, rose rapidly in the 1950s, as the world economy rebounded at the conclusion of World War II. Since then (19601970), a modest increase has continued, essentially parallel to the growth of world population. But, in comparison with population trends, it appears that the great demand has been for manufactured textile fibers. Much of this increase has resulted from an improved standard of living and the absence of major wars.

Fig. 21.4 Production of fibers in the world, 19502000.



A detailed economic examination of the processing of fibers and the changes that have taken place during the last half century would show two rather vivid occurrences. The first of these is a rapid decrease in the prices of the newer fibers as they became established, followed by a leveling out and stabilization. The second is the relative stability of prices of the manufactured fibers on short-term and even long-term bases, as compared with fluctuations in the prices for the natural fibers where governmentally imposed stability has not been in effect. Data are not presented about it in this text, but in the first half of the 20th century there was a saying in the textile industry that the person who made or lost money for the company was the one who was responsible for buying cotton and wool "futures." However, it should also be emphasized that list prices of manufactured fibers are ceiling prices and do not reflect the short-term discounts, allowances, and special arrangements that are given in a free marketplace when the demand for any manufactured fiber softens.

A presentation of complete information about the consumption of raw materials, chemical reactions, reagents and catalysts used, and efficiencies of operation in the production of manufactured fibers undoubtedly would contribute to a better understanding of the industrial chemistry involved. Several factors have prevented this, however. In the first half of the 20th century, a historical belief in the efficacy of trade secrets still permeated the chemical industry. Even with the increased mobility of technical and scientific personnel during and following World War II, the idea still prevailed that if nothing other than patents was allowed to become public knowledge, so much the better. The situation has changed considerably since about 1960, as can be noted from the availability of information contained in the list of suggested readings that follows this chapter; yet, secrecy tends to be maintained despite the fact that key employees move from company to company, and the chemical engineering knowledge available in chemical companies that produce large volumes of fibers permits an almost complete appraisal of a competitor's activities.

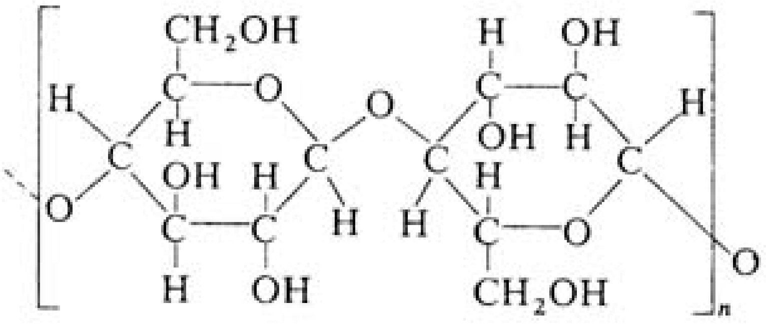
In general, in the early period of production of a fiber, the cost of the original raw material may have had very little bearing on the selling price of the final fiber. A most important factor is the action of the producer's competitors and the conditions of the market and the demand that can be developed. But the complexity of the processes involved in conversion determines the base cost of the fiber at the point of manufacture. As the process becomes older, research reduces this complexity; with simplification, there may be rapid drops in plant cost. If demand remains high, such reductions will not be expected to be reflected in selling prices; rather, profits are high. As more producers enter the field in order to share in those profits, output capacity surpasses demand, and in accordance with classical economic theory, major selling price reductions result. This was happening, in general, in the 1960s and for cellulose-based manufactured fibers in the 1970s. But beginning in 1973, the cost of petroleum-based products started to rise steeply and erratically. This rise was based not on economic considerations alone but, on political considerations among the oil producing and exporting countries (OPEC) as well. Further upward pressure on manufactured fiber prices has resulted from governmental limitations placed on chemical usage and exposure and on amounts of chemicals that can be discharged into the air and water. To meet these limitations, the manufactured fiber industry has had to supply large infusion of capital. In some instances, such expenditures could not be justified, and plant capacity was shut down permanently. This was particularly true in the case of filament rayon. In recent years, the factors of rapidly rising raw material/energy prices and the costs of meeting environmental regulations have not allowed the prices of manufactured fibers to fall as production experience has been gained and technological advances have been introduced. Instead, selling prices have been continually adjusted upward in an effort to pass along unavoidable cost increases so as to maintain profitability. In areas of application in which a manufactured fiber replaced a natural one because of lower prices and stable availability, swings in fashion and increases in imports sometimes have caused a reduction in fiber utilization. This scenario, combined with environmental concerns, is believed to have particularly applied to acrylic fiber, whose production has decreased recently in both Europe and the United States.

The great importance of manufactured fibers in the chemical industry and in the overall economy of the United States (and, in general, the developed countries) becomes apparent when the volume of production of these materials is considered and compared with the market value of even the least expensive of the raw materials used by them. The amounts of oil and natural gas consumed by the manufactured fiber industry represent around 1 percent of national annual usage. Of this amount, about one half is used to produce raw materials from petrochemicals, with the other half used for energy to convert trees to wood pulp for cellulose-based fibers and to convert the wood pulp and petrochemical-derived raw materials to fibers.

21.4 Rayon

Chemical Manufacture

Rayon, the first of the manufactured fibers produced in large volume, is based on the natural polymer cellulose, a repeat unit of which is shown below:

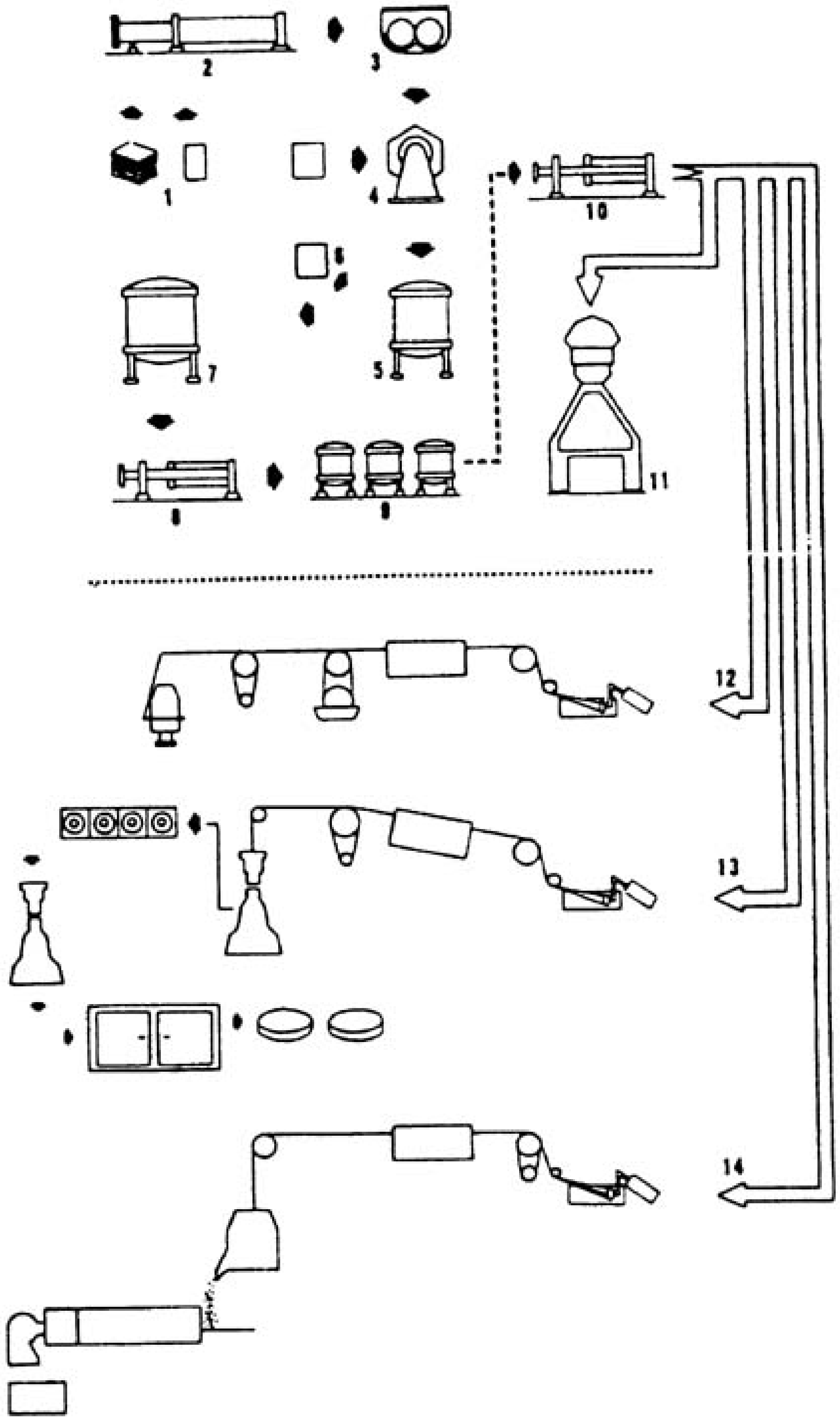


Two anhydroglucose units

Although in the early days the main source of this raw material was cotton linters, a combination of improved technologies for obtaining alpha cellulose from wood and the shortage of cotton linters used for the manufacture of cellulose nitrate during World War II resulted in a shift in raw material to wood pulp. Only certain trees constitute the most economical supply of dissolving pulp, as the final product is called; the process economics depend upon the cost of logs delivered at the pulping mill and the relative yield of alpha cellulose after the unusable lignin and the other components of the wood are discarded.

A general flow diagram for the manufacture of rayon is given in Fig. 21.5. The dissolving pulp is received by the rayon manufacturer in sheet or roll form. In the manufacturing process, impurities are removed, with special attention being given to removal of traces of such metallic elements as manganese and iron, the former having an effect on the manufacturing process (as will be noted later) and the latter an effect on the color of the final product. The production of dissolving pulp involves drastic chemical action at elevated temperatures, which substantially reduces the originally very high molecular weight of the cellulose. The portion not soluble in 1718 percent aqueous caustic, known as alpha cellulose, remains, and the lower-molecular-weight beta and gamma fractions are largely soluble and lost. The composition of the pulp is aimed at high alpha content. A typical economic trade-off is involved. The pulp producers can secure an alpha content of up to 98 percent by means of a cold caustic extraction, or, on the other hand, the rayon manufacturer can use a less expensive, lower alpha content pulp (9096%) and expect to secure a lower yield. The sellers have numerous grades available to meet the specific process needs and end-product requirements of each of the buyers.

Fig. 21.5 Flow diagram for manufacture of viscose yarn: (1) cellulose sheets and caustic soda; (2) steeping press; (3) shredder; (4) xanthating churn; (5) dissolver; (6) caustic supply; (7) ripener; (8) filtration; (9) deaeration; (10) filtration; (11) continuous process; (12) tire cord; (13) pot spinning; (14) staple spinning.

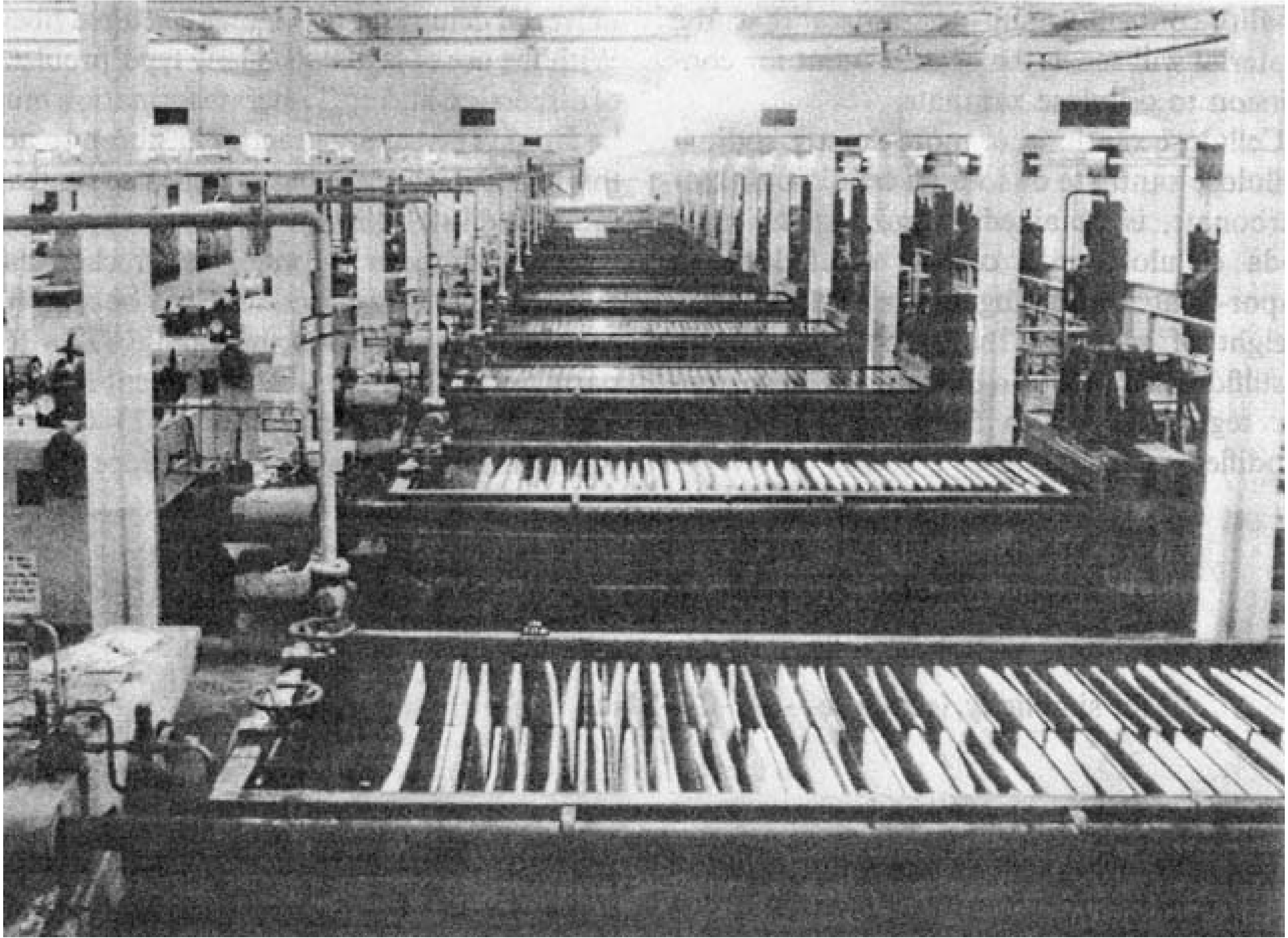


In the manufacture of rayon, it is the usual practice to begin "blending" at the first step, which involves steeping the pulp. Further blending proceeds throughout successive steps. The warehouse supply of pulp consists of numerous shipments, and in making up the batches for the conventional process, a few sheets are taken from each of several shipments. This serves two purposes. It prevents a slight variation in a single pulp lot from unduly affecting any given volume of production, and it provides a moving average so that changes with time are reduced to a minimum.

The cellulose sheets are loaded vertically, but loosely, into a combination steeping bath and press (Fig. 21.6), which is slowly filled with a solution of 1719 percent caustic, where they remain for about 1 hr. In the steeping, the alpha cellulose is converted into alkali or "soda" cellulose; at the same time, as already mentioned, the caustic solution removes most of the beta and gamma celluloses (also called hemicelluloses). The exact chemical composition of the soda cellulose is not known, but there is evidence that one molecule of NaOH is associated with two anhydroglucose units in the polymer chain.

$$(C_6H_{10}O_5)_n\ (\text{cellulose}) + 18\%\ \text{aqueous NaOH} \rightarrow [(C_6H_{10}O_5)_2 \cdot \text{NaOH}]_n\ (\text{swollen, insoluble, soda cellulose I}) + \text{soluble soda cellulose from } \beta \text{ and } \gamma \text{ celluloses}$$

Fig. 21.6 Steeping of cellulose in the manufacture of viscose rayon. (*Courtesy Avtex Fibers, Inc*.)

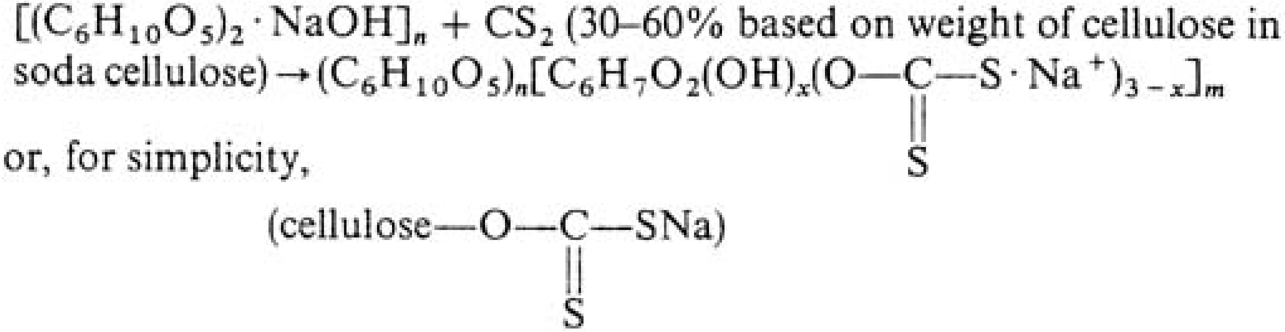


The excess caustic solution is drained off for reuse. Additional amounts are removed by forcing the sheets through a press. The sheets are still in a swollen state and retain from 2.7 to 3.0 parts of the alkali solution. The spent steeping solution squeezed out of the pulp is processed for recovering the caustic from the organic materials.

The sheets of soda cellulose are discharged into a shredder. If blending is desired, the charges from two or more steeping presses are mixed in a single shredder, where the already soft sheets are torn into crumbs; cooling is provided to prevent thermal degradation. Shredding is controlled to produce crumbs that are open and fluffy, and that will allow air to penetrate the mass readily; this is essential in aging.

Soda cellulose is aged by holding it at a constant temperature in perforated containers. The oxygen in the air produces uniform aging accompanied by a reduction in molecular weight and an increase in the number of carboxyl groups present. The target of aging is an average molecular weight high enough to produce satisfactory strength in the final fiber but low enough so that the viscosity of the solution will not be excessively high at the desired concentration for spinning. Each of the various rayon end products has its optimum degree of polymerization or chain length, ranging from about one fourth the original length for regular rayon to one half for certain high-performance fibers. As noted earlier, this optimum size is generally established by effecting a compromise between process economics and desired end-product properties. The aging proceeds for periods of up to two or three days, although the tendency is to speed up the operation by using higher temperatures and traces of metal ions, such as manganese or cobalt, to catalyze the reaction. A combination of experience and constant quality-control testing guarantees that the material will reach the correct point for conversion to cellulose xanthate.

Cellulose xanthate, or more exactly, sodium cellulose xanthate or sodium cellulose dithiocarbonate, is obtained by mixing the aged soda cellulose with carbon disulfide in a vapor-tight xanthating churn. Based upon weight of cellulose, the amount of carbon disulfide used will be in the range of 30 percent for regular rayon to 5060 percent for modified varieties.



$$[(C_6H_{10}O_5)_2 \cdot \text{NaOH}]_n + CS_2\ (30\text{–}60\%\ \text{based on weight of cellulose in soda cellulose}) \rightarrow (C_6H_{10}O_5)_n[C_6H_7O_2(OH)_x(O-\underset{\underset{\displaystyle S}{\|}}{C}-S\cdot Na^+)_{3-x}]_m$$

or, for simplicity,

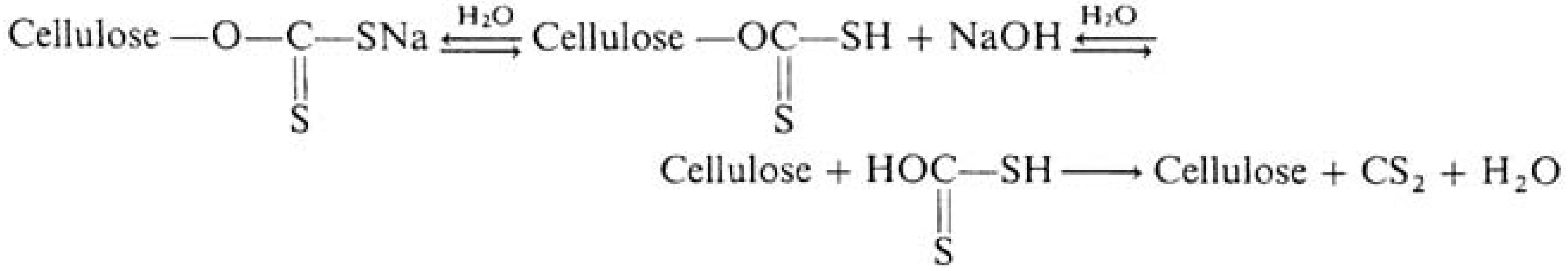
$$(\text{cellulose}-O-\underset{\underset{\displaystyle S}{\|}}{C}-SNa)$$

The xanthate is soluble in a dilute solution of sodium hydroxidea characteristic discovered by Cross and Bevan in 1892and this property makes the spinning of rayon possible. It is a yellow solid; when dissolved in a dilute solution of alkali, it becomes a viscous, honey-colored liquidhence the word "viscose." At this stage, the viscous solution may contain about 7.25 percent cellulose as xanthate in about a 6.5 percent solution of sodium hydroxide, although concentrations of both vary, depending on what end products are desired. The solution is ready for mixing with other batches to promote uniformity, to be followed by filtration, ripening, deaeration, and spinning. The filtration process usually involves several stages so that filters of decreasing pore size may be used to secure a balance of throughput and stepwise particle and gel removal.

Such an operation is a straightforward one for "bright" rayon, but only in the days of "artificial silk" did the shiny fiber alone satisfy the market. After a few years, a dull-appearing fiber also was demanded. At first, fine droplets of oil in the filaments were used to produce dullness until it was discovered that titanium dioxide pigment having a particle size smaller than 1 µm in diameter was even more satisfactory. The latter has since become the universal delustrant for all manufactured fibers. With the use of pigments of any type, problems of dispersion and agglomerate formation must be faced. The usual practice has been to add this pigment when mixing the cellulose xanthate into the dilute solution of caustic.

However, there are many other chemicals and additives that a producer may be required to add to the solution, including: (1) a few parts per million of a tracer element for later identification of the product; (2) coloring pigments for "dope dyed" rayon ("dope dyeing" will be discussed in greater detail under another heading); (3) chemicals for controlling the rate of precipitation and regeneration for obtaining rayon with so-called high performance; and (4) polymers and chemicals to impart specific properties to the fiber. From the standpoint of chemical processing, it is obvious that these additives may also be added when the sodium cellulose xanthate is dissolved in dilute caustic solution, or may be injected into the solution before it enters the spinnerette prior to being extruded. To keep the operations as flexible as possible, the additives should be injected at the last possible moment so that when a changeover is desired, there will be a minimum amount of equipment to be cleaned. On the other hand, the farther along in the operation that additives are placed into the stream, the greater the problem of obtaining uniformity in an extremely viscous medium, and the greater the difficulty in maintaining exact control of proportions before the viscous solution is passed forward and spun. Furthermore, all insoluble additives must be of extremely small particle size, and all injected slurries must be freed of agglomerates by prefiltration; if not, the viscous solution containing the additives must be filtered. Each manufacturer of viscose rayon develops the particular conditions for making additions, depending on a multitude of factors, not the least of which is the existing investment in equipment. All manufacturers must face the universal necessity of filtering the solution with or without pigments or other additives, so that all impurities and agglomerates that might block the tiny holes in the spinnerette are removed.

Although it was known in the years following the discovery by Cross and Bevan that a viscose type of solution could be used in the preparation of regenerated cellulose, the conversion of this solution into useful fibers was not possible until the discovery that the solution required aging until "ripe." Ripening is the first part of the actual chemical decomposition of cellulose xanthate, which, if allowed to proceed unhampered, would result in gelation of the viscose solution.



$$\text{Cellulose}{-}\text{O}{-}\underset{\underset{\text{S}}{\|}}{\text{C}}{-}\text{SNa} \xrightleftharpoons{H_2O} \text{Cellulose}{-}\text{O}\underset{\underset{\text{S}}{\|}}{\text{C}}{-}\text{SH} + \text{NaOH} \xrightleftharpoons{H_2O}$$

$$\text{Cellulose} + \text{HO}\underset{\underset{\text{S}}{\|}}{\text{C}}{-}\text{SH} \longrightarrow \text{Cellulose} + CS_2 + H_2O$$

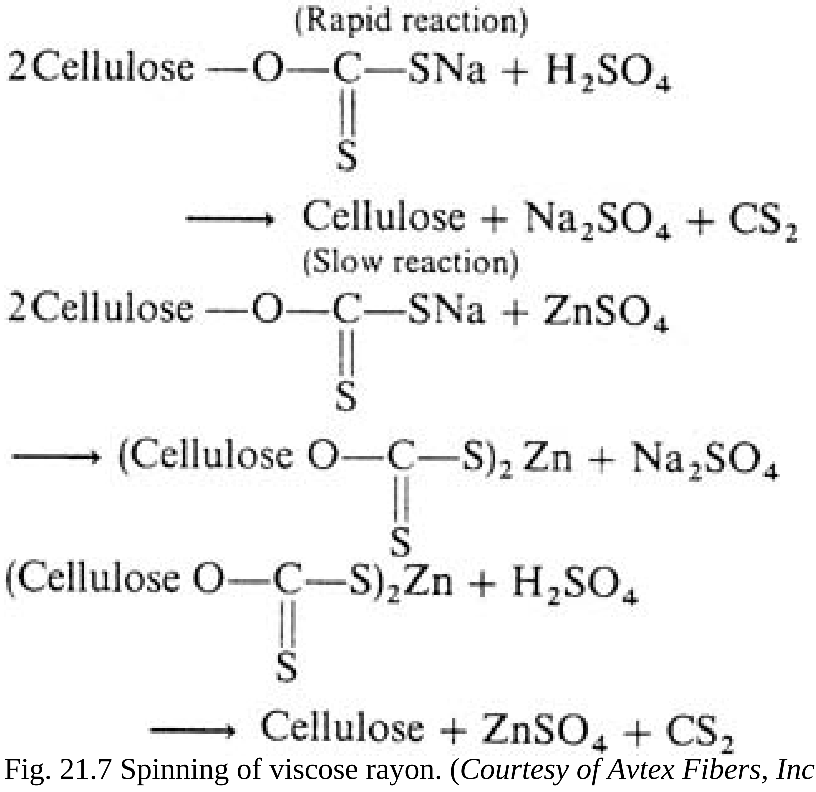
Experience has taught the manufacturer the correct time and conditions for the aging operation, but the requirement of aging itself demands that the entire process be so planned that the viscose solution will arrive at the spinnerette possessing, as nearly as possible, the optimum degree of ripeness, to produce fibers having the desired characteristics. This degree of ripeness is determined by an empirical test made periodically, which is a measurement of the resistance of the solution to precipitation of the soda cellulose when a salt solution is titrated into it. Thus, it is known as the "salt index" or "Hottenroth number" after its originator. An additional step in the overall ripening operation involves the removal of dissolved and mechanically held air by the use of a vacuum on a moving thin film of the viscose solution.

It should be mentioned that so inevitable is decomposition of cellulose xanthate and consequent gelation of the contents of pipes and tanks that all viscose rayon plants must be prepared to pump in-process viscose solutions to other spinning machines or to a waste receiver, purging the entire system with dilute caustic solution, in the event of a long delay in spinning.

Wet Spinning

Spinning a viscose solution into rayon fibers (wet spinning) is the oldest of the three common ways of making manufactured fibers. In this method, the polymer is dissolved in an appropriate solvent, and this solution is forced through fine holes in the face of the spinnerette, which is submerged in a bath of such composition that the polymer precipitates. The pressure necessary for this extrusion is supplied by a gear pump, which also acts as a metering device; the solution is moved through a final or "candle" filter before it emerges from the holes of the spinnerette. There is immediate contact between these tiny streams and the liquid or "wet" bath. As the bath solution makes contact with the material extruded from the holes, chemical or physical changes take place. These changes, whether of lesser or greater complexity, convert the solution of high molecular weight linear polymer first to a gel structure and then to a fiber. As will be observed in what follows, it is an interesting fact that the spinning of viscose rayon, with all of the ramifications made possible by variations in the composition of the solution and the precipitating bath, as well as in the opera-ting conditions, presents the chemist and the chemical engineer with both the oldest and the most complex wet-spinning process.

The formation of rayon fibers from viscose solution is far from being simple, from either a physical or a chemical standpoint. The spinning bath usually contains 15 percent zinc sulfate and 710 percent sulfuric acid, as well as a surface-active agent, without which minute deposits will form around the holes in the spinnerette. Sodium sulfate (1522 percent) is present, formed by the reactions, and as sulfuric acid is depleted and sodium sulfate concentration builds up, an appropriate replenishment of the acid is required. There is a coagulation of the organic material as the sulfuric acid in the spinning bath neutralizes the sodium hydroxide in the viscose solution; at the same time, chemical decomposition of the sodium cellulose xanthate takes place to regenerate the cellulose. If zinc ions are present, which is the usual situation in the production of the improved types of rayon, an interchange takes place so that the zinc cellulose xanthate becomes an intermediate. It reacts at a slower rate, causing slower decomposition to cellulose. This provides conditions for more effective stretching or drawing of the fiber. Chemical additives usually are present to repress hydrogen ion action. The gel-like structure, the first state through which the material passes, is not capable of supporting itself outside the spinning bath. As it travels through the bath, however, it quickly becomes transformed into a fiber that can be drawn from the spinning bath and that can support itself in subsequent operation (Fig. 21.7). The reaction between the bath and the fiber that is forming are paramount in determining the characteristics of the final product; it is for this reason that additives (previously mentioned), as well as zinc ions, may be used to control both the rate of coagulation and regeneration. In this manner, the arrangement of the cellulose molecules may be controlled to produce the conformational structure desired. A practical application of this will be discussed later.



(Rapid reaction)

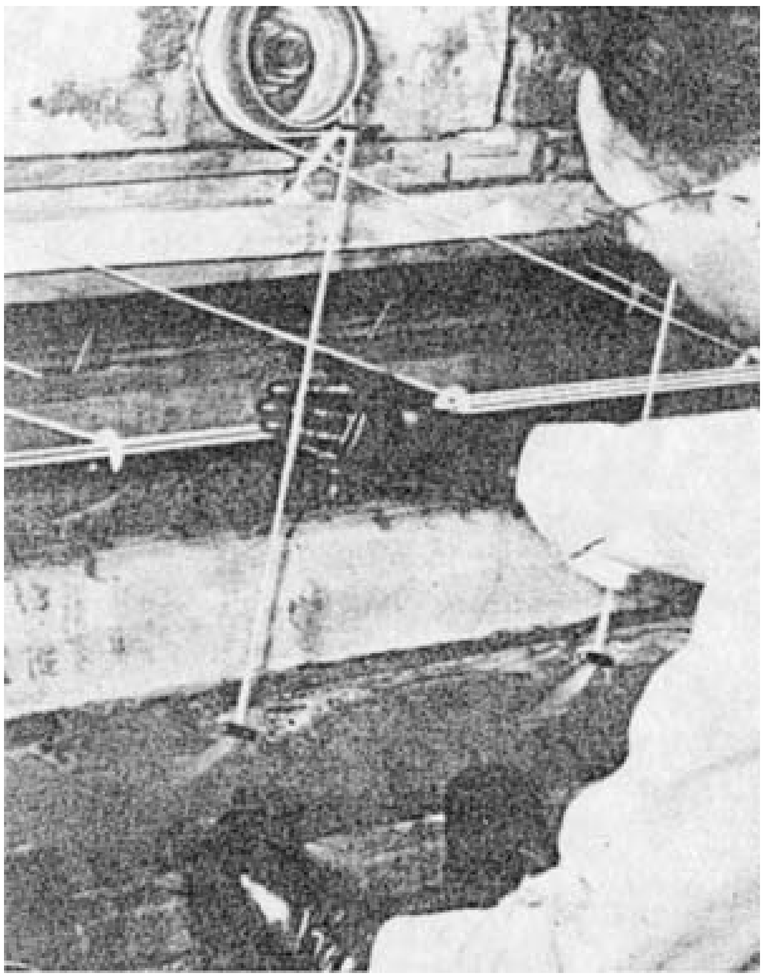
$$2\text{Cellulose}{-}\text{O}{-}\underset{\underset{\text{S}}{\|}}{\text{C}}{-}\text{SNa} + H_2SO_4 \longrightarrow \text{Cellulose} + Na_2SO_4 + CS_2$$

(Slow reaction)

$$2\text{Cellulose}{-}\text{O}{-}\underset{\underset{\text{S}}{\|}}{\text{C}}{-}\text{SNa} + ZnSO_4 \longrightarrow (\text{Cellulose O}{-}\underset{\underset{\text{S}}{\|}}{\text{C}}{-}\text{S})_2\,\text{Zn} + Na_2SO_4$$

$$(\text{Cellulose O}{-}\underset{\underset{\text{S}}{\|}}{\text{C}}{-}\text{S})_2\text{Zn} + H_2SO_4 \longrightarrow \text{Cellulose} + ZnSO_4 + CS_2$$

Fig. 21.7 Spinning of viscose rayon. (*Courtesy of Avtex Fibers, Inc.*)



Because of hydraulic drag, stretching occurs in the bath and also in a separate step after the yarn leaves the bath. In both cases, the linear molecules of cellulose are oriented from random positions to positions more parallel to the fiber axis. If a rayon tire cord is to be the final product, the fibers must be severely stretched to produce a very high orientation of the molecule. This is the basis of the tire cord's high strength and ability to resist stretching, without which growth of the tire body would occur. For regular textile and nonwoven uses, such high strengths are not desired, and the spinning and stretching conditions are controlled to produce rayon of lower strength and greater stretchability under stress.

In order to stretch the yarn uniformly during the manufacturing process, two sets of paired rollers or "godets" are employed, each of the two sets operating at different rotational speeds. The yarn is passed around the first set of godets several times to prevent slippage and is supplied to the stretching area at a constant speed. A second set of godets moves it forward at a more rapid rate, also without slippage. Stretching may range from a few to 100 or more percent. Spinning speeds are of the order of 100 m/min, but may vary with both the size of the yarn and the process used.

Spinning conditions, composition of the spinning bath, and additives to the viscose solution determine the physical characteristics of the rayonits breaking strength and elongation, modulus, ability to resist swelling in water, and characteristics in the wet state as compared with those of the dry material. Not only must the chemical composition of the spinning bath be carefully controlled, but the temperature must be regulated at a selected point, somewhere in the range of 3560°C, to ensure those precipitation and regeneration conditions that are essential to the manufacture of any particular viscose rayon having the properties needed for a selected end use.

After precipitation and regeneration of cellulose have been completed and raw rayon fiber has been formed, the subsequent steps must be controlled so that differences in treatment are minimized; otherwise such sensitive properties as "dye acceptance" will be affected, and the appearance of the final product will vary.

Minute traces of suspended sulfur resulting from the chemical decomposition of cellulose xanthate must be removed by washing with a solution of sodium sulfide. It is expedient to bleach the newly formed fibers with hypochlorite to improve their whiteness; an "antichlor" follows. The chemicals originally present and those used to purify the fibers must be removed by washing. As a final step, a small amount of lubricant is placed on the filaments to reduce friction and improve processibility in subsequent operations.

Several different processes are used for the steps involved in spinning and purifying continuous filament rayon. One of the most common involves the formation of packages of yarn, each weighing several pounds, for separate treatment. After it has been passed upward out of the spinning bath and stretched to the desired degree, the yarn is fed downward vertically into a rapidly rotating canlike container called a spinning pot or "Topham" box (after the man who invented it in 1900). It is thrown outward to the wall of the pot by centrifugal force and gradually builds up like a cake, with excess water being removed by the same centrifugal force. This cake is firm, although it must be handled with care, and is sufficiently permeable to aqueous solution to permit purification.

In another method of package-spinning, the yarn is wound onto a mandrel from the side at a uniform peripheral speed. With this process, the yarn may be purified and dried in the package thus formed. In any of these systems, the spinning and stretching, as well as subsequent steps, may involve separate baths.

The continuous process for spinning and purifying textile-grade rayon yarn merits particular mention from the standpoint of industrial chemistry, as it is rather an axiom that a continuous process is to be preferred over a batch or discontinuous operation. This method employs "advancing rolls" or godets that make it possible for the yarn to dwell for a sufficient length of time on each pair, thus allowing the several chemical operations to take place in a relatively small area. Their operation depends on the geometry existing when the shafts of a pair of adjacent cylindrical rolls are oriented slightly askew. Yarn led onto the end of one of these and then around the pair will progress toward the other end of the set with every pass, the rate of traversing, and therefore the number of wraps, being determined by the degree of skewness.

The production of rayon to be converted to staple fiber also is amenable to line operation. Here, the spinnerette has many thousands of holes, and a correspondingly large number of filaments are formed in the precipitation bath. The resulting tow then is stretched to the desired degree and immediately cut in the wet and unpurified condition. The mass of short lengths can be conveyed through the usual chemical treatments, after which it is washed and dried. It is fluffed to prevent matting and is packaged for shipment in large cases.

Cuprammonium, Nitrocellulose, and Cellulose Acetate Processes for Rayon

Cuprammonium Cellulose

Cellulose forms a soluble complex with copper salts and ammonia. Thus, when cellulose is added to an ammoniacal solution of copper sulfate that also contains sodium hydroxide, it dissolves to form a viscous blue solution, and in this form it is known as Cuprammonium cellulose. The principles on which the chemical and spinning steps of this process are based are the same as those for the viscose process. Cellulose is dissolved, in this case in a solution containing ammonia, copper sulfate, and sodium hydroxide. Unlike the viscose solution, the Cuprammonium solution need not be aged and will not precipitate spontaneously on standing except after long periods. It is, however, sensitive to light and oxygen. It is spun into water and given an acid wash to remove the last traces of ammonia and copper ions. Although this rayon was never manufactured in a volume even approaching that achieved by the viscose process, the smaller individual filaments inherent in it made it useful in certain specialty markets. It no longer is manufactured in the United States but continues to be made abroad.

Nitrocellulose and Cellulose Acetate

Although nitrocellulose and cellulose acetate intermediates have been made and regenerated to form cellulose fibers, neither of these historical processes is still in operation.

Textile Operations

After the filament rayon fiber has been spun and chemically purified, much of it passes through what are known as "textile operations" before it is ready to be knitted or woven. Because these steps of twisting and packaging or beaming are common to the manufacture of all manufactured fibers, it is advisable to review briefly the background and the processes.

Rayon, the first manufactured fiber, not only had to compete in an established field, but also had to break into a conservative industry. Silk was the only continuous filament yarn, and products made from it were expensive and possessed of high prestige, so that they offered a tempting market for rayon. Thus, the new product entered as a competitor to silk and, as already noted, became known as "artificial silk." Under the circumstances, it was necessary for rayon to adapt itself to the then existing silk processing operations and technologies. It was customary to twist several silk filaments together to secure a yarn of the desired size, strength, and abrasion resistance. Because rayon was weaker than silk and its individual filaments were smaller, it required as much twisting as silk or even more.

This twisting could have been carried out in the same plant where the yarn was spun, but the existence of silk "throwsters" (from the Anglo-Saxon *thrawan*, to twist or revolve) made that unnecessary. However, as the rayon industry developed, the amount of yarn twisted in the producing plant or sent forward to throwsters decreased. Over the years, the trend has been to use less twist and to place, instead, several thousand parallel ends directly on a "beam," to form packages weighing as much as 300400 lb, which are shipped directly to a weaving or knitting mill. The advent of stronger rayons, as well as other strong fibers, and a diminishing market for crepe fabrics which required highly twisted yarns, accelerated the trend away from twisting.

In all twisting and packaging operations, the yarn makes contact with guide surfaces and tensioning devices, often at very high speeds. To reduce friction, it is necessary to add a lubricant as a protective coating for the filaments. This is generally true of all manufactured fibers, and it is customary to apply the lubricant or "spinning finish" or "spinning lubricant" as early in the manufacturing process as possible. For those materials that develop static charges in passing over surfaces, this lubricant also must provide antistatic characteristics.

It is difficult to overstate the importance of fiber lubricants to the successful utilization of manufactured fibers. Few problems can be more damaging to a fiber-handling operation than a lubricant upset. A separate chapter could be written on lubricant usage, but some of its more important aspects will be mentioned here. Obviously, lubricants must reduce friction between the fiber and various surfaces to allow movement without excessive damage to the fiber or the surface contacted, the latter being any one of a variety of metals or ceramics. Similarly, a fiber comes in contact with the surface of other fibers in staple fiber processing and in packaging. The lubricant composition must be stable under a variety of storage conditions, without decomposing or migrating within the package or being lost from the fiber surface by adsorption into the fiber, and must be nontoxic and nondermatitic. It also must be compatible with other materials added during textile processes, such as the protective size coat applied to warp yarns before weaving or the wax coat often applied to yarns before circular knitting. Possible metal corrosion must be evaluated for each lubricant composition. Finally, after having performed its function, the fiber finish or lubricant must be completely scoured from the fabric to permit uniform adsorption of dyes and fabric finishes. The application of sewing lubricants to fabric to be cut and sewn is yet another area requiring attention.

Spun Yarn

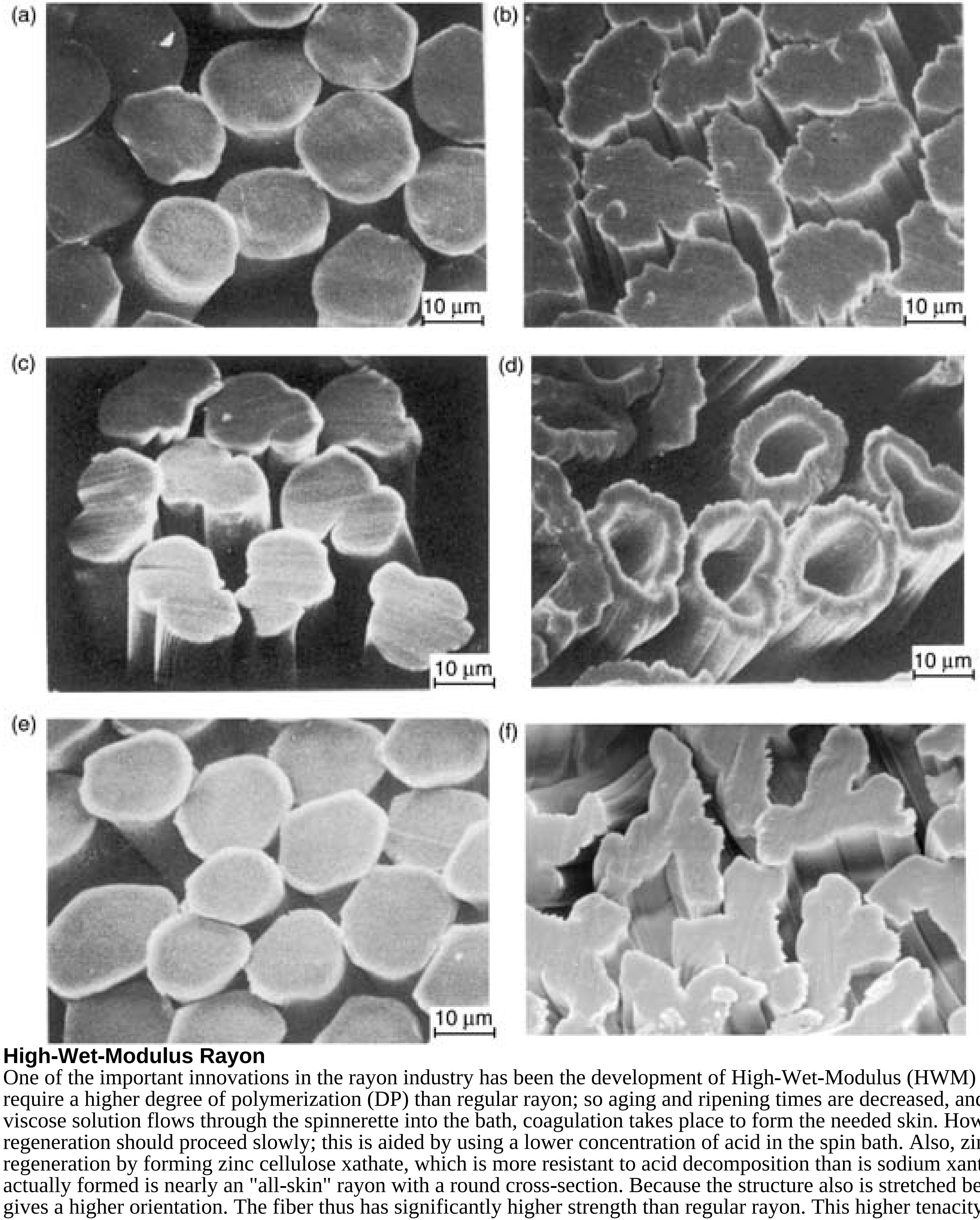
After rayon became established in the textile industry, where it could be used as a silk-like fiber, and its selling price was greatly reduced, other markets for it were developed. The cotton, wool, worsted, and linen systems of converting short discontinuous fibers to yarns were well established, and their products were universally accepted. Here again it was necessary to make rayon fit the requirements of existing equipment and historically acceptable operations. The first of these was that it be cut into the same lengths as those found in cotton and wool. Fortunately, the viscose rayon process was and is eminently suited to the production of tows containing thousands of filaments. The pressure required to force the solution through the holes is so low that neither thick metal sections nor reinforcement of the surface is necessary to prevent bulging, and large spinnerettes containing several thousand holes can be used. Furthermore, the spinning bath succeeds in making contact with all the filaments uniformly. As a result, the spinning of viscose rayon tow is very similar in principle to the production of the smaller continuous filament yarns.

Because both cotton and wool possess distortions from a straight rodlike structure, machinery for their processing was designed to operate best with such crimped fibers. Thus, it was necessary for rayon staple to possess similar lengths and crimpiness in order to be adapted to existing equipment. The crimp, that is, several distortions from a straight path per inch, is produced in rayon "chemically" by modification of the structure. The precipitation-stretching step in spinning is carried out so that the skin and the core of the individual filaments are radially nonuniform and constantly changing over very short lengths along the filaments. Because the skin and the core differ in sensitivity to moisture, the two components shrink differentially, leading to the development of permanent distortions of the filaments. The latent chemical crimp may be enhanced by a thermo-mechanical step. In this, the tow is fed between two wheels, which in turn force it into a chamber, called a "stuffer box," heated with steam, in which it is forced against the compacted material ahead of it, causing the straight filaments to collapse immediately. As the mass of material is pressed forward, it becomes tightly compacted, and it tends to remain axially distorted after it escapes through a pressure-loaded door at the opposite end of the chamber.

Modified Viscose Rayon Fibers

The variations of chemical ingredients, their concentrations, and the temperature of the spinning bath determine the rates of coagulation and regeneration, and thus the relative amounts of "skin" and "core" in the cross-section of the fiber. The skin is known to possess a higher degree of order and better mechanical properties than the core, so an increase in its proportion is desired in higher performance fibers. The degree of orientation is determined by the stretch imparted. Inherent in process variations such as these is what may be called the "art" of viscose rayon manufacturing, whereby great diversity in rayon properties can be obtained. The cross-sectional morphologies of some rayons are illustrated in Fig. 21.8.3,4

Fig. 21.8 Cross-sectional morphologies of some of the rayon fibers. (a) High wet modulus; (b) regular rayon; (c) crimped HWM; (d) hollow; (e) cuprammonium; (f) Trilobal. (Sources: All except *trilobal*: Turbak, A., "Rayon," in *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 14, p. 55, copyright John Wiley & Sons, Inc., New York, 1985 and used with permission of the copyright owner; trilobal photo: Gupta, B. S. and Hong, C. T., *INJ*, 7(1), 38 (1995).)



High-Wet-Modulus Rayon

One of the important innovations in the rayon industry has been the development of High-Wet-Modulus (HWM) rayon. For its manufacture, the cellulose molecules require a higher degree of polymerization (DP) than regular rayon; so aging and ripening times are decreased, and the processing temperature is lowered. As the viscose solution flows through the spinnerette into the bath, coagulation takes place to form the needed skin. However, in order for an increase in tenacity to occur, regeneration should proceed slowly; this is aided by using a lower concentration of acid in the spin bath. Also, zinc is added because zinc ions in the bath slow down regeneration by forming zinc cellulose xathate, which is more resistant to acid decomposition than is sodium xanthate. Because of the slow generation process, what is actually formed is nearly an "all-skin" rayon with a round cross-section. Because the structure also is stretched before it crystallizes, a higher stretch is possible; this gives a higher orientation. The fiber thus has significantly higher strength than regular rayon. This higher tenacity exists not only when the material is dry, but also when it is wet; hence the name high-wet-modulus rayon. Because of its high structural orientation and greater order, the fibers have fewer physically accessible sites for water molecules; thus it is less susceptible to swelling and to the adverse effects of basic cleaning solutions, so that the fiber's launderability is improved. The HWM rayons are used extensively for blending with cotton, wool, silk, and all other manufactured fibers.

High Absorbency Rayons

Over the past years, disposable products have become commonplace, especially in the United States and Europe. Cellulosic fibers, particularly rayons, have served the needs of the disposables industry because of their absorbent qualities. The most useful fibres for disposable/absorbent applications are the rayons with crenulations, crimp, and hollow regions, all of which add to the absorbency of the fiber. These characteristics are achieved in varying degrees by physical and chemical alterations in the spinning process. Crenulations, or random irregularities in the shape of the cross-section, typical for most rayon fibers, are caused by the rapid formation of skin before the dehydration is complete. As the fiber interior loses solvent, it collapses in certain areas and produces the crenulated shape. Furthermore, fabricators have learned how to control the cross-sectional shape of filaments by using spinnerettes containing other than round holes. One example is fibers having Y or trilobal-shaped cross-sections, which have been found to be capable of picking up more water and at a faster rate than possible with fibers of round cross-sections.[4]

Hollow viscose fibers contain gas pockets produced by adding "blowing" agents, such as sodium carbonate, to the viscose. When carbon dioxide is released during regeneration, the fibers inflate, leading to the formation of hollow filaments. The added free volume and decrease in molecular order, increase the ability of the fibers to pick up water.

Other New Developments

A number of other developments are taking place in the rayon industry, the target of one being the manufacture of lint-free rayon for use in products such as circuit boards. In another, graphite particles are blended in with the viscose to reduce static buildup. Production of flame-retardant rayon has received increased attention, in one case being achieved by the addition of phosphorus compounds to the spinning dope. The advanced technology now can produce flame-resistant fibers that, when exposed to high temperatures, will not shrink or emit toxic gases. Other developments are in the area of finding better and environmentally safer solvents for the cellulosic raw material. Searches are under way for solvents that may lead to lyotropic liquid crystalline polymer solutions from which ultrahigh-strength and high-modulus fibers can be spun.

Environmentally Friendly High Wet Strength RayonLyocell

Low wet strength of rayon in general has restricted the application of the fiber to disposable and semi-durable materials. Additionally, the environmental concerns associated with its manufacture have resulted in significant curtailment in production of the fiber during the recent past. Search for ways and means to produce a high wet strength fiber using an environmentally acceptable process have occupied much of the research effort during the past two decades. One of the latest additions to the family of rayon fibers is Lyocell. The fiber has wet strength comparable to that of the natural cellulosic fibers and is manufactured with a solvent that is essentially totally recovered and recycled.[5] The solvent used is N-methyl morphine oxide, $O(C_4H_8)NOCH_3$, popularly known as amine oxide. The manufacturing process involves the dissolving of pulp in hot amine oxide, filtration of the solution, and then spinning into a bath containing a dilute solution of the solvent. The bath removes the amine oxide from the fibers, which are washed and dried, and the removed solvent is almost totally reclaimed for further use. The final fiber is said to have a different molecular structure from that of normal rayon, and a smooth surface and a round cross-section. The fiber is noted to be stronger than cotton and normal rayon in both the dry and the wet states.[6]

< previous page

next page >

21.5 Cellulose Acetate

Historical

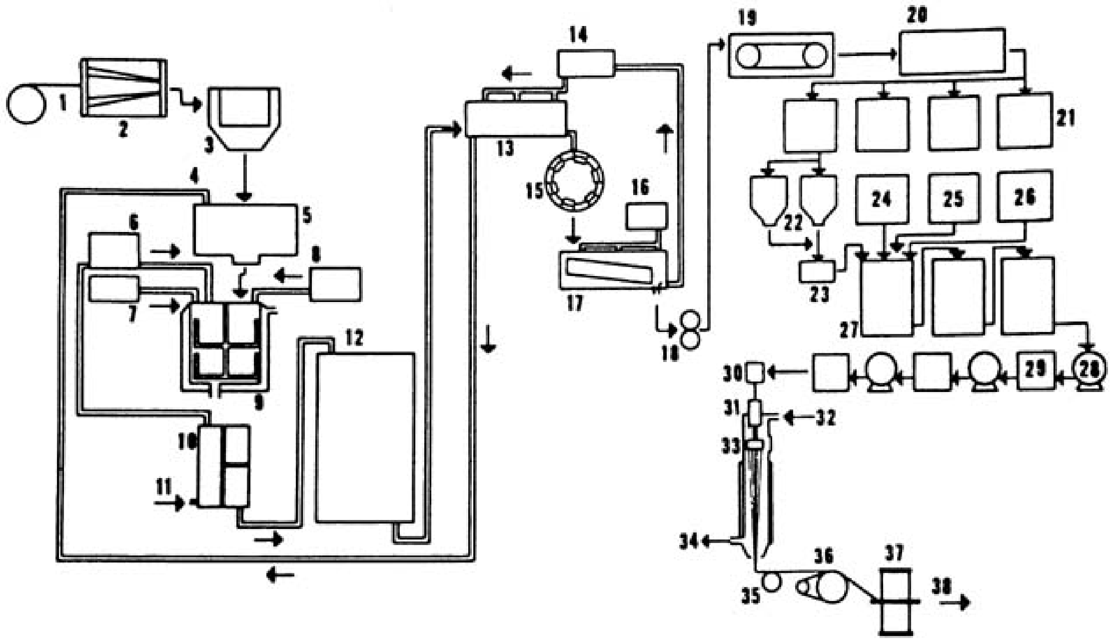
Cellulose acetate was known as a chemical compound long before its potential use as a plastic or fiber-forming material was recognized. The presence of hydroxyl groups had made it possible to prepare cellulose esters from various organic acids, as cellulose consists of a long molecular chain of beta-anhydroglucose units, each of which carries three hydroxyl groupsone primary, the other two secondary. The formula for cellulose (already noted) is [C6H7O2(OH)3]*n*; when this is fully esterified, a triester results. It was learned quite early that although cellulose triacetate is soluble only in chlorinated solvents, a product obtained by partial hydrolysis of the triester to a "secondary" ester (having about 2.352.40 acetyl groups per anhydroglucose unit) was easily soluble in acetone obtaining a small amount of water. Many other cellulose esters have been prepared, but only the acetate has been commercialized successfully as a manufactured fiber. Propionates and butyrates, and mixed esters of one or both the acetate, have applications as plastics. The first acetate fibers were produced in 1921 in Europe and in 1924 in the United States.

Manufacture of Secondary Cellulose Acetate

Cellulose acetate originally was made from purified cotton linters, but this raw material has been entirely replaced by wood pulp. The other raw materials used are acetic acid and acetic anhydride.

Cellulose acetate is manufactured by a batch process (see Fig. 21.9). There has been mention in the patent literature of a continuous system, but its utilization as a production process has not been announced. The "charge" of cellulose, purified, bleached, and shredded, is of the order of 8001500 lb. It is pretreated with about one third its weight of acetic acid and a very necessary amount of water, about 6 percent of its weight. If it is too dry at the time of use, more H2O must be added to the acetic acid. A small amount of sulfuric acid may be used to assist in swelling the cellulose and to make it "accessible" to the esterifying mixture.

Fig. 21.9 Flow diagram for manufacture of cellulose acetate yarn: (1) wood pulp; (2) attrition mill; (3) cyclone; (4) 35% acetic acid; (5) pretreater; (6) magnesium acetate solution; (7) precooled acetylation mix; (8) sulfuric acid; (9) acetylator; (10) ripener; (11) steam; (12) blender; (13) precipitator; (14) dilute acetic acid; (15) hammer mill; (16) water; (17) rotary screen washer; (18) squeeze rolls; (19) drying oven; (20) blender; (21) storage bins; (22) silos; (23) weight bins; (24) acetone; (25) wood pulp; (26) pigment; (27) mixers; (28) hold tanks; (29) filter press; (30) pump; (31) filter; (32) air; (33) jet; (34) acetone recovery; (35) oiling wheel; (36) feed roll; (37) bobbin; (38) inspection.

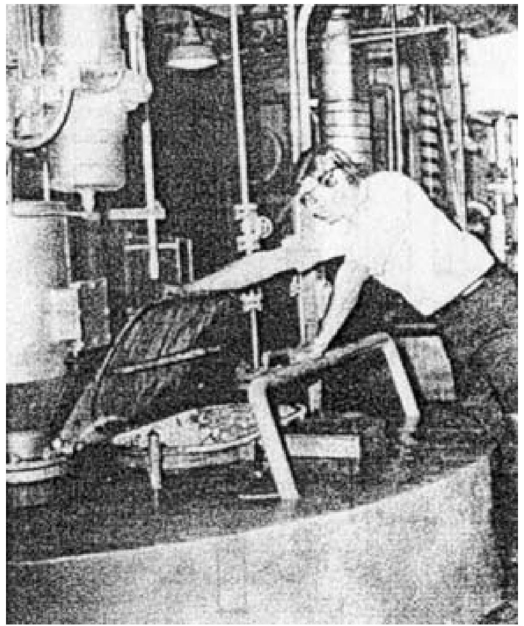


Although there has been much discussion of the chemistry of cellulose acetylation, it is now generally agreed that the sulfuric acid is not a "catalyst" in the normal sense of the word, but rather that it reacts with the cellulose to form a sulfo ester. The acetic anhydride is the reactant that provides the acetate groups for esterification. The acetylation mixture consists of the output from the acetic anhydride recovery unit, being about 60 percent acetic acid and 40 percent acetic anhydride, in an amount 510 percent above the stoichiometric requirement, to which has been added 1014 percent sulfuric acid based on the weight of cellulose used. The reaction is exothermic and requires that the heat be dissipated.

In preparing for acetylation, the liquid reactants are cooled to a point (0°C) where the acetic acid crystallizes, the heat of crystallization being removed by an appropriate cooling system. The slush of acetic acid crystals in the acetic anhydridesulfuric acid mixture is pumped to the acetylizer, a brine-cooled mixer of heavy construction (see Fig. 21.10). The pretreated cellulose is dropped in from the pretreating unit located above. The reaction is highly exothermic, and at the start large amounts of heat are produced. As the temperature of the reaction mixture rises to the melting point of the acetic acid (16.6°C), its large heat of fusion (45.91 cal/g) prevents a dangerous rise in temperature that would degrade the molecular weight of the cellulose chain. As the reaction proceeds, brine in the jacket of the acetylizer provides additional cooling.

$$\text{Cellulose} + (CH_3CO)_2O + H_2SO_4\ (10\text{–}15\%\ \text{based on weight of cellulose}) \xrightarrow[\text{Anhydrous}]{CH_3COOH} [C_6H_7O_2(OSO_3H)_{0.2}(CH_3COO)_{2.8}]_n$$

Fig. 21.10 Process vessel for acetylation of cellulose. (*Courtesy Celanese Fibers Co.*)



The reaction product is soluble in the acetylation mixture; as it is formed and dissolved, new surfaces of the cellulose are presented to the reagents. One variation of this procedure uses methylene chloride, rather than an excess of acetic acid in the reaction mixture. This chemical is used both to dissipate the heat by refluxing (boiling point, 41.2° C) and to dissolve the cellulose ester as it is formed. As the reaction proceeds, the temperature is allowed to rise. Because cellulose is a natural product obtained from many sources, it varies slightly in composition, and at the end of the reaction cannot be predicted exactly; the disappearance of fibers as determined by microscopic examination thus is the usual means of following its progress.

During the acetylation operation, a certain amount of chain fission is allowed to take place in the cellulose molecule. This is to ensure that the viscosity of the cellulose acetate spinning solution will be low enough for ease of handling but high enough to produce fibers with the required strength. The temperature of the reaction controls the rates of both acetylation and degradation of molecular weight.

The next step in the manufacture is "ripening," whose object is to convert the triester, the "primary" cellulose acetate, to a "secondary" acetate having an average of about 2.352.40 acetyl and no sulfo groups (if any sulfuric acid is used in pretreatment) per anhydroglucose unit. While the cellulose sulfo-acetate is still in the acetylizer, sufficient water is added to react with the excess anhydride and start the hydrolysis of the ester. Usually the water is used as a solution of sodium or magnesium acetate, which increases the pH and promotes hydrolysis. The temperature is raised to about 7080°C, by direct injection of steam to speed up the reaction. Hydrolysis is continued until the desired acetyl content is obtained. When this value is reached, an aqueous solution of magnesium or sodium acetate is added to cool the batch and stop the hydrolysis. It is then ready for precipitation. For example,

$$[C_6H_7O_2(OSO_3H)_{0.2}(CH_3COO)_{2.8}]_n$$

$$+ (CH_3COO)_2Mg \xrightarrow{\text{Aqueous conc. } CH_3COOH}$$

$$[C_6H_7O_2(OH)_{0.65}(CH_3COO)_{2.35}]_n + MgSO_4$$

The solution is carried to the verge of precipitation by adding dilute acetic acid. Then it is flooded with more dilute acetic acid and mixed vigorously, so that the cellulose acetate comes out as a "flake" rather than a gelatinous mass or fine powder. The flake then is washed by standard countercurrent methods to remove the last traces of acid, and is dried in a suitable dryer.

Manufacture of Cellulose Triacetate

To obtain completely acetylated cellulose, the reaction requires the use of perchloric acid rather than sulfuric acid as the catalyst. In the presence of 1 percent perchloric acid, a mixture of acetic acid and acetic anhydride converts a previously "pretreated" cellulose to triacetate without changing the morphology of the fibers. If methylene chloride rather than an excess of acetic acid is present in the acetylation mixture, a solution is obtained. However, usually a degree of substitution between 92 and 100 percent is acceptable. For obtaining such a triester, it is possible to use about 1 percent sulfuric acid instead of perchloric acid. When the sulfoacetate obtained from such a reaction is hydrolyzed with the objective of removing only the sulfo-ester groups, the resulting product has about 2.94 acetyl groups per anhydroglucose unit. The preparation, hydrolysis, precipitation, and washing of "triacetate" are in all other respects similar to the corresponding steps in the manufacture of the more common secondary acetate. Cellulose triacetate, formerly produced under the trade name Arnel® by Celanese Corporation, is no longer in production in the United States.

Acid Recovery

In the manufacture of every pound of cellulose acetate, about 4 lb of acetic acid is produced in 3035 percent aqueous solution. The accumulated acid contains a small amount of suspended fines and some dissolved cellulose esters. To remove the suspended material, the acid is passed slowly through settling tanks. Then it is mixed with organic solvents, so that the acid becomes concentrated in an organic layer, which is decanted. Distillation separates the acid from the organic solvent.

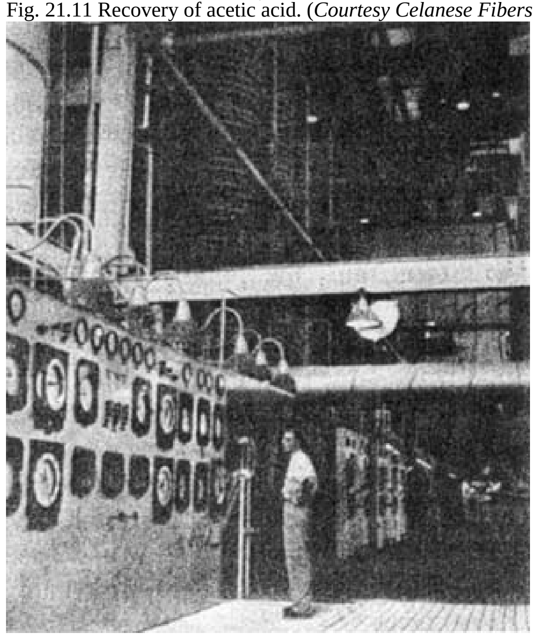
To produce the acetic anhydride, the acid is dehydrated to ketene and reacted with acetic acid using a phosphate catalyst at 500°C or higher in a tubular furnace.

$$CH_3COOH \xrightarrow[\text{Catalyst}]{\text{Heat}} H_2O + CH_2{=}C{=}O$$

$$CH_2{=}C{=}O + CH_3COOH \rightarrow (CH_3CO)_2O$$

The mixture of unreacted acid, water, and anhydride is fed to a still, which yields dilute acetic acid overhead and an anhydrideacetic acid mixture at the bottom (see Fig. 21.11). Conditions are controlled in such a way that the raffinate is about 40 percent anhydride and 60 percent acetic acid. As already mentioned, this is the desired ratio for the reaction mixture used for acetylation of cellulose.

Fig. 21.11 Recovery of acetic acid. (*Courtesy Celanese Fibers Co.*)



Blending of Flake

As in the manufacture of viscose, the products of batch operations are blended to promote uniformity in the manufacture of cellulose acetate. Although a blend of different celluloses is selected in the beginning, the pretreatment, acetylation, and ripening are batch operations with little or no mixing. Before precipitation, a holding tank provides an opportunity for mixing; then precipitation, washing, and dryingall continuouspromote uniformity. The dried cellulose acetate flake moves to holding bins for analysisthe moisture content, acetyl value, and viscosity being especially important. The results of the analyses determine how much further blending is necessary. After blending and mixing of portions of selected batches, the lot is air-conveyed to large storage bins or "silos," which are filled from the center of the top and emptied from the center of the bottom, thus bringing about further mixing.

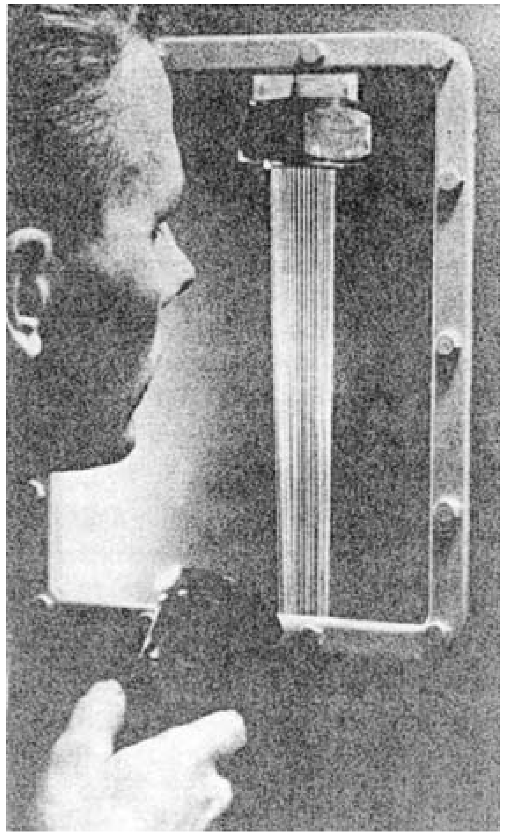
Spinning Cellulose Acetate

Acetone is metered into a vertical tank equipped with a stirrer, and the cellulose acetate flake and filter aid are weighed in an automatic hopper; all operations are controlled by proportioning methods common to the chemical industry. The ratio of materials is about 25 percent cellulose acetate, 4 percent water, less than 1 percent ground wood pulp as a filter aid, and the remainder acetone. The mixture moves forward through two or three stages at the rate at which it is used, the hold time being determined by experience. After dissolution is completed, filtration is carried out in batteries of plate and frame filter presses in three or even four stages, the passage of the "dope" being through presses of decreasing porosity.

Much of the cellulose acetate is delustered by the addition of titanium dioxide pigment, as with viscose rayon. Between filtrations (and after the last filtration), the dope goes to storage tanks that serve to remove bubbles; in this case, a vacuum is not necessary. From the final storage tank, it is pumped into a header located at the top of each spinning machine; then it is directed to a series of metering gear pumps, one for each spinnerette. Because the holes in the cellulose acetate spinning spinnerette are smaller (0.030.05 mm) than those in the corresponding viscose devices, great care must be taken with the final filtration. An additional filter for the removal of any small particles that may have passed through the large filters is placed in the fixture, sometimes called the "candle," to which the spinnerette assembly is fastened. A final filter is placed in the spinnerette-assembly unit over the top of the spinnerette itself.

The method used for spinning cellulose acetate is "dry" spinning. The dope is heated (in some cases above the boiling point of acetone, 56.5°C) to lower its viscosity and thus reduce the pressure required to extrude it, and to supply some of the heat needed for evaporating the acetone solvent (see Fig. 21.12).

Fig. 21.12 Dry spinning of cellulose acetate. (*Courtesy Tennessee Eastman Co.*)



The spinnerette is stainless steel, and because the filaments must be heated and prevented from sticking together, and because space must be allowed for the escape of acetone vapor, the holes must be kept farther apart than those of the spinnerettes used for wet spinning. As the hot solution of cellulose acetate in acetone emerges downward into the spinning cabinet, an instantaneous loss of acetone takes place from the surface of the filaments, which tend to form a solid skin over the still liquid or plastic interior. A current of air, either in the direction the filaments are moving or countercurrent, heats the filaments, and as the acetone is diffused from the center through the more solid skin, each filament collapses to form the indented cross-sectional shape typical of cellulose acetate. The heated air removes the vaporized acetone. Each manufacturer uses a preferred updraft, downdraft, or mixed-draft operation, as needs dictate.

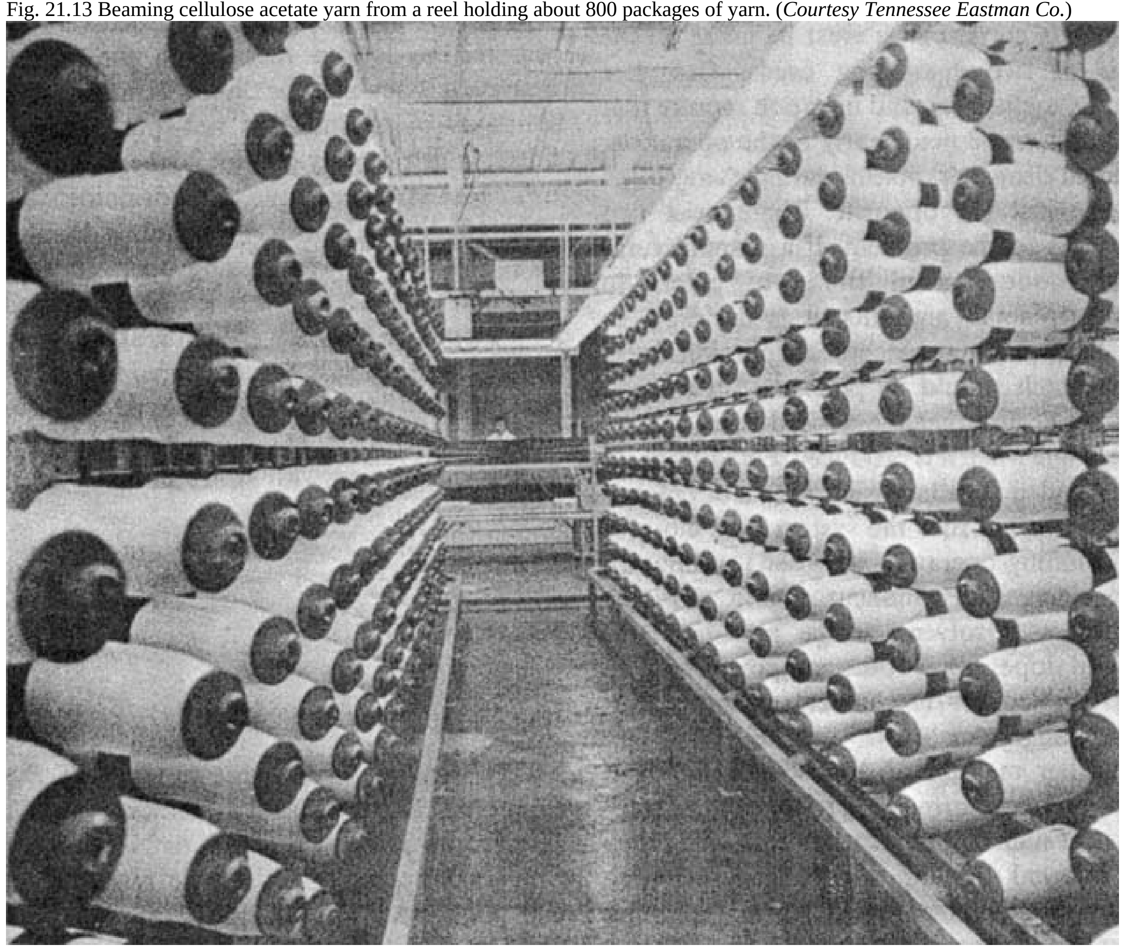
The cabinet through which the yarn passes vertically downward must be long enough to allow sufficient acetone to diffuse outward and evaporate from the surfaces of the filaments so that the latter will not stick to the first surface contacted or fuse to each other. The temperature of the air in the cabinet, the rate of flow, the length of the cabinet, the size and number of filaments, and the rate of travel are all interrelated in the spinning process. Because it is desirable to increase spinning speeds to the limit of the equipment, the tendency has been to construct longer spinning cabinets as each new plant is built. Present spinning speeds are of the order of 600 or more meters per minute, measured as the yarn emerges from the cabinet.

Other dry-spinning operations have followed essentially the same pattern. For example, the dry spinning of cellulose triacetate was identical to that for secondary acetate except that the acetone solvent had been replaced by a chlorinated hydrocarbon such as methylene chloride, the solubility of which was improved by the addition of a small amount of methanol (515%).

The acetate yarn emerging from the cabinet makes contact with an applicator that provides the lubricant required to reduce both friction and static formation in subsequent operations. With its surface lubricated, the yarn passes around a "feed" roll that determines the rate of withdrawal from the spinning cabinet, and then to any of several desired packaging devices.

Unlike the packaging of rayon yarn, cellulose acetates are either "ring" spun or wound into a package called a "disc," "zero twist," or "cam wound." In the ring-spun package, the yarn carries a slight twist of less than one turn per inch, but it requires a relatively expensive bobbin. Since the trend is toward less twisting, such acetate yarn is "beamed" in the producer's plant after little or no twisting, the heavy beams being shipped directly to knitters or weavers (see Fig. 21.13).

Fig. 21.13 Beaming cellulose acetate yarn from a reel holding about 800 packages of yarn. (*Courtesy Tennessee Eastman Co.*)



Filament yarns are twisted for two reasons. One is to supply certain esthetic characteristics such as touch, drapability, and elasticity. The other more fundamental reason is to provide physical integrity to the filament bundle so that it can be warped, woven, and knitted without excessive breakage or fraying of individual filaments.

The yarn just mentioned as having no twist imparted before beaming may have been subjected to intermingling just prior to windup after extrusion. In the intermingling process, yarn with no twist, and usually under low tension, is passed through a zone where it is impinged upon by a jet stream of compressed air. This causes the filaments to interlace or intermingle with each other, and they can become metastable in this configuration when tension is reapplied. In this condition, the yarn has the integrity of twisted yarn and will pass through several textile processing steps without difficulty; but with each handling, some of the intermingling is worked out.

Solvent Recovery

The air containing the acetone vapor is drawn out of the spinning cabinet and passed through beds of activated carbon that sorb the organic solvent. The acetone is recovered by steaming and then by separating it from the water by distillation. The efficiency of recovery is about 95 percent.

Dope-Dying

As with viscose rayon, colored pigments or dyestuffs may be added to the spinning solution so that the yarn will be colored as it is produced, thus eliminating the need for dyeing the final fabric. As mentioned earlier, even in using titanium dioxide, a compromise must be made on the basis of two competing needs. Complete mixing, uniformity, and filtration require that the addition be made early in the operation; minimal cleaning problems during changeovers require just the opposite. There exist two solutions to the problem. If a manufacturer must produce a multitude of colors in relatively small amounts, it is desirable to premix individual batches of spinning dope. Each batch should be pretested on a small scale to ensure that the desired color will be acceptable when it is produced. Facilities must be provided to allow each batch of colored dope to be cut into the system very close to the spinning operation in order to minimize pipe cleaning. Permanent piping must be flushed with solvent or the new batch of colored dope; some of the equipment may be disassembled for mechanical cleaning after each change of color.

Another method of producing spun-dyed yarn involves using a group of "master" dopes of such color versatility that when they are injected by appropriate proportioning pumps into a mixer located near the spinning operation, they will produce the final desired color. The advantages of such an operation are obvious; the disadvantage lies in the public demand for an infinite number of colors. No small group of known pigments will produce final colors of every desired shade.

21.6 Protein Fibers

As previously mentioned, the use of naturally existing polymers to produce fibers has had a long history. In the case of cellulose the results were fabulous. An initial investment of $930,000 produced net profits of $354,000,000 in 24 years for one rayon company.7 On the other hand, efforts to use another family of natural polymersproteinshave thus far resulted in failure or at best very limited production.

These regenerated proteins are obtained from milk (casein), soya beans, corn, and peanuts. More or less complex chemical separation and purification processes are required to isolate them from the parent materials. They may be dissolved in aqueous solutions of caustic, and wet-spun to form fibers, which usually require further chemical treatment as, for example, with formaldehyde. This reduces the tendency to swell or dissolve in subsequent wet-processing operations or final end uses. These fibers are characterized by a wool-like feel, low strength, and ease of dyeing. Nevertheless, for economic and other reasons they have not been able to compete successfully with either wool (after which they were modeled) or with other manufactured fibers.

21.7 Nylon

Historical

Nylon was the first direct product of the technological breakthrough achieved by W. H. Carothers of E. I. duPont de Nemours & Co. Until he began his classic research on high polymers, the production of manufactured fibers was based almost completely on natural linear polymers. Such materials included rayon, cellulose acetate, and the proteins. His research showed that chemicals of low molecular weight could be reacted to form polymers of high molecular weight. By selecting reactants that produce linear molecules having great length in comparison with their cross-section, fiber-forming polymers are obtained. With this discovery, the manufactured fiber industry entered a new and dramatic era.

Manufacture

Nylon 66

The word "nylon" was established as a generic name for polyamides, one class of the new high molecular weight linear polymers. The first of these, and the one still produced in the largest volume, was nylon 66 or polyhexamethylene adipamide. Numbers are used following the word "nylon" to indicate the number of carbon atoms contributed by the diamine and dicarboxylic acid constituents, in this case hexamethylenediamine and adipic acid, respectively.

To emphasize the fact that it does not depend on a naturally occurring polymer as a source of raw material, nylon often has been called a "truly synthetic fiber." To start the synthesis, benzene may be hydrogenated to cyclohexane:

$$C_6H_6 + 3H_2 \xrightarrow{\text{Catalyst}} C_6H_{12}$$

or the cyclohexane may be obtained by fractionation of petroleum. The next step is oxidation to a cyclohexanolcyclohexanone mixture by means of air:

$$xC_6H_{12} + O_2(\text{air}) \xrightarrow{\text{Catalyst}} yC_6H_{11}OH + zC_6H_{10}O$$

In turn, this mixture is oxidized by nitric acid to adipic acid:

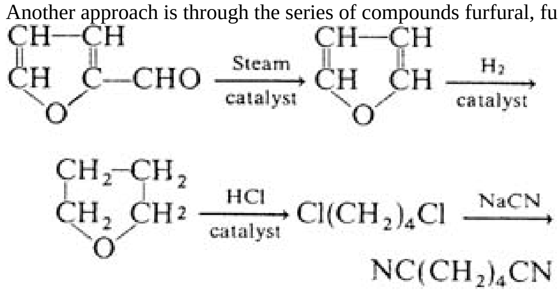
$$C_6H_{11}OH + C_6H_{10}O + HNO_3 \xrightarrow{\text{Catalyst}} (CH_2)_4(COOH)_2$$

Adipic acid so obtained is both a reactant for the production of nylon and the raw material source for hexamethylenediamine, the other reactant. The adipic acid first is converted to adiponitrile by ammonolysis and then to hexamethylenediamine by hydrogenation:

$$(CH_2)_4(COOH)_2 + 2NH_3 \xrightarrow{\text{Catalyst}} (CH_2)_4(CN)_2 + 4H_2O$$

$$(CH_2)_4(CN)_2 + 4H_2 \xrightarrow{\text{Catalyst}} (CH_2)_6(NH_2)_2$$

Another approach is through the series of compounds furfural, furane, cyclotetramethylene oxide, 1,4-dichlorobutane, and adiponitrile, as illustrated below. The furfural is obtained from oat hulls and corn cobs.



Or, 1,4-butadiene obtained from petroleum, may be used as a starting raw material to make the adiponitrile via 1,4-dichloro-2-butene and 1,4-dicyano-2-butene:

$$CH_2{=}CHCH{=}CH_2 \rightarrow ClCH_2CH{=}CHCH_2Cl$$

$$\xrightarrow[\text{Catalyst}]{HCN} NCCH_2CH{=}CHCH_2CN$$

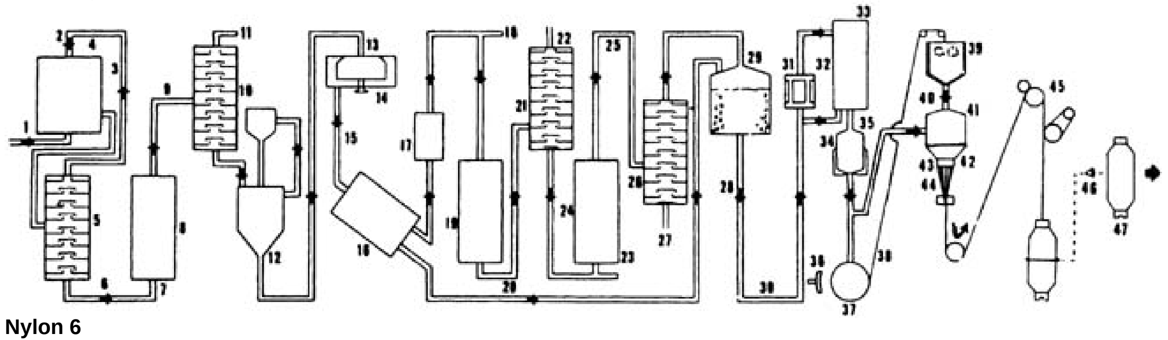
$$\xrightarrow[\text{Catalyst}]{H_2} NC(CH_2)_4CN$$

When hexamethylenediamine and adipic acid are mixed in solution in a one-to-one molar ratio, the "nylon salt" hexamethylenediammoniumadipate, the direct progenitor of the polymer, is precipitated. After purification, this nylon salt is polymerized to obtain a material of the desired molecular weight. It is heated to about 280°C under vacuum while being stirred in an autoclave for 23 h; a shorter holding period follows; and the process is finished off at 300°C. The molecular weight must be raised to a level high enough to provide a fiber-forming material, yet no higher. If it is too high, the corresponding viscosity in the subsequent spinning operation will require extremely high temperatures and pressure to make it flow. Accordingly, a small amount of acetic acid is added to terminate the growth of the long-chain molecules by reaction with the end amino groups.

The polymerized product is an extremely insoluble material and must be melt-spun, as discussed later. Therefore, should a delustered or precolored fiber be desired, it is necessary to add the titanium dioxide or colored pigment to the polymerization batch prior to solidification. For ease of handling, the batch of nylon polymer may be extruded from the autoclave to form a thin ribbon, which is easily broken down into chips after rapid cooling. But, whenever possible, the liquid polymer is pumped directly to the fiber Spinning melt spinning operation (see Fig. 21.14).

Fig. 21.14 Flow diagram for the manufacture of nylon 66 yarn: (1) air; (2) cyclohexane from petroleum; (3) reactor; (4) recycle cyclohexane; (5) still; (6) cyclohexanolcyclohexanone; (7) nitric acid; (8) converter; (9) adipic acid solution; (10) still; (11) impurities; (12) crystallizer; (13) centrifuge; (14) impurities; (15) adipic acid crystals; (16)

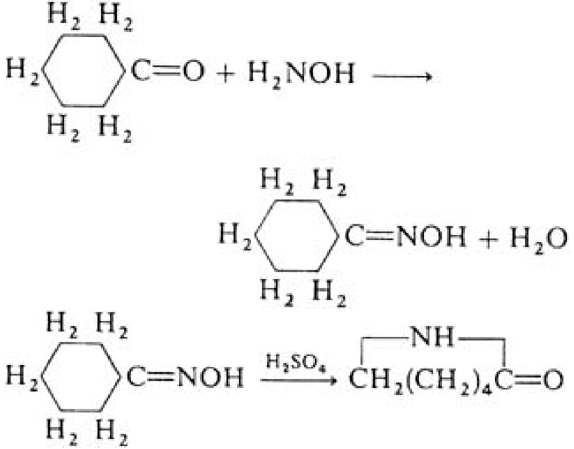
dryer; (17) vaporizer; (18) ammonia; (19) converter; (20) crude adiponitrile; (21) still; (22) impurities; (23) hydrogen; (24) converter; (25) crude diamine; (26) still; (27) impurities; (28) nylon salt solution; (29); reactor; (30) stabilizer; (31) calandria; (32) evaporator; (33) excess water; (34) autoclave; (35) delustrant; (36) water sprays; (37) casting wheel; (38) polymer ribbon; (39) grinder; (40) polymer flake; (41) spinning machine; (42) heating cells; (43) spinnerette; (44) air; (45) draw twisting; (46) inspection; (47) nylon bobbin. (Note: Whenever the demand for liquid polymer at a spinnerette is large, as, for example, in the spinning of tire yarn, it is pumped directly from the autoclave.)



Nylon 6

Nylon 6 is made from caprolactam and is known as Perlon® in Germany, where it was originally developed by Dr. Paul Schlack.8 Its production has reached a very large volume in the United States in recent years.

Like nylon 66, nylon 6 uses benzene as raw material, which is converted through previously mentioned steps to cyclohexanone. This compound is in turn converted to the corresponding oxime by reaction with hydroxylamine, and cyclohexanone oxime is made into caprolactam by the Beckmann rearrangement.



After purification, the lactam is polymerized by heating at elevated temperatures in an inert atmosphere. During self-condensation, the ring structure of the lactam is opened so that the monomer acts as an epsilon-aminocaproic acid radical. Unlike that of nylon 66, the polymerization of caprolactam is reversible; the polymer remains in equilibrium with a small amount of monomer. As with nylon 66, nylon 6 is extruded in thin strands, quenched, and cut into chips for subsequent spinning, or the molten polymer is pumped directly to the spinning equipment.

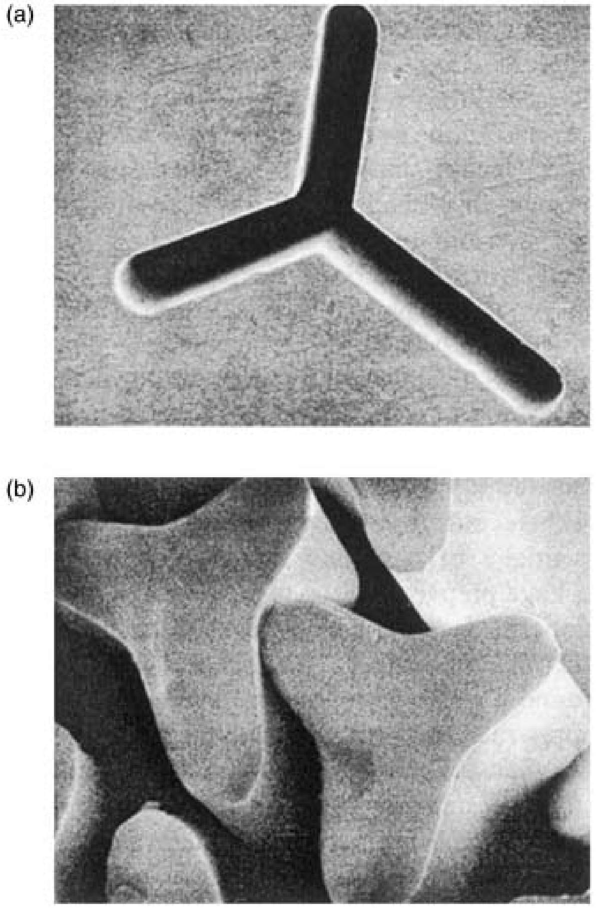
Melt spinning

Because of its extremely low solubility in low-boiling and inexpensive organic solvents, nylon 66 required a new technique for converting the solid polymer into fibers; hence the development of "melt" spinning, the third basic method for producing manufactured fibers. The following description refers essentially to nylon 66 because it was the first to use the method, but the process applies, in general, to all melt-spun manufactured fibers.

In the original production of nylon fiber by melt spinning, the chips of predried polymer were fed from a chamber onto a melting grid whose holes were so small that only passage of molten polymer was possible. Both solid and liquid were prevented from contacting oxygen by maintaining an inert nitrogen atmosphere over the polymer supply. The polymer melted in contact with the hot grid and dripped into a pool where it became the supply for the spinning itself. This melting operation has been entirely replaced by delivery of the molten polymer pumped directly from the polymerization stage or by "screw" melting. In the latter process, the solid polymer in chip form is fed into an extrusion-type screw contained in a heated tube. The depth and the helix angle of the grooves are engineered in such a way that melting takes place in the rear section, and the molten polymer is moved forward under increasing pressure to a uniformly heated chamber preceding the metering pump.

Whatever means is used to secure the molten polymer, it is moved forward to a gear-type pump that provides both high pressure and a constant rate of flow to the final filter and spinnerette. The filter consists of either sintered metal candle filters, several metal screens of increasing fineness, or graded sand arranged in such a way that the finest sand is at the bottom. After being filtered, the molten polymer at a pressure of several thousand pounds per square inch is extruded through the small capillaries in the heavily constructed spinnerette. It is necessary to maintain the temperature of the pool, pump, filter, spinnerette assembly, and spinnerette at about 2030°C above the melting point of the nylon, which is about 264°C for nylon 66 and 220°C for nylon 6. Fibers having desired cross-sectional shapes can be produced by selecting spinnerettes containing holes of appropriate configuration. An example of a trilobal spinnerette capillary and the shape of the resulting trilobal fiber is given in Fig. 21.15.

Fig. 21.15 Scanning electron micrographs of: (a) trilobal-shaped spinerette, and (b) resulting cross-sections of nylon fibers. The pictures were taken at different magnifications. (*Courtesy BASF Corporation*)



The nylon production process requires that the extruded fibers emerge from the spinnerette face into a quench chamber where a cross current of relatively cool quench air is provided to promote rapid solidification. The solid filaments then travel down a chimney to cool further, and a lubricant is applied before they make contact with the windup rolls in order to prevent static formation and to reduce friction in subsequent textile operations. The freshly spun yarn from the spinning chamber is taken up by a traversing winder onto a yarn package and "drawn" in a separate operation. In modern high-speed processes, drawing still is required, but in many cases this is combined with spinning in a single operation, as will be described in what follows.

Drawing

It was learned early that the "as-spun" fibers made from nylon 66 could be extended to about four times their original length with very little effort, but that thereafter a marked resistance to extension took place. It was discovered that during this high extension, the entire length of fiber under stress did not extend uniformly. Rather, a "necking down" occurred at one or more points, and when the entire length under tension had passed through this phenomenon, a high-strength fiber was obtained. It also was found that when more than one necking down was allowed to take place in a given length of fiber, a discontinuity occurred at the point where the two came together. Accordingly, the drawing operation was aimed at forcing the drawing to occur at a single point as the yarn advanced from the supply to the takeup package.

Where still used, cold drawing consists essentially of removing the yarn from the package prepared in the melt spinning operation and feeding it forward at a uniformly controlled rate under low tension. It is passed around a godet or roller that determines the supply rate and prevents slippage; for nylon 66, it then is wrapped several times around a stationary snubbing pin. From there it goes to a second roller that rotates faster than the supply roller to produce the desired amount of stretch, usually about 400 percent. The necking down occurs at the pin. In the case of nylon 6, drawing may be effected satisfactorily without passing the yarn around such a snubbing pin.

The long molecules of the nylon 66 or 6 polymer, which are randomly positioned in the molten polymer, when extruded from the spinnerette tend to form "crystalline" areas of molecular dimensions as the polymers solidify in the form of freshly spun fibers. In the drawing operation, both these more ordered portions as well as the amorphous areas tend to become oriented so that the lengthwise dimensions of the molecules become parallel to the long axis of the fiber, and additional intermolecular hydrogen bonding is facilitated. It is this orientation that converts the fiber having low resistance to stress into one of high strength.

By controlling the amount of drawing as well as the conditions under which this operation takes place, it is possible to vary the amount of orientation and the degree of crystallization. A minimal amount is preferable in the manufacture of yarns intended for textile applications wherein elongation of considerable magnitude and low modulus or stiffness is required rather than high strength. On the other hand, strength and high modulus are at a premium when fibers are to be used in tire cords and other industrial applications. High resistance to elongation is imperative if the tire is not to grow under conditions of use. In this connection, it should be noted that nylon tire cord that has been produced by twisting the original tire yarn and plying the ends of these twisted yarns together is hot-stretched just before use at the tire plant to increase strength and reduce even further the tendency to elongate under tension.

The separate operations of spinning and drawing nylon presented a challenge whose object was combination of the two operations into a single continuous step. But the problem was obvious, for the operating speeds of the two separate steps already had been pushed as high as was thought to be possible. How then would it be possible to combine them into a continuous spin-draw, wherein a stretching of about 400 percent could take place? The answer lay in the manner in which the cooling air was used and in the development of improved high-speed winding devices. By first cooling the emerging fibers by a concurrent flow of air and then cooling them further by a countercurrent flow, the vertical length of the cooling columns can be kept within reason. In-line drawing may occur in one or two stages, and relaxation may be induced if needed. The final yarn is said to be packaged at speeds of 6000 m/min.

Other Nylons, Modifications, and New Developments

Although nylon 66 and 6 account for most of the polyamide fibers produced, a great many others have been experimentally synthesized and have been developed and manufactured in commercial amounts. Of these, some have been made into fibers, some with limited economic success. These nylons are identified by either the same numbering system used for nylon 66 or 6 or by a combination of numbers and letters, as follows:

Nylon 3	$\left(NH-(CH_2)_2CO\right)_n$
Dimethyl nylon 3	$\left(NH-C(CH_3)_2-CH_2-CO\right)_n$
Nylon 4	$\left(NH-(CH_2)_3-CO\right)_n$
Nylon 6T	
Nylon 7	$\left(NH-(CH_2)_6-CO\right)_n$
Nylon 12	$\left(NH-(CH_2)_{11}-CO\right)_n$
Nylon PACM-12	
Nylon 46	$\left(NH-(CH_2)_4-NH-CO-(CH_2)_4-CO\right)_n$
Nylon 610	$\left(NH-(CH_2)_6-NH-CO-(CH_2)_8-CO\right)_n$

Dimethyl nylon 3 is solution-spun because it tends to decompose during melt spinning. Nylon 4 has a moisture regain (mass water per unit mass of dry fiber, under standard atmospheric conditions of 20°C and 65% RH) of 69 percent[9] and therefore is superior to other nylons for textile usages, being comparable to cotton. Nylon 11 was developed in France and has been trademarked as Rilsan®. It has a moisture regain of 1.8 percent and density of 1.04 g/cc as compared with 4 percent and 1.14 g/cc, respectively, for nylon 66. Nylon 7 is made in the former Soviet Union and marketed under the name Enant. The fiber has better stability to heat and ultraviolet light than nylon 66 and 6. Nylon 6T, an aromatic polymer, has a much higher melting point (370°C), a higher density (1.21 g/cc), and slightly higher moisture regain (4.5%) than nylon 66. It also has superior resistance to nylon 66 against heat. This fiber has served as a precursor to the development of aramid fibers. Nylon PACM-12, formerly produced under the trade name Qiana® in the United States, is no longer in production.

A *Chemical and Engineering News* report[10] suggests that the most serious competition to nylon 66 and 6 will be provided by a new, still experimental fiber, nylon 46, being developed by DSM in the Netherlands, Trade-named Stanyl®, this fiber results from the interaction of 1,4-diaminobutane and adipic acid. Better order in the structure in the fiber leads to greater crystallinity and, thus, to greater density (1.18 g/cc). The fiber has a melting point of about 300°C, and a breaking stress or tenacity of 9.5 gram force/denier (•1 GPa), modulus at 120°C of 20 gram force/denier (2.1 GPa), and shrinkage in gas at 160°C of 3 percent.

Some of the outstanding characteristics of nylon that are responsible for its many uses in apparel, home furnishing, and industrial products are its high strength and toughness, elastic recovery, resilience, abrasion resistance, and low density. Among many applications of the fiber are such products as intimate apparel and foundation garments, sportswear, carpets, parachutes, tents, sleeping bags, and tire cords.

The world production of nylon has doubled in recent years, increasing from 3.8 billion lb in 1970 to about 7.5 billion lb in 1990. The fiber accounts for about 24 percent of the synthetic fibers produced worldwide. In the United States, the production of the fiber also doubled in two decades, increasing from 1.2 billion lb in 1970 to about 2.4 billion lb in 1990.

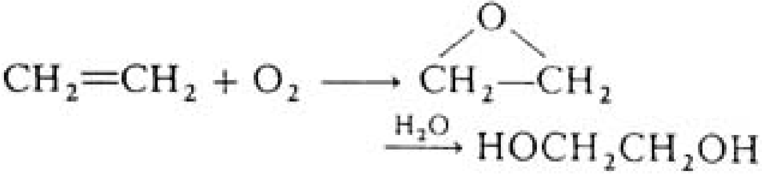
21.8 Polyesters

Historical

The stimulus for the development of polyester, as for nylon, was provided by the fundamental work of Carothers. Although his team's initial work was directed toward this material, because of greater promise shown by polyamides at the time, the developmental work on polyesters was temporarily set aside. The polymer, however, attracted interest in Great Britain, where J. T. Dickson and J. R. Whinfield experimented with it and developed a successful polyester fiber.[7] They found that a synthetic linear polymer could be produced by condensing ethylene glycol with terephthalic acid or by an ester-exchange between the glycol and pure dimethyl terephthalate. The polymer thus obtained could be converted to fibers having valuable properties, including the absence of color. Like nylon, this material has been popularized under its generic name, polyester or just "poly." Those persons working with it commonly refer to it as PET. It first appeared under the trade name Terylene® (Imperial Chemical Industries, Ltd.,) in England, and was first commercialized in the United States in 1953 as Dacron® (E. I. duPont de Nemours & Co).

Manufacture

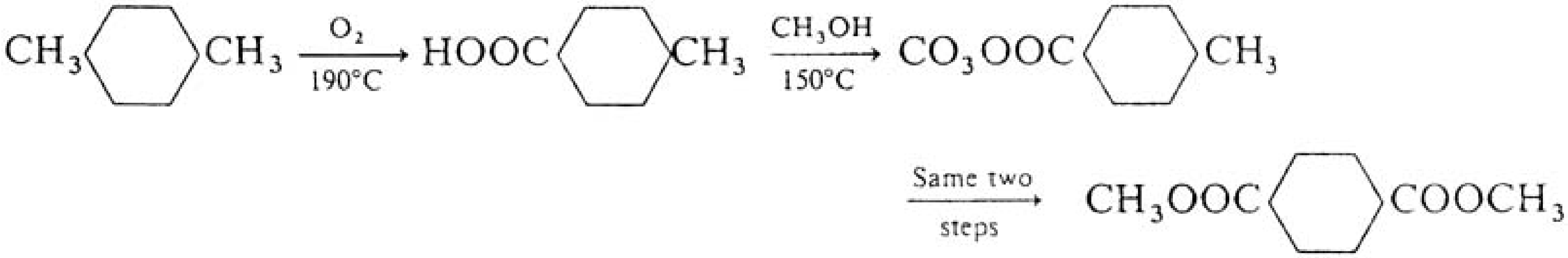
When the development of polyethylene terephthalate (PET) occurred, ethylene glycol already was being produced in large amounts from ethylene, a by-product of petroleum cracking, by the oxidation of ethylene to ethylene oxide and subsequent hydration to ethylene glycol, which, in a noncatalytic process, uses high pressure and temperature in the presence of excess water.



$$CH_2{=}CH_2 + O_2 \longrightarrow CH_2{-}CH_2 \xrightarrow{H_2O} HOCH_2CH_2OH$$

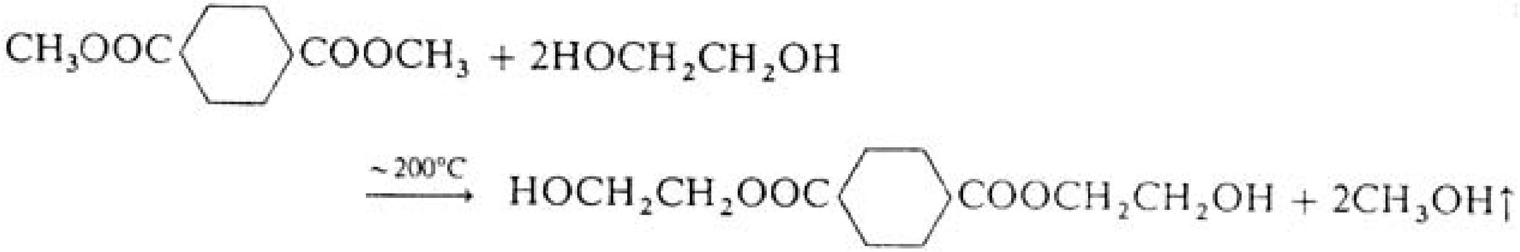
On the other hand, although *o*-phthalic acid, or rather its anhydride, had long been produced in enormous amounts for use in the manufacture of alkyd resins, the *para* derivative was less well known and not available on a large scale. The synthesis is a straightforward one, however, from *p*-xylene, which is oxidized to terephthalic acid, either by means of nitric acid in the older process or by air (catalyzed) in the newer one. In the early years this compound then was converted to the easily purified dimethyl ester in order to obtain a colorless polymer adequate for the manufacture of commercially acceptable fibers.

Several other methods were developed for producing the desired dimethyl terephthalate. The Witten (Hercules) process goes from *p*-xylene to toluic acid by oxidation of one of the methyl groups on the ring, following which the carboxyl group is esterified with methanol. This process then is repeated with the second methyl group to secure the dimethyl ester of terephthalic acid.



Either phthalic anhydride or toluene, both in ample supply as raw materials, can be used in the Henkel processes. Use of phthalic anhydride depends only upon dry isomerization of the potassium salt of the *ortho* derivative to the *para* form at about 430°C and 20 atmospheric pressure; or toluene is oxidized to benzoic acid, whose potassium salt can be converted to benzene and the potassium salt of terephthalic acid by disporportionation.

The first step in the reaction of dimethyl-terephthalate and ethylene glycol is transesterification to form bis(*p*-hydroxyethyl) terephthalate (bis-HET) and eliminate methanol.



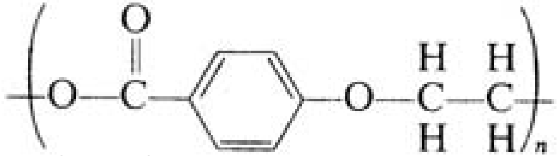
This product then is polymerized in the presence of a catalyst to a low molecular weight compound and the by-product glycol is eliminated. In a second stage, at a temperature of about 275°C and under a high vacuum, the molecular weight is raised to secure the melt viscosity desired for the particular material involved. Like nylon, this final material may be extruded, cooled, and cut into chips for storage and remelting, or it may be pumped directly to the spinning machines.

From the beginning, it was obvious that there would be considerable progress in industrial chemistry, to say nothing of cost reduction, if the process could be simplified by making it unnecessary to go through the dimethyl derivative to secure a product of adequate purity. This was accomplished in the early 1960s when methods of purifying the crude terephthalic acid were developed, and conditions and catalysts were found that made possible the continuous production of a color-free polymer. It is said that the selection of the catalyst is especially aimed at the prevention of ether linkages in the polymer chain due to intracondensation of the glycol end groups.

Two additional rather similar routes are known. Both depend upon the reaction between ethylene oxide, rather than ethylene glycol, and terephthalic acid to form the bis-HET monomer already mentioned. The difference between the two methods lies in the point where purification is done: in one case, it is the crude terephthalic acid; in the other, it is the bis-HET monomer. In both cases this monomer is polymerized by known procedures to form a fiber-grade polyester. The titanium dioxide delustrant is added, as might be expected, early in the polymerizing process.

Another polyester that has reached long-term commercialization is now produced in limited volume as Kodel 200® by Tennessee Eastman Co. and is considered to be 14 cyclohexylene dimethylene terephthalate. The glycol that is used instead of ethylene glycol in this process exists in two isomeric forms, one melting at 43°C and the other at 67°C. This makes possible their separation by crystallization, to secure the desired ratio of the two forms for conversion to the polymer. This ratio determines the melting point of the polymer, a most important property for a material that is to be melt-spun. The polymer from the 100 percent *cis* form melts at 275°C, and that from the 100 percent *trans* form at 318°C. Indications are that the commercial product is about 30/70 *cis-trans*.

In 1973, the Federal Trade Commission modified the generic definition of polyester to include in the polyester category materials that previously were polyester ethers or benzoate polyesters. As a result, the fiber known as poly (ethylene oxybenzoate) or PEB and manufactured under the trade name A-Tell in Japan came to be known as polyester. This material is made by reacting parahydroxy-benzoic acid and ethylene oxide to give paraoxyethylenebenzoic acid, which is then polymerized to obtain PEB:



The fiber softens at about 200°C and melts at 225°C. It is said to have a silk-like hand and appearance and other properties comparable to those of other polyesters.

Polyesters are melt-spun in equipment essentially the same as that used for nylon, already described. Wherever the volume is large and the stability of demand is adequate, the molten polymer is pumped directly from the final polymerization stage to the melt-spinning machine. The molten polymer is both metered and moved forward at high pressure by use of an extruder coupled with a gear-type pump, through filters to the spinnerette, which contains capillaries of about 9 mils (230 µm) diameter. Great care is taken to eliminate moisture and oxygen from the chips, if they are used, and from the spinning chamber. When the polyester fibers are destined to become staple, the emerging filaments from a number of spinnerettes are combined to form a tow, which can be further processed as a unit. Continuous filament yarn is packaged for further processing such as drawing or texturing. Spin-drawing, described later, has become commonplace today and represents major cost savings to the fiber manufacturer.

Drawing

Unlike nylon, which in the as-spun state contains a high amount of crystalline component, PET fibers are essentially amorphous as spun. In order to secure a usable textile yarn or staple fiber, this product must be drawn under conditions that will result in an increase in both molecular orientation and crystallinity. This is done by drawing at a temperature well above the glass transition point, Tg, which is about 80°C. Conditions of rate and temperature must be selected so that the amorphous areas are oriented, and crystallization will take place as the temperature of the drawn fibers drops to room temperature. An appropriate contact-type hot plate or other device is used, and about 300400 percent extension is effected. Figure 21.16 shows a drawn single filament. As with nylon, the conditions of draw, especially the amount, determine the forceelongation properties of the product. Industrial-type yarns, such as those intended to be used as tire cord, are more highly drawn than other yarns and have higher strength with less elongation. The fibers develop the much desired oriented fibrillar crystalline morphology for such applications (see Fig. 21.17).

Fig. 21.16 Drawnecking in polyester single filament. (*Courtesy E. I. DuPont de Nemours & Co.*)

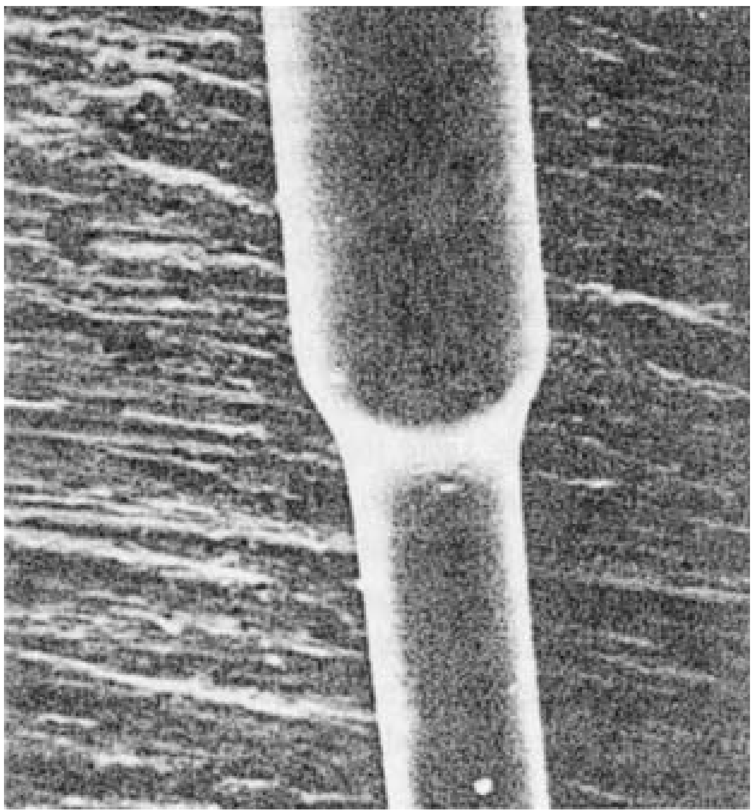
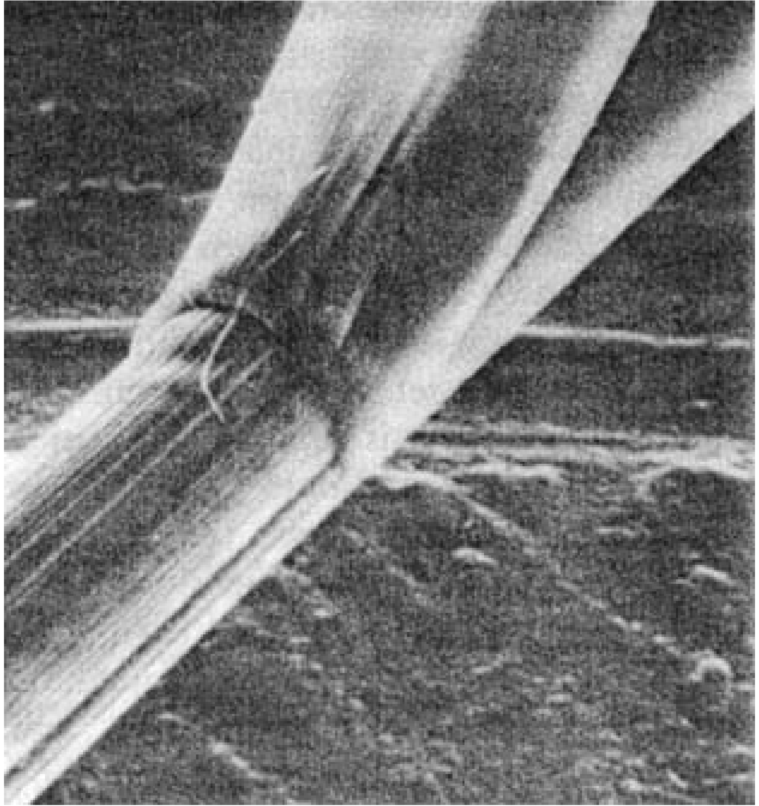


Fig. 21.17 Skin-peeling in polyester showing fibrillar structure. (*Courtesy E. I. DuPont de Nemours & Co.*)



Heat Setting

The ability of textile fibers to be "set" is not characteristic of manufactured fibers alone. Aided in many cases by the presence of starch, cotton fabrics can be ironed to a smooth and wrinkle-free condition; also, a sharp crease in wool trousers has been commonplace for generations. In other words, these fabrics were exposed to moisture at elevated temperatures while being held or pressed into desired geometrical configurations and then allowed to cool before being released from constraint. Such fabrics tend to remain unchanged while cool and dry, even though the fibers from which they are formed carry internal stresses; but reversion takes place upon washing or exposure to high relative humidity.

With the development of nylon, and especially polyesters, a durable kind of setting has become possible. When fabrics made from these fibers are shaped and then exposed to elevated temperatures either in the dry condition or, in the case of nylon particularly, in the presence of water vapor, thermoplastic relaxation of induced stresses in the fiber takes place and configurations at the molecular level adjust to a new and lower energy level. This depends on not only the temperature used but also the duration of the exposure. Thus a few seconds at 230°C will produce the same results as exposure for a considerably longer period at a temperature 5075° lower. The permanency of the setting, that is, the ability of a fabric or garment to return to its original configuration after temporary distortion even while exposed to moisture and raised temperatures, is a function of the severity of the heat setting. To impart true permanence, it is essential that the internal crystalline structure be annealed.

It is this property of polyamides and polyesters that has been the main factor contributing to "ease of care" and the "wash and wear" characteristics of garments made from these polymers. In turn, these garments have revolutionized both the textile and the apparel industries.

Textured Yarns

Fundamentally, the manufacture of "textured" yarns is closely related to the heat setting of fabrics, which must be composed of thermoplastic fibers such as nylon or polyester, the difference being that the individual filaments or bundle of filaments in textured yarns are distorted from an essentially straight rodlike form and then heat-set. In some instances, the fibers are distorted in a more or less random way; at other times, a regular pattern is introduced.

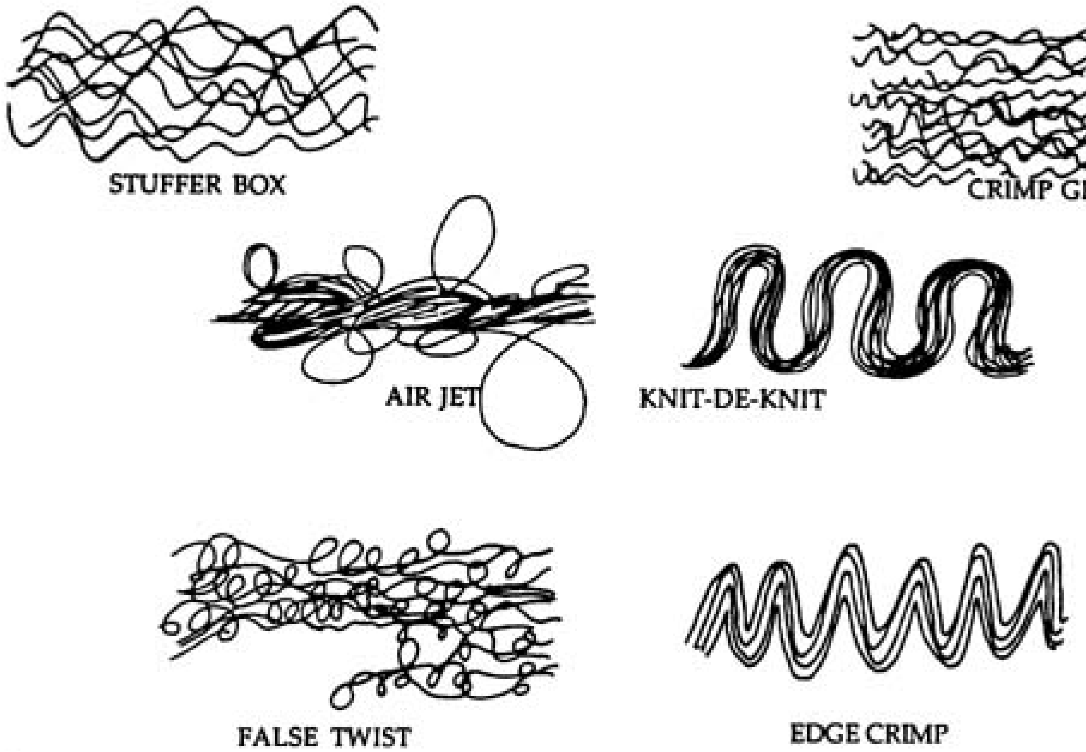
The first commercially successful textured yarn was produced by highly twisting nylon 66, heat-setting it as a full package of yarn, and then untwisting it through zero and a small amount of twist in the opposite direction. This process changed yarn from a close-packed structure to one that was voluminous because of mutual interference of distorted filaments. The technique of heat-setting the twisted yarn as a batch-unit operation now has been replaced by a continuous operation, using what is known as a "false twisting" process. This is based upon the principle that if a length of yarn is prevented from rotating at both ends but is rotated on its axis at its center point, the resulting two sections will contain both "Z" and "S" twists in equal amounts. When this occurs with a moving yarn, any element in it will first receive a twist in one direction, but after passing the false twisting point must revert to zero twist. If it is then made to pass over a hot plate while in the twisted state and is heat-set in that configuration, even after returning to the untwisted condition, the individual filaments will tend to remain distorted when lengthwise stress is released. Because of the low mass and diameters of textile yarns or monofilaments, it is possible to false-twist them at extremely high rotational speeds. Yarn forward speeds of about 1000 m/min are currently obtainable by passing the yarn between, and in contact with, high-speed-friction twisting discs. (When attempts are made to secure higher rates, problems of twist control develop.) The same technique is now more commonly applied to unoriented (undrawn) or partially oriented yarn (POY) at the draw-texturing machine. The resulting yarn may be heat-set as part of the same continuous operation by passing it through a second heater under conditions of overfeed or little or no tension in order to secure both thermally stable geometric configurations in the individual distorted filaments that comprise the yarn, and the degree of "stretchiness" and bulk desired in the final product.

Because these yarns are being made in one less step and also within the plants spinning the parent product, this latest development may be said to constitute another advance in the industrial chemical technology of manufactured-textile products. This draw-texturing appears to be especially applicable to polyester yarns intended for fabrics known as "double knits" and "textured wovens."

Yarn can be forced forward by means of "nip" rolls, although this may seem to be quite contrary to the old adage that one cannot push on an end of string. When this is done so that the yarn is jammed into a receiver (stuffer-box) already full of the preceding materials, it collapses with sharp bends between very short lengths of straight sections. In this condition heat is applied, usually in the form of superheated steam, to set it. In practice, the mass of such yarn is pushed through a heated tube until it escapes at the exit past a spring-loaded gate. During this passage it is heat-set in a highly crimped configuration; then it is cooled before being straightened and wound onto a package. In another continuous process, the yarn or monofilament is pulled under tension over a hot sharp edge so that it is bent beyond its elastic limit and is heat-set in that condition. The process is known as "edge crimping," and the result is not unlike that produced by drawing a human hair over the thumbnail. The process is not used much today, but a yarn with similar crimp is produced by bicomponent spinning.

When such yarns are knitted or woven into fabric, the filaments tend to return to the configurations in which they were originally heat-set. Contraction takes place in the direction of the yarn axis, and this in turn converts the smooth flat fabric into a "stretch" fabric and gives the surface a textured appearance. These fabrics or the garments made from them, whatever the process used to produce the yarns, may be given additional heat treatment to secure stability in a desired geometric configuration. A degree of stretch may be retained, or a flat and stable textured surface may be produced. There are a number of variations of the texturing process, which, combined with the many possibilities of heat setting, impart considerable versatility to the final product (see Fig. 21.18). The growth in the use of these products in the 1960s is well known. Carpeting also provides a significant market for them, as texture is one of the most important characteristics of soft floor coverings. Such products have been important to the successful use and expanded development of nylon and polyester yarns.

Fig. 21.18 Schematic comparison of the shapes of some textured yarns.



Staple Process

Unlike nylon, which is used mostly in the form of continuous filament yarn, polyester is utilized both in staple and in continuous filament form in large volumes. For producing staple fiber, the spinning machines consist of a series of packs, 1030, each with 10003000 holes. The extrudates from different packs after solidification and application of finish are combined to form a subtow and collected in a can. Then several such cans are brought together in a creel area, and subtows from these cans are combined to feed a staple drawline. The latter may involve one or more stages of drawing and relaxation, one or more stuffer-box type crimpers, a drying unit, and a cutter. A baling unit following the cutter collects and bales the cut fiber.

Continuous Filament Yarn Process Variants

For producing continuous filament yarn, several process routes are available,11 each of which involves the basic processes of polymer synthesis, extrusion, quenching, stretching, and winding. In one process, specially meant for textile-type uses, spinnerettes with 20100 holes typically are used. After solidification and finish application, the filaments may be split into smaller bundles, depending upon end-use applications, for downstream processing. The drawing can also be done in a single integrated process (spin-draw) immediately after finish application to yield a fully oriented yarn (FOY), which having no bulk or texture, is referred to as a flat yarn. Spinning speeds of the order of 1000 m/min or more and winding speeds of about 4000 m/min are used. Similar yarn may be produced on a separate drawing unit, but this processcalled draw-twisting because a certain amount of twist is also inserted prior to windingis less commonly used today. It is more usual to see the separate drawing process done in conjunction with a texturing processmost commonly false twisting, and called draw-texturingto give a textured yarn. The process leads to orientation and crystallization of the filament structure and bulking of the yarn. The feed yarn for such processes usually is produced at 30004000 m/min spinning speed as POY. The latter is designed to be run on draw-texturing machines at speeds ranging from 800 to 1200 m/min.

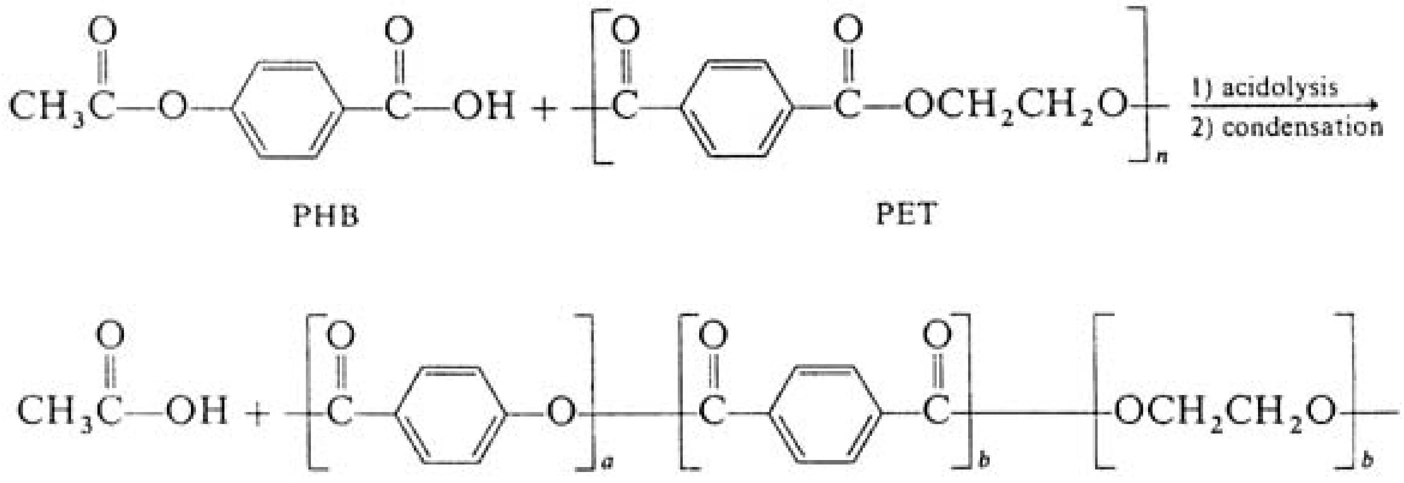
For industrial applications, polyester filaments having high strength low shrinkage (HSLS), low creep properties, and high glass transition temperatures are targeted. To produce such a filament, more severe processing conditions and higher molecular weight polymer are generally used. The filament is spun at low speeds (5001000 m/min), sometimes with retarded quench, to obtain minimum orientation. The drawing can be achieved in the more common integrated, or spin-draw, process or in a separate draw-twist operation. High strength is achieved by drawing the filament to several times its length over very high-temperature rolls and then heat-setting and relaxing the structure prior to winding. Low shrinkage properties are obtained with a relaxing step at high temperature.

Modifications and New Developments

As was the case with the nylon fibers, the potential polyester fibers offered in apparel, home furnishing, and industrial applications was judged to be enormous. For this potential to be realized in practice, however, some characteristics had to be improved, and others had to be engineered for specific end uses. Thus, fibers of different cross-sectional shapes were developed in order to impart anti-soiling, reflective, and resilient characteristics for rug and carpet applications. A difficulty associated with the early polyester fiber that restricted its applications was its lack of ability to take on dyes through one of many methods available for dyeing. This problem was overcome by introducing chemicals that added sulfonate groups to the molecule and by substituting in some cases isophthalic acid for a small portion of terephthalic acid. These changes allowed fibers to be dyed by cationic and disperse dyes, the dyes most frequently used for polyester. Another area of modification has been the development of inherently flame-resistant fiber. One process involves copolymerizing a derivative of phosphoric acid with PET. An exciting new development in polyester filament yarn for apparel uses is the production of microdenier fiber (denier per filament less than one), discussed later in a separate section. The introduction of finer-denier yarns opened up a whole new field for developing fabrics with special esthetic and performance characteristics that were not possible earlier.

With the use of fibers in conveyor belts, tires, and composites, fibers of greater strength and modulus and lower extensibility have been needed. Much effort was directed in the 1970s, and later, to developing such fibers from polyester. The composition and the properties of wholly aromatic polyamides or aramids are discussed in a later section. When both the diacid and diamine components are *para*-substituted aromatic compounds, the resulting polymer is capable of forming lyotropic liquidcrystalline solutions. These solutions can be dry- or wet-spun into fibers with unusually high tensile strength and tensile modulus. When a similar strategy is tried to make polyester fiber from a homopolymer of a *para*-substituted aromatic diacid and a *para*-substituted aromatic diol, only infusible and intractable materials are obtained. A solution to this problem has been found in the development of polyester copolymers that give thermotropic liquidcrystalline melts over a useful temperature range and have viscosities suitable for melt extrusion into fibers or films having high levels of orientation. Spin-line stretch factors of the order of several hundred percent are used to achieve orientation, and physical properties are developed further by heat treatment at temperatures approaching melting conditions.

The first fibers from a thermotropic liquid crystalline melt whose properties were reported were spun from a copolyester of *para*-hydroxybenzoic acid (PHB) and PET by workers at Tennessee Eastman Co. The preparation of the copolymer proceeds in two stages. First, *para*-acetoxybenzoic acid is reacted with PET in an acidolysis step to give a copolyester prepolymer, which in the second step is condensed further to a higher degree of polymerization suitable for fiber formation.



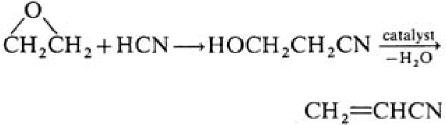
When the mol. percent of PHB in the copolymer exceeds about 3040 percent, a liquidcrystalline melt is obtained. Up to about 60 mol. percent, order in the melt increases and melt viscosity decreases. Compositions containing about 60 mol. percent PHB can be melt-spun into fibers using standard extrusion techniques. It is the unusual combination of properties that makes this class of materials valuable for the formation of high-strength fibers and plastics.

Among melt-spun fibers, those based on thermotropic liquidcrystalline melts have the highest strength and rigidity reported to date, and appear comparable to polyamides spun from lyotropic liquidscrystalline solutions. This was a very active field of research in the 1970s and later, and many comonomers have been reported. Obviously, these compositions must contain three components at a minimum, but many have four or five components. Some frequently used constituents, in addition to those mentioned above, are 2,6-naphthalene-dicarboxylic acid, hydro-quinone, 4,4′-biphenol, isophthalic acid, and 4,4′-dihydroxy-diphenyl ether.

21.9 Acrylics

Polymer Manufacture

Acrylic fibers are spun from polymers that are made from monomers containing a minimum of 85 percent acrylonitrile. This compound may be made from hydrogen cyanide and ethylene oxide through the intermediate ethylene cyanohydrin:



$$\overset{O}{CH_2CH_2} + HCN \longrightarrow HOCH_2CH_2CN \xrightarrow[-H_2O]{\text{catalyst}} CH_2{=}CHCN$$

It also may be made directly from acetylene and hydrogen cyanide:

$$CH - CH + HCN \rightarrow CH_2{=}CHCN$$

But the reaction that currently is preferred uses propylene, ammonia, and air:

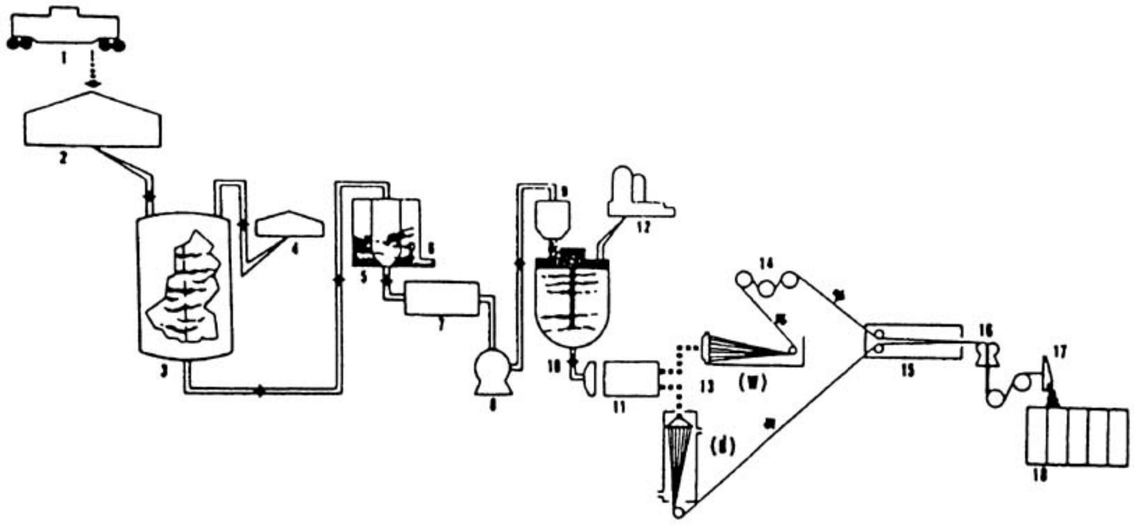
$$3CH_2{=}CHCH_3 + 3NH_3 + 7O_2\,(\text{air}) \xrightarrow[<500°C]{\text{Catalyst}} CH_2{=}CHCN + 2CO + CO_2 + CH_3CN + HCN + 10H_2O$$

Pure acrylonitrile may polymerize at room temperature to polyacrylonitrile (PAN), a compound that, unlike polyamides and polyesters, does not melt at elevated temperatures but only softens and finally discolors and decomposes. Nor is it soluble in inexpensive low-boiling organic solvents. Because fibers made from it resist the dyeing operations commonly used in the textile industry, the usual practice is to modify it by copolymerization with other monomers, for example, vinyl acetate, styrene, acrylic esters, acrylamide, or vinyl pyridine in amounts up to 15 percent of the total weight (beyond which the final product may not be termed an acrylic fiber). The choice of modifier depends on the characteristics that a given manufacturer considers important in a fiber, the availability and cost of the raw materials in the manufacturer's particular area of production, and the patent situation.

In copolymerizing acrylonitrile with another monomer, conditions must be controlled in such a way that the reaction produces a polymer having the desired chain structure and length. The reaction takes place in the presence of substances capable of producing free radicals. In addition, certain trace metals that have been found to increase reaction rates offer a means of controlling chain length. When polymerization is carried out in solution, after an induction period, the reaction is rapid and liberates a considerable amount of heat. Furthermore, because the polymer is not soluble in the monomer, a thick paste is formed. These facts limit the usefulness of such a process. Carrying out the polymerization in the presence of a large amount of water (water/monomer of 2/1 to 3/1) is a convenient method and the one most generally used. In this case the polymer forms a slurry, and the water provides a means for removing the heat from the site of the reaction. Moreover, most of the common redox-catalyst systems are water-soluble. Polymerization may be carried out batchwise or by a continuous process.

In the batch method, the monomers and catalyst solutions are fed slowly into an agitated vessel containing a quantity of water. The heat of reaction is removed either by circulating cold water through the jacket surrounding the vessel or by operating the reaction mixture at reflux temperature and eliminating the heat through the condenser water. The monomer and catalyst feeds are stopped when the desired amounts have been added, and polymerization is allowed to continue until there is only a small amount of monomer remaining in the reaction mixture. Then the slurry is dumped from the reaction vessel, filtered, washed, and dried (see Fig. 21.19).

Fig. 21.19 Flow diagram for the manufacture of acrylic fiber: (1) acrylonitrile; (2) tank farm; (3) polymerizer; (4) comonomer and catalyst; (5) centrifuge; (6) waste liquid; (7) dried polymer; (8) grinding; (9) polymer storage; (10) dissolver; (11) filter; (12) solvent plant; (13) spinnerette; (13w) wet spinning; (13d) dry spinning; (14) roller dryer; (15) additional treatment; (16) crimper; (17) cutter; (18) acrylic fiber bale.



In the continuous-overflow method, rather than stopping the monomer and catalyst feed when the reaction vessel is full, the slurry is simply allowed to overflow; the solids are removed by filtration, washed, and dried. The filtrate contains a certain amount of unreacted monomer, which is recovered by steam distillation after the trace metal present has been chelated to stop the redox reaction and thus further polymerization. The dried polymer is the raw material from which fibers are spun.

As mentioned, PAN polymer has two major weaknesses: (1) extremely poor dyeing characteristics due to highly ordered structure, and (2) very low solubility in most of the solvents. To overcome these problems, comonomers are incorporated into the chains. Comonomers have been used to improve hydrophilicity, dyeability, flame retardancy, heat resistance, and so on. For obtaining hydrophilic acrylic fiber, comonomers containing hydrophilic functional groups such as hydroxyl, ester, carboxyl, amide, and substituted amide are used. To make the fiber base-dyeable, comonomer containing carboxylic and sulfonic acid groups is introduced; and to obtain acid-dyeable fiber, comonomers containing pyridine, aliphatic amine, or quaternary ammonium salt are used. Antistatic acrylic fibers can be made by incorporating in the chain polyethylene oxide, polyalkylene derivatives, polyethylene glycol, acrylates, or methacrylates as the comonomers. For improved flame retardancy, halogen-containing vinyl comonomers are used.

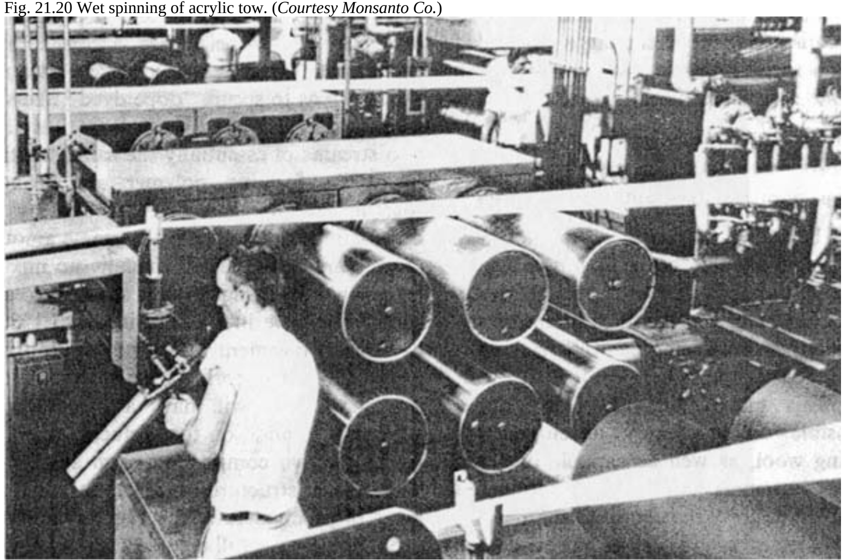
Spinning

As already indicated, pure PAN softens at elevated temperatures, and thermal decomposition starts before the molten state is reached. The same is true of the copolymers commonly used to produce fibers. Accordingly, melt spinning is impossible; spinning must be done from a solution of the polymer. Both dry and wet spinning are carried out in current commercial operations.

The operations used to either wet- or dry-spin acrylics are essentially the same as those already described for rayon and acetate, respectively. The polymer must be completely dissolved in solvent and the solution filtered to remove any impurities that would cause spinnerette blockage. Because acrylic polymers are not soluble in common nonpolar solvents, polar substances such as dimethylformamide, dimethylacetamide, or aqueous solutions of inorganic salts such as zinc chloride or sodium thiocyanate are required. Only wet spinning is possible with the latter. Dimethyl formamide boils at 152.8°C and exerts a vapor pressure of 3.7 mm of Hg at 25°C compared with acetone (used in dry spinning of cellulose acetate), which has a vapor pressure of 228.2 mm of Hg at 25°C. It follows that, unlike acetone which requires an activated-carbon system for recovery, dimethylformamide may be condensed directly from the gas stream used to evaporate the solvent from the forming fiber.

In order to obtain the desired characteristics of modulus, rupture tenacity, and rupture elongation, acrylics, like rayon, require stretching which is usually carried out after the fiber has been spun, either as part of the fiber washing operation or after the fiber has been dried. These same properties are influenced by spinning speeds, and the temperature of the drying air, if they are dry-spun, or the temperature and the composition of the bath, if wet-spun. The multitude of combinations made possible by the use of various comonomers and the flexibility of the fiber-forming operations furnish the different manufacturers with versatility and the users with a variety of acrylic fibers. Figure 21.20 shows a wet-spinning operation for acrylic tow.

Fig. 21.20 Wet spinning of acrylic tow. (*Courtesy Monsanto Co.*)



Acrylic fibers possess a property that made it possible for them in the late 1950s and early 1960s to find immediate, even spectacular, acceptance in the knitted sweater field, until then dominated by wool. When acrylic fibers, normally in the form of a heavy tow, are hot-stretched (e.g., by being drawn over a hot plate and then cooled under tension), they are converted to a labile state. Upon immersion in hot water, such fibers will contract considerably, but not to their prior unstretched length. In practice, this characteristic is used to produce a bulky yarn resembling the woolen yarns long accepted for use in sweaters. The process is described briefly below.

Using "stretch-break" equipment, the stretched labile fibers are further cold-stretched to the breaking point so that the fiber breaks at different points leading to a distribution of fiber lengths, similar to the lengths found in wool. These are crimped and then mixed with thermally stable acrylic fibers that have been stretched and relaxed and have about the same length and degree of crimp. The blend is converted to a spun yarn by the same process used in making woolen yarns, and in turn this yarn is knitted into sweaters and other similar products. When such garments are dyed in hot water, the labile fibers, intimately blended with stable ones, contract lengthwise individually. In the process, segments of the stable units tend to be carried along physically by entrapment and friction; but because such fibers do not change their overall length, the yarn as a whole decreases in length. Lateral displacement of the large volume of stable fibers results in the formation of a more voluminous structure known as "hi-bulk" yarn.

Bicomponent or Conjugate Spun Fibers

As will be shown, it should be theoretically possible to make any of the common manufactured fibers in bicomponent forms. However, acrylics have received the most attention for quite good reasons. Their general characteristics have tended to make them competitive with wool. This means that they should be processible on machinery developed for handling wool, as well as capable of being accepted into markets previously dominated by an animal hair fiber. It follows that because the natural fiber possesses crimp which produces the cohesion that determines its behavior in processing and in part its appearance and "hand" in usage, a similar crimp was desired for acrylics.

The principle that is the basis for bicomponent fibers usually is likened to that which underlies the bicomponent metal strips often used in temperature controllers. With the latter, differential-thermal expansion of the two joined components results in a bending of the thermal element. With fibers, moisture usually is the agent that acts upon the two side-by-side portions. Differential swelling or shrinkage causes the fiber to be brought into a crimped, or preferably, a spirally distorted condition. As such, the side-by-side structure exists naturally in wool.

The combination of small size and large number of holes in a spinnerette might lead one to conclude that it would be almost impossible to design a spinnerette assembly that could bring two streams of polymer or polymer solutions together at each such hole and extrude them side-by-side to form a single filament. Such designs have, in fact, been made; but solutions of fiber-forming polymers fortunately possess properties that encourage laminar flow and thus make other approaches possible. This phenomenon was remarked upon earlier in connection with dope dyeing; when a suspension of a colored pigment is injected into a dope stream, a considerable problem must be overcome to achieve adequate mixing so as to secure "dope dyed" fibers of a uniform color. Thus, it was known that when two streams of essentially the same solution of a fiber-forming polymer are brought together, side-by-side, and moved forward down a pipe or channel by the same amount of pressure behind each, virtually no mixing takes place. By bringing these streams to each spinnerette hole in such an individual side-by-side arrangement and using appropriate mechanical separators, the extruded filament from each hole will have a bicomponent structure. In addition to producing fibers in which the two components form a bilateral symmetrical structure, an ingenious arrangement of predividers of the two streams can produce from the full complement of holes in a single spinnerette a selected group of fibers wherein the amount and the position of each of the two components are randomly distributed throughout their cross-sections. It follows that curls of uniform or random geometry may be produced to meet the required needs.

21.10 Vinyl and Modacrylic Fibers

Vinyls

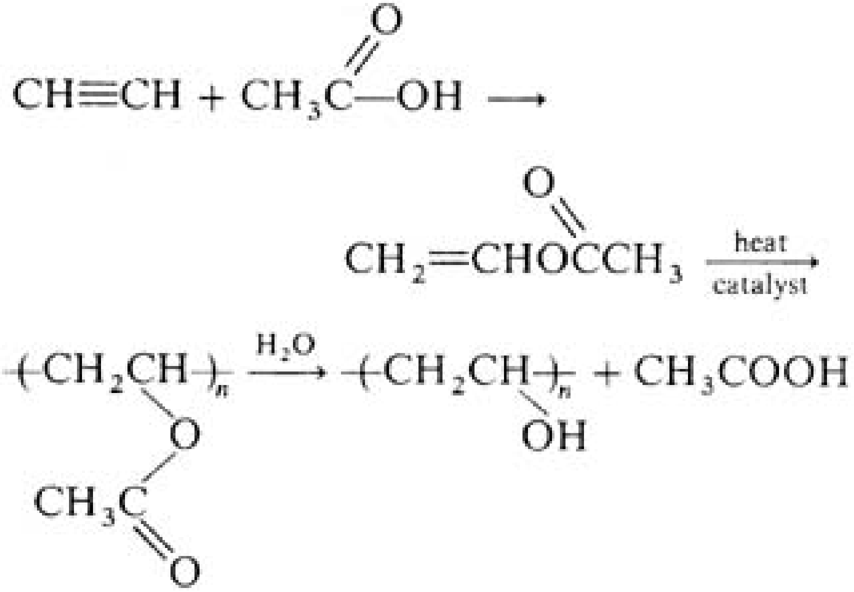
When nylon 66 was developed, it was described as being "synthetic" or "fully synthetic" in order to differentiate it from rayon and acetate. This was no small act of courage, as the word "synthetic," in that period just following the repeal of Prohibition in the United States, was often associated in the public mind with the least palatable kind of alcoholic beverages. In due time, what is known in the advertising business as "puffing" led it to be known as the "first fully synthetic fiber," which was an anachronism. It so happens that fibers based upon polyvinyl chloride (PVC) predated nylon by several years.

About 1931, the production of fibers from PVC was accomplished by dry spinning from a solution in cyclohexanone. But by chlorinating the polymer, it was possible to secure solubility in acetone, which has the advantage of possessing a boiling point about 100°C lower than that of cyclohexanone. Several million pounds per year of this fiber were produced in Germany during World War II to relieve the shortages of other materials. Unfortunately, PVC begins to soften at about 65°C, and in the fibrous state, it shrinks disastrously upon heating. Because of its low softening point, it cannot be dyed at the temperatures commonly used for this purpose, and, furthermore, it resists dyeing.

Modifications of PVC have been produced by copolymerization with other monomers. The first successful one consisted of 90 percent vinyl chloride copolymerized with 10 percent vinyl acetate. It was dry-spun from acetone and given the trade name Vinyon by its producer, Union Carbide Corporation. (In 1960, vinyon was accepted as a generic name for fibers containing not less than 85 percent vinyl chloride.) It has never been produced in large volume; it is used for heat-sealable compositions.

A copolymer of vinyl chloride with vinylidene chloride was used for a number of years to produce melt-spun, heavy monofilaments, which found use in heavy fabrics, where the chemical inertness of the polymer was needed, in outdoor furniture, and in upholstery for seats in public-transportation vehicles.

Another vinyl-based fiber, polyvinyl alcohol, or vinal, was developed in Japan but has not been produced or used in the United States. As such, it illustrates the importance of both relative availability of raw materials and differences in markets, in the success of a chemical product. Acetylene made from calcium carbide is converted to vinyl acetate, which, following polymerization, is saponified to polyvinyl alcohol.



The polyvinyl alcohol is soluble in hot water, and the solution is wet-spun into a coagulating bath consisting of a concentrated solution of sodium sulfate. The fibers are heat-treated to provide temporary stability so that they may be converted to the formal derivative by treatment with an aqueous solution of formaldehyde and sulfuric acid. This final product resists hydrolysis up to the boiling point of water. It seems reasonable to assume that it contains hemiacetal groups and some unreacted hydroxyls on the polymer chain as well as cross-linking acetyl groups between the adjacent molecules.

Under the trade name Kuralon® (Kuraray Co., Ltd), it achieved a production level of about 180 million lb in 1970, but production dropped to 16 million lb of continuous filament and 87 million lb of staple in 1980. The former has been mainly used in industrial rubber products, and the latter has been used mostly for uniforms, nonwoven and coated fabrics, and filters.

Modacrylics

In the United States, the modification of PCV has moved in the direction of copolymerizing vinyl chloride with acrylonitrile, or perhaps it should be said that PAN has been modified by copolymerizing the acrylonitrile with chlorine-containing vinyl compounds. In any case, one modacrylic fiber is currently produced in the United States, a modacrylic being defined as containing at least 35 percent but not over 85 percent acrylonitrile.

The first two modacrylic fibers ever introduced in the United States were Dynel® (by Union Carbide) in 1949 and Verel® (by Tennessee Eastman) in 1956. The former was a copolymer of 60 percent vinyl chloride and 40 percent acrylonitrile, and the latter was said to be a 5050 copolymer of vinylidene chloride and acrylonitrile with perhaps a third component graft-copolymerized onto the primary material to secure dyeability. SEF® and its version for wigs, Elura®, were introduced by Monsanto Fibers in 1972. A few foreign manufacturers are making modacrylic fibers, but the only modacrylic fiber currently in production in the United States is SEF®.

Modacrylic fibers, like acrylic, require after-stretching and heat stabilization in order to develop the necessary properties. It is thought that the stretching is of the order of 9001300 percent, and that, in a separate operation, shrinkage of about 1525 percent is allowed during the time that the fibers are heat stabilized.

The modacrylic fibers, like vinyon and unlike the acrylic fibers, have not become general purpose fibers. They can be dyed satisfactorily and thus are acceptable in many normal textile products; but their non-flammability tends to place them in uses where that property is important, even vital. Blended with other fibers, they are used in carpets; but their largest market is in deep-pile products, such as "fake-furs," or in doll hair, where a fire hazard cannot be tolerated.

21.11 Elastomeric Fibers

The well-known elastic properties of natural rubber early led to processes for preparing it in forms that could be incorporated into fabrics for garments. One such process uses standard rubber technology. A raw rubber of high quality is compounded with sulfur and other necessary chemicals, calendered as a uniform thin sheet onto a large metal drum, and vulcanized under water. The resulting skin is spirally cut into strips that may be as narrow as they are thick, for example, 0.010 in. by 0.010 in.2 in cross-section. These strips are desulfurized, washed, dried, and packaged. Larger cross-sections are easier to make. This product, coming out of the rubber rather than the textile industry, is known as a thread.

Another method produces a monofilament known as a latex thread. As the name would indicate, rubber latex is the raw material, and because extrusion through small holes is required, the purity of the material must be of a high order. With proper stabilization, the latex solution may be shipped from the rubber plantation to the plant, where it is compounded with sulfur and other chemicals needed for curing, as well as with pigments, antioxidants, and similar additives. This is followed by "precuring" to convert the latex to a form that will coagulate upon extrusion into a precipitating bath of dilute acetic acid and will form a filament having sufficient strength for subsequent operations. It passes out of the bath and is washed, dried, vulcanized in one or two stages, and packaged.

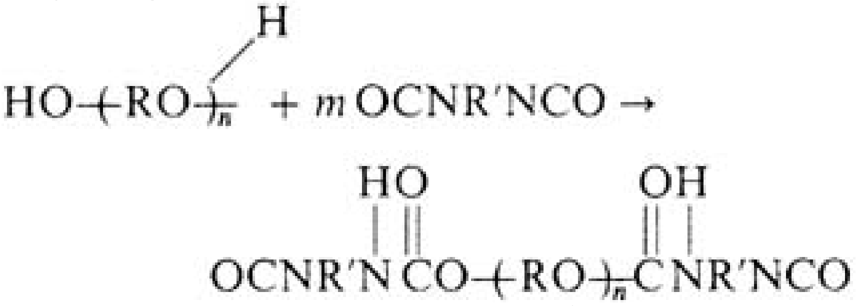
The rubber threads manufactured by either process can be used as such in combination with normal nonelastomeric yarns in fabrics made by weaving or knitting; but most of them, especially those made by a latex process, first are covered by a spiral winding of natural or manufactured yarns. Often two layers are applied in opposite directions to minimize the effects of torque. Such coverings have two purposes. The first is to replace the less desirable "feel" of rubber on human skin by that of the more acceptable "hard" fiber. The second concerns the engineering of desired properties into the product to be woven or knitted into fabric. As an elastomeric material begins to recover from a state of high elongation, it supplies a high stress; but as it approaches its original unstretched condition, the stress drops to a very low order. When wound in an elongated state with a yarn having high initial modulus and strength, the elastomeric component cannot retract completely because its lateral expansion is limited, and jamming of the winding yarn occurs. Thus, the combination of such materials can be made to provide stretch and recovery characteristics needed for a broad spectrum of applications.

The traditional elastomeric threads have been subject to certain inherent limitations, however. The presence of unreacted double bonds makes them sensitive to oxidation, especially with exposure to the ultraviolet radiation of direct sunlight. They also have low resistance to laundry and household bleaches and dry-cleaning fluids.

During recent years, elastomeric yarns or threads have been used to impart comfort, fit, and shape retention to a variety of garments such as women's hosiery and swimwear. Such garments must be thin and highly effective per unit of weight. The materials of which they are composed must be compatible with these requirements. Thus, it was not unexpected that the producers of manufactured fibers, already eminently successful in meeting the needs of the marketplace, should look to the field of elastomeric fibers for new possibilities. Given the limitations of rubbers, both natural and synthetic, as well as the relationships between molecular structure and behavior of fiber-forming linear polymers, the scientists faced new challenges.

As an oversimplification, it can be said that within limits a rubber-like material can be stretched relatively easily but reaches a state where crystallization tends to occur. The structure produced in this manner resists further extension, and the modulus rises sharply. In contrast to the conditions that occur when the manufactured fibers discussed earlier such as nylon or polyester are drawn to form fibers of stable geometry in the crystalline and oriented states, the crystalline state of the elastomeric fibers is labile unless the temperature is lowered materially. Thus, to improve on the chemical sensitivity of rubber, new approaches were necessary. The solution was found in developing linear block copolymers containing "soft" liquid-like sections that impart elasticity, connected with "hard" components that act as tie points to hold the structure together.

The soft, flexible, and low-melting part is commonly an aliphatic polyether or a polyester with hydroxyl end groups and molecular weight in the range of 5004000. The hard portion is derived from an aromatic diisocyanate supplied in an amount that will react with both end groups of the polyether or polyester to form urethane groups. The product, an intermediate known as a pre-polymer, is a thick liquid composed essentially of molecules carrying active isocyanate groups at each end. For example:



$$\mathrm{HO{-}(RO)_n{-}H} + m\,\mathrm{OCNR'NCO} \rightarrow \mathrm{OCNR'}\overset{\mathrm{H}}{\mathrm{N}}\overset{\mathrm{O}}{\mathrm{C}}\mathrm{O{-}(RO)_n{-}}\overset{\mathrm{O}}{\mathrm{C}}\overset{\mathrm{H}}{\mathrm{N}}\mathrm{R'NCO}$$

where(RO)is an aliphatic polyether chain, R′ is one of several commonly available ring structures, *n* • 1030, and *m* • 1.52.

The elastomeric polymer is obtained by "extending" the prepolymer through its reaction with short-chain diols such as butanediol or diamines such as ethylene diamine, thus completing the formation of hard groups between soft, flexible chains. When amines are used, the final step is typically done in a polar solvent such as dimethyl acetamide. The conversion of these polymers into usable fibers may be accomplished by wet-, dry-, or melt-spinning operations, depending on the polymer. Additives to impart whiteness or improve resistance to ultraviolet radiation and oxidation may be incorporated in the spinning solutions or in the melts.

The development of elastomeric fibers has resulted in a variant of wet spinning called "reaction" or "chemical" spinning. In point of fact, rayon, the first wet-spun material, might properly be said to be produced by "reaction wet spinning" or "chemical wet spinning" because complex chemical reactions always have been involved in that operation. In any case, it has been found that the prepolymer of an elastomeric fiber may be extruded into a bath containing a highly reactive diamine so that the chemical conversion from liquid to solid occurs there.

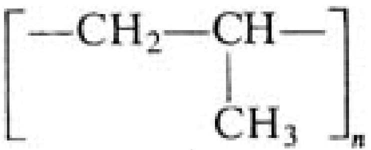
The elastomeric fibers produced in this fashion are based upon segmented polyurethanes and by definition are known generically as spandex yarns. Each manufacturer uses a trade name, for the usual commercial reasons. Perhaps the most noteworthy aspect from the standpoint of industrial chemistry is the multitude of options available to the manufacturer through the ingenious use of various chemicals for soft segments, hard units, chain extenders, and conditions of chemical reaction, followed by numerous possibilities for extrusion and after-treatments. In the United States, there are two main producers of spandex fibers: DuPont (Lycra®) and Globe Rubber Co. (Cleerspan®, Glospan®). There are numerous worldwide producers, including: Bayer, Germany (Dorlastan®); Asahi, Japan (Roica®); Nisshinbo, Japan (Mobilon®); and Tae Kwang, Korea (Acelan®).

21.12 Polyolefin Fibers

Polypropylene

Although polyethylene was considered a source of useful fibers at an early date, its low melting point (110120°C) as well as other limitations precluded active development during the period when production of other fibers based upon the petrochemical industry expanded enormously. The higher melting point of high-density polyethylene gave some promise, but it was overshadowed by the introduction of polypropylene (PP) around 19581959. Great expectations were held for the latter as a quick competitor with the polyamides and the polyesters, already successful, as well as the acrylics, which then were entering the fiber field in volume. PP was thought to have several advantages. The raw material costs were low, only a few cents a pound; also there was a high level of sophistication in the spinning and processing of fibers, and a presumption that this would readily lead to the development of means for converting the polymer to fibers; and, finally, there was the belief that the American consumer would be ready to accept, and perhaps even demand, something new and different, which this polymer offered. However, the limitations of PP fibers, such as lack of dyeability, low melting temperature, low heat stability, and poor light stability, combined with the lower prices and the greater versatility of the already established fibers, dashed the hopes for quick success. However, all of these deficiencies except the low melting temperature and lack of dyeability now have been overcome. The fiber has found an increasingly important place, and its properties have led to new techniques of manufacture and specialized uses.

The structural formula of PP is as given below, where $100,000 < n < 600,000$ for chips or granules, and $50,000 < n < 250,000$ for fibers:



$$\left[-CH_2-\underset{\displaystyle CH_3}{\underset{|}{CH}}-\right]_n$$

The steric configuration is extremely important in the polymer. Only isotactic polypropylene (iPP) has the properties necessary for forming fibers. The molecules are cross-linked only by Van der Waals forces, so it is important that they pack as closely as possible. The isotactic molecules form a 31 helix, as shown in Fig. 21.21,12 and exhibit a high crystallization rate. The atactic molecules, shown in the figure, do not pack well, and although the syndiotactic molecules can pack better and crystallize, this configuration is not a normal product of commonly used catalyst systems. Some properties of isotactic, syndiotactic, and atactic PP are listed in Table 21.2.13 The insolubility of iPP in hydrocarbon solvents at room temperature often is used to separate iPP from atactic polypropylene (aPP).

Fig. 21.21 Diagrams showing (a) irregular atactic, (b) stereoregular isotactic, and (c) stereoregular syndiotactic configurations in polypropylene polymer. (Source: Ahmed, M., "Polypropylene FibersScience and Technology," *Textile Science and Technology*, 5, 16, Elsevier Science Publishers SV, Academic Publishing Div., New York, Amsterdam (1982).)

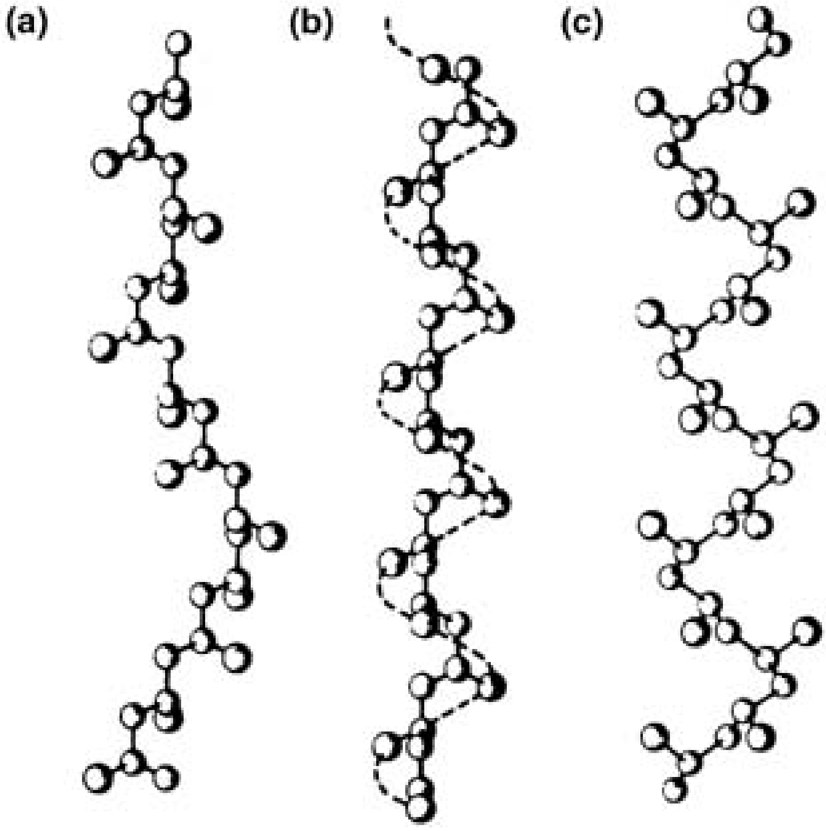


TABLE 21.2 Properties of Isotactic, Syndiotactic, and Atactic Polypropylene

Property	Isotactic	Syndiotactic	Atactic
Density, g/cm^3	0.920.94	0.890.91	0.850.90
Melting Point, °C	165	135	
Solubility in hydrocarbons at 20°C	None	Medium	High
Yield strength	High	Medium	Very low

Source: Lieberman, R. B., and Barbe, P. C., "Propylene Polymers," in *Concise Encyclopedia of Polymer Science and Engineering*, J. I. Kroschwitz (Ed.), p. 916, 1990. Copyright © John Wiley and Sons and reproduced by permission of the copyright owner.

Early in the manufacture of PP, a concept was developed for dry spinning directly from the solution obtained in the polymerization operation. Had it been feasible, it would have been the realization of a chemical engineer's dream: the gaseous olefin fed into one end of the equipment, and the packaged fiber, ready for shipment to a textile mill, coming out the other end. But it did not turn out that way, and today melt spinning is the accepted technique for the production of staple fibers, monofilament, and multifilament yarns. To this usual method have been added the fibrillation and the "slit film" procedures for producing yarns.

The PP materials are completely resistant to bacterial attack, are chemically inert, and are unaffected by water. Monofilaments can be produced that possess high strength, low elongation under stress, and dimensional stability at normal atmospheric temperatures. PP monofilaments have found broad application in cordage and fishing nets (which float), and if highly stabilized they are woven into fabrics used for outdoor furniture, tarpaulins, and similar applications. Large filament denier staple is used widely in "indooroutdoor" carpets. Also, staple fibers have found major applications in tufted indoor carpets and nonwovens used for diaper, filtration, and civil engineering fabrics.

Synthesis

The early PP plants used a slurry process adopted from polyethylene technology. An inert liquid hydrocarbon diluent, such as hexane, was stirred in an autoclave at temperatures and pressures sufficient to keep 1020 percent of the propylene monomer concentrated in the liquid phase. The traditional catalyst system was the crystalline, violet form of $TiCl_3$ and $AlCl(C_2H_5)_2$. Isotactic polymer particles that were formed remained in suspension and were removed as a 2040 percent solid slurry while the atactic portion remained as a solution in the liquid hydrocarbon. The catalyst was deactivated and solubilized by adding HCl and alcohol. The iPP was removed by centrifuging, filtration, or aqueous extraction, and the atactic portion was recovered by evaporation of the solvent. The first plants were inefficient because of low catalyst productivity and low crystalline yields. With some modifications to the catalyst system, basically the same process is in use today.

In 1963, liquid polymerization was introduced in which liquid propylene, catalysts, and hydrogen were pumped continuously into the reactor while polypropylene slurry was transferred to a cyclone separator. The unconverted monomer gas was removed, compressed, condensed, and recycled, and the polymer was treated to reduce the catalyst residue. This system also suffered from a poor catalyst yield, and the polymer produced lacked the required stereospecificity, so that it was necessary to remove the atactic portion of the polymer.

In the mid-1960s, a gas phase process was introduced for production of the polymer. The monomer was pumped over adsorbing beds and entered the reactor with the catalyst system. These feed streams of monomer and catalyst, together with a mechanical stirrer, created a turbulent bed of powdered polymer. Periodically the polymer powder was vented off in a carrier gas to extrusion storage hoppers. Meanwhile, the heat of polymerization was removed by condensing the unreacted monomer in a cooling loop and returning it to the reactor, where it immediately vaporized. This process eventually led to the production of highly crystalline products and was adopted by several companies in the United States.

Most processes in use today rely on a combination of these technologies. Montedison's introduction in 1975 of third-generation catalysts gave high yields and allowed polymerization to take place at 6080°C and 2.53.5 MPa (362507 psi). This was welcome news during the energy crisis, but the resulting polymer was not stereospecific enough to eliminate the need for removal of aPP. Real progress came with the discovery of superactive third-generation catalysts, which gave both the optimal yield and stereospecificity.14

Production

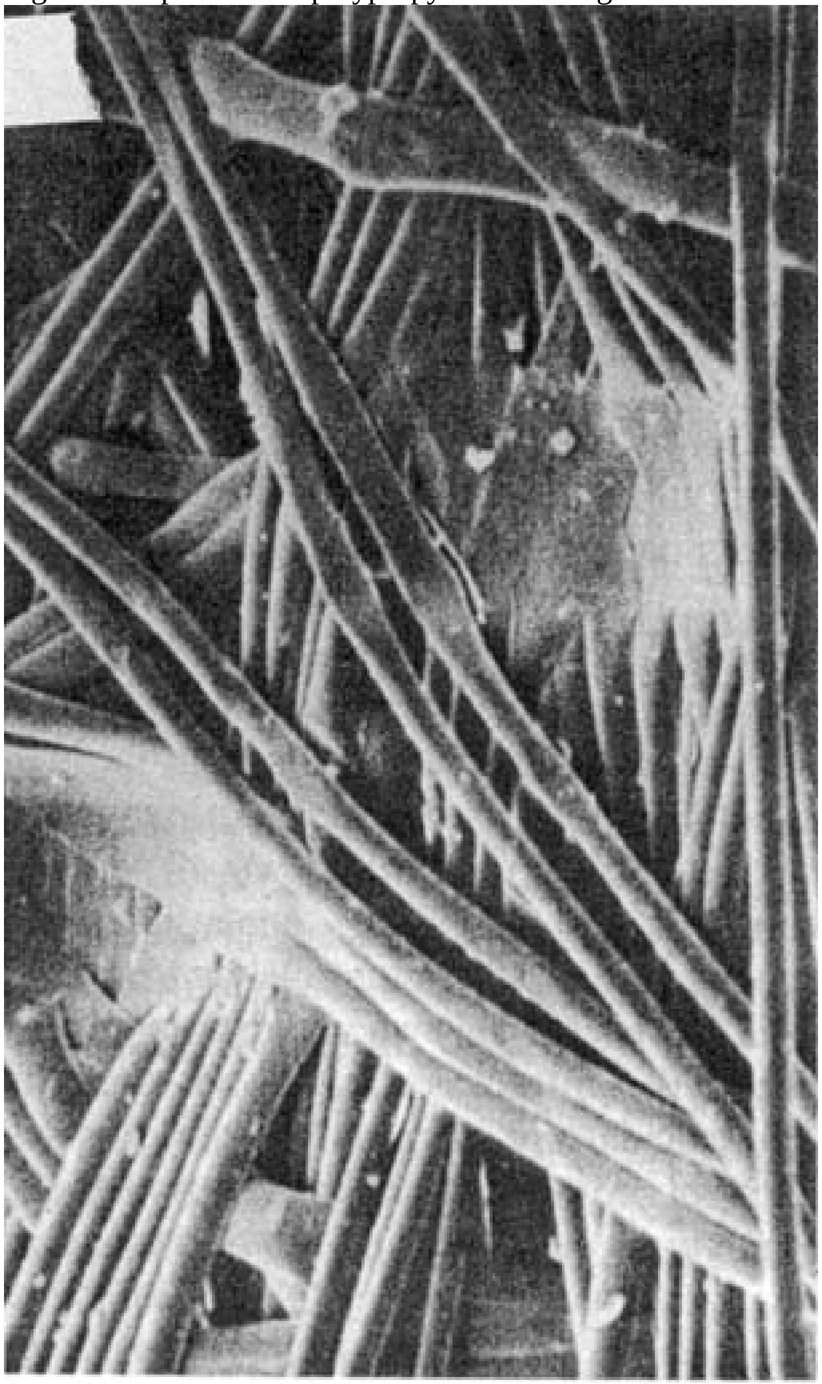
Classical melt spinning, which was developed for the production of nylon filaments, is widely used to produce PP fibers today. It involves a high-speed process (20003000 m/min) that is particularly suitable for long production runs. The average molecular weight of polypropylene polymer, like that of other addition (olefin) polymers, is relatively high compared to that of other polymers. This results in a high melt viscosity; so, unlike the case of other polymers, its extrusion temperatures are 70100°C above its melting point. Single-screw extruders are used for melting and homogenizing the polymer. The screw diameters are from 45 to 200 mm and screw lengths are 24, 30, or 36 times the diameter. The polymer granules are fed into the extruder hopper, where they are melted and homogenized. Chips carrying pigments can be fed into and blended with the main charge of the extruder if colored fibers are desired. The molten polymer is forced through the spinnerette via a screen pack to eliminate any contaminant particles. The spinnerette hole determines the shape of the filaments, and the flow rate and the takeup speed determine the size. The polymer has high specific heat and low thermal conductivity, so the cooling zones must be longer than those for polyester and nylon. For filament yarn production, the filaments are drawn at high speed and wound on packages. For staple fiber production, the filaments are collected in the form of tow and then are drawn, crimped, and cut. The multifilament yarns are often textured to improve bulk and appearance. The false twist method is generally used to texture finer yarns, with the stuffer box used for coarser yarns.

The short spinning method used to produce staple fiber is considerably slower than high-speed spinning. The lower spinning speeds (30150 m/min) would have a negative effect on productivity, but this is counteracted in industry by the use of spinnerettes with a large number of holes (up to 55,000). The required cooling zones are much shorter because of the lower speeds and the use of higher volumes of quench air, which gives this method its name. Because the drawing units can match these low speeds, the two can be fed directly and continuously from the spinning machine to the draw frame to the texturing chamber or the crimper, to produce bulked continuous filament (BCF) yarn or staple fiber, respectively. The short spinning method is used to produce high-tenacity fibers. Many PP yarns are produced using the slit film method. The film extruder is almost identical to a filament extruder. In it, the molten polymer is forced through a film die that converts the melt into film, where the thickness of the film can be controlled by adjusting one of the die lips. The takeoff unit is either a chill roll that removes the film uniformly and cools it below Tm or a water bath followed by nip rolls. The cooled film is slit into separate tapes using a slitter bar that contains a large number of special knives separated by spacers. The film tapes are heated and drawn to their final length.

Several methods are used to produce fibrillated film. They are produced in much the same way as slit film, but these techniques take advantage of the tendency of PP to fibrillate. In one method, a profiled tape is extruded and hot-drawn. It is drawn again to achieve a 10:1 ratio, and the film splits into separate filaments. In another method, called roll embossing, the film is hot-drawn, and then embossed using profiled rollers. The profiled film is drawn again, and fibrillation results. In the pin-roller technique, the film is drawn and cut by knives or pins on a rotating cylinder. This method can produce individual fibers or a controlled web network.

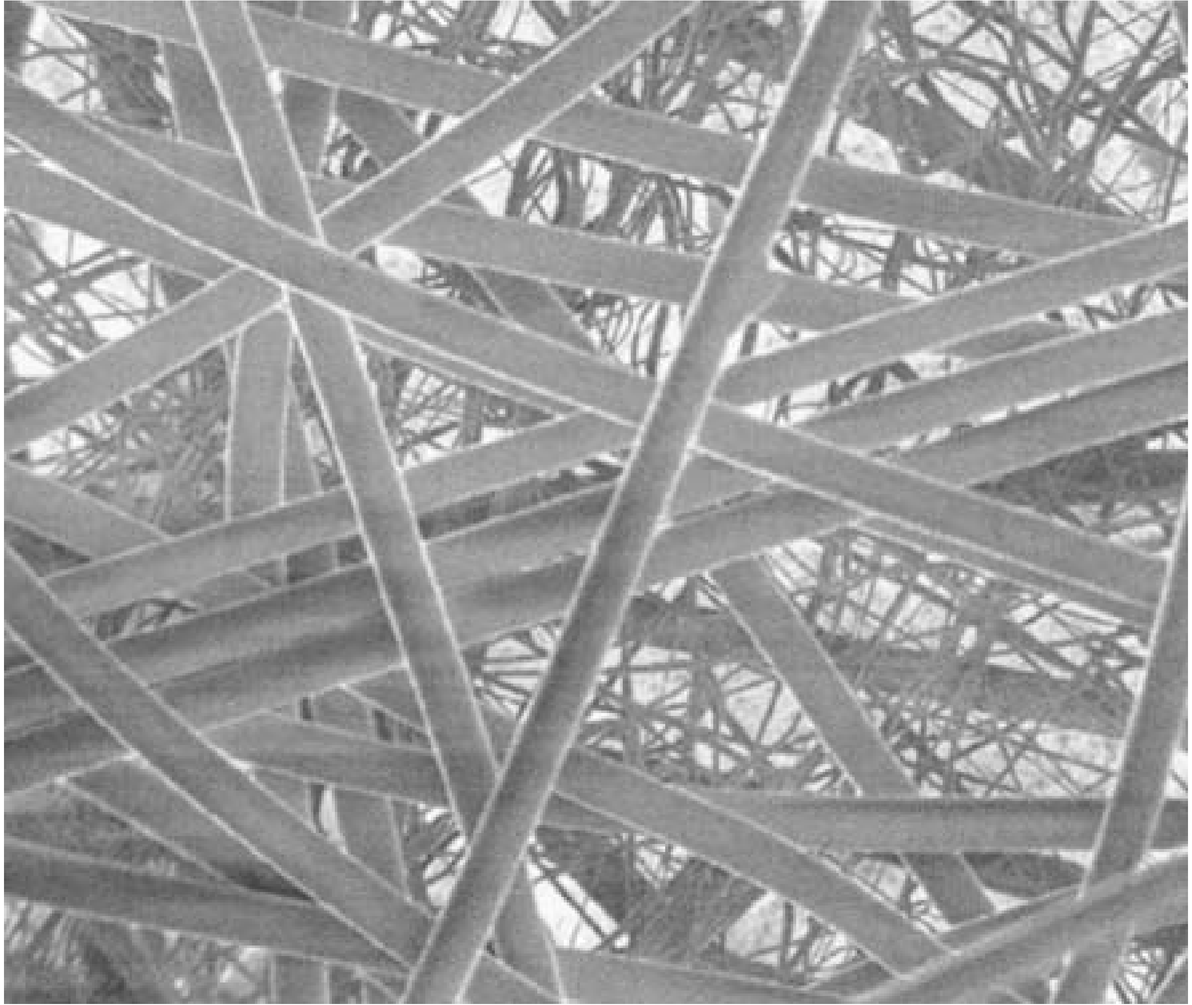
PP nonwovens are created by forming a staple fiberweb and then consolidating it into a fabric. The fibers can be entangled by a needling machine, an air jet, or a water jet. Another method uses a single-stage process in which melt-spun fibers are drawn through an air aspirator jet and deposited randomly on a conveyor. The fibers then are bonded by fusion under heat and pressure to give a "spunbonded" nonwoven fabric (see Fig. 21.22). Yet another method, known as melt blowing, can be used in producing a nonwoven web having special characteristics. The melt-blown (MB) fibers are characterized as ultrafine fibers because of their size relative to other fibers. In the process, molten polymer is forced through a melt blowing die and die tip orifice that are finer than most orifices used to make manufactured fibers. As the polymer emerges, it is attenuated by a jet of high-velocity hot air. This allows the polymer to stay in a molten state but attenuate until broken. The fibers then come in contact with cool quench air, which causes the former to solidify. The fibers deposit on a collector screen and form a MB web. These webs, composed of very fine size fibers, can be engineered for applications in areas where barrier (such as against bacteria), insulative, and absorptive (such as for oil) characteristics are important.

Fig. 21.22 Spunbonded polypropylene showing interfiber bonding, which binds the structure. (*Courtesy E. I. DuPont de Nemours & Co.*)



One of the most recent developments in the use of polyolefin fibers is in composites of spunbond (SB) and melt-blown systems. Some examples of structures being made are SB/MB, known as SM, SB/MB/SB, or SMS, and other combinations of SB and MB. The production and properties of these are particularly enhanced by the use of PP/polyethylene bicomponent fibers in the preparation of the MB layers. Such components are particularly suited for use as light weight coversheets or facings for diapers and sanitary products. The spunbound layers on the top and bottom provide the needed high strength whereas the layers of MB microfibers sandwiched between them provide the greatly enhanced visual uniformity and barrier properties.15 Figure 21.23 illustrates a low weight (15 g/m2) SMMS fabric comprised of MB fibers accounting for less than 15 percent of the structure by weight.16

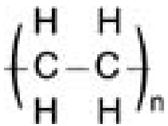
Fig. 21.23 Spunbond/melt-blown composite, SMMS, containing two layers of melt-blown microfibers sandwiched between two layers of spunbonded material and bonded by hot calendaring. (*Source:* Madsen, J. B., *Nonwovens World*, p. 69, 2001, AugustSeptember.)



The production figures of PP show impressive growth. The volume, which was less than 10 million lb in 1965, reached over 1.5 billion lb in 1990. This included approximately 17 percent filament yarn, 23 percent staple fiber, 18 percent spunbonded fabrics, and 38 percent split film products.

High Performance Polyethylene

Polyethylene is probably the material one encounters most abundantly in daily life, such as in the form of grocery bags, shampoo bottles, toys, but now one also finds it as the material in such super high performance products as bullet-proof vests. For such a versatile fibrous material, it has the simplest structure of all known polymers:



$$\left(-\overset{H}{\underset{H}{C}}-\overset{H}{\underset{H}{C}}-\right)_n$$

However, during the general addition polymerization process, some of the carbon atoms, instead of having hydrogen attached to them, have segments of polyethylene chains grown on them, leading to a branched or low-density polyethylene (LDPE). If branching is eliminated or greatly minimized, such as found with the use of the ZieglerNatta polymerization process, utilizing special catalysts, one can get linear chains capable of greater packing. This provides a high-density polyethylene (HDPE) material suitable for fiber use. Molecular weights of the order of one million have been achieved resulting in ultrahigh molecular weight polymer (UHMWPE). Accordingly, a recent addition to the group of high performance fibers is the ultrahigh molecular weight, extended-chain, linear polyethylene. Although fundamental work in the area of developing fully oriented and crystalline structures in polyethylene polymer had been going on since the mid-1960s, it was not until the late 1970s that the possibility of producing such materials on a commercial scale became evident. Presently, three companies worldwide are manufacturing extended-chain polyethylene fibers, which have very high strength and high modulus characteristics.

Routes to High Performance

The achievement of ultimate strength and high modulus in fibers has been a subject of great interest to material scientists and fiber producers. In an attempt to identify the preferred structures for high performance, many theoretical analyses have been conducted to calculate the limiting values. The basis of such calculations is the assumption that in a fiber in which all polymer chains of infinitely long dimensions are extended and oriented parallel to the axis, rupture will occur only when the stress exerted exceeds the intramolecular bond strength. Such calculations show that the limiting values of conventional fibers usually are several times higher than those obtained in actual practice. The reasons for this difference lie in one or more of the following explanations: (1) the molecular weights are not high enough; (2) the chains are not fully extended; (3) the chains are not fully oriented. For a given polymer system and molecular weight, tensile strength and modulus can be enhanced by extending and orienting the chains. Practically, many modern techniques, including zone stretching, multiple-step stretching, and state-of-the-art high-speed extrusion methods, have been adopted to achieve such results.

Still, the results of studies on conventional fibers show that modifications in physical processing alone cannot lead to values that even approach the theoretical maximum within an order of magnitude. Flexible chains, which characterize the bulk of the commercial polymers, tend to conform to a random coil or folded chain structure in an as-spun material, and are very difficult, to reorganize into an extended-chain structure by known methods. Accordingly, in order to achieve ultimate properties, either novel spinning methods are needed, which allow flexible chains to be fully extended, or the chains used must be so configured that they have high intrinsic stiffness and would remain extended in solution or a melt. Both of these goals have been met and are exemplified in the production of (1) high-strength polyethylene fibers, discussed in this section, and (2) aramid fibers, discussed in the next section.

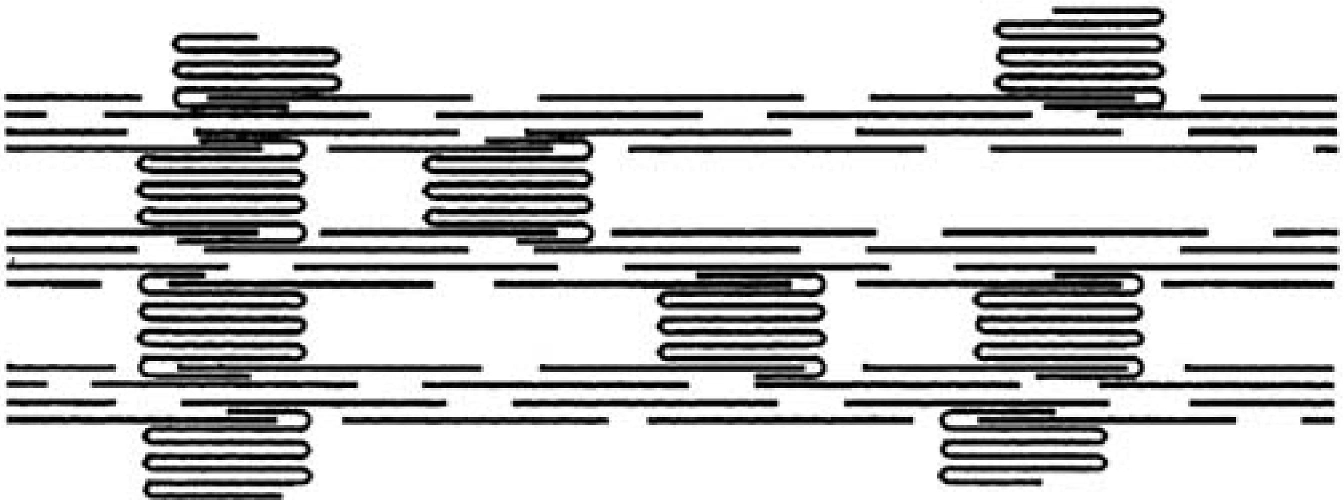
Extended-chain polyethylene fiber became available commercially in 1984 when DSM, a Dutch firm in the Netherlands, introduced Dyneema®, and Mitsui Petrochemicals in Japan announced Tekmilon®. Allied Signal of the United States entered the field in 1985 when it introduced Spectra® fibers. These materials are characterized by very high strength and modulus, which are achieved by the use of ultrahigh molecular weight polyethylene spun by the gel spinning method into fibers having extended-chain structures and near perfect orientation.

Gel Spinning

In general, the purpose for which the gel spinning method is used is to produce an as-spun fiber that contains a loose network of chains with few entanglements, which then can be drawn out to ultrahigh levels to yield a highly oriented structure. The surface-growth method of Pennings,17 which uses a Couette type apparatus, and the gel spinning method of SmithLemstra,18 which uses a more conventional spinning apparatus, led to the achievement of this goal. In the former method, a polymer solution is stirred between two counterrotating cylinders that provides the elongational flow necessary for initial chain alignment. Essentially the same result is achieved by passing the polymer solution through a constriction prior to spinning in the latter method.

In the surface-growth technique, polymer solution between the inner rotating cylinder and the outer stationary cylinder is maintained within a certain temperature range above the polymer crystallization temperature. A fiber seed is immersed through an opening in the outer cylinder, and its tip is made to attach to the polymer layers absorbed on the inner rotating cylinder. By pulling on the other end of the fiber and winding it onto a bobbin, while at the same time replenishing the solution in the gap between the cylinders, a fiber can be produced continuously. This process results in a fibrous precipitate with a "shish-kebab" morphology (Fig. 21.24). One can conclude, then, that the flow field that is developed extends the chains, which then crystallize in a fibrillar form. Chains that are left unextended, those that are below a certain length for the speed, use the preformed clusters (shishes) as nuclei and crystallize as overgrowths in the form of chain-folded lamellae (kebabs). Fibers formed in this way showed remarkable mechanical properties with modulus in excess of 1200 gram force/denier (102 GPa) and tensile strength in the neighborhood of 30 gram force/denier (2.6 GPa).

Fig. 21.24 "Shish-kebab" structure showing extended-chain crystals with lamellar overgrowths.



Production

Recognition that the shish-kebab fibers produced by the surface-growth procedure result from the deformation of a gel-like entangled network layer at the rotor surface led to the development of gel-spun polyethylene fibers. The fiber is made by the solution spinning method. The polymer is dissolved in a high-temperature solvent, such as decalin, at a low concentration, 10 percent or less, and extruded into cold water. Initially the fiber is formed as a gel that still contains the spinning solvent. It can be oriented by drawing at remarkably high draw ratios of 30 or greater, either before or after solvent removal. A method of spinning polyethylene fiber is described in the patent by Kavesh and Prevorsek.19 The factors that govern the properties of fibers produced by this method are the polymer molecular weight, the concentration of polymer in the solvent, the type of solvent, the solution temperature, and the stage and the extent of drawing.

The high-performance behavior in these fibers is obtained from their having a very high molecular weight (15 million) combined with a very high degree of extended-chain crystal continuity. The fiber can aproach a 100 percent crystalline structure with a theoretical maximum density (•0.97 g/cc). Their use of course, is, limited by the melting point of the polymer, which even in the extended form is only about 150°C. Although higher than that of ordinary polyethylene, it is still much lower (80°C or more) than those of commercial textile and other higher performance industrial fibers. There also has been concern about the creep that occurs in these fibers, although significant improvement has been made in this regard since the introduction of the first fiber. However, these fibers have such a unique combination of strength and lightness that they have proved highly successful in a number of applications, such as sailcloth, body armor, medical implants, fishing net, and sports equipment.

21.13 Aramids

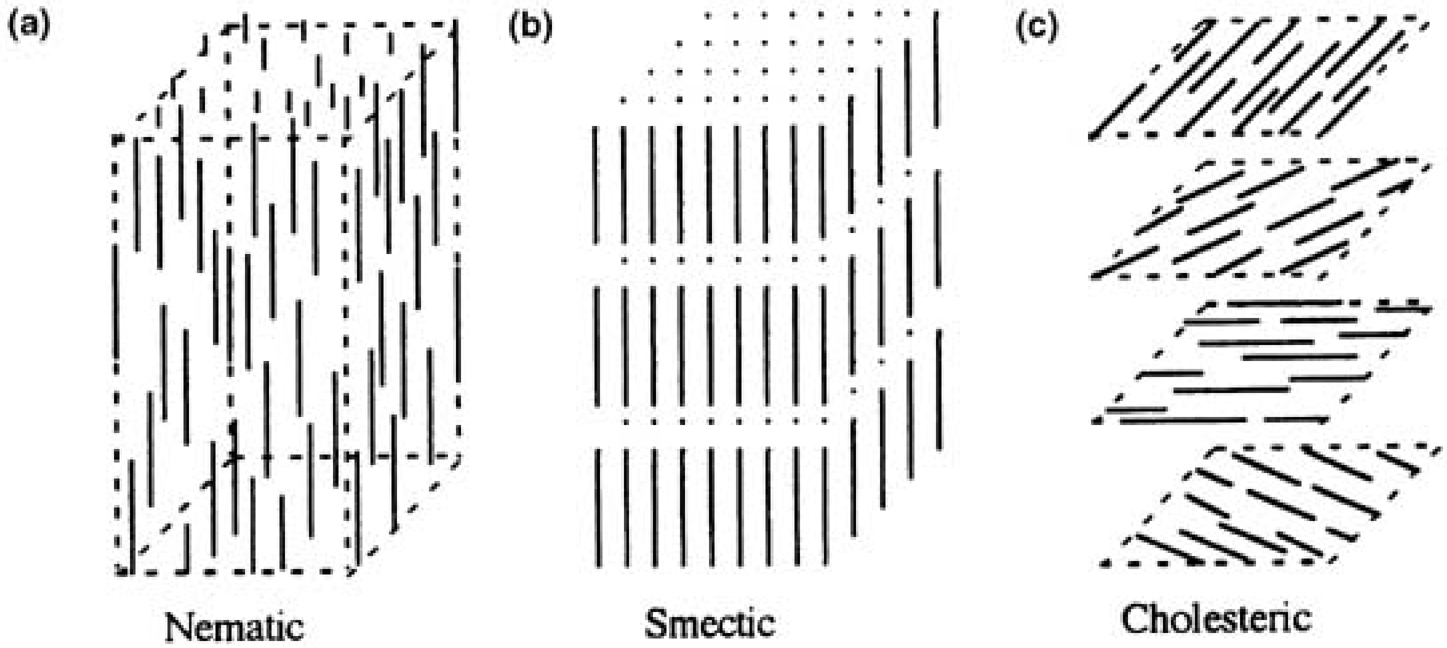
Introduction

As pointed out in the preceding section, a second route for developing fibers having properties approaching the ultimate is the use of polymer chains that have high intrinsic stiffness and will remain extended in solution or melt. The development of aramid organic fibers based on aromatic polyamides met these requirements and added another chapter to the history of the development of synthetic fibers. Nomex® aramid, a thermally resistant fiber based on a *meta*-oriented structure, was commercialized by the DuPont company in 1962.

Following the technological breakthroughs which led to the discovery of (1) the liquid crystalline behavior of *para*-oriented aramids20 and (2) a novel method for spinning anisotropic liquid crystalline polymer solutions,21 Kevlar® aramid fiber was produced and commercialized by the DuPont company in 1972. Other fibers based on aromatic polyamide compositions, which were produced and commercialized by other companies, were Technora® (Teijin, Japan), Teijinconex® (Teijin, Japan), and Twaron® (Akzo, The Netherlands). Additionally, SVM is a fiber produced in the Former Soviet Union and it was announced in 1990 that a new aramid fiber had been introduced by Hoechst, in Germany.

The rigid rod chains in *para*-aramids tend to form so-called liquid crystals when they are dissolved in polar solvents or heated to certain temperatures. The polymers showing liquid crystalline behavior in melts are called the thermotropic liquid crystalline polymers, and those showing similar properties in solution are called the lyotropic liquid crystalline polymers. These liquid crystals exist in three distinct phases according to their specific structures (Fig. 21.25).

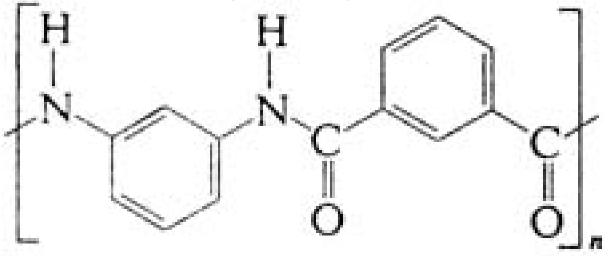
Fig. 21.25 Structure of liquid crystalline polymers showing (a) nematic, (b) smectic, and (c) cholesteric orders.



A one-dimensional arrangement of rigid rods is called a nematic structure; a two-dimensional ordered arrangement represents a smectic structure; and the cholesteric structure is formed in terms of rotating oriented sheets so that rigid rods align parallel to each other in every layer, but the directional vectors in each layer are different. The preferential phase considered for fiber spinning is nematic. Nematic solutions or melts are easy to develop into oriented structures through shearing and elongational flow during extrusion because of the rigidity of the polymer chains. Thus high orientation can be obtained in the as-spun fibers without much post-treatment.

Manufacture

In 1973, the Federal Trade Commission recognized aramid as a distinctly different generic material and defined it as "a long-chain synthetic polyamide in which at least 85 percent of the amide linkages are attached directly to two aromatic rings." This distinguishes aramids from nylon, which was redefined as a polyamide with less than 85 percent of the amide linkages attached to two aromatic rings. The first aramid fiber produced in the United States was Nomex®, the reactants being *m*-phenylene diamine and isophthaloyl chloride to give poly(*m*-phenylene-isophthalamide) (MPD-I).

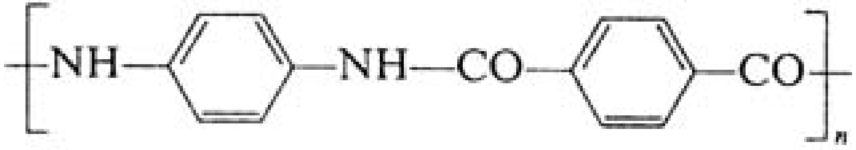


MPD-1

This polymer could not be melted without decomposition, so the preferred fiber formation route was solution spinning. Patent literature suggests that the fiber is spun from a solvent system composed of dimethylformamide and lithium chloride. The final properties are achieved by stretching in steam after washing to remove residual solvent.

The physical and chemical properties of this fiber are not remarkably different from those of other strong polyamides, but it does have excellent heat and flame resistance that makes it particularly suited for use in protective clothing and in specified industrial end uses. Military flight suits, fire-fighter uniforms, and hot gas filtration are a few of its many possible applications.

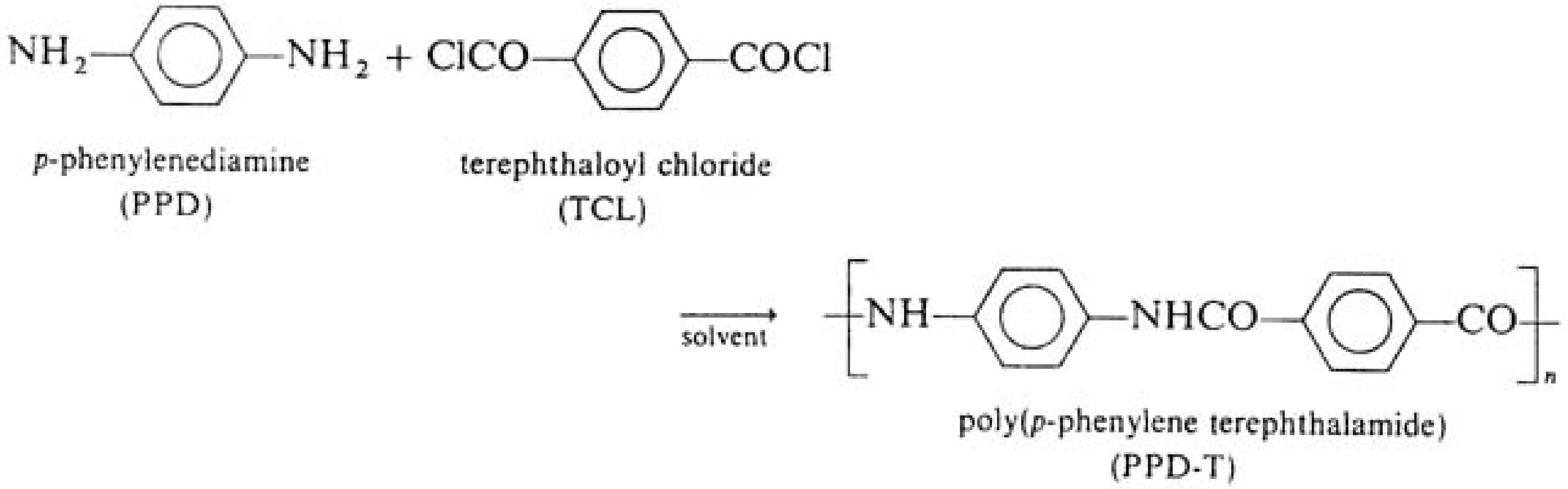
The other important fiber in this category, which also was first produced commercially in the United States, is Kevlar®, introduced in 1971 as fiber B and later coded as Kevlar® 29. It was produced from poly(*p*-phenylene terephthalamide).



poly(*p*-phenylene terephthalamide)

Later, a higher modulus version, Kevlar® 49, believed to be made by heat anealing of Kevlar® 29, was introduced.

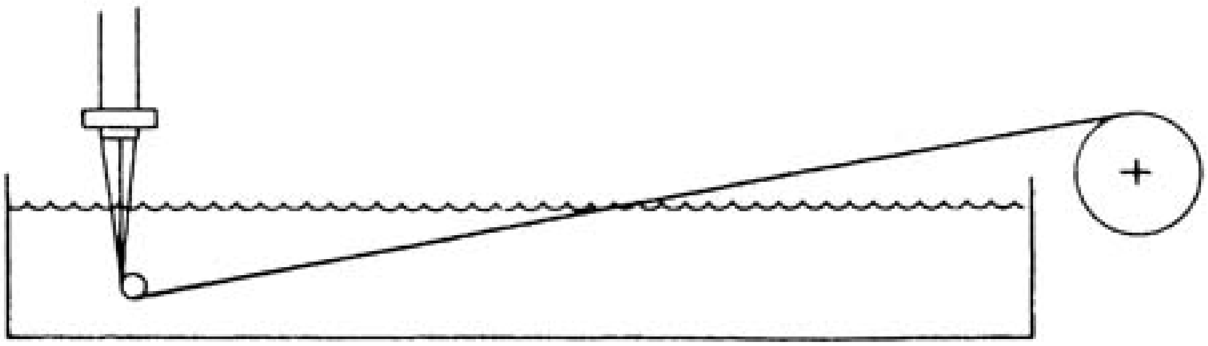
Poly(*p*-phenylene terephthalamide) PPD-T, can be polymerized to a fiber-forming molecular weight by polycondensation of terephthaloyl chloride and 1,4-phenylene diamine.



Poly(1,4-phenylene terephthalamide) (PPD-T) of high molecular weight can be prepared by low-temperature solution polymerization techniques. This polymer is less soluble in amide-type solvents than is poly(*p*-benzamide). The most successful conditions required hexamethylene phosphoramide (HMPA), the original solvent, alone or mixed with N-methyl-2-pyrrolidone (NMP), although other mixtures such as NMP containing CaCl2 also could produce a fiber-forming polymer. During polymerization, the molecular weight increases rapidly within the first few seconds of the reaction. The critical molecular weight or viscosity is exceeded, and the stir opalescence typical of lyotropic solutions is observed. Although gelation of the reaction mixture occurs quickly, polymerization continues, but at a greatly reduced rate. With the choice of a suitable solvent system, gelation can be delayed until the desired higher molecular weight is reached. Large-scale manufacture of a polymer requires continuous polymerization to minimize cost. In the case of PPD-T, special problems that had to be accommodated included the rapid gelation of the reaction solution with increasing molecular weight, the need to control the temperature of a vigorous exothermic reaction, and the handling of solvent HMPA, which is suspected to be a carcinogen.

Although melt spinning would be preferred from the standpoint of process simplicity and conversion cost, aramids must be spun from solutions, by wet, dry, or dry jetwet methods because they decompose before or during melting, ruling out melt spinning. Dry spinning is used to produce Nomex® fiber, where a dope (20%) of the polymer in solvent is converted to yarn. In wet spinning, the polymer dope is extruded into a nonsolvent where the fiber coagulates. The coagulated fiber then is washed and often drawn to develop desired fiber properties. In wet spinning, the spinnerette is in the coagulation bath, hence the dope temperature and the coagulation temperature are the same. In dry jetwet spinning (Fig. 21.26), as used for Kevlar®, on the other hand, the spinnerette is separated from the coagulation bath, allowing independent control of the dope extrusion and coagulation temperatures. The extrusion jet is placed a small distance above the coagulation bath, and the nascent fibers descend into the liquid, pass under a guide, and proceed in the bath while undergoing stretch; then they are withdrawn from the bath and wound up. A subsequent washing step may be required to remove residual acid solvent. In systems where the dope must be hot to lower viscosities to extrudable levels, and the bath cold for developing specific structures during coagulation, as is preferred with Kevlar®, dry jetwet spinning is the option. For production of Kevlar® fibers, a PPD-T/concentrated sulfuric acid solution, containing approximately 20 percent polymer, is extruded at 90°C into a cold water bath (•1°C). Following extrusion, the fiber is washed, dried, and given post-treatment, depending upon the properties desired.

Fig. 21.26 Schematic of dry jetwet spinning method.



Kevlar® is reported to have about twice the breaking strength (2327 gram force/denier or 2.93.4 GPa) of high-tenacity nylon and polyester, but its most outstanding physical property is its high stiffness (5501000 gram force/denier or 70127 GPa), which can range up to an order of magnitude greater than that of standard polyester. This property has led to high-volume usage of the fiber as reinforcement in composite materials such as belts in radial tires and aerospace structures. The world production figure (1999) of this fiber was estimated to be over 60 million lb.

21.14 High-Temperature-Resistant Fibers

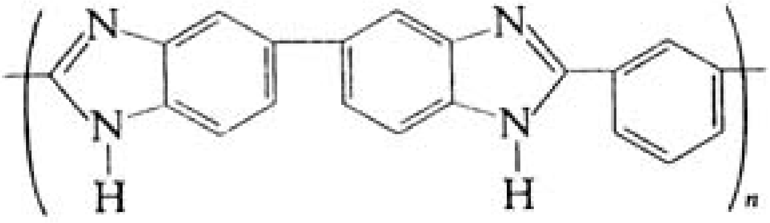
The need for high-temperature-resistance fibers has arisen from demands of a number of industrial applications, as well as applications in aerospace programs. In many of these applications, the usual characteristics of organic-based fibers are desired, but the high temperature resistance of inorganic fibers is required. Thus, the fibers are expected to retain their structural integrity at temperatures of 300°C and above for considerable periods of time, but otherwise their properties should resemble those of the more common manufactured textile fibers. The two leading groups of fibers in this area are the *meta*-aramids and PBI.

***Meta*-Aramid**

The major fibers in this group, based on sales volume, are the products made from poly(*m*-phenylene-isophthalamide), which were introduced by DuPont in 1962 as HT-1 nylon (later known as Nomex®) and by Teijin in 1972 as Conex®. General manufacture of the fiber was described earlier under the heading "Aramids." Other manufacturers now are entering the field with products of similar chemical structure. Although these products are made and spun by different processes, their chemical and physical properties are similar. The fiber is usually utilized in the form of cut staple, which are amenable to conversion on traditional spinning and weaving machinery. Its mechanical properties (tenacity 45 gram force/denier, ultimate strain 2530 percent and initial modulus 90100 gram force/denier) compare favorably with those of other textile fibers. The main utility of the fiber lies in the resistance it offers to combustion: it has a limiting oxygen index (LOI) of about 0.29, a melting point above decomposition temperature, an ignition point above 600°C and a flash point about 800°C.22 The LOI gives a relative measure of flame resistance; the higher the number, the lower the flammability. The fiber Kermel® from Rhone Poulenc also is classified as an aramid. It is, chemically, a polyamide-imide fiber and has an LOI of about 0.31. A major application of these fibers is in protective clothing. In order to reduce their cost, they often are blended with other, less expensive, flame-retardant fibers such as those based on cotton, rayon, and wool.

PBI

Poly-2-2′-(*m*-phenylene)-5,5′-bibenzimidazole, commonly called polybenzimidazole (PBI), was developed under the aegis of the U.S. Air Force Materials Laboratory in cooperation with the then-existing Celanese Corporation. The fiber went into commercial production in the United States in 1983. It is a condensation polymer obtained from the reaction of tetra-aminobiphenyl and diphenylisophthalate in a nitrogen atmosphere at temperatures that may reach 400°C in the final stages.23 The structure of a repeating unit is shown below.



The polymer is dissolved at a high temperature under nitrogen pressure in dimethylacetamide, to which a small amount of lithium chloride may be added to increase the stability of the solution. Then, it is dry-spun in an atmosphere of heated nitrogen (about 200°C), from which the solvent is recovered; next it is stretched slightly in steam and washed. Drawing and relaxing are done in an inert atmosphere, as might be expected, because temperatures up to 250°C or higher are used. The fiber then is given a stabilization treatment in a sequence of steps and made into staple fibers using conventional crimping and cutting techniques. The stabilization treatment involves reaction with sulfuric acid and heating at high temperatures (•475°C) for short periods of time. The process, known as sulfonation, yields a product that has a significantly lower shrinkage than the unstabilized material.

The final yarn is golden yellow, and because this color appears to be an intrinsic property of the polymer, it may have some limitation as far as the civilian market is concerned. This material originally suffered from high shrinkage on exposure to flame; however, further developments including the sulfonation treatment have reduced the shrinkage to only about 510 percent at 600°C.24 The fiber is capable of retaining about one half of its original strength (•3 gram force/denier) upon exposure to air for 18 hr at 350°C or 1 hr at 425°C, it has an LOI of 0.41, which is well above that of the aramids, and it has high resistance to inorganic acids and bases and organic chemicals. Further, with tensile and moisture regain properties comparable to those of many textile fibers, the PBI fiber is well suited for blending with other fibers and conversion into final products using conventional spinning and weaving or non woven equipment. Applications include high-performance protective apparel, flight suits, and aircraft furnishings.

21.15 Polytetrafluoroethylene

Historical

Polytetrafluoroethylene (PTFE) was discovered in 1938 when Dr. Roy J. Plunkett and his assistants working on new nontoxic, nonflammable refrigerants at E. I. duPont de Nemours and Co. found that one cylinder, which was supposed to contain tetrafluoroethylene (TFE), ceased to release the gaseous material. Upon opening the cylinder, they discovered that the inside was covered with a white powder. Polymerization of TFE had taken place. The result was PTFE or Teflon®, which is the trade name applied to that polymer by duPont. The company tested the polymer and found that it was virtually inert to all known solvents, acids, and bases, a characteristic that was unique at that time. It also found that this material was resistant to high temperature and had the lowest coefficient of friction of any known solid. The research on PTFE was intensified during World War II because of military demands for improved materials for products such as gaskets, packing, and linings for containers for handling corrosive materials. When the war was over, the release of classified information caused a booming interest in the polymer. Since that time, PTFE has found itself in many different applications.

Manufacture

In one process, the manufacture of the monomer, TFE, involves the following reactions: Hydrogen fluoride is made by reacting calcium fluoride with sulfuric acid:

$$\mathrm{CaF_2 + H_2SO_4 \rightarrow CaSO_4 + 2HF}$$

Chloroform is reacted with hydrogen fluoride in the presence of antimony trifluoride as the catalyst:

$$\mathrm{CHCl_3 + 2HF \rightarrow CHClF_2 + 2HCl}$$

TFE is obtained by the thermal decomposition of this monochlorodifluoromethane (known as Freon) in a continuous noncatalytic gas-phase reaction, carried out at or below atmospheric pressure at temperatures from 600 to 900°C:

$$\mathrm{2CHClF_2 \rightarrow C_2F_4 + 2HCl}$$

Numerous side-products are generated in this process. Many of them are present in trace amounts, but the highly toxic perfluoroisobutylene, CF2=C(CF3)2, requires special precautions.

TFE also may be manufactured by the reaction of zinc and tetrafluorodichloroethane:

$$\mathrm{ClF_2C{-}CF_2Cl + Zn \rightarrow C_2F_4 + ZnCl_2}$$

or by the reaction of tetrafluoromethane molecules in an electric arc furnace:

$$\mathrm{2CF_4 \rightarrow C_2F_4 + 2F_2}$$

TFE is a colorless, tasteless, odorless, and nontoxic gas. To avoid any undesired reactions during storage, inhibitors must be added. The polymerization is carried out by an addition-type reaction in an aqueous emulsion medium and in the presence of initiators such as benzoyl peroxide, hydrogen peroxide, and persulfates. The monomer is fed into a cooled emulsion medium and then heated to a temperature of 7080°C, at which the polymerization takes place.

$$\mathrm{CF_2{=}CF_2 \rightarrow ({-}CF_2{-}CF_2{-})}$$

The pressure may range from 40 to 100 atmospheres. After removal of the unreacted material, the polymer is washed, pressed, and dried. The degree of polymerization can be quite high, of the order of 50,000. PTFE is manufactured in four different forms: granular, fine powder, aqueous dispersion, and micro powder, and in a variety of grades, each differing in properties. Fillers such as glass fibers, asbestos, graphite, or powdered metals may be added to the granules in order to modify properties.

Fiber Manufacture

Because Teflon® is not soluble, it cannot be wet- or dry-spun, and because it is thermally unstable at its melting point of about 400°C, this combination would seem to pose an impossible problem for the production of fibers. Research into the fundamental characteristics of the polymer, however, revealed that the submicroscopic particles precipitated from the polymerization reaction were about 100 times as long as they were thick.

In one manufacturing process, an aqueous dispersion of PTFE is mixed with a solution of ripened cellulose xanthate, from which a fiber is obtained by a wet-spinning process, after which the cellulose is completely decomposed by heating. The remaining PTFE is sintered into continuous fibers by transporting them over heated metal rolls, followed by stretching to achieve the desired diameters and physical structure. In a process known as "paste extrusion," the powder is mixed with an organic plasticizer and compressed at 300500 psi to make a preform or billet. The latter is extruded into filaments which are then dried to evaporate the lubricant. In another process, films are produced that are slit into strips of very small widths, which then are stretched and sintered. In yet another, tapes are extruded, which are converted to fibers by stretching while being twisted to a very small cross-sectional area.

Properties

The unique combination of properties, including chemical resistance, thermal oxidation resistance, high lubricity, electrical and thermal insulation, low flammability and excellent weatherability, are derived primarily from two factors, namely the molecular structure and the molecular weight of PTFE. The structure consists of a core chain of carbon atoms with a fluorine sheath, which essentially completely shields the core. The fluorine atoms are so tightly packed that steric interactions cause a slight rotation of the carbon chain from the normal planar zigzag to a helical conformation. This dense shield has a low surface energy and a very smooth surface with no side chains or imperfections. As a result, the interchain forces are low and individual molecules are able to slide past each other with relative ease. Compared with other polymers, therefore, creep tends to be high.

The polymer chains pack themselves very closely and regularly to give cylindrical packing which consequently leads to very high crystallinity (•90%). The material has a high melting point which is about 330°C. The smooth surface when combined with low surface energy makes the surface so neutral that it resists sticking to any material. Accordingly, PTFE has the unique property of the lowest coefficient of friction (•0.007), with essentially no stick-slip character. The material has no affinity for water and, therefore, is totally hydrophobic.

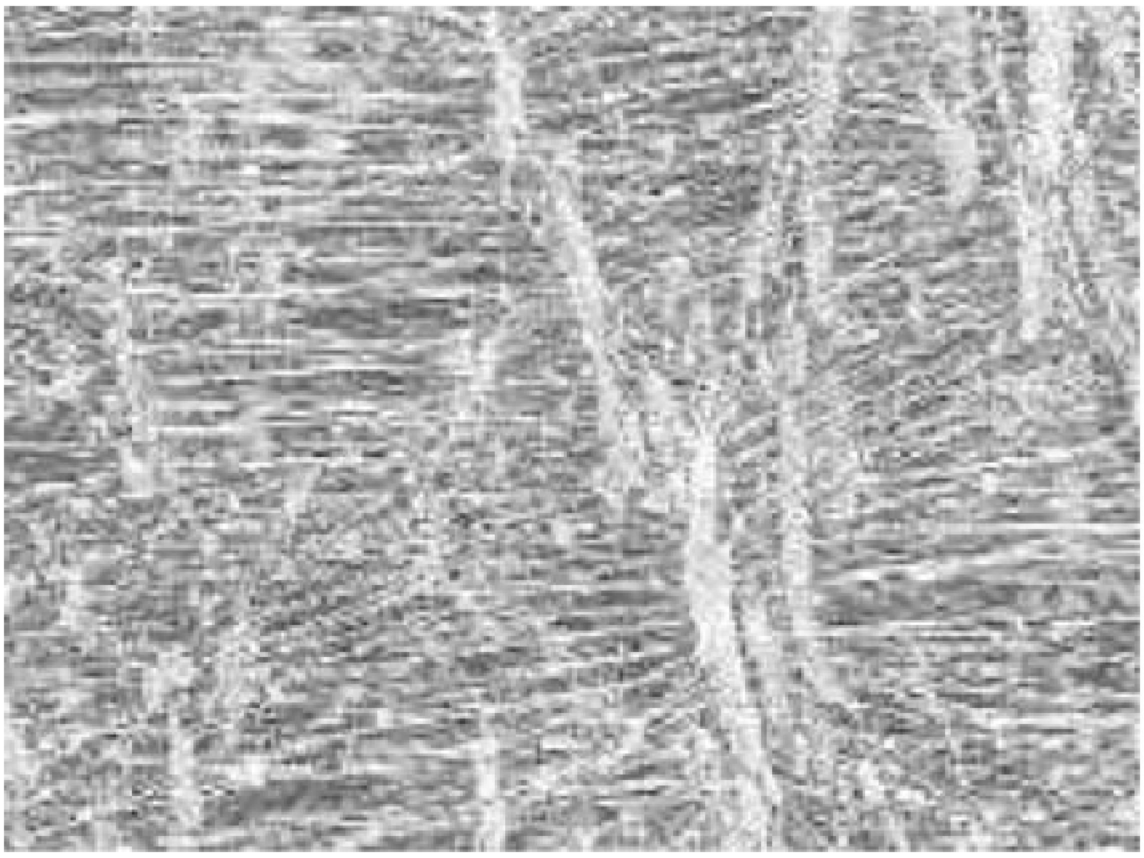
The chemical inertness and the thermal stability of this polymer are so great that in spite of its high price ($410/lb, depending on the resin type) it is used in chemical operations where drastic conditions exist and no other organic material is suitable. Its low friction allows it to be used as non-stick coating for metals, work surfaces, and cooking utensils. The polymer is regarded as biocompatible and tends to be accepted by the body. This has allowed PTFE to be explored as a material for surgical implants.

Expanded PTFE (ePTFE)

In early 1970s, Dr. Robert W. Gore invented a process by which PTFE could be expanded and gave the trade name of Gore-Tex® to the product obtained. In this process, the specific gravity is reduced by the introduction of micropores but much of the original properties of the polymer are retained. The new material is a hydrophobic but porous membrane of PTFE that is used as a protective layer in a number of applications.

In the process used, in general, a paste is formed of the PTFE polymer and a plasticizer and shaped into an article. It is then expanded by stretching in one or more directions, and while it is held in the stretched form, it is stabilized by heating to high temperature (•327°C) and cooling. The porosity that is produced by the process is retained in the final product.25 The structure formed by the process consists of "nodes" and "fibrils." The nodes, that vary in size from 5 to 500 µm and are always found perpendicular to the direction of expansion, are interconnected by fibrils. This is seen in Fig. 21.27, which shows a micrograph of Mikrotex®, an ePTFE product.26 An ePTFE may have as many as a billion or more randomly spaced pores per square centimeter. These pores are unique in size: they are three or more orders of magnitude smaller than the size of a water droplet but two or more orders of magnitude larger than a water vapor molecule. Thus, when used as a rain wear, the product allows the perspiration vapors to escape but blocks out the liquid water from penetrating. The product can likewise serve as a barrier against chemicals and microbes.

Fig. 21.27 Expanded PTFE micrograph (Microtex® Membrane at 500×) showing a porous structure containing nodes interconnected by fibrils. (*Courtesy Menardi-MikroPul, LLC.*)



< previous page

next page >

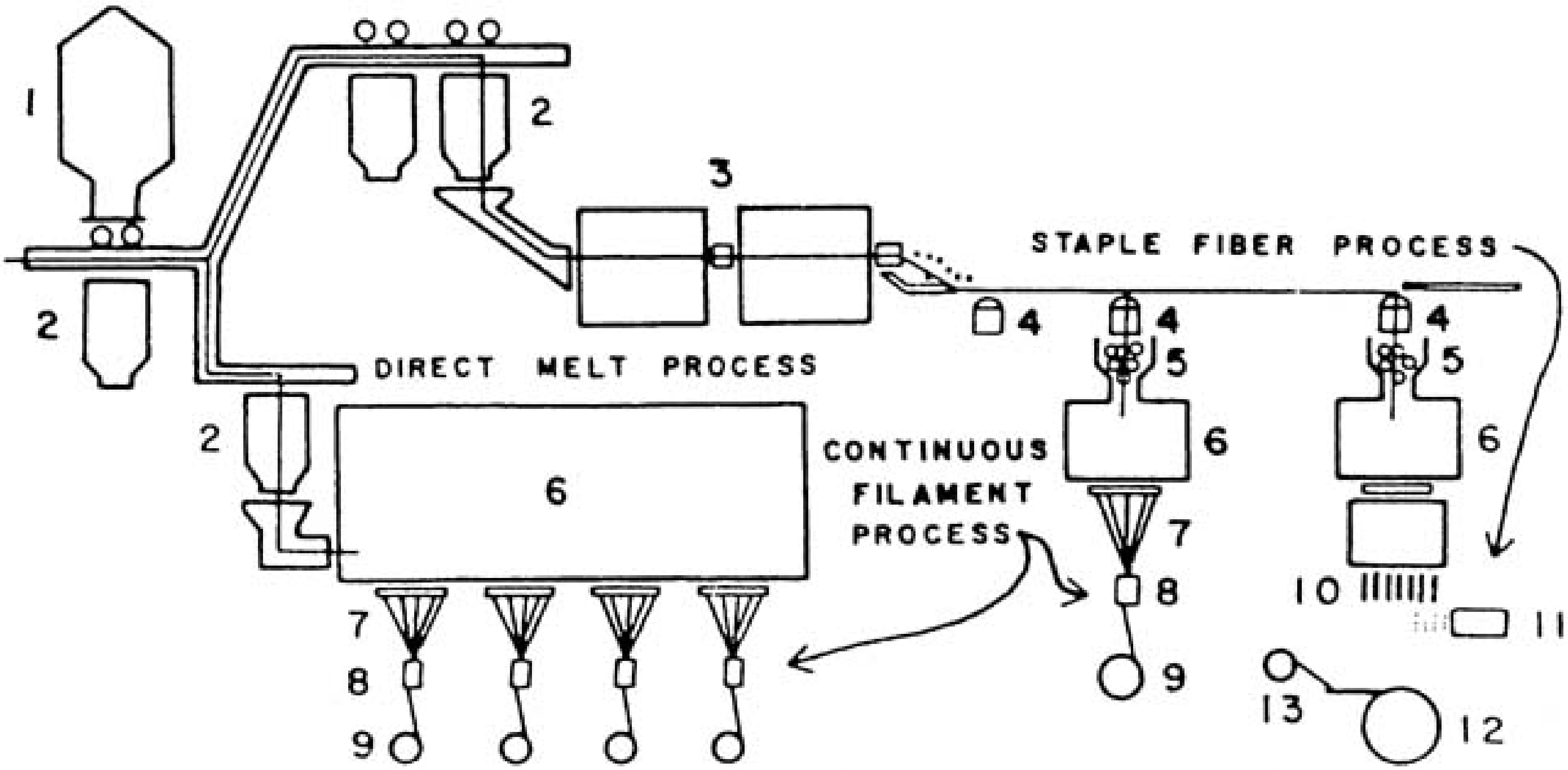
21.16 Glass and Carbon Fibers

Glass

Among the manufactured inorganic fibers. glass is produced in by far the largest volume. There has been a rapid increase in the use of textile grades of these fibers, and outside the textile field enormous quantities of glass fibers are used in air filters, in thermal insulation (glass wool), and for the reinforcement of plastics.

Glass possesses obvious and well-known characteristics which have largely determined the methods used to form it into large objects. It flows readily when molten and can be drawn into filaments, whose extreme fineness appears to be limited only by the drawing speed. The method used in producing textile-grade glass fibers follows this principle (see Fig. 21.28.)

Fig. 21.28 Flow diagram for manufacture of textile glass fiber: (1) glass batch; (2) batch cans; (3) marble forming; (4) cullet cans; (5) marbles; (6) melting furnaces; (7) filament yarn formation; (8) gathering and sizing; (9) yarn packaging; (10) air jets; (11) lubricant spray; (12) collection for staple fibers; (13) staple fiber packaging. (*Courtesy Owens-Corning Fiberglass Corp.*)



In the commercial operation, the molten glass, produced either directly from raw materials or by remelting of marbles, is held at a uniform temperature in a vessel, whose bottom carries a bushing containing small uniform holes. The molten glass flows through these holes as tiny streams that are attenuated into filaments at speeds on the order of 3000 m/min; these filaments are coated with a lubricant, gathered into groups to form yarns, and wound up. For a particular glass viscosity, the size of the individual filaments is determined by the combination of the hole size and the speed of attenuation.

Because of the inherently high modulus of glass, very fine filaments are required in order to approach the required properties of textile materials. Thus, the diameter of glass filaments falls in the range of 3.87.6 µm whereas the average diameter of the finest organic fibers is about twice as great. The fiber and yarn numbering system is based on nomenclature used in the glass industry and differs from the traditional systems accepted in the textile and organic fiber industries.

The method of manufacture of glass staple fibers differs from those used to produce the corresponding organic materials, all of which are based upon cutting the continuous filament product. Air jets, directed in the same line of flow as the emerging streams of glass, attenuate the streams, and break the solid glass into the lengths desired for further processing, which are gathered on an appropriate vacuum drum and delivered as slivers or a matte. To produce fibers that may be coarser and considerably less uniform in length, to be used for the production of filters, paper, or thermal insulation, large streams of molten glass are cross-blown by blasts of hot air, steam, or burning gas.

As might be expected from the nature of glass, the conversion of glass fibers into final products has required the development of new lubricants, finishes, and processing techniques. For example, because glass fabrics cannot be dyed directly or printed with the colors demanded for their acceptance as draperies, the colorant must be applied to a resin coating. But before applying the coating it is necessary to remove the lubricant that was placed on the fibers to permit their conversion into a fabric, which is done by burning. The elevated temperature resulting from this operation also relaxes the internal strains developed in the glass fibers during the steps of the textile operations and sets the yarns in the required geometry. The fabric then is resin-treated, cured, and dyed or printed.

Another inherent property of glass is the tendency of unprotected fiber surfaces to abrade each other to destruction under the action of very little mechanical working. When it was first considered for rubber-reinforcing purposes, its poor adhesion to rubber and the inadequacy of the then-existing bonding agents frustrated attempts by manufacturers to take advantage of the very high tensile strength, completely elastic behavior, high modulus, and lack of moisture sensitivity of glass fibers. However, it has been possible to modify the fiber surfaces so that satisfactory adhesion is achievable, and the impregnant can be applied in such a way that fiber-to-fiber contact is prevented. With the adhesion problems solved, glass in cord form could effectively enter markets in belt-type tire construction and in all kinds of power-transmitting rubber belts. It is estimated that the total glass fiber production in the United States in 2000 was approximately 2.5 billion lb.

Carbon and Graphite

Following World War II, the development of jet aircraft and rockets created demands for fibers having thermal resistance, strength, and modulus far beyond what could be obtained in existing organic fibers. Much of this need was for reinforcing materials that could be embedded in matrices of one type or another. As a result, techniques have been developed for preparing fibers from a good many metals and refractory inorganic compounds. Although these materials are essential for certain uses, the volume of production still is low, and the prices are correspondigly high (as much as $1000/lb).

Carbon and graphite fibers are made from rayon and acrylic precursors by driving off virtually all of the hydrogen and oxygen contained in them. The principle is essentially the same as that which brought about the formation of coal, or, citing a more recently discovered and dramatic example, the conversion of the original wooden beams of buildings in Herculaneum, buried by a flow of mud from Vesuvius in A.D. 67, to what appears to be charcoal. In the present commercial process, the starting material is selected so as to produce a final product of the desired size and properties. Cotton, bamboo, and other natural fibers were the earliest materials used as precursors. In the 1950s, rayon was used for this purpose, but the first attempts led to fibers with poor strength. Union Carbide, in the mid-1960s, made the first strong carbon fiber from rayon and extended its work to using polyacrylonitrile as the pecursor. The use of mesophase pitch as the starting material began in the 1970s. Carbon fibers may be divided into three types: (1) highly graphitized with a high modulus, (2) heat-treated at a lower temperature to produce fibers with lower modulus but high strength, and (3) randomly oriented crystallites with relatively low modulus, and low strength, but most important, low cost.

The preparation of fibers generally consists of heat-treating the precursor at a low temperature (200350°C), usually in air, which gives a stable fiber for higher temperature processing. During this step, extensive decomposition occurs, and a percentage of the initial weight is lost, which is related to the cleavage of the CC and C=O bonds and expulsion of H_2O, CO, and CO_2 (additionally HCN in the case of PAN precursor) as gases. This step is followed by carbonization in an inert atmosphere at 10002000°C, which is said to collapse the cyclized structure into a stacked ring carbon fiber structure. Almost all noncarbon elements are evolved as volatiles. This is followed by graphitization, which usually is carried out at temperatures above 2500°C for short periods in argon or nitrogen. The process increases the purity, removes the defects, and further improves the order in the structure. Thus, the difference between the so-called carbon and graphite fibers lies in differences in the ranges of temperatures at which the last step is carried out, the degree of carbon content (97% for carbon and 99.6% for graphite), and the mechanical properties, which are superior in the graphite fibers. In usual discussions, the term "carbon fibers" covers both materials. The majority of carbon fibers produced today are made from a PAN precursor. PAN fibers are fine, and have a higher degree of molecular order and a higher decomposition point than those from rayon. This precursor also leads to a greater carbon yield (45%, as opposed to 24% from rayon), but the fibers are more expensive to produce than those based on rayon. Recent commercial developments have allowed the production of carbon fibers from low-cost petroleum or coal-tar pitch, instead of synthetic fibers. These precursors lead to a higher yield (90%) of carbon, improved lubricity of fiber products, and higher production rates. However, the pitch-based fibers may be more brittle and harder to handle, have a higher specific gravity (2.0, as opposed to 1.8 for those from PAN and 1.66 from rayon) and lower compatibility with some matrix materials.

The fibers have a diversity of applications. One major application is in composites, where they are used for reinforcing resins and metals to provide structural materials with high strength, high modulus, and light weight. The resulting composites are used in the aircraft and aerospace (the largest users), automotive, and sports industries. The fibers also find uses in protective garments, electrical devices, insulation, and filtration. The use of the fibers in the world was about 25 million lb in 2000. The prices dropped significantly during the past decades, but in 2000 the majority remained in the $1570/lb range. The price of the ultrahigh-modulus carbon fibers, however, can be as high as $1500/lb, or greater.

21.17 Sulfar

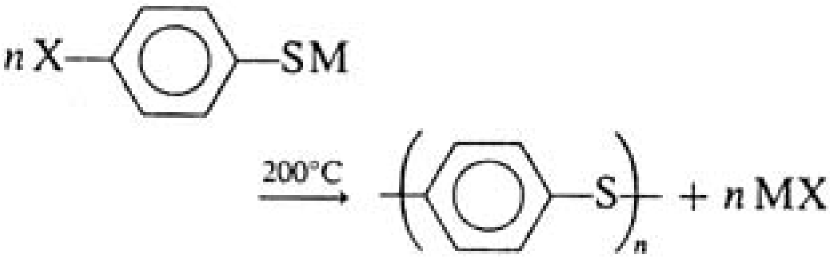
Historical

Sulfar fibers are extruded from poly(phenylene sulfide) or PPS by the melt-spinning process. The first PPS polymer was made in 1897 by the FriedelCrafts reaction of sulfur and benzene. Researchers at Dow Chemical, in the early 1950s, succeeded in producing high-molecular weight linear PPS by means of the Ullmann condensation of alkali metal salts of *p*-bromothiophenol.

In 1973, Phillips Petroleum Company introduced linear and branched products under the trade name Ryton® by reacting 1,4-dichlorobezene with sodium sulfide in a dipolar aprotic solvent. In 1983, the same company succeeded in stable melt-spinning of PPS. In 1986, the Federal Trade Commission gave the fiber the generic name Sulfar, defined as "a manufactured fiber in which the fiber forming substance is a long chain synthetic polysulfide in which at least 85 percent of the sulfide (S) linkages are attached directly to two aromatic rings."

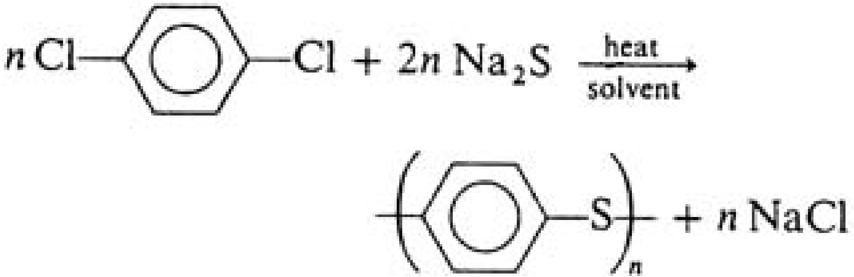
Manufacture

In one process, synthesis was carried out by self-condensation of a metal salt of a *p*-halothiophenol:



where *X* is a halogen, preferably Br, and *M* is a metal (Na, Li, K, or Cu, preferably Cu). This reaction was carried out under nitrogen in the solid state or in the presence of materials such as pyridine as reaction media. Considerable difficulty was encountered in removing the by-product, copper bromide, from polymers made by this process. The current commercial synthesis of the polymer is carried out by reaction between *p*-dichlorobenzene and sodium sulfide in a polar solvent.

The process discovered by workers in the laboratories of Phillips Petroleum Co. marked a significant departure from prior processes, and made it possible to prepare a variety of arylene sulfide polymers from the readily available starting materials.27



Melt spinning of PPS involves problems such as plugging of the filter and the spinnerette. In order to prevent cross-linking and gel formation during the process of spinning, the company introduced cure retarders comprising Group IIA or Group IIB metal salts of fatty acids, which improved the heat stability of PPS:

$$\left(CH_3(CH_2)_nCOO\right)_2 M$$

where *M* is a Group IIA or IIB metal, and *n* is an integer from 8 to 18. Representative compounds of the type described above include calcium stearate, calcium laurate, calcium caparate, and calcium palmitate. In general, the cure-retardant additives are employed in an amount within the range, of about 0.15, preferably about 0.52, weight percent based on the weight of the PPS.

Sulfar fibers, sold under the trade name Ryton® in the United States, are characterized by high heat resistance, inherent flame retardancy, excellent chemical resistance, low friction coefficient, good abrasion resistance, and good electrical properties.28 Physical characteristics include medium tenacity (3.5 gram force/denier or 423 MPa) and elongation (2540%) and low shrinkage (<5% at 100°C). The fibers find application in a number of industrial products, including filter fabric for coal-fired boiler bag houses, paper maker's felts, materials for electrical insulation, high-performance composites, gaskets, and packings.

Several Japanese and European companies have begun the production of PPS, some with a U.S. partner. The new decade should see the introduction of a number of new fibers and fiber products based on the PPS polymer.

21.18 Microdenier Fibers

One of the more important developments in the field of fiber technology in recent years has been the production of fine denier fibers, with worldwide activity in developing new products and outlets for them. Improvements in the quality of polymers, coupled with new technology for extrusion of fibers, have led to the production of fibers with sizes ranging from 1 to 0.1 denier or even lower. Fabrics produced from such fibers have novel and unique properties, and they are finding applications in a wide variety of apparel, and industrial products, including high-fashion fabrics with silk-like texture, synthetic suede, breathable porous but rainproof or bacteria-proof fabrics, wipes for oil and other spills, clean room materials, cloths for cleaning camera, microscope, and spectacle lenses, and for compact discs.

Unfortunately, there is no universally accepted definition for microdenier, and companies have been free to use terms they choose. In general, there seems to be a consensus that the term "fine" may be used for a denier of 1 or less and "micro" for a denier less than 0.5. The production of such fibers is difficult and expensive because the throughput rate must be reduced in order to obtain fine denier filaments, and there is a limit to how fine in denier a regular process could be made to go. Generally, the technology of production of fine and microdenier filaments can be divided into five categories. First, conventional spinning technology can be used, by using fine-size dies and adjusting the throughput rate and quenching and drawing parameters to obtain fine denier polyester filaments. This method has been used by most fiber manufacturers to produce such fibers in limited quantities. In another method, alkali reduction or surface etching is used to dissolve the surface layers, in an effort to reduce the cross-sectional size of filaments in a polyester fabric. A weight loss of as much as 25 percent has been achieved by this process. In the MB process, molten polymer is forced through a melt blowing die and die tip orifices, and the emerging stream is attenuated by a jet of high-velocity hot air until broken. Then the broken fiber is forced into a stream of cold air, where the fiber solidifies. The fiber is collected on a wire screen or apron with other fibers, and a homogeneous MB web is produced. This method is used extensively with PP materials, but other polymers such as nylon, polyester, and polyethylene also have been used. Fibers as fine as 0.5 µm and finer have been produced.

In another approach, filaments containing two polymers that do not adhere to each other are spun and then split. One may, for example, spin a bicomponent fiber of nylon that has several filaments of polyester embedded. After a fabric containing bicomponent filaments has been woven, it is treated to split the components, thus converting the original filament to several smaller filaments. Deniers of the order of 0.1 can be achieved by this ingenious method. In the last method, instead of splitting the two components as in the previous example, one component is dissolved away chemically, leaving bundles of very fine fibers in the fabric.

The majority of the technology for producing fine and microdenier fibers is new and thus expensive; so efforts in the future can be expected to be directed toward optimizing process parameters in existing methods and discovering faster and cheaper methods for manufacturing these esthetically very pleasing and functionally very promising fibers.

21.19 Fiber Variants

Introduction

In a previous section, data and plots were given showing the rapid rise in consumption and production of manufactured fibers at the expense of natural fibers. The principal reason for this has been the wide range of manufactured fiber variants that can be produced from a single fiber-forming polymer. The wide range of polymers available, each with its particular properties, adds yet another dimension. This is not to say that there is only one type of cotton, wool, silk, or asbestos fiber; there are many varieties of natural fibers, but their supply is limited by natural factors such as climate and genetics. The relative availabilities of manufactured fiber types can be altered by controlled chemical-process changes, whereas the amount and the quality of a desired cotton type that can be grown is determined to a great extent by climatic conditions, which humans have not yet learned to control. Another factor that has aided the growth of manufactured fibers is their consistent quality and properties. Again, the grade and the quality of natural fibers are subject to the vagaries of nature.

For the purposes of this discussion, fiber variants will be divided into two types: chemical and physical. Chemical variants will be those involving a small but significant change in composition, whereas physical changes will be those involving a change in either the dimensions of the fiber or its stress/strain or stability features. The definitions of the two variants also could be based on modification of either esthetics or functionality.

Physical Variants

Most manufactured fibers are available as staple, tow, and filament. Natural fibers are available only in the characteristic forms in which they occur, with filament silk and cotton staple as examples. All manufactured fibers are formed initially as filament yarns. The German adjective *endlos* (literally translated "endless") is very descriptive, as filament yarns are continuous strands consisting of one or more members that for most practical purposes are infinite in length. Fine filament yarns (40100 denier) are used in producing lightweight apparel fabrics, whereas coarse filament yarns (8001200 denier) are found as reinforcement in tires or conveyor belts. These examples are chosen as extremes to show the range or applicability of manufactured fibers, and represent only a small fraction of the actual range of end uses.

When many filament yarns are collected into a bundle immediately after formation, the resulting structure is called a tow. Tows may range from 10,000 to over a million in total denier. In a next step, the tow may be crimped by the process previously described, which imparts what is usually a sawtooth appearance along the length of all the filaments. For some end uses, the crimped tow itself is provided by the fiber maker, as are, for example, the previously discussed acrylic tows, which are converted to staple as part of the spinning of yarns with a wool-like character. Another example is the cellulose acetate tows used to form cigarette filters or the ink reservoirs for marker pens. In the latter case, the compact tow bundle first is treated to separate the individual filaments, giving a voluminous structure, which then is gathered into a continuous rod, wrapped in paper, and cut into appropriate lengths.

In the fiber-making plant, tow may be cut into short lengths of staple, ranging in length from $\frac{1}{16}$ to 6 in., depending upon the end use. For instance, the very short staples are used either in making flocked structures or in the production of papers containing blends of natural cellulose and manufactured fiber. The longest lengths are used in spinning heavy yarns for carpets or cordage. But most staple is $1\frac{1}{2}-3$ in. length and is used to form blends with cotton, rayon, or wool in the yarns employed in standard apparel fabrics. The staple length of the manufactured fiber is chosen to match that of the other blend component; otherwise, uneven yarn of poor quality results.

The size of a manufactured fiber can be altered by changing the size of the hole through which it is extruded and maintaining a constant takeup speed. Commercial fibers range from about 1.25 to 25 denier, corresponding to average diameters of about 550 µm. The very large ones would be used to make doll's hair or wigs, but the majority of manufactured fiber staples are made in the 1.56.0 denier range, corresponding to cotton blending fiber at the low denier end and coarse wool blending fiber at the high end. Staple and filament yarns used in carpets are in the 1216 denier range, whereas industrial filament yarns such as tire cord are about 6 denier per filament. The size of a fiber is a determinant of its stiffness, which in turn influences the draping, quality, and the surface feel (often called hand in the textile industry) of a fabric made from it.

Although fibers normally are extruded through circular jet holes, the use of noncircular holes in the jets has led to the availability of a wide variety of cross-sectional shapes. In the case of fibers dry- or wet-spun from solutions, most of the mass exiting from the jet hole is not the polymer. For example, cellulose acetate fiber is made from a 25 percent solution of cellulose acetate in acetone containing a small amount of water. After the fiber leaves the jet face, solvent begins to evaporate, and as a result the area of the fiber cross-section decreases. The final result is a fiber of roughly circular cross-section, but with a serrated edge, and much smaller in area than the parent jet hole. When cellulose acetate is extruded through a triangular jet hole, the end result is a fiber of "Y" cross-section, due to shrinkage from the original triangular shape.

In the melt-spinning process, there is no solvent loss to influence final cross-sectional shape. In the case of a triangular jet hole in melt spinning, the molten fiber leaves the jet face with a triangular cross-section but, being molten, immediately tends to return to a circular cross-section due to surface tension forces. It is necessary, therefore, to quench or cool the fiber as soon as possible in order to maintain the desired cross-sectional definition. Some loss of this definition is unavoidable during the drawing step that normally follows extrusion in a melt-spinning process.

Fibers of noncircular cross-section can modify and change both functional and esthetic properties in textile structures. The triangular cross-section is typical in those respects; its shape leads to a stiffer fiber than circular fiber of the same cross-sectional area, and in a fabric this results in less drapability and a crisper surface feel. Also, the flat surfaces reflect light in a different way than do curved surfaces and can create desirable lustrous effects. These optical effects are subject to many subtle influences having to do with the size of the reflecting surface and the amount of internal reflection that takes place. Triangular or "Y" cross-sectional fibers have greater specific surface areas per unit weight than their circular counterparts, which accounts for their use in aerosol filtration, where surface area is a major factor in efficiency. Yarns made with triangular cross-section fibers are more voluminous than those from round cross-section fibers. Thus when the two types are made into fabrics of equal weight, the variant cross-sectional fabric will transmit less light and be less permeable to air than that made with round cross-section fibers. In the spinning of blended staple yarns, maximum strength is obtained if the blend components have similar load/elongation characteristics. The stiffness of polyester staple fibers can be varied mainly by changing the draw ratio used to orient the fiber after extrusion and thus producing fibers suitable for blending with cotton, rayon, acrylic, or wool fibers, which possess markedly different properties. In many industrial applications, fiber and yarns are used under conditions where they bear a load while hot, as for example in tires and power transmission belts, where it is important that they do not grow or stretch significantly under these conditions. Accordingly, industrial yarns are drawn to a greater extent than other yarns to reduce as much as possible the stretch of the final products. Relatively, these yarns have high strengths, high stiffnesses, and low elongations at rupture.

Although in most instances it is desirable to have fibers that are dimensionally stable, in some structures it is advantageous to use mixtures of fibers that are stable with fibers that do shrink upon exposure to heat or steam. High shrinkage potential usually is built into a fiber by stretching it and not giving it a stabilization treatment. The yarn bulking that occurs with blend yarns of high- and low-shrinkage acrylic staples has already been described, in the section on that fiber. Using the same principle, felt-like structures can be made by heat treatment of nonwoven battings containing low and high-shrink staple.

Chemical Variants

The use of titanium dioxide as an additive in the delustering or dulling of manufactured fibers was discussed in the section on viscose rayon. The addition of pigment particles influences the processing and performance of the fiber, along with changing its appearance. Because of the whiteness of the delustered fiber, it requires more dyestuff to reach a given shade than that needed in the case of a bright fiber. The sliding friction of a delustered fiber is lower because the pigment particles protruding from the fiber surface reduce the contact area between a fiber and the surface it slides againsta guide, for example. By the same token, some pigments can accelerate the wear of contacted surfaces. It has been suggested that the drawing and consequent orientation of delustered fibers proceeds more smoothly because the pigment particles act as nucleation sites where molecular motion is initiated. Unless specially treated, the surface of anataseone of the crystalline forms of titanium dioxidecan accelerate the ultraviolet light degradation of acetate or nylon fibers. It is postulated that the crystal surface catalyzes the formation of peroxides from the water and oxygen under the influence of ultraviolet light, and that peroxides are the active species in the resulting polymer degradation. For this reason, rutilethe other common crystalline form of titanium dioxideis used to deluster fibers when improved sunlight resistance is needed.

The degree of polymerization or the molecular weight of the polymers on which manufactured fibers are based can be controlled as part of one of the early steps in the process. The polymer molecular weight chosen for a fiber has a strong influence on process economics, ease of conversion to fiber, and end-use performance. Commercially acceptable fibers are based on the best balance of these factors. However, in fabrics that are open in texture and made from standard polyester staple spun yarns of low twist, a condition known as "pilling" will develop as a result of wearing. The pills are made up of fiber ends that have worked loose from the yarn bundles as a result of surface rubbing and have wrapped around themselves. In the case of fabrics from natural fibers, which are generally less wear-resistant than manufactured fibers, these pills or fiber bundles will be lost by attrition with continuing wear. Because the wear resistance of a manufactured fiber is related to its molecular weight in a general way, the pilling tendency of a polyester staple can be reduced by lowering its degree of polymerization. This compromises the tensile characteristics of the fiber only to a small extent, and all the other desirable properties such as minimum care characteristics are essentially unaffected. When the molecular weight of a polyester or a nylon polymer is increased above the standard, the resultant fibers will have increased tensile properties and fatigue resistance. The filament polyester and the nylon yarns used in end products such as ties and conveyor belts are based on such polymers.

The technology for "dope dyeing" or mass coloration of fiber as part of the fiber manufacturing operation was described in the sections on viscose rayon and cellulose acetate. The use of this technology has decreased, largely because of problems of profitably managing the required inventory of colors in rapidly changing fashion markets, and for this reason it is used extensively only if the fiber cannot be dyed by any other means. PP fiber for use in outdoor carpets is a good example of such a situation. However, if one regards "white" as a color, one finds that a substantial portion of the polyester staple fiber produced for blending with cotton or rayon contains an optical brightener or a fluorescing agent. This is needed to overcome the yellowing tendency of polyester following the absorption of hydrophobic soils. The cotton or rayon fibers in a blend are continually rewhitened by the fluorescing agents added to laundry detergents for that purpose, but these agents are without effect on the polyester blend component. Because the polyester component usually is at least half of the blend, the spun-in optical brightener it contains is vital for the maintenance of overall whiteness.

In the cases of polyester, nylon, and acrylic fibers, their manufacturers have developed fiber variants with a wide range of dyeing behavior, referred to as dye variant fibers. Polyester fibers usually are dyed with what are described as disperse dyes. These dyes are only slightly soluble in boiling water and are used in the form of dispersions. The dye in aqueous solution is assumed to be in monomolecular form, and is absorbed from the dyebath into the polyester fiber by a process often called solid solutioning. As this occurs, more solid dye is dissolved to replenish that entering the fiber. The overall rate of dyeing is a very complex phenomenon but in part is determined by the molecular structure of the fiber through which the dye must diffuse. Generally, the more that the polymer molecules have been organized by drawing and annealing into more geometrically perfect domains or crystallites, the slower is the dyeing rate. The ability of the molecules to be thus organized into compact ordered structures can be reduced by polymerizing a small amount of a foreign dibasic acid or glycol into the polymer. Usually, 510 mol. percent is sufficient to prevent this regularity. Adipic acid, isophthalic acid, and polyethylene glycols are used to produce some of the comonomers for fast- or deep-dyeing polyesters. In this way, these fibers are made more economical to dye or print because special dyebath additives, high dyeing temperatures, and high-pressure steam-print fixation usually are not required.

Polyester fibers can be given an additional mechanism for dyeing if an ionic comonomer is added during polymerization. A common additive is an alkali metal salt of dimethyl-5-sulfo-isophthalate, which gives sulfonic (anionic) groups as part of the polymer structure. These groups allow the fiber to absorb basic (cationic) dyes by a specific ionic mechanism. The amount of cationic dye that can be absorbed by the fiber is stoichiometrically related to the number of anionic sites present in the fiber; this is quite distinct from the general solid solutioning that takes place with disperse dyes and polyester. A cationic dyeable polyester is useful for two main reasons in fabric coloration. First, cationic dyes give brighter, clearer shades than disperse dyes, and this can be important for both solid-dyed shades and prints during ever changing fashion cycles. Second, fabrics containing arrangements of unmodified and cationic dyeable polyesters can be dyed in the piece to a variety of color/white combinations by selection of dyestuffs, as day-to-day changes in demand may require. This is more economical than dyeing yarns to different colors, holding them in inventory, and then weaving them into fabrics, if and when they are required.

Nylon 66 can be dyed either with disperse dyes or with acid (anionic) dyes, the former carried out by essentially the same steps as described above for polyester, except that nylon absorbs these dyes much more readily. The ability of nylon to absorb acid (anionic) dyes is the result of a significant number of accessible free amine (cationic) end groups being present in the polymer. The dyeing of nylon with acid dyes is analogous to the dyeing of wool or to the dyeing of modified polyester with basic dyes, except that the polarities of the interacting groups are reversed in the latter case.

For best fastness to light and washing, nylon 66 is dyed with acid dyes, and nylon dye variants thus are based on manipulation of the level of acid dye uptake. By adding a monobasic acid such as acetic acid to the reaction mix near the end of the polymerization process, the amino end groups are converted to amide groups which have no affinity for acid dyes under normal dyeing conditions. This technique creates light acid dyeing or acid reserve dye variants that can have some capacity to absorb basic dyes at the carboxyl end groups. Nylon dye variants with increased acid dye uptake can be made by using a slight excess of diamine in the polymerization. In this way, there are no free carboxylic acid end groups. Nylon dye variants have found the greatest acceptance in floor coverings where attractive patterns can be piece-dyed using controlled dyebath conditions and selected acid dyestuffs.

Acrylic fibers are dyed most frequently with basic dyes. This is made possible by copolymerizing acrylonitrile with an acidic monomer such as styrene-*para*-sulfonic acid. Acrylic fiber suitable for acid dyeing can be made by using a basic comonomer such as a vinyl pyridine or a vinyl pyrrolidone.

Fibers not having inherent flame resistance often can be given this property by incorporation of a suitable additive. This may be done by copolymerization of the additive into the polymer reaction of the additive with the polymer after polymerization, or by applying a polymeric or monomeric noninflammable finish to the surface as a coating. These additives usually contain bromine, nitrogen, or phosphorus, or a combination of these elements. Great care must be taken in choosing the additive and its level of addition in order to prevent loss of other desirable fiber properties and to avoid any harmful effects to processors or ultimate consumers.

A wide variety of special durable surface treatments have been used on manufactured fibers. These include treatments for imparting such characteristics as soil resistance, antistatic behavior, and wearer comfort through moisture wicking and transport. Fiber finishes also have been used successfully in promoting adhesion between two materials, as, for example, between polyester tire cord and rubber, and between glass fiber and polyester resin.

21.20 References

Time-Life Books, Seven Centuries of Art, New York, 1970.
Mark, H., and Whitby, G. S., Collected Papers of W H. Carothers, John Wiley & Sons, New York, 1940.
Turbak, A., "Rayon," in Encyclopedia of Polymer Science and Engineering, 2nd ed., Vol. 14, A. Klingsberg and T. Baldwin (Eds.), p. 55, John Wiley & Sons, New York, 1985.
Gupta, B. S., and Hong, C. J., INJ, 7(1), 38 (1995).
Davis, S., Textile Horizons, 9(2), 62 (1989).
Albrecht, W., Reintjes, M., and Wulfhorst, B., Chem. Fibers Int., 47, 298 (1997).
Markham, J. W., Competition in the Rayon Industry, p. 16, Harvard University Press, Cambridge, MA, 1952.
Robinson, J. S., Fiber-Forming Polymers: Recent Advances, Noyes Data Corp., Park Ridge, NJ, 1980.
Barnes, C. E., "Nylon 4-Development and Commercialization," Lenzinger Berichte, 62, 6266, March 1987.
O'Sullivan, D., Chemical and Engineering News, 62(21), 33 (1984).
Davis, G. W., Everage, A. E., and Talbot, J. R., Fiber Producer, 12(6), 45 (1984).
Ahmed, M., Polypropylene Fibers-Science and Technology, Textile Science and Technology, Vol. 5, p. 16, Elsevier, New York, 1982.
Lieberman, R. B., and Barbe, P. C., "Propylene Polymers," in Concise Encyclopedia of Polymer Science and Engineering, J. I. Kroschwitz (Ed.), p. 916, John Wiley & Sons, New York, 1990.
Hogan, J. P., and Banks, R. L., "History of Crystalline Polypropylene," in History of Polyolefins, R. B. Seymour and T. Cheng (Eds.), p. 103, D. Reidel Pub. Co., Boston, 1986.
Gupta, B. S., and Smith, D. K., "Nonwovens in Absorbent Materials," in Absorbent Technology, P. K. Chatterjee and B. S. Gupta (Eds.), p. 378, Elsevier, Amsterdam, 2002.
Madsen, J. B., Nonwovens World, 69 (2001, AugustSeptember).
Zwijnenburg, A., and Pennings, A. J., Colloid Polym. Sci., 259, 868 (1978).
Smith, P., and Lemstra, P. J., J. Mater. Sci., 15, 505 (1980).
Kavesh, S., and Prevorsek, D., U.S. Patent 4,413,110, to Allied Chemical, 1983.
Kwolek, D. L., U.S. Patent 3,600,350, to E. I. duPont de Nemours and Co., 1971.
Blades, H., U.S. Patent 3,767,756, to E. I. duPont de Nemours and Co., Inc., 1973.
McIntyre, E., Textile Horizons, 8(10), 43 (1988).
Chenevey, E. C., and Conciatori, A. B., U.S. Patent 3,549,603, to Celanese Corp., 1970.
Coffin, D. R., Serad, G. A., Hicks, H. L., and Montgomery, R. T., Textile Res. J., 52, 466 (1982).
Gore, R. W., U.S. Patent 3,953,566, to W. L. Gore & Associates, Inc., 1973, April 27.
Menardi-MikroPul L. L. C., www.mikropul.com/products/media/mikrotex.html
Edmonds, J. T., Jr., and Hill, H. W., Jr., U.S. Patent 3,354,129, to Phillips Petroleum Co., 1967.
Scruggs, J. G., and Reed, J. O., "Polyphenylene Sulfide Fibers," in High Technology Fibers, Part A, M. Lewin and J. Preston (Eds.), Marcel Dekker, Inc., New York, 1985.

21.21 Suggested Reading

The reader is referred to the four encyclopedias listed below for additional information. They contain enormous quantities of information on manufactured fibers as well as comprehensive bibliographies.

Concise Encyclopedia of Polymer Science and Engineering, John Wiley & Sons, New York, 1990.
Encyclopedia of Polymer Science and Engineering, 2nd ed., John Wiley & Sons, New York, 1985. (17 volumes, index volume, and supplement volume.)
Encyclopedia of Polymer Science and Technology, Interscience Publishers, New York. (16 volumes.)
Kirk-Othmer Encyclopedia of Chemical Technology, 3rd ed., Interscience Publishers, New York, (21 volumes and a supplement, 3rd ed.; to date, 16 volumes.)

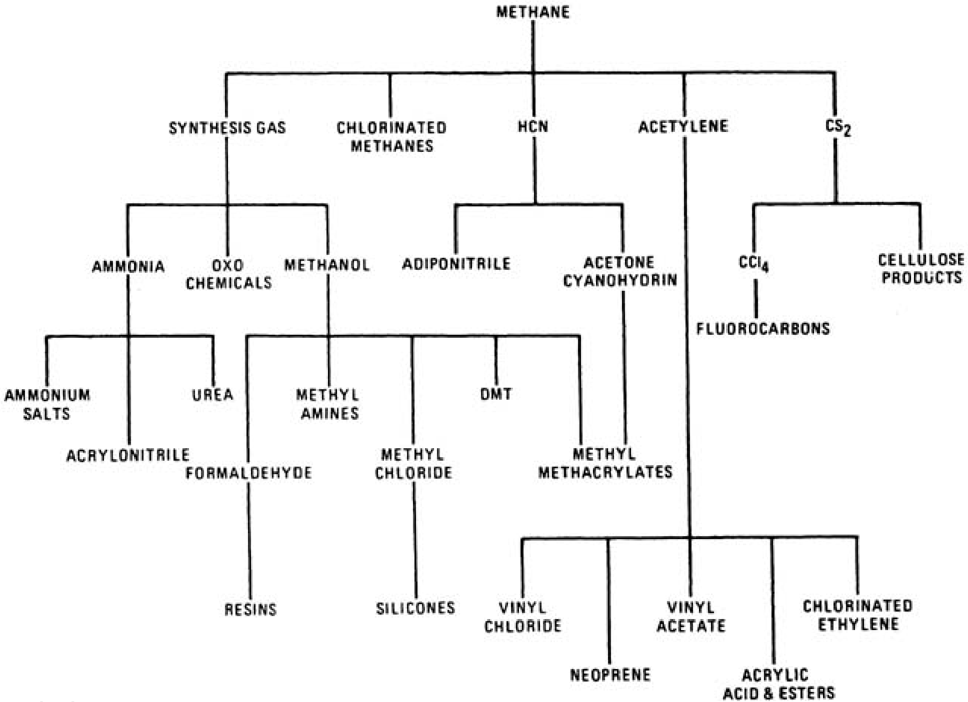
The following books contain broad discussions of manufactured textile fibers.

Baer, E. and Moet, A. (Eds.), High Performance Polymers, Hanser Publishers, New York, 1991.
Billmeyer, F. W., Textbook of Polymer Science, John Wiley & Sons, New York, 1984.
Ciferri, A., and Ward, I. M. (Eds.), Ultra-High Modulus Polymers, Applied Science Publishers, London, 1979.
Datye, K. V., Chemical Processing of Synthetic Fibers and Blends, John Wiley & Sons, New York, 1984.
Hearle, J. W. S., and Peters, R. H. (Eds.), Fibre Structure, The Textile Institute, Manchester, Butterworths, London, 1963.
Lewin, M., and Preston, J. (Eds.), Handbook of Fiber Science and Technology: High Technology Fibers, Vol. III, Marcel Dekker, Inc., New York, 1983.
Mark, H. F., Atlas, S. M., and Cernia, E. (Eds.), Man-Made Fibers; Science and Technology, Vols. I, II, and III, John Wiley & Sons, New York, 1967, 1968, and 1968.
Moncrieff, R. W., Man-Made Fibres, 6th ed., John Wiley & Sons, New York, 1975.
Morton, W. E., and Hearle, J. W. S., Physical Properties of Textile Fibres, The Textile Institute, Manchester, Butterworths, London, 1993.
Peters, R. H., Textile Chemistry; The Chemistry of Fibers, Vol. I and Impurities in Fibers; Purification of Fibers, Vol. II, Elsevier, New York, 1963 and 1967.

22.1 Chemicals Derived from Methane

Methane is readily available as the major component of natural gas, and its primary use is as fuel. It is also found in coal mines and as a product of anaerobic biological decay of organic materials in swamps and landfills. It is possible to make many chemicals from methane in a laboratory. However, methane is relatively inert chemically and is truly useful as a raw material for only a few commercial chemicals. These conversions generally require high temperatures and pressures or very aggressive chemicals such as chlorine, and usually are operated on a very large scale. Here, only those materials that are currently made from methane in commercial quantities are considered. The most important of these are shown in Fig. 22.2.

Fig. 22.2 Some important synthetic chemicals derived from methane.



Synthesis Gas

The most important route for the conversion of methane to petrochemicals is via either hydrogen or a mixture of hydrogen and carbon monoxide. The latter material is known as synthesis gas. The manufacture of carbon monoxidehydrogen mixtures from coal was first established industrially by the well-known watergas reaction:

$$C + H_2O \rightarrow CO + H_2$$

Two important methods presently are used to produce the gas mixture from methane. The first is the methanesteam reaction, where methane and steam at about 900°C are passed through a tubular reactor packed with a promoted iron oxide catalyst. Two reactions are possible, depending on the conditions:

$$CH_4 + H_2O \quad \rightarrow \quad CO + 3H_2$$

$$CO + H_2O \quad \rightarrow \quad CO_2 + H_2$$

The second commercial method involves the partial combustion of methane to provide the heat and steam needed for the conversion. Thus the reaction can be considered to take place in at least two steps, the combustion step:

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$$

followed by the reaction steps:

$$CH_4 + CO_2 \quad \rightarrow \quad 2CO + 2H_2$$

$$CH_4 + H_2O \quad \rightarrow \quad CO + 3H_2$$

The process usually is run with nickel catalysts in the temperature range of 8001000°C. Steam reforming usually is used on the lighter feedstocks, and partial oxidation is used for the heavier fraction.

Synthesis gas is the starting material for the manufacture of ammonia and its derivatives and also for methanol, as well as for other oxo-synthesis processes. It also is a source of carbon monoxide in the manufacture of such chemicals as acetic acid. And it is also a source of hydrogen for petroleum refining processes. However, shortages and corresponding high prices of natural gas and naphtha have generated interest in other synthesis gas feedstocks such as coal and residual oil.

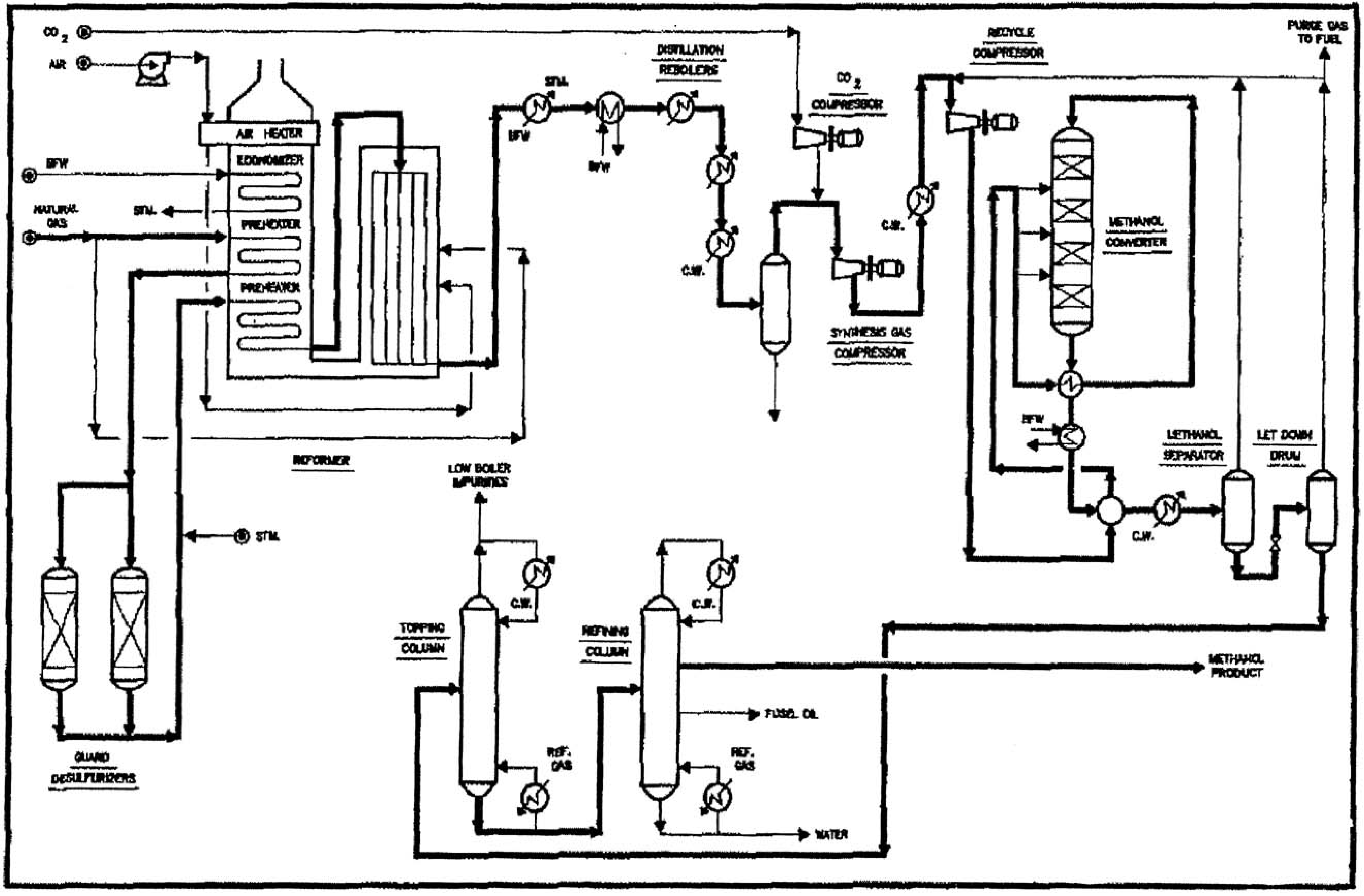
Ammonia

Although ammonia is not an organic chemical, it is one of the largest-volume synthetic petrochemicals. From it many reactive organic chemicals are derived such as urea, acrylonitrile, caprolactam, amines, and isocyanates. Almost all of the 40.6 or more billion lb of ammonia produced in 1999 in the United States was based on hydrogen from petroleum and natural gas. Detailed descriptions of ammonia processes are found in Chapter 29.

Methanol

Methanol was once known as wood alcohol because it was a product of the destructive distillation of wood. All American methanol was produced in that way before 1926. That year, however, marked the first appearance in the world of German synthetic methanol. Today, almost all of the approximately 10 billion lb/year of methanol made in the United States comes from large-scale, integrated plants for conversion of natural gas to synthesis gas to methanol (Fig. 22.3). The world consumption has reached 57 billion lb in 2000.

Fig. 22.3 Methanol via natural gas reforming. (*Chem Systems Report No. 93-1. Copyright Nexant Chem Systems, Inc. and used by permission of the copyright owner*.)



In principle, methanol, because it is derived from synthesis gas, can be made not only from convenient natural gas but also from any source of reduced organic carbon such as coal, wood, or cellulosic agricultural waste. It then can be used as a readily stored fuel or shipped for use as fuel or raw material elsewhere. As petroleum and natural gas become more difficult to recover, alternative carbon sources such as these will be used more. An example involving methanol is a coal-based acetic anhydride facility started in the 1980s in the United States. Another example is the use of some of the natural gas formerly wasted during recovery of Middle Eastern oil to make methanol and other chemicals. Direct use of methanol as a motor vehicle fuel is being studied, but it is not known when or even if such use will be significant in terms of methanol usage or decreased petroleum usage.

Methanol synthesis resembles that of ammonia in that high temperatures and pressures are used to obtain high conversions and rates. Improvements in catalysts allow operation at temperatures and pressures much lower than those of the initial commercial processes. Today, "low-pressure" CuZnAlminium oxide catalysts are operated at about 1500 psi and 250°C. These catalysts must be protected from trace impurities that the older "high-pressure" (5000 psi and 350°C) and "medium-pressure" (3000 psi and 250°C) catalysts tolerate better. Synthesis gas production technology has also evolved so that it is possible to maintain the required low levels of these trace impurities.

Methanol is used as a solvent, an antifreeze, a refrigerant, and a chemical intermediate. The greatest chemical uses for methanol as of 1998 were formaldehyde, 33 percent; MTBE, 27 percent; acetic acid, 7 percent; and chloromethane, 5 percent. Other chemicals derived from methanol include methyl methacrylate, methylamines, and dimethyl terephthalate.

Formaldehyde

Formaldehyde may be made from methanol either by catalytic vapor-phase oxidation:

$$CH_3OH + \tfrac{1}{2}O_2 \rightarrow CH_2O + H_2O$$

or by a combination oxidationdehydrogenation process:

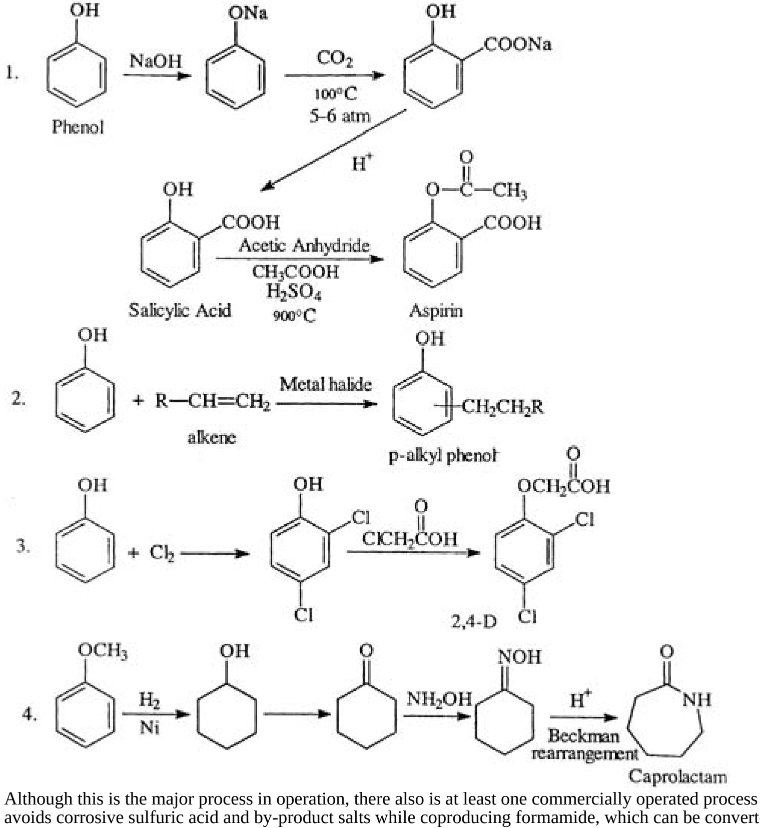
$$CH_3OH \rightarrow CH_2O + H_2$$

It also can be produced directly from natural gas, methane, and other aliphatic hydrocarbons, but this process yields mixtures of various oxygenated materials. Because both gaseous and liquid formaldehyde readily polymerize at room temperature, formaldehyde is not available in pure form. It is sold instead as a 37 percent solution in water, or in the polymeric form as paraformaldehyde [HO(CH2O)*n*H], where *n* is between 8 and 50, or as trioxane (CH2O)3. The greatest end use for formaldehyde is in the field of synthetic resins, either as a homopolymer or as a copolymer with phenol, urea, or melamine. It also is reacted with acetaldehyde to produce pentaerythritol [C(CH2OH)4], which finds use in polyester resins. Two smaller-volume uses are in ureaformaldehyde fertilizers and in hexamethylenetetramine, the latter being formed by condensation with ammonia.

U.S. production of formaldehyde in 2000 was approximately 10 billion lb of 37 percent formaldehyde, amounting to about 85 percent of capacity. Usage is expected to grow at about 3 percent/year through 2005.

Methyl Methyacrylate

Methyl methacrylate is formed in a three-step process from methanol, acetone, and HCN:



Although this is the major process in operation, there also is at least one commercially operated process for converting isobutylene and/or tert-butanol to methacrylic acid followed by esterification with methanol. A new process based on acetone cyanohydrin also has been reported. It avoids corrosive sulfuric acid and by-product salts while coproducing formamide, which can be converted to HCN for recycle to the process.

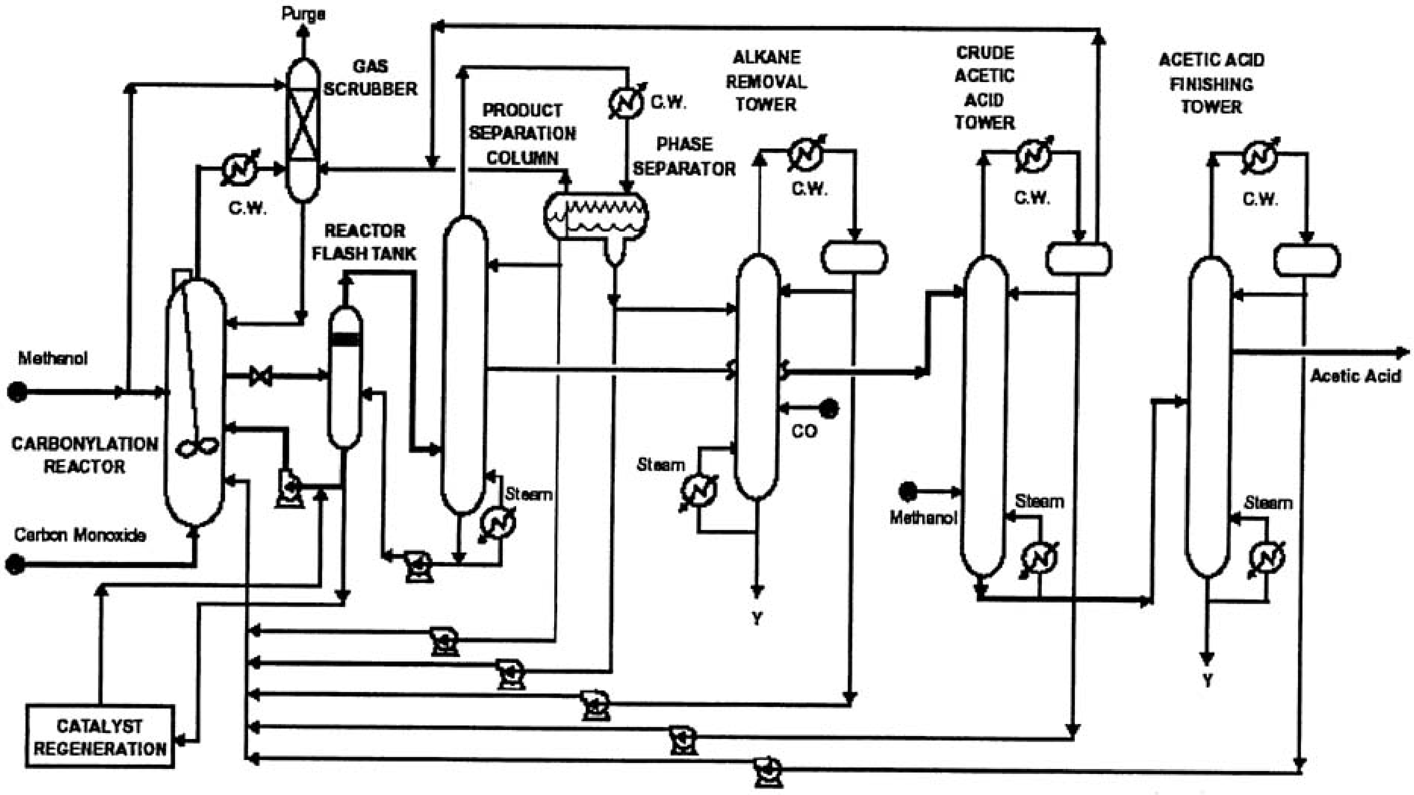
U.S. production of methyl methacrylate in 1999 totaled 1.4 billion lb, which is about 82 percent of world capacity. Its uses are almost exclusively based on polymerization to poly(methyl methacrylate), which, because of its physical strength, weathering resistance, optical clarity, and high refractive index, has major uses in cast and extruded sheet (33%), molding powders and resins (16%), and surface coatings (22%).

Acetic Acid

The worldwide production of acetic acid was reported to be 15.7 billion lb in 1998. Acetic acid is a global product with about one third of production capacity now outside the United States, Western Europe, and Japan. The majority buildup is in Asia. In the future, the capacity in Asia will continue to increase substantially.

Acetic acid is produced by methanol carbonylation (the dominant process) as well as by acetaldehyde oxidation, ethanol oxidation, and light hydrocarbon oxidation. When methanol carbonylation was first practiced in the United States and West Germany, a cobalt iodide catalyst was used, and the process required up to 10,000 psia pressure. The technological breakthrough that allowed methanol carbonylation to become the leading acetic acid process was the discovery of rhodiumiodine catalysts, which can be operated at moderate pressure (500 psia) and at a methanol selectivity of 99 percent to acetic acid. Figure 22.4 is a Schematic of the process. The recent advancement is the implementation of the low water technology. It significantly reduces the production cost by increasing productivity and lowering utility and capital cost.

Fig. 22.4 Acetic acid via methanol carbonylation. (*Chem Systems Report No. 99/00 S5. Copyright Nexant Chem Systems, Inc. and used by permission of the copyright owner*.)



In the United States, 5.3 billion lb of acetic acid was consumed in 1998. The applications, in decreasing order were: vinyl acetate (which alone accounts for more than 40 percent of U.S. acetic acid consumption), dimethyl terephthalate/terephthalic acid, acetate esters, cellulose acetate, other acetic anhydride uses, textiles, monochloroacetic acid, and several smaller uses. Growth projections are close to 3 percent from 1999 to 2003. The growth is tied largely to vinyl acetate monomer manufacture and to a lesser extent terephthalic acid manufacture.

Methyl t-Butyl Ether (MTBE)

In 1980, MTBE was the fastest-growing derivative of methanol. This is a result of its only significant use, which is as an antiknock agent replacing lead in gasoline. In 1990, it was the fastest-growing chemical in the world. The world production has reached 48 billion in 1999. The U. S. production was about 30 billion lb, which was 63 percent of world capacity.

Because of its high miscibility in water and its increased use over the last several years, MTBE is now being found in many areas of the United States in groundwater reservoirs. This problem received national attention in June 1996, after MTBE was discovered in the drinking water supply of the city of Santa Monica, California. This incident led to many legislative initiatives in California and culminated in an Executive Order issued by California Governor Davis on March 25, 1999 to remove MTBE from all gasoline sold in California at the earliest possible date, but not later than December 31, 2002. Large declines in MTBE production and use are expected in the next few years.

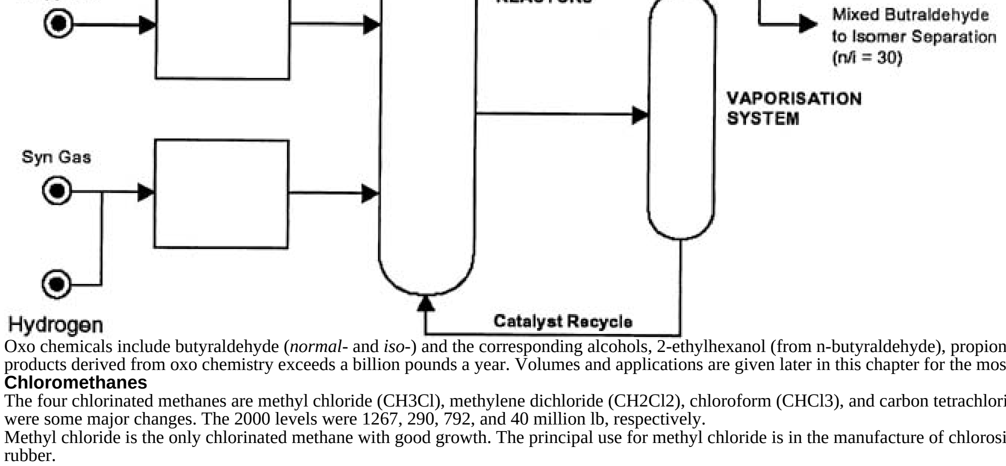
MTBE is made by reacting methanol with the isobutylene contained in mixed-C4 refinery streams. This is possible because butanes, the other butanes, and butadiene are inert under the mild conditions used. The process is catalyzed by acidic ion exchange resins.

Oxo Chemicals

The so-called oxo process combines carbon monoxide and hydrogen with olefins to make saturated aldehydes having one more carbon atom than the olefins have. The earliest such reaction studied used ethylene to produce both an aldehyde and a ketone. Thus, the reaction was named "oxo" after the German *oxierung*, meaning "ketonization."

The low-pressure oxo process based on rhodium complex catalysts has largely replaced the older, high-pressure process, which used cobalt carbonyls as catalyst. The low-pressure process is operated at about 100°C and 200 psig. A new generation oxo process with bisphosphite modified rhodium catalyst is shown schematically in Fig. 22.5.

Fig. 22.5 Low-pressure oxo process. (*Chem Systems Report No. 98/99 S13. Copyright Nexant Chem Systems, Inc. and used by permission of the copyright owner*.)



Oxo chemicals include butyraldehyde (*normal*- and *iso*-) and the corresponding alcohols, 2-ethylhexanol (from n-butyraldehyde), propionaldehyde, and n-propyl alcohol, and lesser amounts of higher aldehydes and alcohols derived from C5 through C17 olefins. The total volume of products derived from oxo chemistry exceeds a billion pounds a year. Volumes and applications are given later in this chapter for the most important products.

Chloromethanes

The four chlorinated methanes are methyl chloride (CH3Cl), methylene dichloride (CH2Cl2), chloroform (CHCl3), and carbon tetrachloride (CCl4). The U.S. production levels were about 520, 500, 450, and 750 million lb respectively, in 1990. Due to environmental regulation, there were some major changes. The 2000 levels were 1267, 290, 792, and 40 million lb, respectively.

Methyl chloride is the only chlorinated methane with good growth. The principal use for methyl chloride is in the manufacture of chlorosilanes (89%) for the silicone industry. Other smaller uses are for methyl cellulose ether, quaternary ammonium compounds, herbicides, and butyl rubber.

Methyl chloride is produced by two methods: by the reaction of hydrogen chloride and methanol and by the chlorination of methane. Due to increasing demand for methyl chloride, the more selective methanol hydrochlorination has become increasingly important, whereas the nonselective methane chlorination route has declined. The hydrochlorination process also has the advantage that it utilizes, instead of generating, hydrogen chloride, a product whose disposal has become increasingly difficult. The methanol hydrochlorination process can be carried out in either liquid or gas phase. The gaseous phase reaction is carried out at 250280°C. It uses a smaller reactor but requires extra energy to vaporize aqueous HCl.

$$\mathrm{CH_3OH + HCl \rightarrow CH_3Cl + H_2O}$$

Methylene chloride and chloroform can be made, along with the other products, by the direct chlorination of methane. It is much more common, however, to produce them by the chlorination of methyl chloride. This can he done either thermally (350450°C) or photochemically. The HCl by-product can be recycled back into a hydrochlorination process for production of the methylene chloride starting material.

Methylene chloride is used primarily as a solvent for degreasing and paint removal, and it is also used in aerosols and foam-blowing agents. Since 1985, new environmental regulation has a major impact on this chemical. The consumption of methylene chloride has reduced by 60 percent due to recycling and product substitutions.

Chloroform is used to produce chlorodifluoromethane (HCFC-22), which is used as a refrigerant (70%) and to synthesize the monomer tetrafluoroethylene (30%). Fluoropolymers that use HCFC-22 as a feedstock are strong. But the uses for refrigerant will be phased out beginning 2010. The major uses for carbon tetrachloride were to make aerosol propellants such as dichlorodifluoromethane (CFC-12) and trichlorofluoromethane (CFC-11). The volume of carbon tetrachloride decreases to almost zero as CFC-11 and 12 will be phased out.

Acetylene

In the early days of the chemical industry, acetylene was a key starting material for many important products. Initially it was obtained for chemical purposes by reaction of calcium carbide with water; but that practice has given way to acetylene recovery from hydrocarbon cracking, so that now 86 percent of acetylene used in chemical manufacturing is made in this way. Owing to difficulty in its safe collection and transport, it is almost always used where it is prepared.

Acetylene still is a preferred raw material for some products, but it has been largely replaced by ethylene for many others. Chemicals once produced from acetylene by processes now considered outdated include: vinyl chloride, vinyl acetate, acetaldehyde, acrylonitrile, neoprene, and chlorinated solvents.

1,4-Butanediol

In 2000, 164 million lb of 1,4-butanediol were made from acetylene in the United States. It is the largest consumer of acetylene.

The Reppe process is used to make 1,4-butanediol from acetylene. In this process, acetylene and formaldehyde are reacted in the presence of a copperbismuth catalyst. The resulting intermediate, 2-butyne-1,4-diol is hydrogenated over a Raney nickel catalyst:

$$\mathrm{HC} - \mathrm{CH} + 2\mathrm{HCHO} \rightarrow \underset{\text{2-butyne-1,4-diol}}{\mathrm{HOCH_2C} - \mathrm{CCH_2OH}}$$

$$\underset{\text{2-butyne-1,4-diol}}{\mathrm{HOCH_2C} - \mathrm{CCH_2OH}} + \mathrm{H_2} \rightarrow \underset{\text{1,4-butanediol}}{\mathrm{HO(CH_2)_4OH}}$$

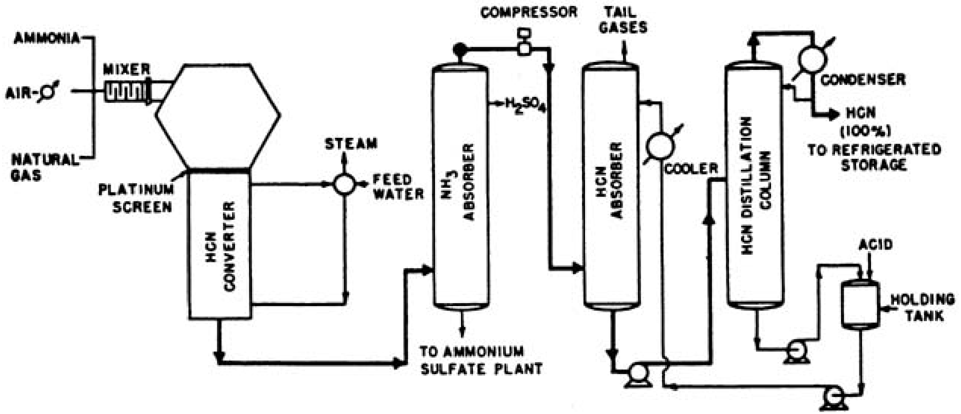
The use of acetylene to make vinyl chloride (VCM) is now considered outdated and it is a minor process compared to the production of VCM from ethylene. Only 120 million lb of 16 billion lb VCM was made from acetylene.

Applications of 1,4-butanediol include the manufacture of THF (tetrahydrofuran), used as a solvent, and of poly(butylene terephthalate), used in thermoplastic resins.

Hydrogen Cyanide

Most of the hydrogen cyanide used today is prepared, as illustrated in Fig. 22.6, by ammoxidation of methane over a platinum catalyst. Absorption in water and distillation give pure hydrogen cyanide. Although all new hydrogen cyanide capacity is based on this technology, it also can be made from coke-oven gas, from acidification of inorganic cyanides, and by dehydration of formamide. A significant amount of material is obtained as a by-product of ammoxidation of propylene in acrylonitrile manufacture. This amounts to about 25 percent of the demand. Because hydrogen cyanide is very toxic, the producers use much of it on site to minimize the potential for human exposure during shipping.

Fig. 22.6 The hydrogen cyanide process. (*Reprinted from Ind. Eng. Chem.*, 51, *no. 10, 1235, 1959; Copyright 1959 by the American Chemical Society and reprinted by permission of the copyright owner*.)



The U.S. demand in 2000 was about 1.6 billion lb. Uses for hydrogen cyanide include: adiponitrile (for nylon 6/6), 47 percent; methyl methacrylate, 27 percent; sodium cyanide, 8 percent; methionine, 6 percent; and chelating agents, 2 percent.

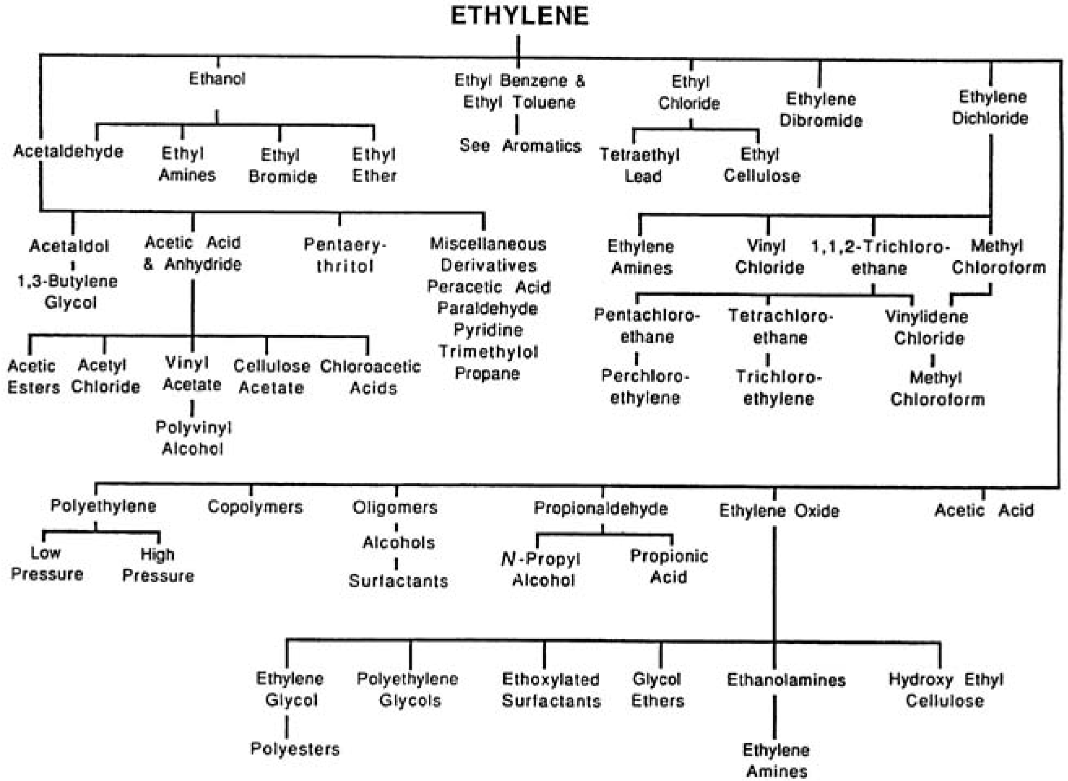
Carbon Disulfide

Carbon disulfide is made by the catalytic reaction of methane with sulfur vapor. This can be thought of as analogous to oxidation of methane with oxygen, giving carbon dioxide. Carbon disulfide is used as raw material for making rayon (43%), agricultural chemicals (36%) and cellophane (3%). The total U.S. consumption in 2000 was about 160 million lb.

22.2 Chemicals Derived from Ethylene

Ethylene surpasses all other organic petrochemicals in production and in the amount sold. It is used as raw material for a greater number of commercial synthetic organic chemical products than is any other single chemical. Figure 22.7 shows the more important derivatives of ethylene. Ethylene consumption has grown explosively since 1940 when 300 million lb were used, mostly for making ethanol and ethylene oxide. During World War II, styrene use grew markedly, and polyethylene was developed as insulation for the then-new radar electronics. These materials later found a multitude of applications, which were responsible in large part for ethylene consumption reaching nearly 5 billion lb in 1960. Strong growth in ethylene dichloride and ethylene oxide contributed to over 18 billion lb of ethylene consumption in 1970. Continued growth raised this figure to 27 billion lb in 1978 and over 58 billion lb in 2000, which is almost two hundred times the 1940 volume. The overall growth for 48 years averaged a remarkable 10 percent/year.

Fig. 22.7 Important derivatives of ethylene.



Ethylene is manufactured by cracking hydrocarbons. A discussion of hydrocarbon cracking as a route to ethylene is found in Chapter 15.

Polyethylene

The largest consumers of ethylene are the various types of polyethylene: Low Density Polyethylene (LDPE), High Density Polyethylene (HDPE), and Linear Low Density Polyethylene (LLDPE). Chapter 19 gives detailed discussions of preparation of the various types of polyethylene.

LDPE is produced by high-pressure, high-temperature radical polymerization of pure ethylene. When improved properties are required, copolymers with one or more other vinyl monomers such as ethyl acrylate, vinyl acetate, or acrylic acid are used. LDPE has a relatively branched molecular structure, and the branches are relatively long. It is used for a multitude of purposes because of its properties and economics. Some important uses include: films for packaging of food and other merchandise; shipping trays and pallets; lightweight, flexible water- and chemical-resistant containers or barriers; and temporary coverings as in construction and agriculture.

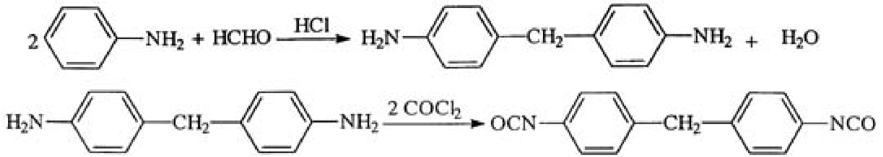
HDPE is produced by a low-pressure polymerization process in which highly pure gaseous ethylene is converted by proprietary catalysts to solid polyethylene particles. It has a very linear molecular structure. HDPE is stronger, tougher, and more rigid than LDPE, so it is used where such properties are advantageous. The major uses for HDPE are in blow-molded bottles, cans, and tanks for products such as milk, bleach, detergent, and fuel and in grocery sacks and other paper-replacement markets.

LLDPE is made by a catalytic process very similar to that for HDPE, but it is a softer poly-ethylene than HDPE with properties similar to those of LDPE. Its properties are achieved by inclusion of comonomers such as butene or hexene. A relatively disordered crystalline state is obtained by introducing many short branches into an otherwise highly linear molecule. Thus, the less expensive equipment of the HDPE process can be used to make a product having the greater flexibility and impact strength characteristic of LDPE.

In 2000, world capacity for polyethylene was nearly 112 billion lb.

Ethylene Oxide

Ethylene oxide was discovered in 1859 by Wurtz. He stated that ethylene oxide could not be made by direct oxidation of ethylene, and it was nearly 80 years before this was disproved. Wurtz made ethylene oxide by the method known today as the chlorohydrin process, in which ethylene is reacted in turn with hypochlorous acid and base. This process was commercialized during World War I in Germany, and until 1985 was still used commercially in the United States.



Since 1985, processes for the direct oxidation of ethylene using either air or oxygen and a silver catalyst have been the only remaining commercial processes for ethylene oxide production in the United States. Figure 22.8 illustrates an air-based process, and Fig. 22.9 an oxygen-based process. In this very exothermic conversion, oxygen and ethylene combine on a silver catalytic surface to make ethylene oxide. Oxygen and ethylene concentrations are controlled at low levels to avoid creating explosive mixtures.

Fig. 22.8 Air-based direct oxidation process for ethylene oxide. (*Encyclopedia of Chemical Technology, Kirk and Othmer, web site ed., ethylene oxide, manufacture, 2002. Copyright by John Wiley & Sons, Inc. and reproduced by permission of the copyright owner.*)

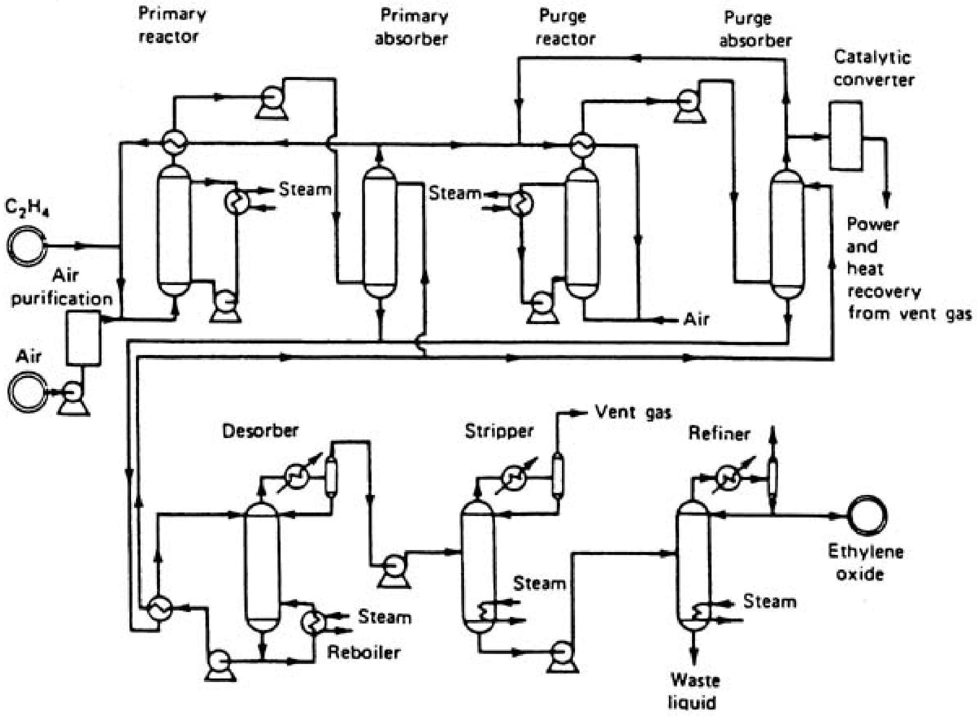
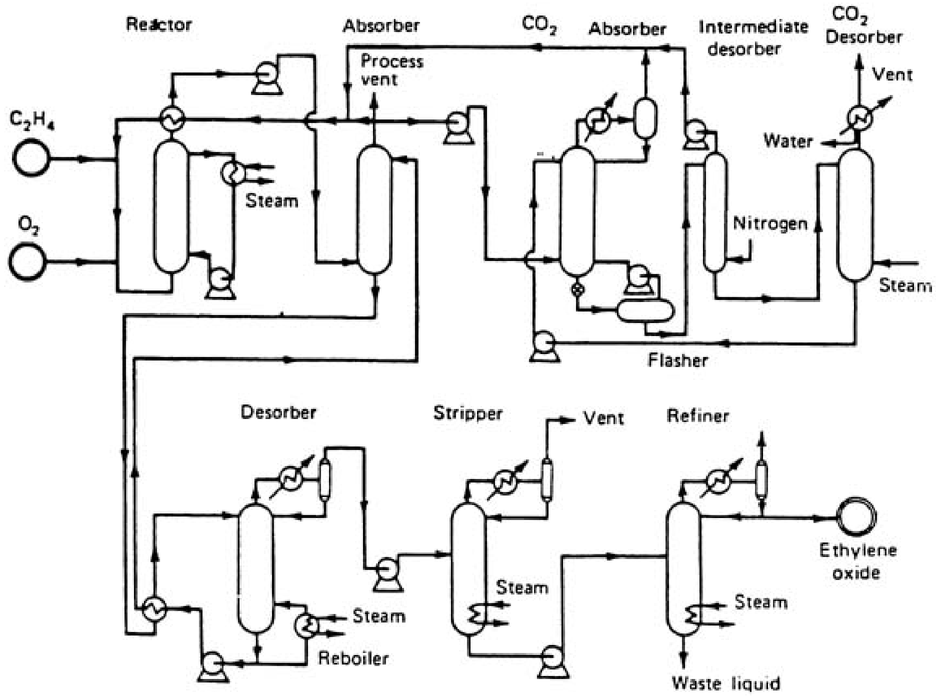


Fig. 22.9 Oxygen-based direct oxidation process for ethylene oxide. (*Encyclopedia of Chemical Technology, Kirk and Othmer, web site ed., ethylene oxide, manufacture, 2002. Copyright by John Wiley & Sons, Inc. and reproduced by permission of the copyright owner.*)

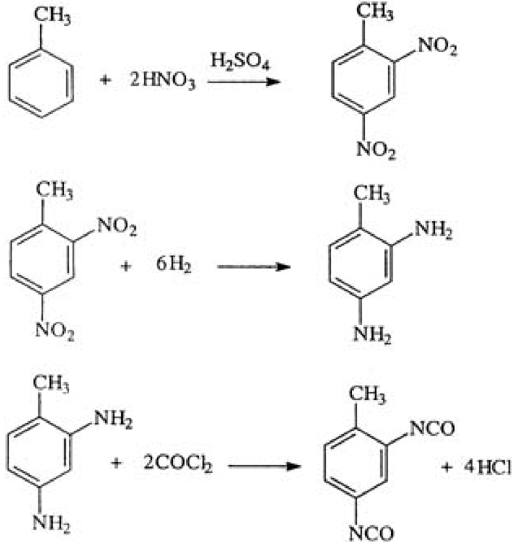


The competing reactions of total combustion to carbon dioxide and isomerization must be avoided. Ethylene oxide plants in which air is used as source of oxygen require additional investment for purge reactors and associated absorbers. This investment is offset by the need, in the oxygen-based process, for an oxygen production plant and a carbon dioxide removal system. In general, the oxygen-based process is thought to be more economical, and all the plants built since the mid-1970s have been oxygen-based.

In 1999, U.S. ethylene oxide capacity was 9.1 billion lb with production of 8.2 billion lb. Major uses in that year were: ethylene glycol, 57 percent; nonionic surfactants, 11 percent; ethanolamines, 11 percent; glycol ethers, 7 percent; diethylene glycol, 5 percent; and triethylene glycol, 2 percent. The remaining 7 percent of ethylene oxide consumption included PEGs (poly(ethylene glycol)), urethane polyols, and exports. In the following sections, several of these derivatives are discussed in more detail.

Ethylene Glycols

Monoethylene glycol or ethylene glycol is the major derivative of ethylene oxide. Ethylene glycol was initially made commercially by hydrolysis of ethylene chlorohydrin. Today, hydrolysis of ethylene oxide is the preferred route.



Reaction of ethylene oxide with water is accomplished using a large molar excess of water to favor ethylene glycol formation over that of diethylene glycol and triethylene glycol, as the glycols are more reactive with ethylene oxide on a molar basis than is water. A multiple-effect distillation scheme is used to recycle the unreacted excess water. The major product, ethylene glycol, is recovered between light impurities and the heavier glycol by-products.

In 1999, U.S. consumption of ethylene glycol totaled 5.5 billion lb. Of that, 1.5 billion lb (28%) was used in the production of polyester bottles, primarily for soft drinks. Polyester fiber applications accounted for 1.4 billion lb (26%), primarily for the textile industry. Polyester film and miscellaneous applications consumed another 0.4 billion lb (7%). Anti-freeze applications have held steady at approximately 1.6 billion lb over the last 20 years, and have become relatively less important with time than the polyester applications. This trend is expected to hold in the future. Increased demand for polyester bottles is expected to fuel growth in the United States, while bottle and textile applications are expected to fuel growth in other areas of the world.

Diethylene glycol usage is about 800 million lb/year in the United States. Major uses are unsaturated polyester resins (21%), polyurethane resins (21%), and antifreeze blending (10%). Other applications include use as raw materials for triethylene glycol (7%) and for morpholine (7%). Diethylene glycol is also used for dehydration of natural gas and in textile conditioning.

Triethylene glycol consumption is approximately 115 million lb/year in the United States. The major use, natural gas drying, depends on the low volatility and strong affinity of triethylene glycol for water. Lesser amounts are used as intermediate for vinyl plasticizers, polyester resins, and polyols. Additional direct applications include solvent and humectant uses.

Polyethylene glycols are produced by base-catalyzed addition of ethylene oxide to a low molecular weight glycol such as diethylene glycol. These glycols are higher-molecular-weight analogs of mono-, di-, and triethylene glycol but differ from the latter compounds in that they are not pure substances but rather consist of distributions of low molecular weight polymers. With average molecular weights beginning at about 200 (that of tetraethylene glycol) and going up to about 1000, these materials are liquids at ambient temperature. They are used as plasticizer intermediates, dispersant media, lubricants, and humectants. Above an average molecular weight of 1000, the polyglycols become waxy solids and find use in ointments, cosmetics, and lubricants taking advantage of their oil and water compatibility and low toxicity. At very high molecular weights, homopolymers of ethylene oxide are used for thickening, for water-soluble films, and for reducing friction in, for example, water delivery in fire hoses. Their value in this last application is that a given size hose can be made to deliver a greater flow of water.

Surfactants

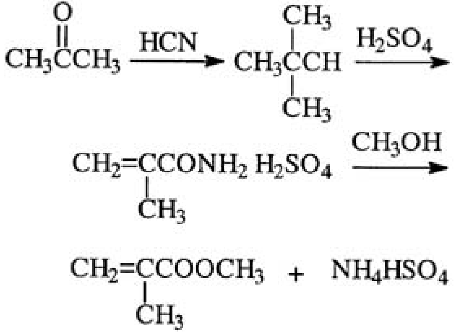
Ethylene oxide-containing surfactants are generally of nonionic or anionic classes. The nonionic materials are made by base-catalyzed addition of ethylene oxide to either fatty alcohols or alkylphenols. Sulfation can be used to convert these compounds to the sulfated anionic surfactants. The products contain from a few to many ethylene oxide molecules per alcohol. The chain of poly(ethylene oxide) in a nonionic product acts as the hydrophile, and the alkyl or alkaryl residue is the hydrophobe. A sulfate salt group adds to the hydrophilicity of an anionic surfactant.

Surfactants based on aliphatic alcohols are used as cleaners in both domestic and industrial applications. They provide excellent properties such as wetting, dispersion, and emulsification. The ethoxylates derived from alkylphenols are chemically stable and highly versatile, finding more use in industrial practice than in domestic applications. They are used both as processing aids and as components in various products. Their applications include metal cleaning, hospital cleaners and disinfectants, agricultural chemical formulation surfactants, insecticides and herbicides, oil-well drilling fluids, and many others.

In the United States in 1999 for surfactant applications, 340 million lb of ethylene oxide was consumed in the production of alkylphenol ethoxylates, and 600 million lb was consumed in the production of ethoxylates of aliphatic alcohols.

Ethanolamines

Ethanolamines are manufactured by reacting ethylene oxide and ammonia. The



relative amounts of the three amines will depend primarily on the ammonia-to-oxide feed ratio. The three products are separated by distillation. Over the years, the relative demand for the three products has varied greatly. Thus, operational flexibility must be maintained.

The ethanolamines are water-miscible bases from whose properties stem their major uses as neutralizes in aqueous formulations such as metalworking fluids. Monoethanolamine is used in detergents, in "sweetening" (removing carbon dioxide and hydrogen sulfide from) natural gas, for removing carbon dioxide from ammonia during its manufacture, and as a raw material for producing ethyleneamines by reductive amination. Diethanolamine finds use in detergents and as an absorbent for acidic components of gases, as well as its major use as a raw material for surfactant diethanolamides of fatty acids. Triethanolamine's main end uses are in cosmetics and textile processing.

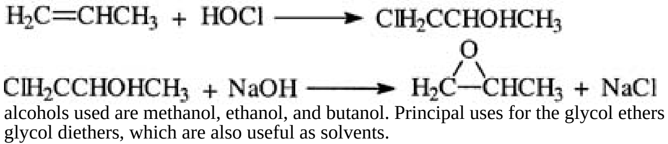
In the United States in 1999, ethylene oxide consumed in the production of ethanolamines totaled approximately 900 million lb. Approximately 300 million lb went into each of monoethanolamine, diethanolamine, and triethanolamine.

Glycol Ethers

In the same way that water reacts with one or more molecules of ethylene oxide, alcohols react to give monoethers of ethylene glycol, producing monoethers of diethylene glycol, triethylene glycol, and so on, as by-products.

$$CH_2CH_2 + H_2O \longrightarrow HOCH_2CH_2OH$$

Since their commercial introduction in 1926, glycol ethers have become valuable as industrial solvents and chemical intermediates. Because glycol monoethers contain a OCH2CH2OH group, they resemble a combination of ether and ethyl alcohol in solvent properties. The most common



$$H_2C{=}CHCH_3 + HOCl \longrightarrow ClH_2CCHOHCH_3$$

$$ClH_2CCHOHCH_3 + NaOH \longrightarrow H_2C{-}CHCH_3 + NaCl$$

alcohols used are methanol, ethanol, and butanol. Principal uses for the glycol ethers are as solvents for paints and lacquers, as intermediates in the production of plasticizers, and as ingredients in brake fluid formulations. Condensation of the monoethers produces glycol diethers, which are also useful as solvents.

Solvent characteristics of glycol ethers are enhanced by esterifying with acetic acid. The resulting acetate esters are used extensively in coating formulations, especially those formulations in which their high solvent power allows a decreased total solvent usage in compliance with volatile organic compound (VOC) emission standards.

In the United States in 1999, ethylene oxide consumption for production of glycol ethers was approximately 560 million lb.

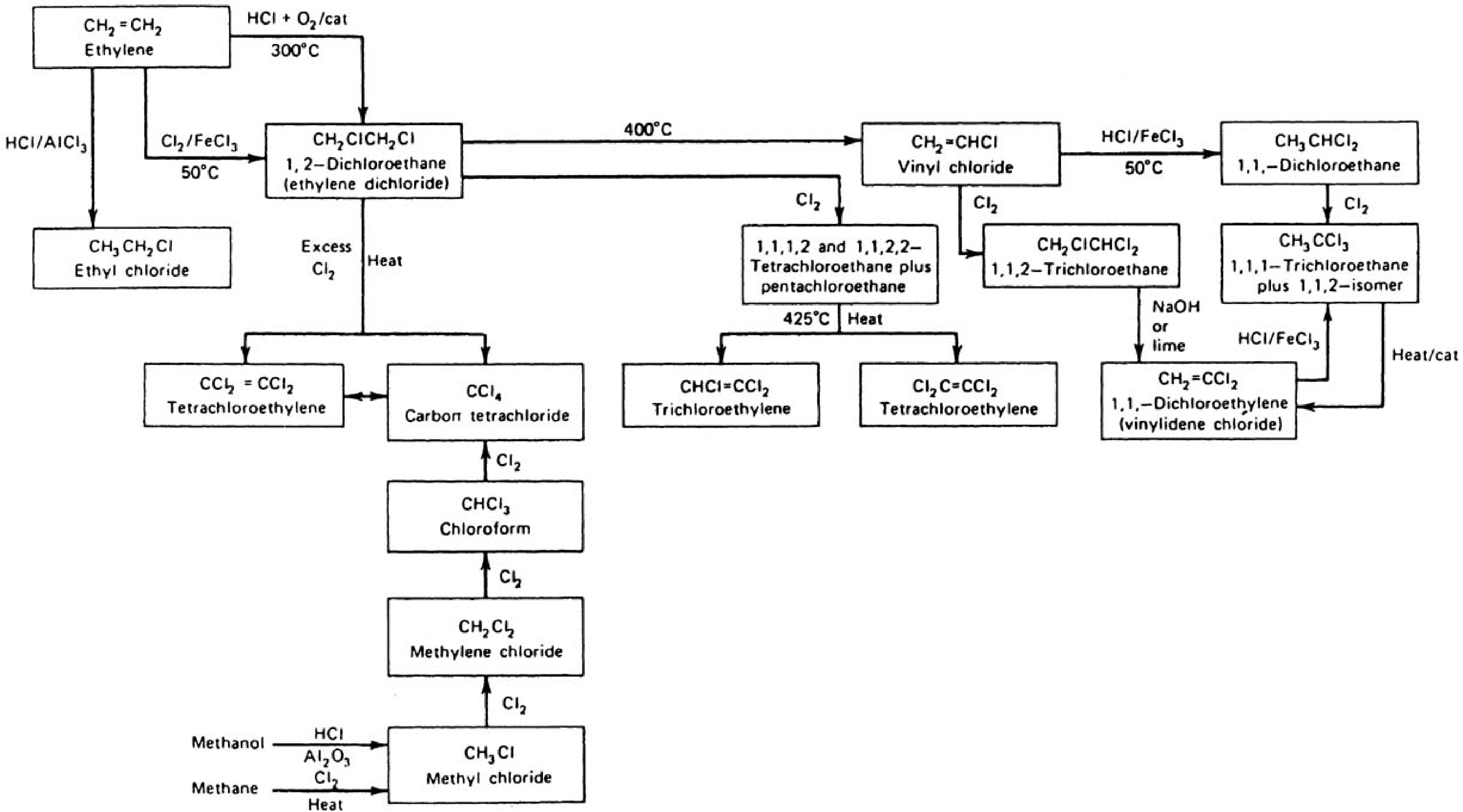
Other Uses of Ethylene Oxide

About 2 percent of ethylene oxide is consumed in miscellaneous applications, such as its use as a raw material in manufacture of choline, ethylene chlorohydrin, hydroxyethyl starch, and hydroxyethyl cellulose and its direct use as a fumigant/sterilant. Production of 1,3-propanediol via hydroformylation of ethylene oxide was begun on a commercial scale in 1999. 1,3-Propanediol is a raw material for polytrimethylene terephthalate, which finds uses in fibers, injection molding, and in film. Use of ethylene oxide in making 1,3-propanediol is expected to be as much as 185 million lb by 2004, up from 12 million lb in 1999.

Chlorinated Ethanes and Ethylenes

A number of important large-volume petrochemicals are obtained through the chlorination of ethane and ethylene. The largest-volume chlorinated derivative is 1,2-dichloroethane (18 billion lb/year), most of it is used to make vinyl chloride. It has about 4 percent growth rate in the past decade. Because of their unique solvent and chemical intermediate properties, the market for chlorinated ethanes and ethylenes (exclude 1,2-dichloroethane) grew steadily until it reached a peak in 1980. Owing to environmental problems, particularly in the solvent area, the demand for some of the end uses has been declining steadily since then. The 1999 annual production for tetrachloroethylene, trichloroethylene, and 1,1,1-trichloroethane are at about 300, 210, and 250 million lb. Figure 22.10 shows the possible production routes to the major chlorinated derivatives.

Fig. 22.10 Manufacturing processes for C1 and C2 chlorocarbons. (*Encyclopedia of Chemical Technology, 3rd ed., Vol. 5, p. 671. Copyright by John Wiley & Sons, Inc. and reproduced by permission of the copyright owner.*)



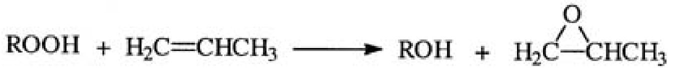
Chlorinated Ethanes

Of the nine possible chlorinated derivatives of ethane, only three are of commercial importance: ethyl chloride, 1,2-dichloroethane (ethylene dichloride), and 1,1,1-trichloroethane (methyl chloroform). The other compounds have no important end uses and are produced either as intermediates or as unwanted by-products. They normally are converted to useful materials by a cracking process (for trichloroethylene) or by perchlorination (for carbon tetrachloride and tetrachloroethylene).

Ethyl Chloride

Most of the ethyl chloride is made by the exothermic hydrochlorination of ethylene, in either the liquid or the vapor phase:

$$\mathrm{RH} + \mathrm{O_2} \longrightarrow \mathrm{ROOH}$$



$$\mathrm{ROOH} + \mathrm{H_2C{=}CHCH_3} \longrightarrow \mathrm{ROH} + \mathrm{H_2C\overset{O}{-}CHCH_3}$$

A much smaller amount is produced by the thermal chlorination of ethane. This direct chlorination may be run in conjunction with another process, such as oxychlorination, which can use the by-product HCl as feed.

Ethyl chloride rose to commercial importance because of the automotive industry. It was the starting material for tetraethyllead, at one time the most commonly used octane booster. Demand has been cut drastically because of the conversion from leaded to unleaded gasoline for environmental reasons. Other uses for ethyl chloride are in the production of ethyl cellulose, as an ethylating agent, as a blowing agent, and in solvent extraction.

1,2-Dichloroethene (Ethylene Dichloride)

Ethylene dichloride (EDC) is one of the truly major petrochemicals. The U.S. production exceeded 18 billion lb in 1999. Almost 95 percent of this went into the manufacture of vinyl chloride monomer.

Figure 22.11 shows an integrated plant for producing EDC and vinyl chloride from ethylene, chlorine, and air. In this process, vinyl chloride (VCM) is produced by the thermal cracking of EDC. The feed EDC may be supplied from two sources. In the first source, ethylene and chlorine are reacted in essentially stoichiometric proportions to produce EDC by direct addition. In the second source, ethylene is reacted with air and HCl by the oxychlorination process. Ideally, both processes are carried out in balance, and the oxychlorination process is used to consume the HCl produced in the cracking and direct chlorination steps. The chemical reactions are as follows:

$$\mathrm{CH_2{=}CH_2 + 2HCl + \tfrac{1}{2}O_2 \rightarrow ClCH_2{-}CH_2Cl}$$

$$\mathrm{CH_2{=}CH_2 + Cl_2 \rightarrow ClCH_2{-}CH_2Cl}$$

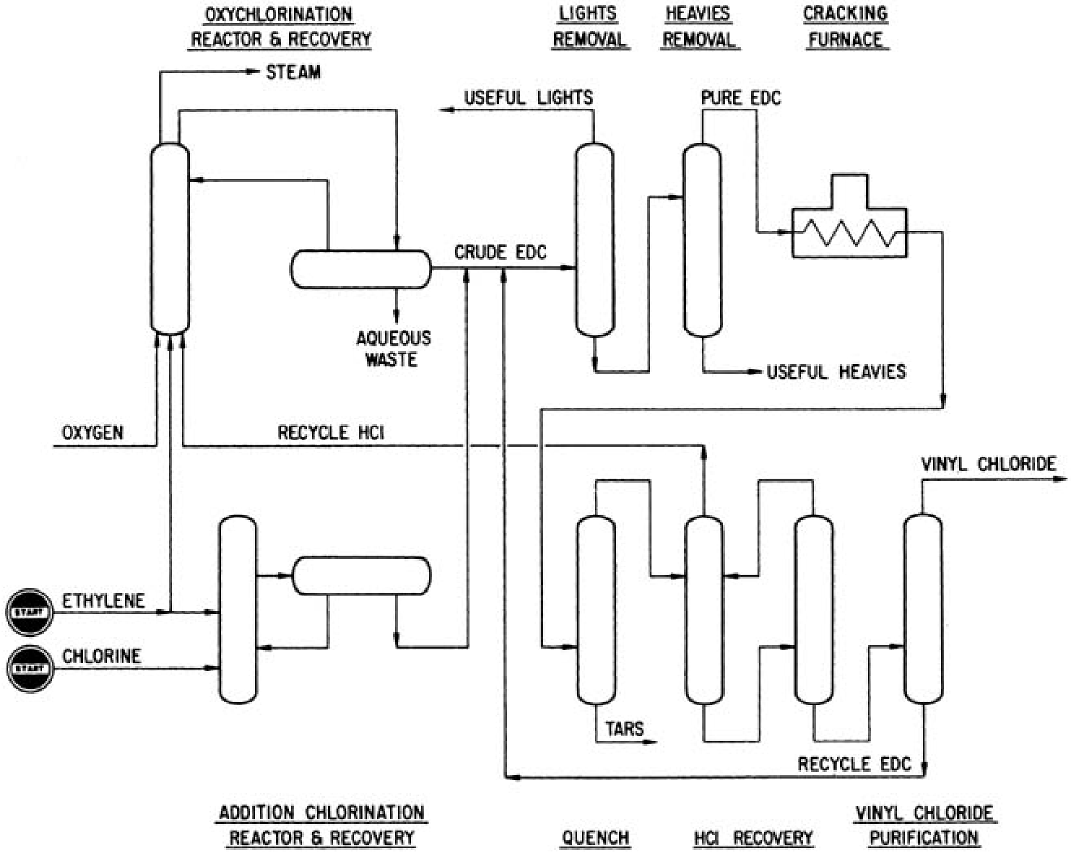
$$\mathrm{ClCH_2{-}CH_2Cl \rightarrow CH_2{=}CHCl + HCl}$$

Thus the overall reaction for the integrated plant is:

$$\mathrm{4CH_2{=}CH_2 + 2Cl_2 + O_2}$$

$$\mathrm{\rightarrow 4CH_2{=}CHCl + H_2O}$$

Fig. 22.11 Integrated EDC/vinyl chloride process. (*Hydrocarbons Processing, p. 174, 1985 November. Copyright Gulf Publishing Company and used by permission of the copyright owner*.)



The direct chlorination of ethylene usually is run in the liquid phase and is catalyzed with ferric chloride. High-purity ethylene normally is used to avoid product purification problems. The cracking (pyrolysis) of EDC to VCM typically is carried out at temperatures of 430530°C without a catalyst. The hot gases are quenched and distilled to remove HCl and then VCM. The unconverted EDC is returned to the EDC purification train. The oxychlorination step is the heart of the process and has two major variables, the type of reactor and the oxidant. The reactor may be either a fixed bed or a fluidized bed, and the oxidant is either air or oxygen. The temperature is in the range of 225275°C with a copper chloride-impregnated catalyst.

Profitable disposal of the by-product HCl once was the major restriction to the growth of EDC. The advances in the oxychlorination, which uses the HCl and air to produce ethylene dichloride, opened the door for the rapid replacement of the acetylene-based routes.

Almost 95 percent of all EDC goes to make VCM. Of that, less than 20 percent actually is isolated as EDC. Smaller uses are as a solvent and as a raw material for other chlorinated hydrocarbons such as trichloroethylene and perchloroethylene. Also a small amount is used to produce ethylene diamines.

Vinyl Chloride

Approximately 16.5 billion lb of VCM were produced in the United States in 1999, making it one of the largest-volume petrochemicals. It has been reported that more than 35 percent of the global production of chlorine goes to the manufacture of VCM. Although most of the VCM comes from EDC by the route described previously, it can be obtained from other sources, including its production in the catalytic hydrochlorination of acetylene and as a by-product in the synthesis of other chlorinated hydrocarbons.

More than 95 percent of all VCM is used to produce polyvinyl chloride (PVC), an important polymer for the housing and automotive industries. (A detailed description of PVC is included in Chapter 19.) The rest of the VCM goes into the production of chlorinated solvents.

1,1,1-Trichloroethane (Methyl Chloroform)

1,1,1-Trichloroethane was a major solvent, particularly for cold and vapor degreasing. It was phased out for emissive uses in the United States in 1996 because of its ozone depletion potential. The only application left is as chemical precursor for HCFC-141b and HCFC-142b. However, both are subject to phaseout schedule of the Montreal Protocol, and their production has been frozen at the 1996 level. The U.S. consumption has fallen from 700 million lb in 1988 to about 200 million lb in 1999.

1,1,1-Trichloroethane can be produced by three methods: by chlorination of 1,1-dichloroethane, from 1,1,2-trichloroethane via 1,1-dichloroethylene, and by direct chlorination of ethane. In the United States the first route produces about 70 percent. In this process the EDC feedstock is rearranged to 1,1-dichloroethane via cracking to VCM, followed by addition of HCl in the presence of a catalyst. For the final step, the dichloroethane is thermally or photochemically chlorinated. The reactions are as follows:

$$ClCH_2{-}CH_2Cl \rightarrow CH_2{=}CHCl + HCl$$

$$CH_2{=}CHCl + HCl \rightarrow CH_3{-}CHCl_2$$

$$CH_3{-}CHCl_2 + Cl_2 \rightarrow CH_3{-}CCl_3 + HCl$$

Chlorinated Ethylenes

VCM is by far the largest-volume chlorinated ethylene derivative. The others of commercial interest are tetrachloroethylene (perchloroethylene), trichloroethylene, and 1,1-dichloroethylene (vinylidene chloride).

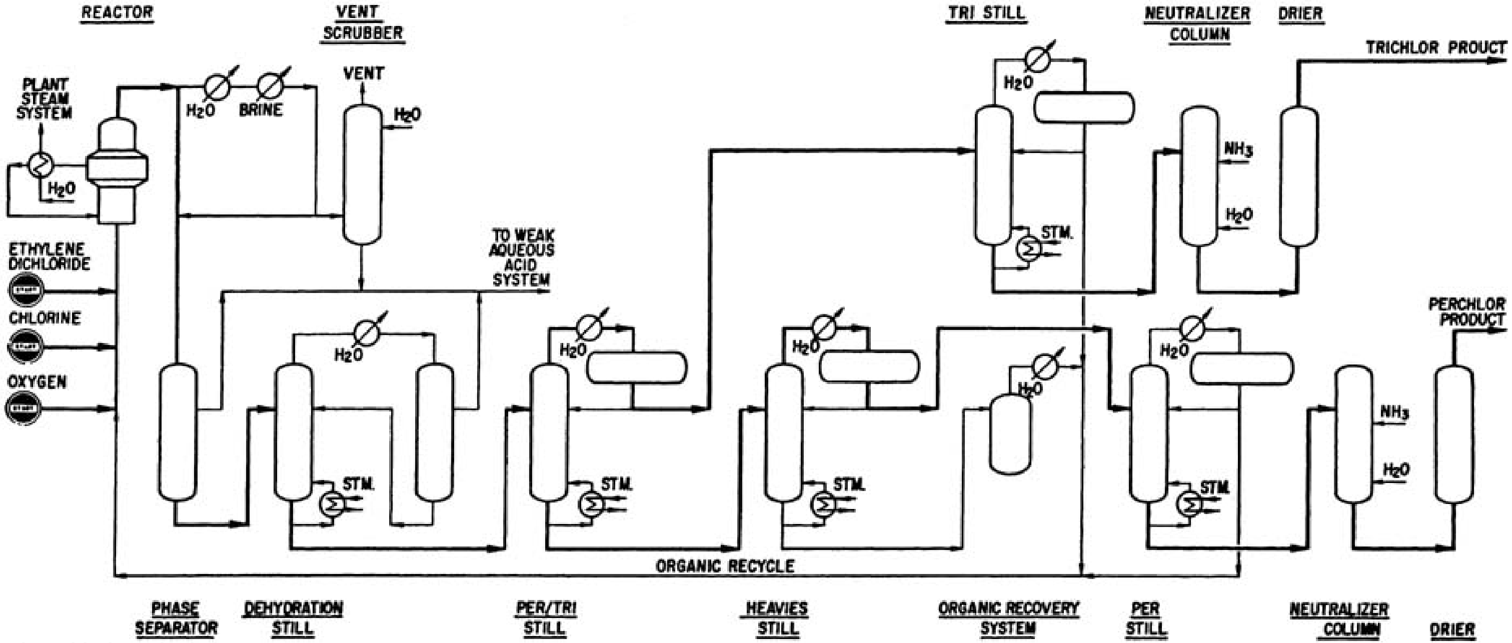
Tetrachloroethylene (Perchloroethylene)

Perchloroethylene historically has been the dominant solvent in the dry-cleaning industry because of its good stability and low flammability. Environmental concerns reduced its usage in dry-cleaning from 500 million lb in 1988 to less than 100 million lb in 1999. However, increasing quantities of perchloroethylene are being used to make alternative chlorofluorohydrocarbons, such as HCFC-123 and HCFC-134a. The total production volume in 1999 was 318 million lb.

Most perchloroethylene has been coproduced with carbon tetrachloride by the chlorination of propylene and/or chloropropanes. After the phaseout of CFC-11 and -12, the market for carbon tetrachloride disappeared. Producers have modified their units to shift the production to perchloroethylene.

An oxychlorination/oxyhydrochlorination process for the production of perchloroethylene and trichloroethylene is shown in Fig. 22.12. The process can accept a wide range of low-cost feedstocks, such as ethylene, chlorinated C2 hydrocarbons, and by-product streams from VCM, chloromethanes, methyl chloroform, and EDC plants. The product ratio of trichloroethylene to perchloroethylene can be adjusted over a wide range.

Fig. 22.12 Perchloroethylene/trichloroethylene plant. (*Reproduced from Hydrocarbon Processing, p. 154, 1985 November. Copyright by Gulf Publishing Co. and used by permission of the copyright owner.*)



Other Chlorinated Ethylenes

Trichloroethylene was a major solvent for degreasing in the late 1960s and early 1970s. Since that time, its production has decreased from 500 million lb to 100 million lb in 1993 because of environmental pressures on the solvent users and replacement by 1,1,1-trichloroethane. Recently, trichloroethylene has recovered market share in metal cleaning due to the phasing out of 1,1,1-trichloroethane in 1996. Also, the use as precursor for HFC-134a synthesis continues to increase. The production volume in 1998 was 245 million lb.

Although 1,1-dichloroethylene (vinylidene chloride) is a relatively small-volume product, it provides a way of upgrading the unwanted 1,1,2-trichloroethane by-product from the manufacture of EDC and 1,1,1-trichloroethane. Its major use is as an intermediate for polyvinylidene chloride and its copolymers, which are important barrier materials for food packaging.

Ethanol

Ethanol is made by both ethylene hydration and fermentation of starches and sugars. In this section the synthetic route will be discussed. The fermentation route is covered in Chapter 24.

In the World War II era, 72 percent of U.S. ethanol was derived from molasses fermentation. By 1978 the balance was 90 percent from direct catalytic hydration and the rest from fermentation. In 1998 the balance had returned to the dominance of fermentation, with 83 percent of the 10 billion lb of U.S. ethanol made in this way. The recent swing toward fermentation is due to the use of 90 percent of the fermentation ethanol as motor fuel, as a result of post-oil-embargo U.S. government policy.

Direct hydration of ethylene is by far the major route to synthetic ethanol. It is accomplished under pressure at 250300°C over an acidic catalyst. Ethylene and high temperature steam are mixed and passed over an acidic catalyst, usually phosphoric acid on a support. A modest conversion is achieved even with the severe conditions. Cooling of the exit stream and passage through a separations system give ethylene and water for recycle. Ethanol is made either as a 95 percent azeotrope with water or as an anhydrous material from a drying system.

Synthetic ethanol has the following uses: as a chemical intermediate (for ethyl acetate, ethyl acrylate, glycol ethers, ethylamines, etc.), 30 percent; in toiletries and cosmetics, 20 percent; as a coatings solvent, 15 percent; as a raw material for vinegar, 10 percent; in household cleaners, 7 percent; in detergents, 5 percent; in pharmaceuticals, 5 percent; in printing inks, 3 percent; and in miscellaneous uses, 5 percent.

Ethylbenzene

Ethylbenzene is used almost exclusively (99%) as a raw material for producing styrene. The remainder is used as solvent and in the manufacture of diethylbenzene. The world and U.S. demand were 44.7 and 12.6 billion lb, respectively. A growth of about 3 percent/year was expected for the next few years.

Over 90 percent of all ethylbenzene is produced by alkylation of benzene with ethylene in the presence of an acidic catalyst such as aluminum chloride or an acidic zeolite. Figure 22.13 shows a liquid phase alkylation process with zeolite catalyst.

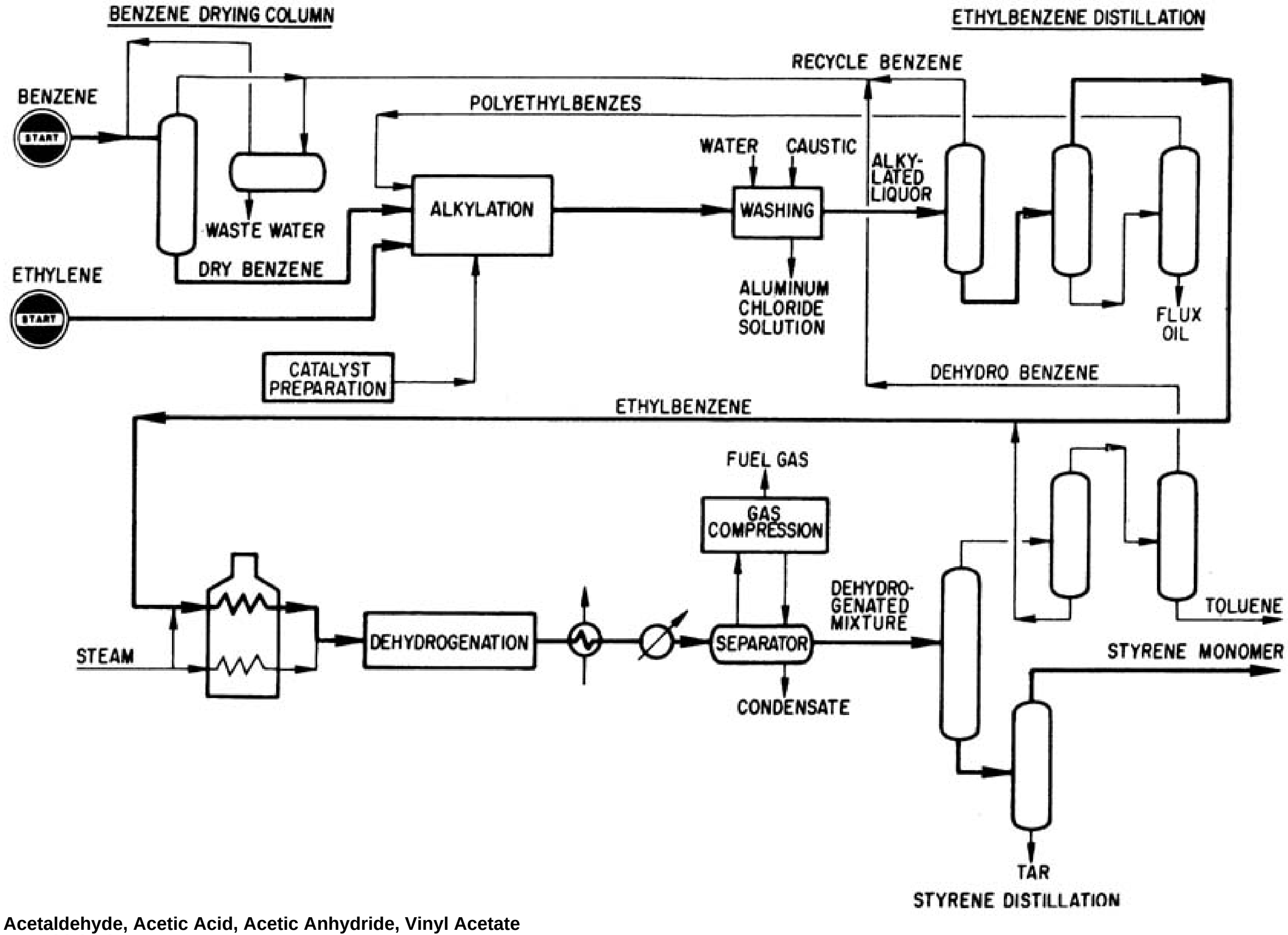
$$C_6H_6 + CH_2{=}CH_2 \rightarrow C_6H_5CH_2CH_3$$

Conversion to styrene is accomplished either by dehydrogenation:

$$C_6H_5CH_2CH_3 \rightarrow C_6H_5CH{=}CH_2 + H_2$$

or by a sequence of oxidation to ethylbenzene hydroperoxide, reduction to methyl phenyl carbinol (by a process that also oxidizes propylene to propylene oxide), and dehydration of the methyl phenyl carbinol to styrene.

Fig. 22.13 Integrated plant for manufacture of ethylbenzene and styrene. (*Reproduced from Hydrocarbons Processing, Petrochemical Handbook, p. 169, 1985 November. Copyright Gulf Publishing Co. and used by permission of the copyright owner.*)

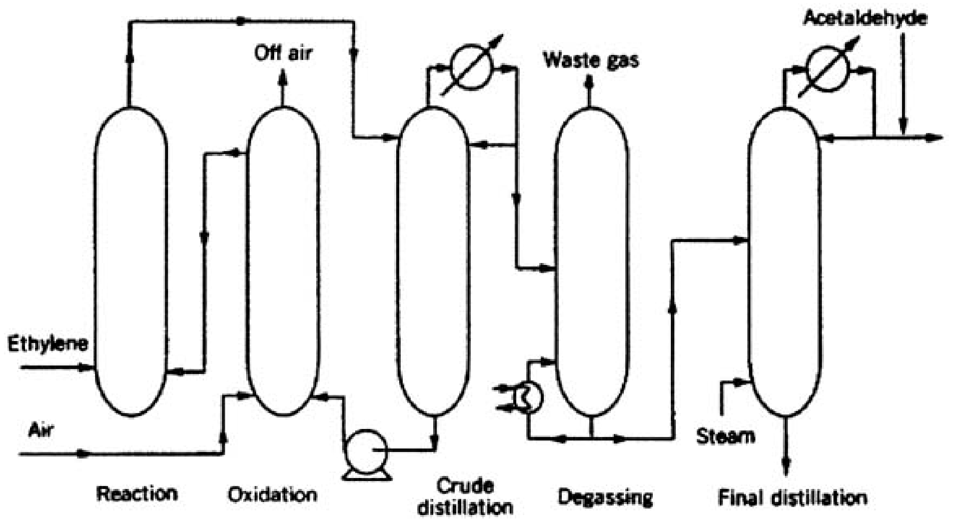


Acetaldehyde, Acetic Acid, Acetic Anhydride, Vinyl Acetate

Acetaldehyde

Acetaldehyde has been made from ethanol by dehydrogenation and by catalytic hydration of acetylene. Today direct oxidation of ethylene in the liquid phase catalyzed by palladium and copper has replaced these earlier methods. Figure 22.14 shows an ethylene-to-acetaldehyde unit based on this last route.

Fig. 22.14 Two-stage acetaldehyde process. (*Encyclopedia of Chemical Technology, Kirk and Othmer, web site ed., acetaldehyde, manufacture, 2002. Copyright by John Wiley & Sons, Inc. and reproduced by permission of the copyright owner*)

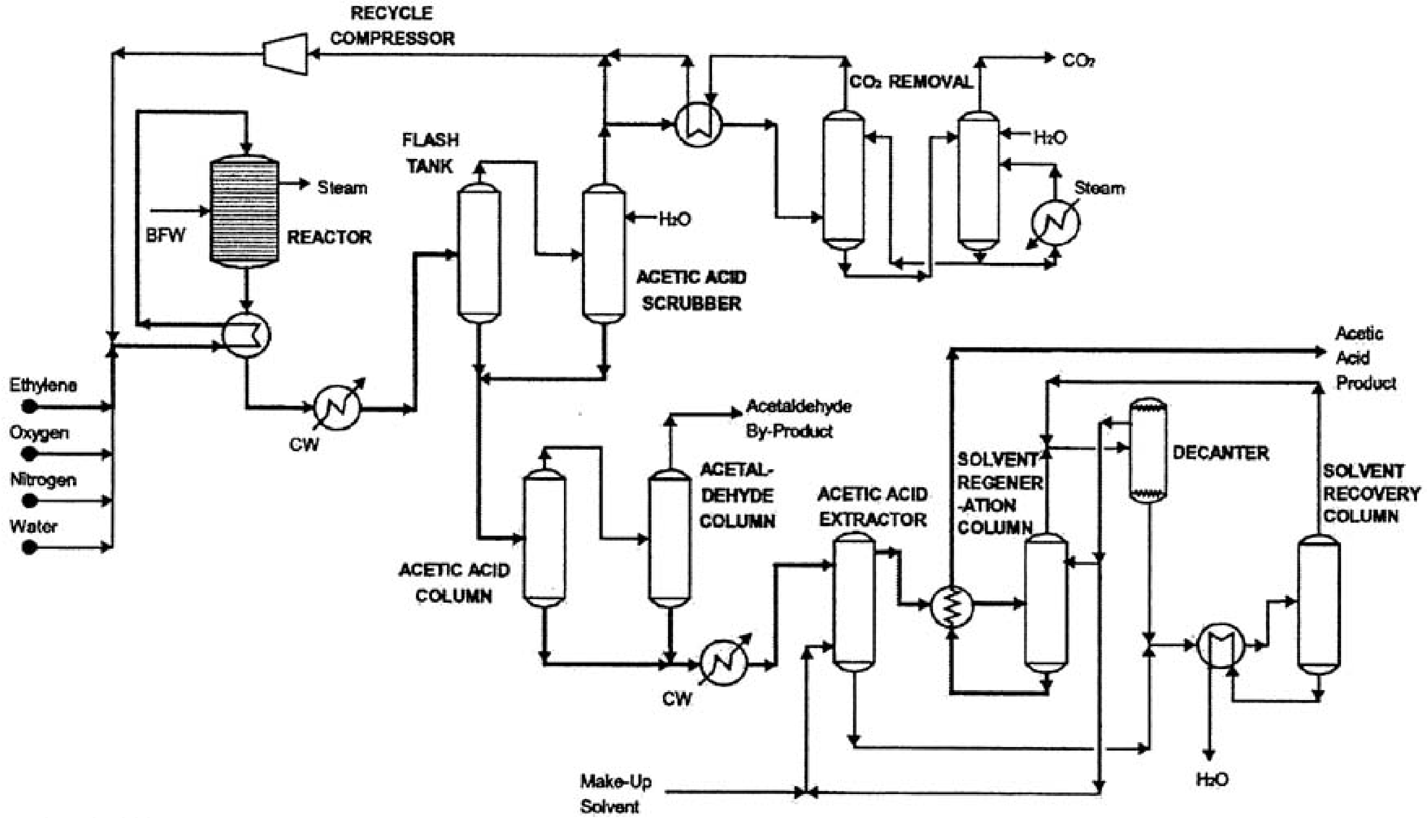


Acetaldehyde once was widely used as raw material for a variety of large-volume chemical products such as acetic acid and butanol. U.S. usage peaked in 1969 at 1.65 billion lb. Today, most of the former uses have been superseded by routes based on C1 or other chemistry such as methanol carbonylation to acetic acid and butanol from propylene by oxo chemistry. Of the remaining uses, which totaled about 400 million lb in the United States in 2000, pyridine and substituted pyridines are the major consumers at 40 percent. It is also used as a raw material for peracetic acid, pentaerythritol, and 1,3-butylene glycol.

Acetic Acid

Acetic acid used to be derived from ethylene with acetaldehyde as an intermediate. The relatively high price of acetaldehyde compared to methanol and carbon monoxide, however, caused a shift away from this route. While most acetic acid is currently produced by methanol carbonylation, as discussed earlier, a new route directly from ethylene was commercialized in 1997. This route employs a palladium-containing catalyst and combines ethylene directly with oxygen to produce acetic acid with approximately 86 percent selectivity. Figure 22.15 is a schematic diagram of the process.

Fig. 22.15 Acetic acid via direct ethylene oxidation. (*Chem Systems Report No. 99/0055. Copyright Nexant Chem Systems, Inc. and used by permission of the copyright owner*.)

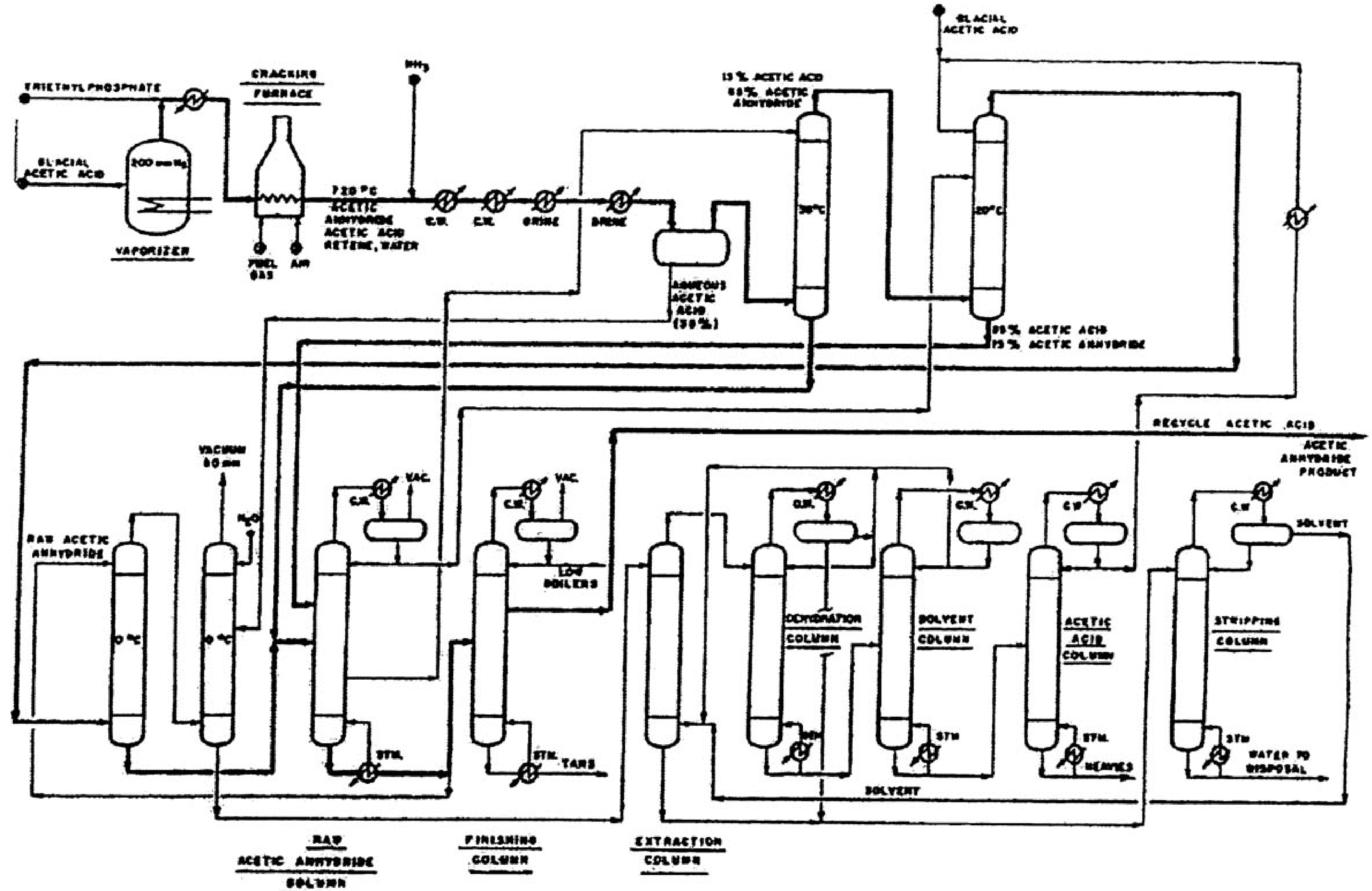


Acetic Anhydride

A total of 1.9 billion lb of acetic anhydride was produced in the United States in 1999. Commercial production of acetic anhydride is currently accomplished through two routes, one involving ketene and the other methyl acetate carbonylation. A former route based on liquid phase oxidation of acetaldehyde is now obsolete.

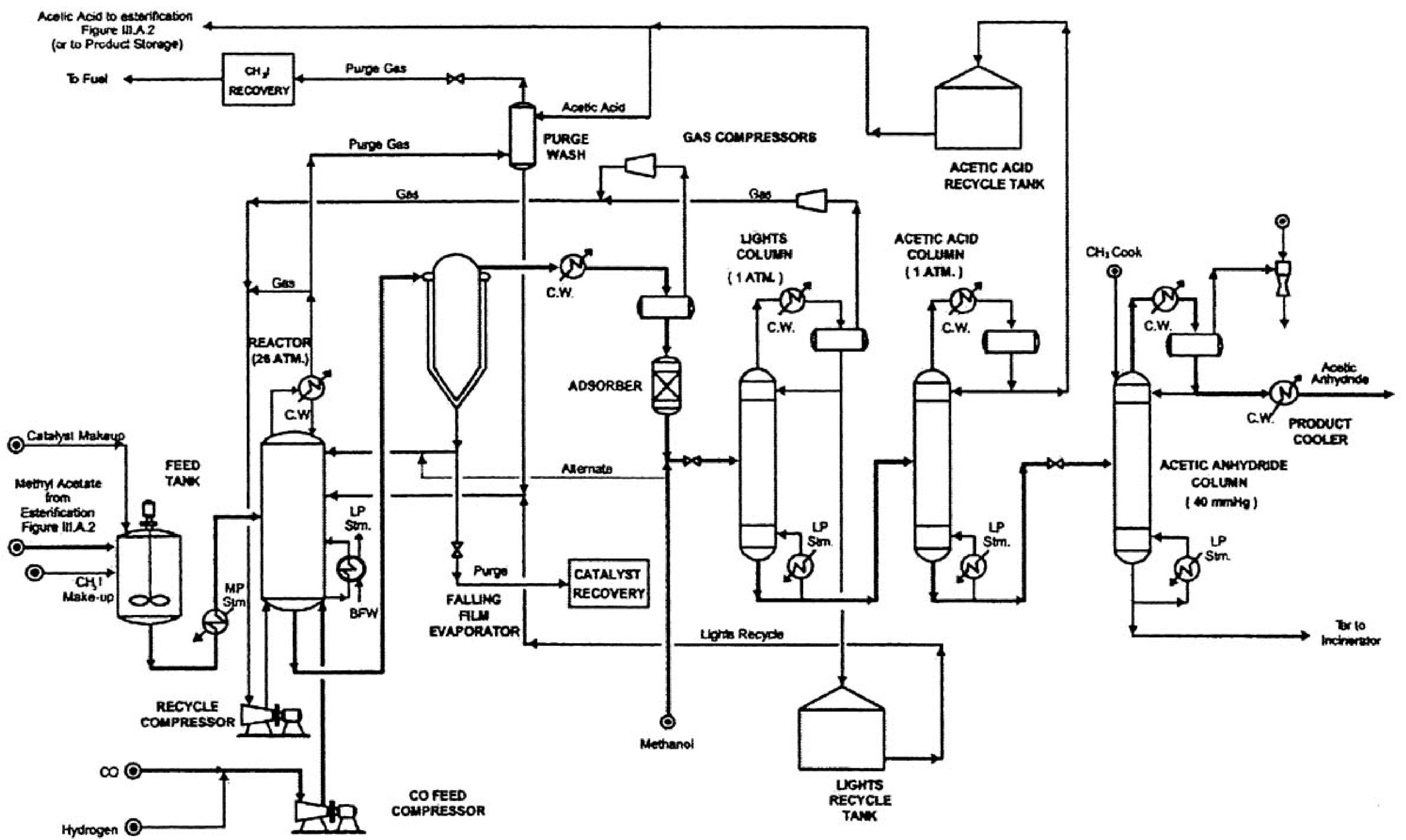
In the ketene process, acetic acid is thermally dehydrated at 750°C to ketene. The ketene is separated from by-product water and reacted with another mole of acetic acid to produce acetic anhydride. Figure 22.16 is a schematic diagram of this process.

Fig. 22.16 Acetic anhydride from acetic acid ketene process. (*Chem Systems Report No. 97/98-1. Copyright Nexant Chem Systems, Inc. and used by permission of the copyright owner*.)



The methyl acetate carbonylation process was successfully started and operated in the early 1980s. In this process, methyl acetate, itself the product of a one-step esterification of acetic acid and methanol, is reacted with carbon monoxide in the presence of a promoted rhodium-iodide catalyst. Figure 22.17 illustrates this process exclusive of esterification to make methyl acetate.

Fig. 22.17 Acetic anhydride by carbonlyation of methyl acetate. (*Chem Systems Report No. 97/98-1. Copyright Nexant Chem Systems, Inc. and used by permission of the copyright owner*.)



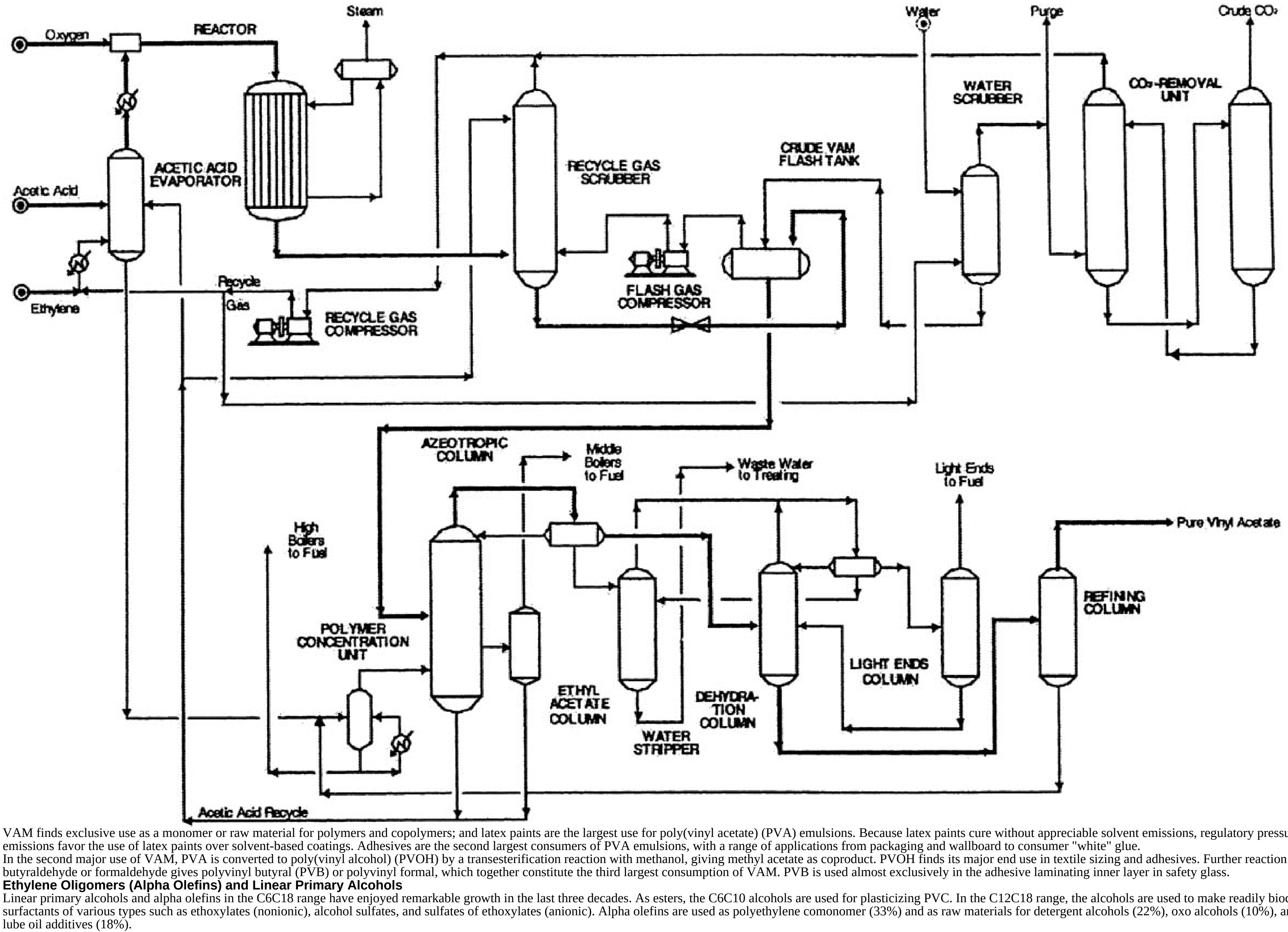
The greatest use of acetic anhydride is in esterifying cellulose to cellulose acetate for application as cigarette filter tow and in textiles. In the United States, acetic anhydride is manufactured by cellulose acetate manufacturers and largely used internally. Other products using acetic anhydride as a raw material are mostly mature with low growth rates; they include triacetin, plastic modifiers and intermediates for pharmaceuticals, herbicides, pesticides, and dyes for polyolefins. Growth in production was projected to be 0.8 percent/year in the United States through 2003.

Vinyl Acetate

Vinyl acetate (VAM, for vinyl acetate monomer) production is the largest consumer of acetic acid worldwide. In North America, vinyl acetate production in 2000 was 1.7 billion lb. Growth in North America in the period 20002005 is expected to be 1.0 percent/year and for the world, 2.4 percent/year.

Production of vinyl acetate is based primarily on vapor phase oxidative addition of acetic acid to ethylene. Figure 22.18 illustrates the process.

Fig. 22.18 Vapor phase vinyl acetate from ethylene process. (*Chem Systems Report No. 98/99-S3. Copyright Nexant Chem Systems, Inc. and used by permission of the copyright owner*.)



VAM finds exclusive use as a monomer or raw material for polymers and copolymers; and latex paints are the largest use for poly(vinyl acetate) (PVA) emulsions. Because latex paints cure without appreciable solvent emissions, regulatory pressures against such emissions favor the use of latex paints over solvent-based coatings. Adhesives are the second largest consumers of PVA emulsions, with a range of applications from packaging and wallboard to consumer "white" glue.

In the second major use of VAM, PVA is converted to poly(vinyl alcohol) (PVOH) by a transesterification reaction with methanol, giving methyl acetate as coproduct. PVOH finds its major end use in textile sizing and adhesives. Further reaction of PVOH with butyraldehyde or formaldehyde gives polyvinyl butyral (PVB) or polyvinyl formal, which together constitute the third largest consumption of VAM. PVB is used almost exclusively in the adhesive laminating inner layer in safety glass.

Ethylene Oligomers (Alpha Olefins) and Linear Primary Alcohols

Linear primary alcohols and alpha olefins in the C6C18 range have enjoyed remarkable growth in the last three decades. As esters, the C6C10 alcohols are used for plasticizing PVC. In the C12C18 range, the alcohols are used to make readily biodegradable surfactants of various types such as ethoxylates (nonionic), alcohol sulfates, and sulfates of ethoxylates (anionic). Alpha olefins are used as polyethylene comonomer (33%) and as raw materials for detergent alcohols (22%), oxo alcohols (10%), and lubricants and lube oil additives (18%).

Production of linear primary alcohols and production of alpha olefins are accomplished by similar reactions in which ethylene is oligomerized by organometallic catalysts based on aluminum alkyls such as triethylaluminum. The two processes are distinguished by the way in which the growing hydrocarbon chain is removed from the catalyst center. In the case of the alpha olefin products, ethylene growth to a hydrocarbon chain of a few or many two-carbon units is interrupted when the hydrocarbon on aluminum is displaced as an alpha olefin by exchange with fresh ethylene. Thus the aluminum alkyl is regenerated, and chain growth starts again. In production of alpha alcohols, the hydrocarbon group on the aluminum catalyst is cleaved by oxygen at the sensitive carbonaluminum bond to give aluminum oxide and an alcohol. Figures 22.19 and 22.20 illustrate production schemes for alpha alcohols and alpha olefins, respectively.

Fig. 22.19 Flow diagram of a process for primary alcohols from ethylene. (*Encyclopedia of Chemical Technology, Kirk and Othmer, web site ed., alcohols, higher aliphatic, synthetic processes, manufacture, 2002. Copyright by John Wiley & Sons, Inc. and reproduced by permission of the copyright owner*.)

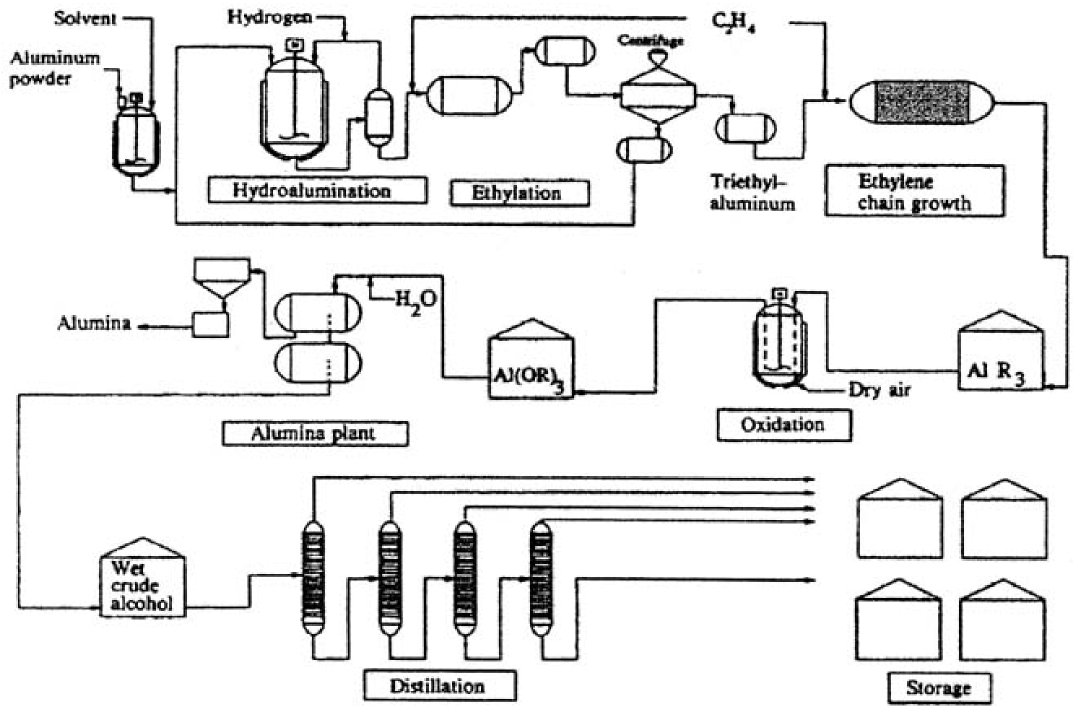
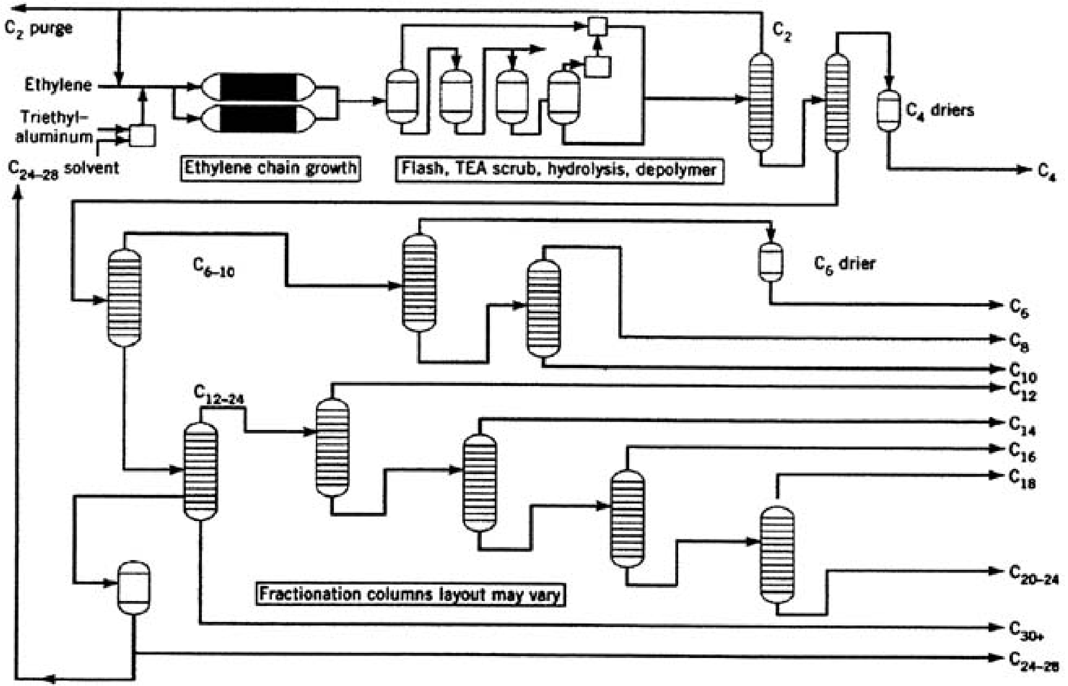


Fig. 22.20 Flow scheme of a process for alpha-olefins from ethylene. (*Encyclopedia of Chemical Technology, Kirk and Othmer, web site ed., olefins, higher, manufacture, 2002. Copyright by John Wiley & Sons, Inc. and reproduced by permission of the copyright owner*.)



Demand for alpha olefins in North America in 1999 totaled nearly 2.7 billion. The rate of growth through 2004 is expected to be 5.7 percent/year.

EthylenePropylene Elastomers

Ethylene propylene copolymer and terpolymer rubbers (EPRs) are produced at the rate in excess of two billion pounds per year worldwide. Of this, 41 percent is in North America, 27 percent in Western Europe, and 23 percent in Japan.

EPR is produced by polymerization of a mixture of ethylene and propylene and optionally a small amount of a nonconjugated diene such as ethylidene norbornene, norbornene, 1,4-hexadiene, or dicyclopentadiene. Two processes, one a solution and the other a suspension process, are employed. They use organometallic catalysts, the most common being products of combining (organo)vanadium halides with alkyl aluminum or alkyl aluminum halides. The resulting catalysts are deactivated by water and alcohols.

The comonomer diene confers sulfur vulcanizability on the elastomer. Otherwise, a peroxide cure is required for cross-linking. The polymers are readily oil-extended with 2050 percent oil for many applications. Some uses result from the ability of these products to resist oxidation by ozone.

Applications of ethylenepropylene copolymers and terpolymers include: automotive (the major use area), thermoplastic olefin elastomers, single-ply roofing, viscosity index improvers for lube oils, wire and cable insulation, hose, appliance parts, and polymer modification.

Propionaldehyde

Propionaldehyde is produced by the oxo reaction of ethylene with carbon monoxide and hydrogen. n-Propyl alcohol is produced by hydrogenation of propionaldehyde, and propionic acid is made by oxidation of propionaldehyde.

n-Propyl alcohol is used as solvent in printing inks and as an intermediate in the preparation of agricultural chemicals. Propionic acid is used as a grain preservative as, for example, in preventing spoilage of wet corn used as animal feed. The use of propionic acid as a grain preservative is an alternative to drying by heating, which consumes fuel, and is considered mostly when fuel is expensive.

Other Ethylene Uses

Some lesser-volume ethylene uses are in:

1. *agriculture*, as a ripening agent for fruits and vegetables
2. *vinyl toluene* for use in unsaturated polyester resins
3. *aluminum alkyls* used in making organometallic catalysts and as initiators for processes such as ethylenepropylene rubber, polybutadiene, low-pressure polyethylene, and ethylene oligomerization to make alpha-olefins and C6C18 alcohols
4. *diethyl sulfate* made from sulfuric acid and ethylene and used as an alkylating agent in many applications
5. *alkylation of anilines* for chemical intermediates used in pesticides, pharmaceuticals, dyes, and urethane comonomers

22.3 Chemicals Derived from Propylene

Propylene consumption for chemical synthesis in 1998 in the United States was 30 billion lb. This demand was exceeded by that of only one other synthetic organic chemical, ethylene. The demand was projected to grow at about 4.7 percent/year through 2003. Major uses of propylene are in polypropylene, acrylonitrile, propylene oxide, and cumene. A breakdown of propylene consumption by product is:

Polypropylene	45%
Acrylonitrile	12
Propylene oxide	11
Cumene	8
Oxo alcohols	8
Isopropyl alcohol	4
Oligomers	4
Acrylic acid	5
Export, other	3

Propylene is produced as a coproduct of ethylene cracking and is a product of petroleum refinery operations.

Polypropylene

More than 40 years after its introduction, polypropylene is the largest chemical consumer of propylene in the United States. It is produced primarily by a bulk or gas phase process, with the older slurry process still used by some. Because of its greater stiffness relative to polyethylene, polypropylene is used for more demanding applications. This stiffness is not without a drawback in the form of increased brittleness, which can be moderated through incorporation of ethylene by copolymerization. Its major uses are in fibers and injection molding. Chapter 19 reviews polypropylene in some detail.

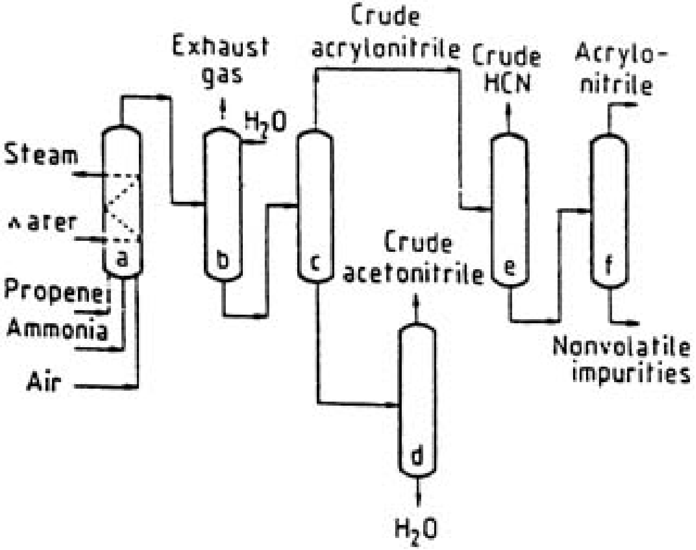
Acrylonitrile

In 1960, almost all of the 260 million lb annual production of acrylonitrile was based on acetylene. Ten years later, the volume had risen to 1.1 billion lb, which was based almost entirely on an ammoxidation process with ammonia, propylene, and air as feeds. However, in the latter 1980s the growth rate had slowed considerably.

The air, ammonia, propylene process for acrylonitrile is shown in Fig. 22.21. The main reaction is given below.

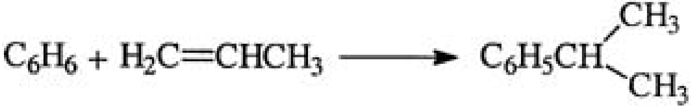
$$2H_2C{=}CHCH_3 + 2NH_3 + 3O_2$$
$$\rightarrow 2H_2C{=}CHCN + 6H_2O$$

Fig. 22.21 Simplified diagram of the Sohio acrylonitrile process. (a) Fluidized-bed reactor; (b) absorber column; (c) extractive distillation column; (d) acetonitrile stripping column; (e) lights fractionation column; (f) product column. (*Ullmann's Encyclopedia of Industrial Chemistry, W. Gerhartz (Ed.), 5th ed., Vol. Al, p. 179. VCH Verlagsgesellschaft, Weinhein, Federal Republic of Germany, 1985.*)



In this process the highly exothermic oxidation is performed in a fluidized bed to facilitate heat removal. Note also that acetonitrile and hydrogen cyanide are by-products. In the case of hydrogen cyanide, this source is of major commercial importance.

Acrylic fibers are by far the major end use for acrylonitrile. They find use primarily in fabrics for clothing, furniture, draperies, and carpets. The second largest consumer of acrylonitrile is acrylonitrilebutadienestyrene (ABS) and styrene acrylonitrile (SAN) resins. ABS is useful in industrial and construction applications, and the superior clarity of SAN makes it useful in plastic lenses, windows, and transparent household items.



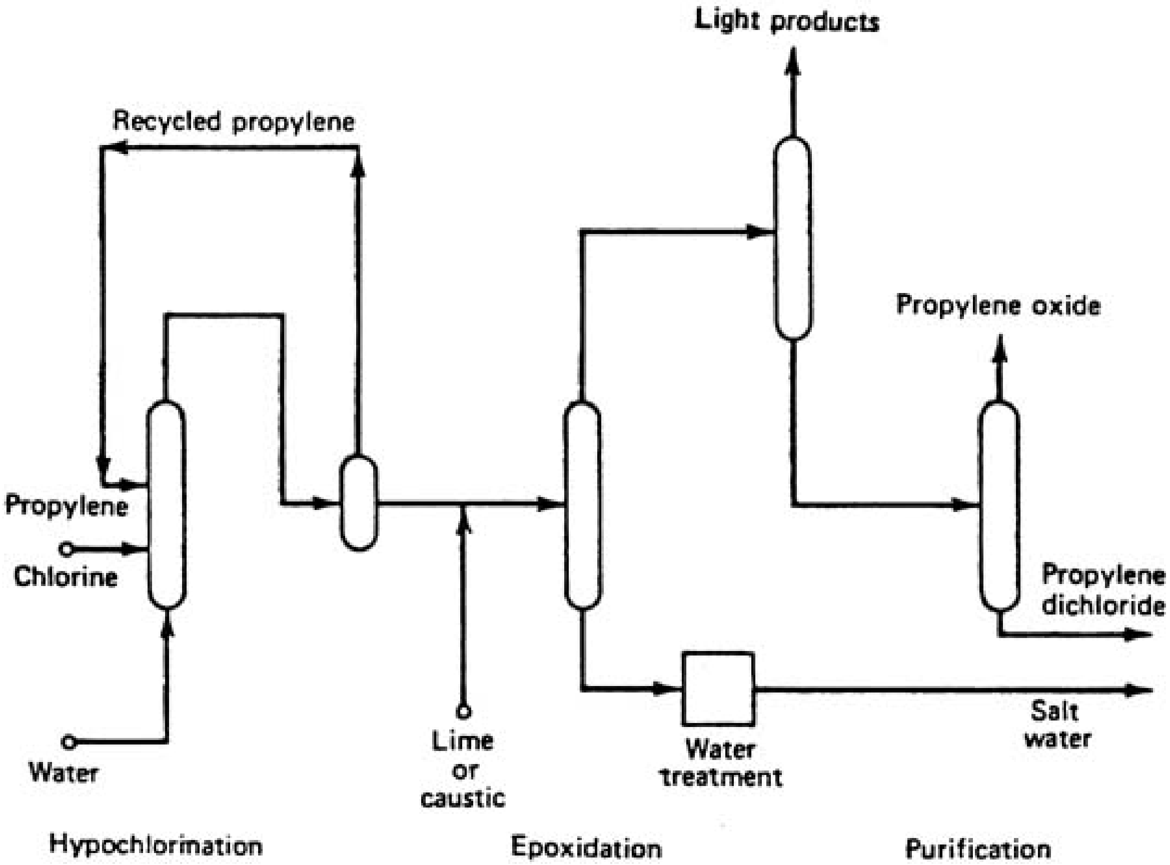
$$C_6H_6 + H_2C{=}CHCH_3 \longrightarrow C_6H_5CH(CH_3)_2$$

Propylene Oxide

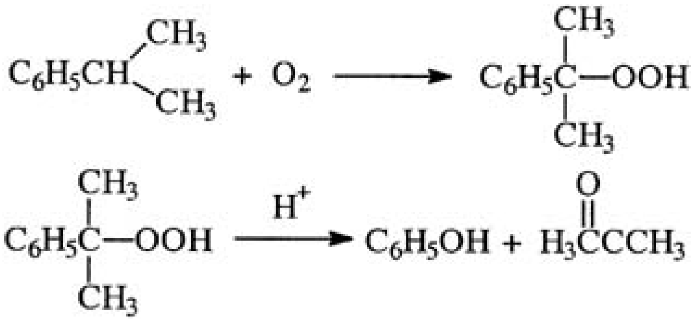
Propylene oxide (PO) is one of most important organic chemical from the propylene family. The global capacity was almost 11 billion lb in 1998. The PO production in the United States reached 4.2 billion lb in 1998. Polyurethane polyether polyols are the largest usage, which contributed to about 60 percent of domestic PO consumption. It is also used to make propylene glycol, glycol ethers, polyglycols, glycerine, surfactants, and amino propanols.

PO was manufactured by the chlorohydrin route first during World War I in Germany by BASF and others. This route (below) involves reaction of propylene with hypochlorous acid followed by treatment of the resulting propylene chlorohydrin with a base such as caustic or lime. The products of the second reaction are PO and sodium or calcium chloride (Fig. 22.22).

Fig. 22.22 Chlorohydrin process. (*Encyclopedia of Chemical Technology, Kirk and Othmer. 3rd ed., Vol. 19, p. 255, 1980. Copyright by John Wiley & Sons, Inc. and used by permission of the copyright owner.*)



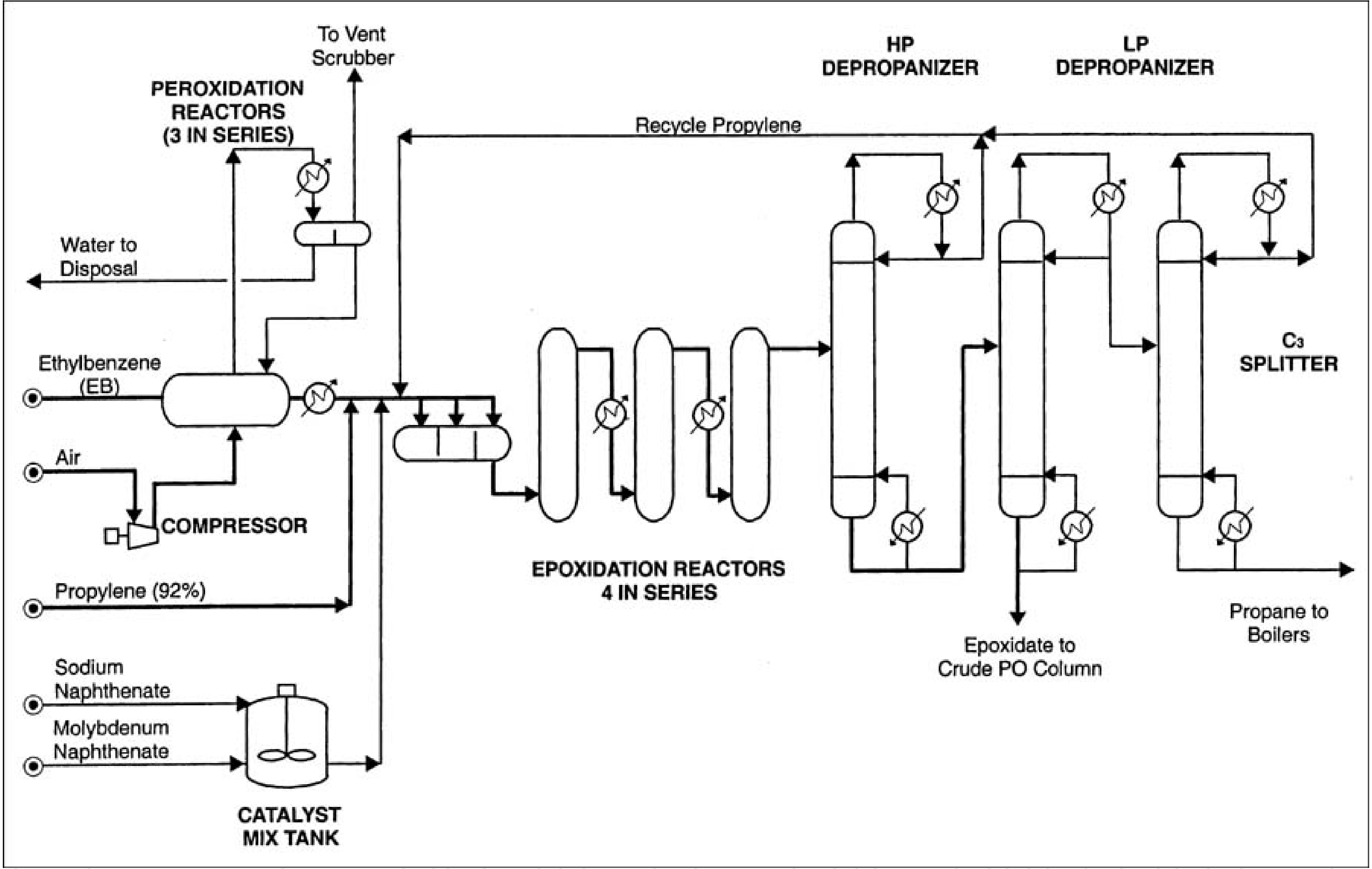
Until 1969, the chlorohydrin process was the only PO process, and The Dow Chemical Company was the largest producer. In that year, Oxirane brought on stream the first peroxidation process involving catalyzed epoxidation of propylene with tert-butyl hydroperoxide. In 1977, Oxirane (later Arco Chemical) commercialized a process which employed ethylbenzene hydroperoxide as the epoxidizing agent and produced PO and styrene (Fig. 22.23).



$$C_6H_5CH(CH_3)_2 + O_2 \longrightarrow C_6H_5C(CH_3)_2{-}OOH$$

$$C_6H_5C(CH_3)_2{-}OOH \xrightarrow{H^+} C_6H_5OH + H_3CC(=O)CH_3$$

Fig. 22.23 Styrene/PO via peroxidation and epoxidation. (*Chem Systems Report No. 97/98-7. Copyright Nexant Chem Systems, Inc. and used by permission of the copyright owner*.)



The peroxide processes convert propylene to its epoxide while reducing the hydroperoxide to the corresponding alcohol (e.g., tert-butyl alcohol or phenyl methyl carbinol). Because the processes produce the alcohols in larger amounts than PO, their success depends upon finding uses for the alcohols. tert-Butyl alcohol can be dehydrated to isobutylene and hydrogenated to isobutane for recycle to the PO process. It can also be converted to MTBE. Phenyl methyl carbinol can be dehydrated to styrene, making this process a more involved conversion of ethyl benzene to styrene than is direct dehydrogenation.

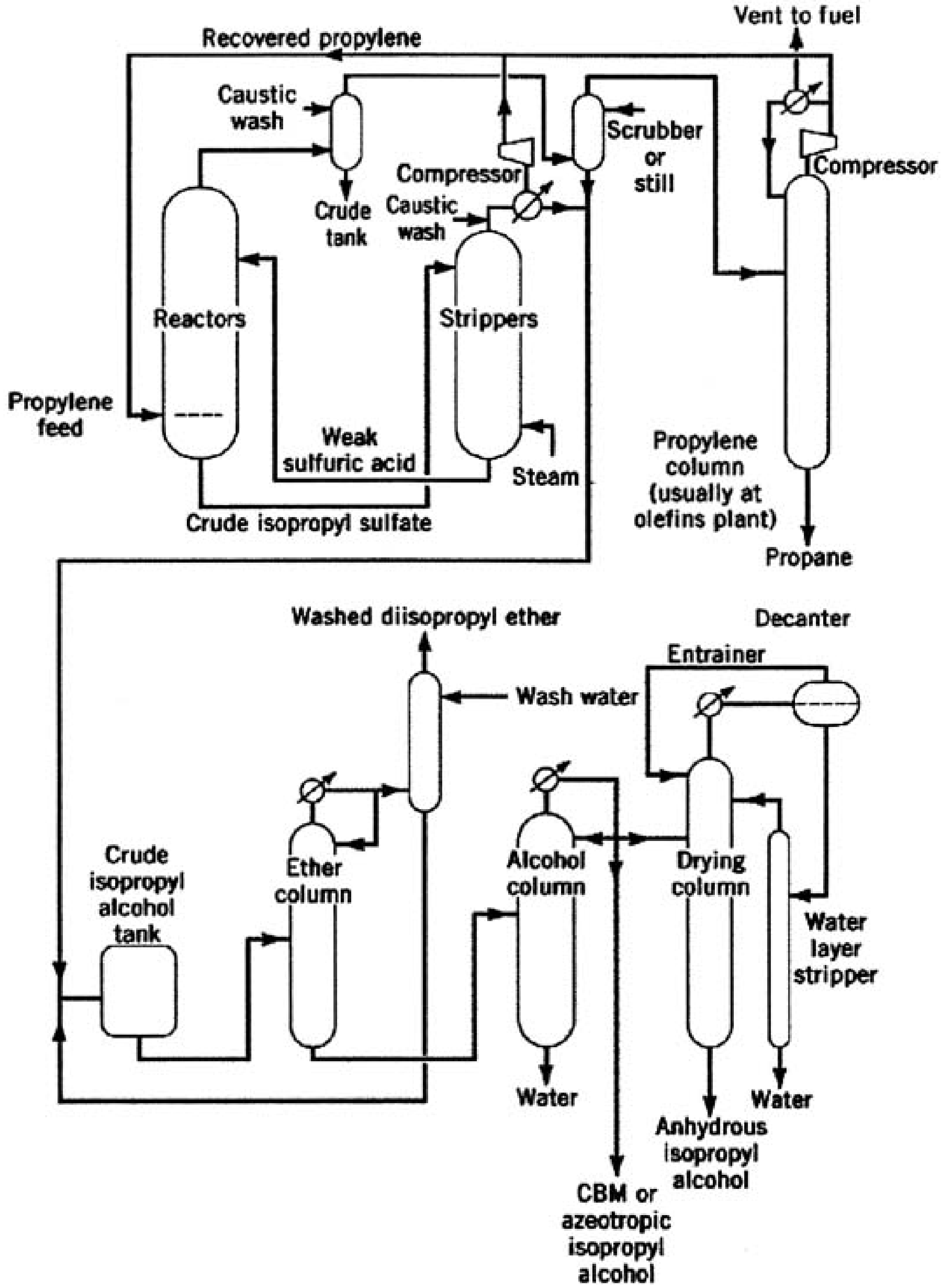
Propylene Glycols

PO is converted to mono-, di-, and tri-glycols by a hydrolysis. It is similar to the hydrolysis of ethylene oxide to mono- and di-ethylene glycol. The propylene glycols are used for many of the same applications as the corresponding products derived from ethylene oxide. Because of their very low toxicity, they also can be used for pharmaceutical, cosmetic, food applications, liquid detergent, tobacco humectant, deicing fluid and antifreezes. In 1999, the U.S. consumption for mono-, di-, and tri-propylene glycols were 1.08 billion lb, 125, and 16 million lb, respectively.

Isopropyl Alcohol

Isopropyl alcohol (IPA) has been called the first petrochemical. Both historically and today, it is prepared by sulfuric acid-mediated indirect hydration of propylene (see Fig. 22.24). Originally it was the source of most of the acetone used in the world. Now, this route must compete with acetone derived from the cumene oxidation process, in which cumene is converted to equimolar amounts of phenol and acetone. The amount of IPA used for producing acetone declined from 47 percent in 1978 to 7 percent in 2001. IPA can also be made by hydrogenation of acetone, but the large capacity for IPA by indirect hydration is a disincentive for this application. Direct hydration of propylene in a vapor-phase, catalytic process is also commercially practiced. This is similar to hydration of ethylene to make ethanol. Relative to the sulfuric acid-mediated process, it offers the advantage of decreased corrosion. However, it suffers from a requirement for a pure propylene feed, whereas the former process can be used with a dilute, refinery stream.

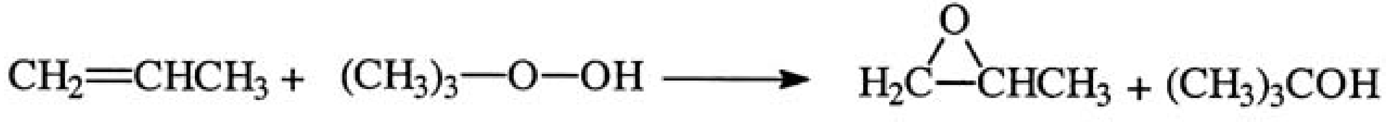
Fig. 22.24 Flow scheme of a process for isopropanol. (*Encyclopedia of Chemical Technology, Kirk and Othmer, web site ed., isopropyl alcohol, manufacture, 2002. Copyright by John Wiley & Sons, Inc. and reproduced by permission of the copyright owner*.)



Isopropyl alcohol is an excellent solvent with a blend of polar, nonpolar, and hydrogen-bonding character that makes it useful in a broad spectrum of applications. Its moderate volatility makes it convenient for uses involving evaporation or recovery by distillation. Thus, it is no surprise that much of IPA's consumption is for solvent uses. In North America in 1999 about 1.2 billion lb of IPA were consumed. Major uses were: solvent applications, 47 percent; isopropylamines, 15 percent, esters and ketones, 20 percent; and others, including pharmaceuticals, 18 percent. The total demand in 1999 was significantly less than the 1.9 billion lb recorded for 1978. This downward trend is not unique to IPA and is primarily a result of regulatory pressure in the United States to decrease emissions of VOCs in coating and other applications.

Cumene

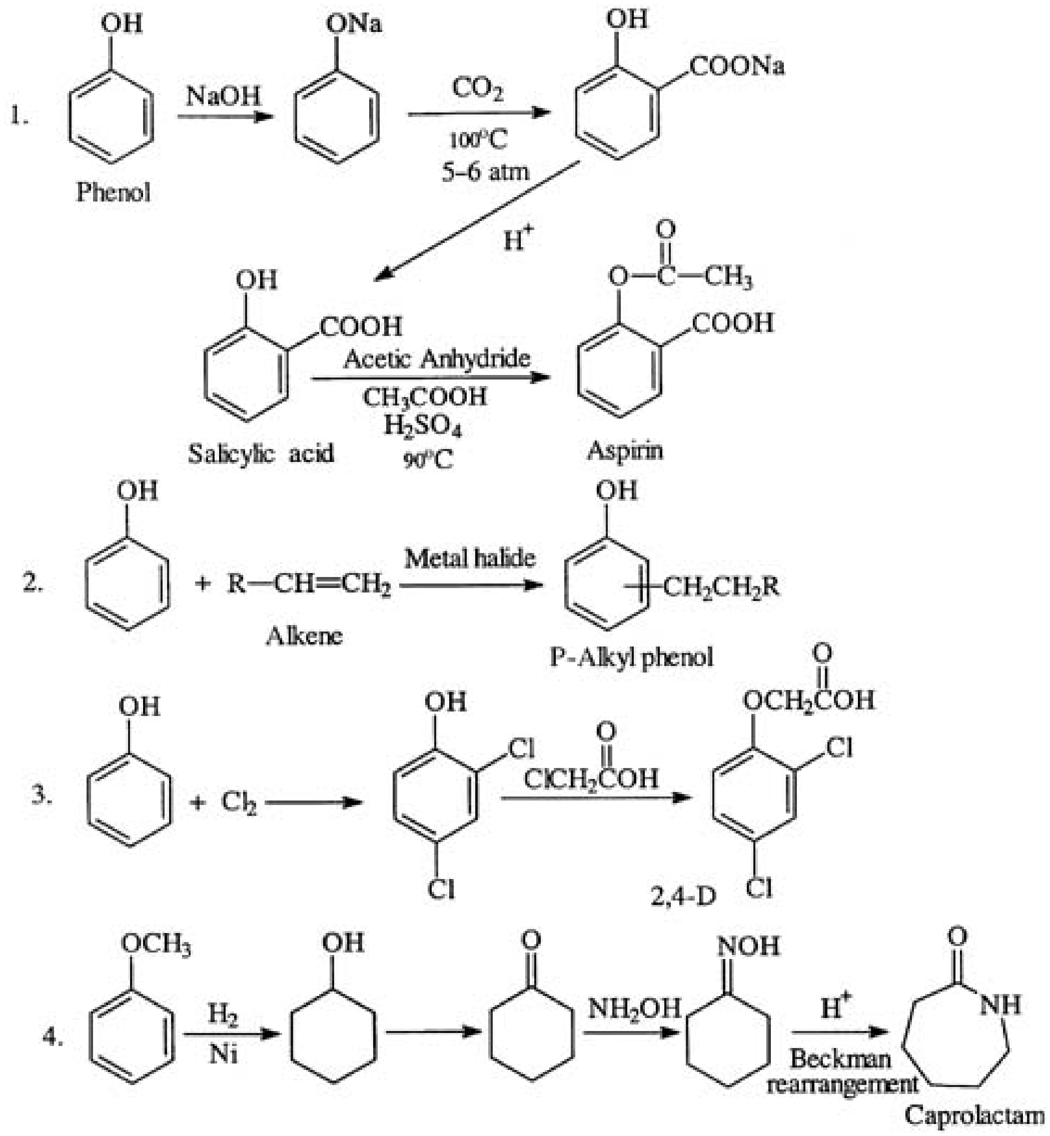
Cumene manufacture consumed about 10 per-cent (2.2 billion lb) of the propylene used for chemicals in the United States in 1998. It is prepared in near stoichiometric yield from propylene and benzene with acidic catalysts (scheme below). Many catalysts have been used commercially, but most cumene is made using a "solid phosphoric acid" catalyst. Recently, there has been a major industry shift to zeolite-based catalyst. The new process has better catalyst productivity and also eliminates the environmental waste from spent phosphoric acid catalyst. It significantly improves the product yield and lowers the production cost. Cumene is used almost exclusively as feed to the cumene oxidation process, which has phenol and acetone as its coproducts.



$$\underset{\text{propylene}}{CH_2{=}CHCH_3} + \underset{\text{t-butyl hydroperoxide}}{(CH_3)_3{-}O{-}OH} \longrightarrow \underset{\text{propylene oxide}}{H_2\overset{O}{C{-}}CHCH_3} + \underset{\text{TBA}}{(CH_3)_3COH}$$

Acetone

Acetone in commerce is derived mostly from cumene oxidation. This is a two-step process involving oxidation of cumene to the hydroperoxide followed by acid catalyzed decomposition to acetone and phenol:



The 1998 U.S. use of acetone was about 2.6 billion lb. Its major uses are methacrylic acid and esters (44%); solvent (17%); bisphenol-A (20%); and aldol chemicals (such as methyl isobutyl ketone) (13%).

Oxo Chemicals

About 8 percent of the propylene converted into chemicals is used to make oxo alcohols such as 1-butanol and 2-ethylhexanol, which are called oxo alcohols because they are derived from olefins by the oxo process, which converts them to aldehydes. (The oxo process was described earlier in this chapter.)

Butyl Alcohols and Aldehydes

Hydroformylation of propylene gives a mixture of n-butyraldehyde and isobutyraldehyde. This mixture is formed approximately in the ratio of 2:1 from the high-pressure, cobalt-catalyzed oxo process. There has always been a much greater demand for the linear n-butyraldehyde than the iso product, so it has been necessary to find uses for the latter.

The low-pressure, rhodium-catalyzed oxo process has made the product mix conform to the relative demand for the two aldehydes. This process gives a 10:1 ratio of n-butyraldehyde to isobutyraldehyde.

Each aldehyde can be hydrogenated to the corresponding alcohol for use as a solvent or an intermediate for plasticizers and resins. n-Butyraldehyde is also converted to 2ethylhexanol by sequential condensation and hydrogenation. 2-Ethylhexanol is used to make the phthalate ester, which finds wide use as a plasticizer of PVC.

Propylene Oligomers: Dodecene and Nonene

The manufacturing processes for these materials are very similar to the one for cumene. When nonene is the desired product, additional fractionation is required, the extent of which is determined by product specifications.

In the reactor portion of this process, the olefin stock is mixed with benzene (for cumene) or recycle lights (for tetramer). The resulting charge is pumped to the reaction chamber. The catalyst, solid phosphoric acid, is maintained in separate beds in the reactor. Suitable propane quench is provided between beds for temperature control purposes because the reaction is exothermic.

Dodecene is an intermediate for surfactants, mainly through two routes. One, the larger user, produces dodecylbenzene sulfonate for anionic detergents. The other goes through the oxo process to tridecyl alcohol, which then is converted into a nonionic detergent by the addition of alkylene oxides.

Nonene has two major outlets, the larger being the oxo production of decyl alcohol which is used in the manufacture of esters, and so forth, for plasticizers. The other significant use for nonene is in the manufacture of nonylphenol, an intermediate for the important series of ethoxylated nonylphenol nonionic surfactants.

Acrylic Acid and Esters

Acetylene once was the raw material for commercial production of acrylic acid and esters, but in 1970 production of acrylic acid by oxidation of propylene was first practiced commercially. In a few years, the new process had essentially replaced the old. In 2000, acrylic acid production in the U.S. was of the order of 2.0 billion lb, and that of acrylate esters was of the order of 1.8 billion lb.

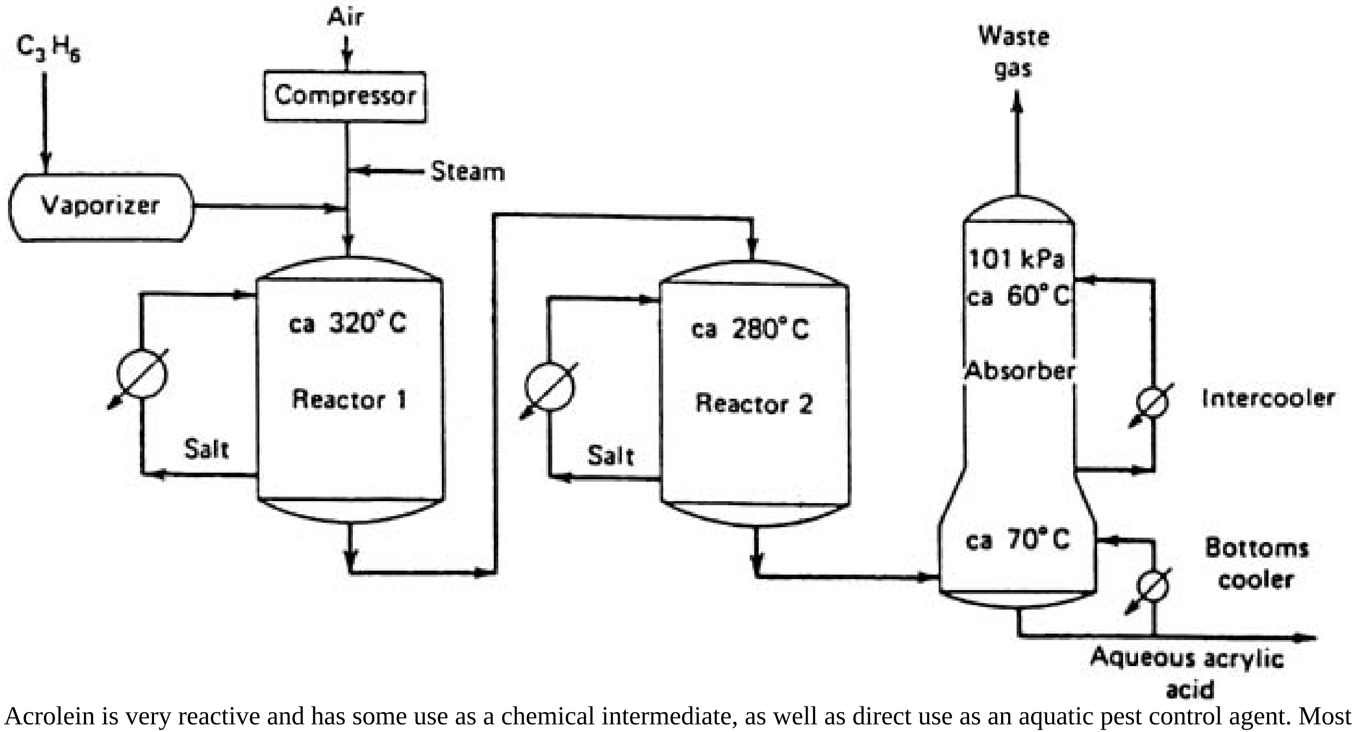
The oxidation of propylene is carried out in two stages (Fig. 22.25). Acrolein exiting the first-stage converter can be isolated, or it can be further oxidized to acrylic acid in the second converter. The process is operated with two reaction stages to allow optimum catalyst and process conditions for each step.

$$\mathrm{H_2C{=}CHCH_3 + O_2 \rightarrow H_2C{=}CHCHO + H_2O}$$

$$\mathrm{H_2C{=}CHCHO + \tfrac{1}{2}O_2 \rightarrow H_2C{=}CHCO_2H}$$

$$\mathrm{H_2C{=}CHCO_2H + ROH}$$
$$\mathrm{\rightarrow H_2C{=}CHCO_2R + H_2O}$$

Fig. 22.25 Acrylic acid from oxidation of propylene. (*Encyclopedia of Chemical Technology, Kirk and Othmer, 3rd ed., Vol. 1, p. 339, 1980. Copyright by John Wiley & Sons, Inc. and reproduced by permission of the copyright owner.*)



Acrolein is very reactive and has some use as a chemical intermediate, as well as direct use as an aquatic pest control agent. Most acrolein is converted without isolation to acrylic acid.

Acrylic acid is almost exclusively used directly, or after conversion to an ester, as a monomer. Acrylate esters are produced by normal esterification processes. However, in dealing with acrylic acid, acrolein, or acrylates, unusual care must be taken to minimize losses due to polymerization and other side reactions such as additions of water, acids, or alcohols across the reactive double bond. Polyacrylic acids find use in superabsorbers, dispersants, and water treatment. The polyesters are used in surface coatings, textile fibers, adhesives, and various other applications.

Epichlorohydrin

Epichlorohydrin (ECH) is made from propylene, the majority is via allyl chloride intermediate. Total consumption in 1999 was about 600 million lb. Uses for ECH include epoxy resins (65%), synthetic glycerin (22%), and others (paper treatment, specialty ionic exchange resin, glycerol, and glycidol derivatives).

The key reaction in this manufacturing process is the hot chlorination of propylene, which fairly selectively gives substitution to methyl group rather than the addition to the double bond. In this chlorination step, fresh propylene is first mixed with recycle propylene. This mixture is dried over a desiccant, heated to 650700°F, and then mixed with chlorine (C_3H_6 to Cl_2 ratio is 4:1) and fed to a simple steel tube adiabatic reactor. The effluent gases (950°F) are cooled quickly to 120°F and fractionated. The yield of allyl chloride is 8085 percent.

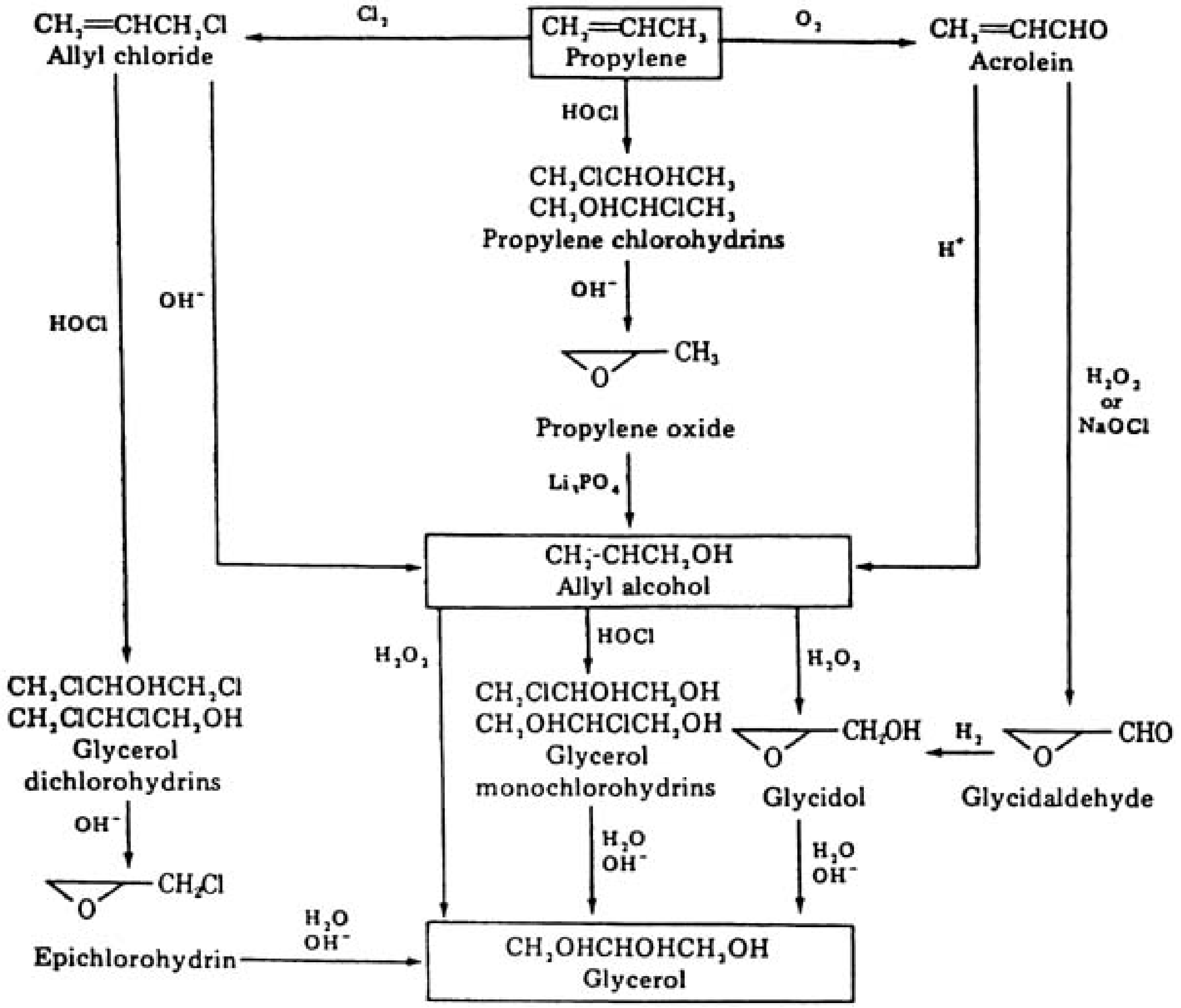
Hypochlorous acid is then reacted with the allyl chloride at 85100°F to form a mixture of dichlorohydrins. The reactor effluent is separated, the aqueous phase is returned to make up the hypochlorous acid, and the nonaqueous phase containing the dichlorohydrins is reacted with caustic or a lime slurry to form ECH which is steam-distilled out and given a finishing distillation.

ECH is used to manufacture epoxy resins for surface coating, castings, and laminates. It is hydrolysed in 10 percent caustic to make synthetic glycerin (see "Glycerin"). ECH is also employed as a raw material for the manufacture and glycidol derivatives used as plasticizers, stabilizers, surface active agents, and intermediates for further synthesis. The polyamide/ECH resin (which is used in the paper industry to improve the wet strength) has had very good growth in the past few years. The average growth rate for ECH for the period 19901999 is about 3 percent and the trend is expected to continue.

Glycerin

Glycerin can be prepared from propylene (via ECH) or as a by-product from fat and oil hydrolysis of the soap industry. Before 1949 all glycerin was obtained from hydrolysis of fatty triglycerides. In the past 50 years, the synthetic glycerin is to serve the portion of demand not satisfied by natural glycerin. In 1998, the U.S. production for natural versus synthetic is about 2.4:1. The diagram of different routes for the manufacture of glycerin is in Fig. 22.26.

Fig. 22.26 Routes for the manufacture of glycerin. (*Encyclopedia of Chemical Technology, 3rd ed., Vol. 11, p. 923, 1980. Copyright by John Wiley & Sons, Inc. and reproduced by permission of the copyright owner*.)



In 1998, glycerin consumption in the United States was about 380 million lb. The average annual growth for 19901998 is about 2.3 percent/year. About 80 percent of U.S. usage of glycerin is in foods, pharmaceuticals, personal care, cosmetics, tobacco, and similar applications. This reflects its extremely low toxicity, sweet taste, and moisturizing and lubricating properties. Chemical uses for glycerin include use as a "starter" alcohol for polyols made by alkoxylation with propylene oxide and ethylene oxide, and as raw material for alkyd polymers, plasticizers, and explosives.

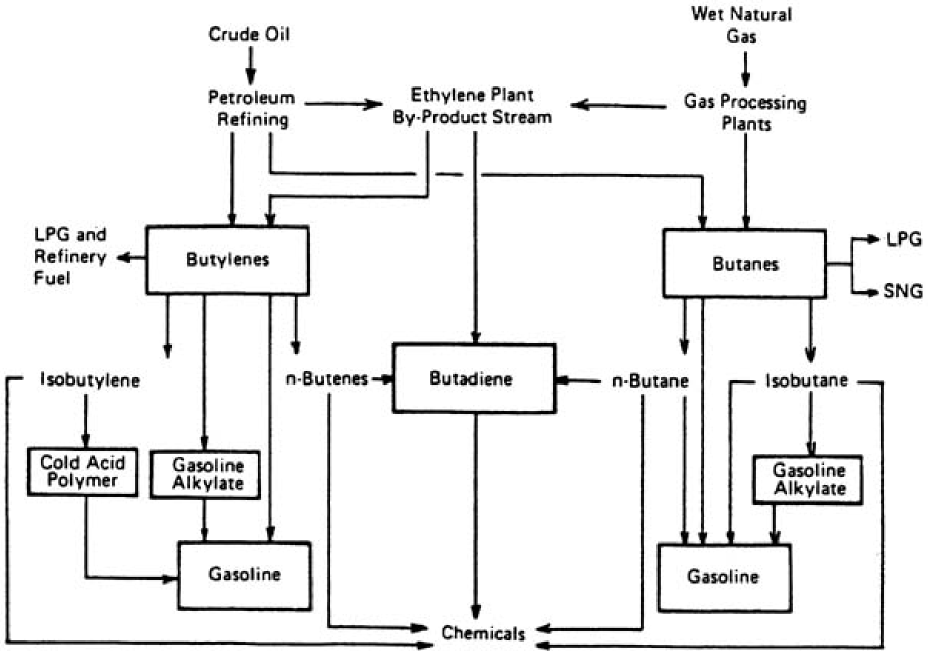
Glycerin by the Epichlorohydrin Process

In the ECH process, synthetic glycerin is produced in three successive operations, the end products of which are allyl chloride, ECH, and finished glycerin, respectively. Glycerin is formed by the hydrolysis of ECH with 10 percent caustic. Crude glycerin is separated from this reaction mass by multiple-effect evaporation to remove salt and most of the water. A final vacuum distillation yields a 99+ percent product.

22.4 Chemicals Derived from Butanes and Butylenes

Saturated four-carbon hydrocarbons (butanes) occur in natural petroleum products such as crude oil and the heavy vapors in wet natural gas. The saturated C4s are also produced from other hydrocarbons during the various petroleum refining processes. The butylenesunsaturated C4sdo not occur in nature, but are derived from butanes or other hydrocarbons either deliberately or as by-products. The complex interrelationships of C4 hydrocarbons, including their production and use, are described in Fig. 22.27.

Fig. 22.27 Origins, interrelationships, and end uses of C4 hydrocarbons. (*Reproduced from Chemical Economics Handbook, p. 300200A*, Stanford Research Institute, Menlo Park, CA, 1980, March.)

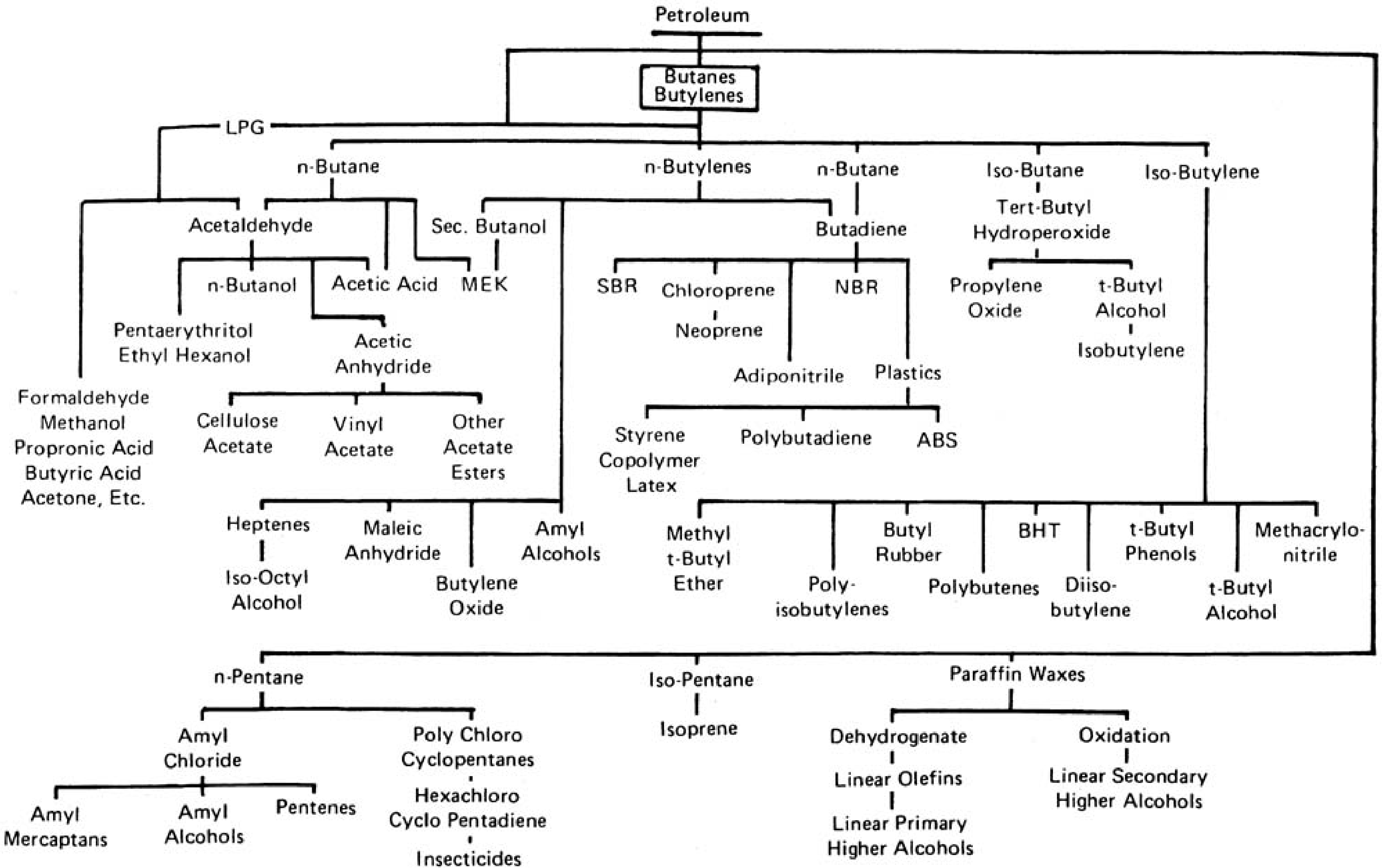


The chemical uses of the C4 hydrocarbons still account for only a small fraction of the available material. To put the volume of C4s used in chemical manufacture in perspective with the amount used for fuel, one finds that approximately 12 percent of the butanes and about 30 percent of the butylenes were used as chemical raw materials. The trends that affect availability of C4 hydrocarbons for chemical and energy end uses are determined by the natural gas processors, petroleum refiners, and, to a growing extent, ethylene manufacture.

Changes in technology and in the availability of optimum feedstocks have far-reaching effects on the entire product mix. For example, when the availability of LPG and ethane for ethylene manufacture has decreased, n-butane and the higher crude cuts have been used, and the proportion of by-product buta-diene has increased.

The spectrum of products which can be derived from the four-carbon hydrocarbons is shown in Fig. 22.28. Several of these can also be produced from other raw materials and are described elsewhere in this chapter.

Fig. 22.28 Chemicals from butanes, butylenes, LPG, and higher aliphatic hydrocarbons.



n-Butane Derivatives

n-Butane can be obtained from natural gas and from refinery hydrocracker streams. Most of the n-butane goes into fuel additive uses. The major chemical use is as a feedstock for ethylene production by cracking. The other important chemical uses for butane are in oxidation to acetic acid and in the production of maleic anhydride. In the past, butane also was the main feedstock for the production of butadiene by dehydrogenation, but it has been replaced by coproduct butadiene obtained from ethylene production.

Ethylene

The largest potential chemical market for n-butane is in steam cracking to ethylene and coproducts. n-Butane is a supplemental feedstock for olefin plants and has accounted for 14 percent of total ethylene production for most years since 1970. It can be used at up to 1015 percent of the total feed in ethane/propane crackers with no major modifications. n-Butane can also be used as a supplemental feed at as high as 2030 percent in heavy naphtha crackers. The consumption of C4s has fluctuated considerably from year to year since 1970, depending on the relative price of butane and other feedstocks. The yield of ethylene is only 3640 percent, with the other products including methane, propylene, ethane, and butadiene, acetylene, and butylenes. About 23 billion lb of butane are consumed annually to produce ethylene.

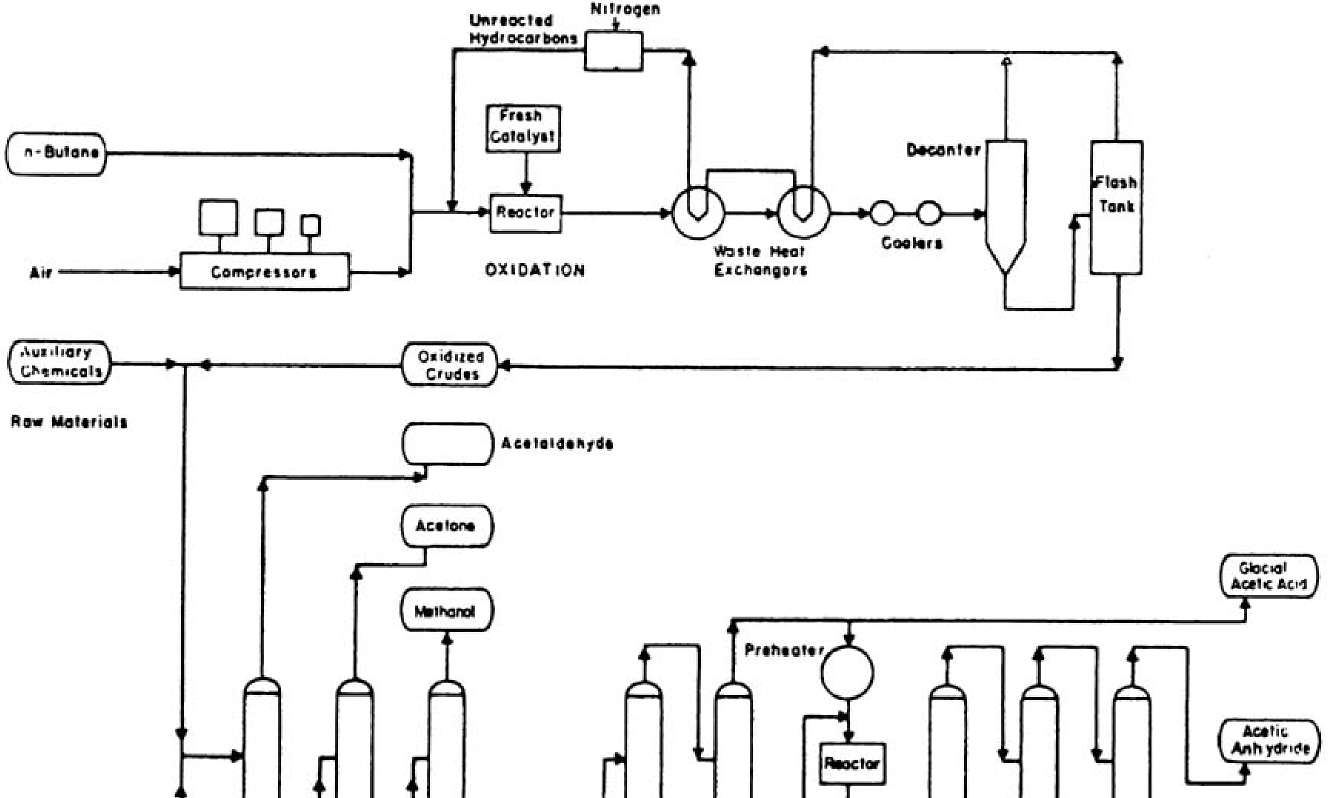
Acetic Acid

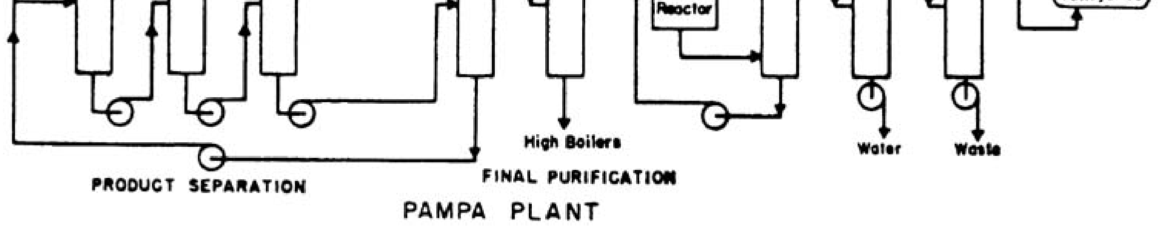
Acetic acid is the most important carboxylic acid produced industrially. The annual production in the United States in 1999 was almost 15.7 billion lb. As with many compounds produced on a large scale, acetic acid has several different commercial processes. The carbonylation of methanol is now the dominant route. (This process was described earlier in this chapter in the section "Methanol".) The oxidation of acetaldehyde, ethanol, and butane are also important. The percent world capacity for virgin acetic acid by different starting materials are: methanol (60%), acetaldehyde (18%), ethanol (10%), and butane (8%).

The liquid-phase oxidation (LPO) of light saturated hydrocarbons yields acetic acid and a spectrum of coproduct acids, ketones, and esters. Although propane and pentanes have been used, n-butane is the most common feedstock because it can ideally yield two moles of acetic acid. The catalytic LPO process consumes more than 500 million lb of n-butane to produce about 500 million lb of acetic acid, 70 million lb of methyl ethyl ketone, and smaller amounts of vinyl acetate and formic acid. The process employs a liquid-phase, high-pressure (850 psi), 160180°C oxidation, using acetic acid as a diluent and a cobalt or manganese acetate catalyst.

Figure 22.29 shows the flowsheet for this process. From the reactor, the product mixture is passed through coolers, and then through separators where the dissolved gases are released. The major components of the oxidized crude are acetic acid, methyl ethyl ketone, and various alcohols and acids. The initial reaction involves formation of butane-2 hydroperoxide, which is not isolated. Further oxidation and decomposition of the resultant radicals produce acetic acid.

Fig. 22.29 Oxidation of butane. (*Pet. Ref. 38, no. 11, 234, 1959. Copyright 1959 by Gulf Publishing Co.*)





Concurrently, the hydroperoxide may be converted to methyl ethyl ketone (MEK). If the initial radical attack is at the primary rather than the secondary carbon, the process makes propionic and formic acids. Reaction conditions can be changed to produce more MEK at the expense of some acetic acid. The maximum acetic acid/MEK ratio is 6.57 on a weight basis. If ethyl acetate is also formed, the ratio can go down to acetic acid/(ethyl acetate + MEK) of 3.64, with MEK being about 55 percent of the by-product.

A portion of the acetic acid, which is the major product, can be converted in a separate unit to acetic anhydride. Acetic anhydride may be produced from acetic acid, acetone, or acetaldehyde. With both acetic acid and acetone the initial product is ketene. The ketene is highly reactive and reacts readily with acetic acid to form acetic anhydride. All this takes place at 700800°C in the presence of a triethyl phosphate catalyst. With acetic acid, the reactions are:

$$\mathrm{CH_3COOH} \rightarrow \mathrm{CH_2{=}C{=}O} + \mathrm{H_2O}$$

$$\mathrm{CH_3COOH} + \mathrm{CH_2{=}C{=}O} \rightarrow \mathrm{(CH_3CO)_2O}$$

Acetic anhydride is used to make acetic acid esters. It is especially effective in difficult acetylations, such as in the manufacture of aspirin and cellulose acetate.

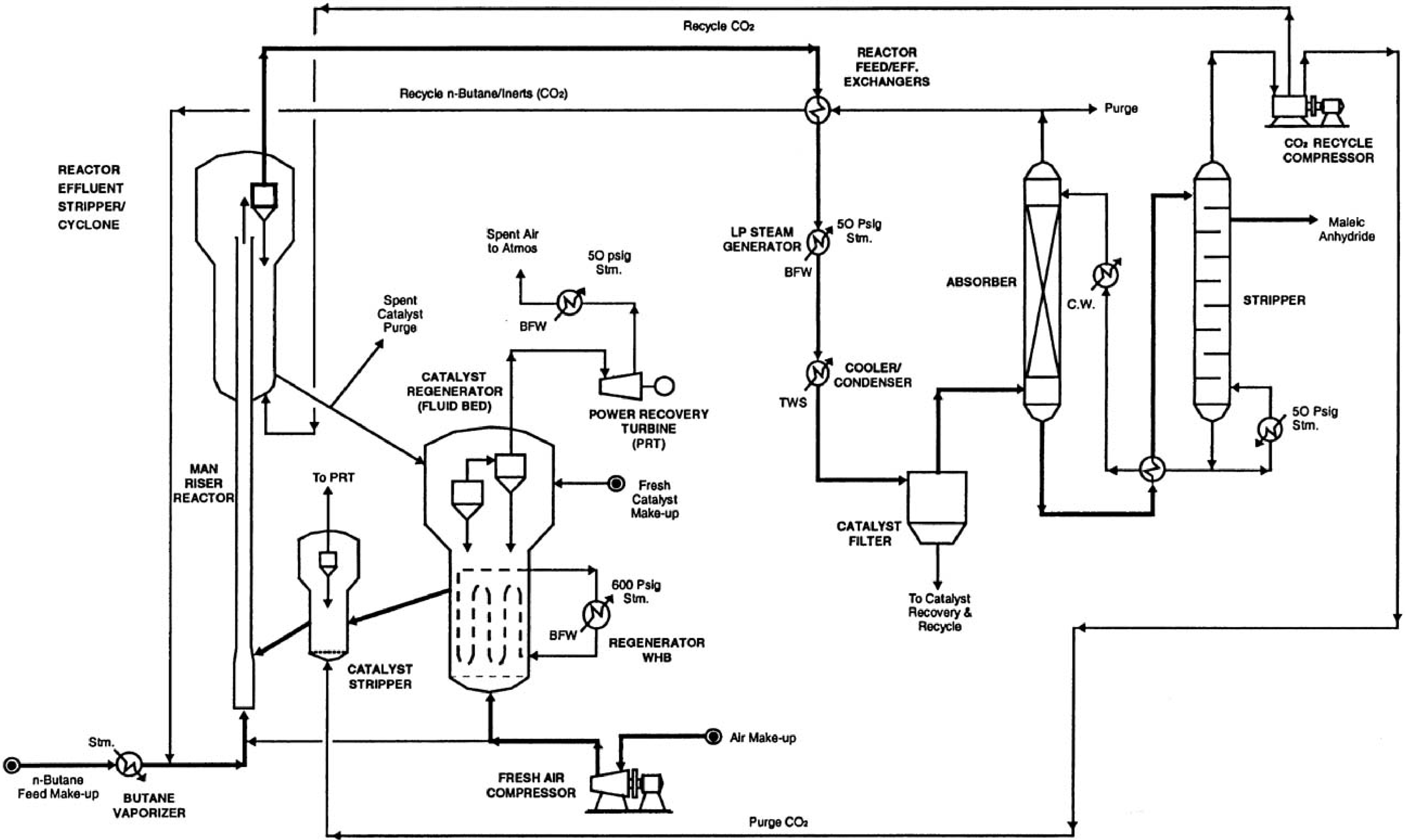
Maleic Anhydride

Maleic anhydride is one of the fastest-growing chemical end-uses for butane. The demand in the United States was about 500 million lb in 2000. About 60 percent of the maleic anhydride produced goes into the manufacture of unsaturated polyester resins, used primarily in fiber-reinforced plastics for construction, marine, and transportation industries. It is also used to make lube oil additives, alkyd resins, fumaric and malic acids, copolymers, and agricultural chemicals.

Essentially all maleic anhydride is manufactured by the catalytic vapor-phase oxidation of hydrocarbons. Prior to 1975, benzene was the feedstock of choice. By the early 1980s, however, many producers had switched to n-butane for economic and environmental reasons. Although benzene as a feedstock for maleic anhydride is no longer used in the United States, it is still used in older plants in Latin America, Europe, and East Asia. The oxidation reaction to produce maleic acid from n-butane is as follows:

Until recently (late 1990s), nearly all butane oxidation to maleic anhydride was conducted in a fluid bed or in a fixed bed multitubular, tube-shell heat exchanger type of reactors. After over a decade of intensive R&D efforts, DuPont was successful in commercializing a Circulating Fluid Bed Reactor (CFBR) catalyst system. CFBR has the advantage of providing 1015 percent higher selectivity even at higher conversions, thereby significantly reducing raw material costs. A diagram of the fluid bed process in presented in Fig. 22.30.

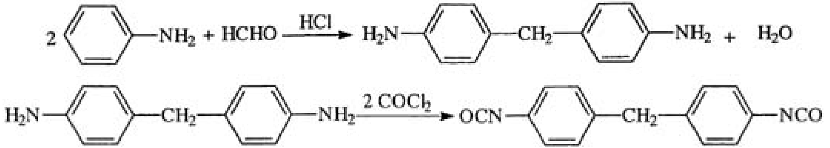
Fig. 22.30 Maleic anhydride DuPont process. (*Chem Systems Report No. 99/00-5. Copyright Nexant Chem Systems, Inc. and used by permission of the copyright owner.*)



Isobutanes

t-Butyl Alcohol/Propylene Oxide

An important use for isobutane is in the peroxidation of propylene with t-butyl hydroperoxide. The feedstocks are propylene and isobutane, and the process is similar to the PO/styrene plant. (See the section "Chemicals from Benzene" below.) In the two-stage conversion route, oxidation of isobutane with air yields a mixture of t-butyl hydroperoxide and t-butanol in a liquid-phase reaction at 135144°C. After separation of products, a molybdenum-catalyzed reaction of the hydroperoxide with propylene at 110°C yields PO and t-butyl alcohol (TBA).



The ratio of TBA to PO, and thus the isobutane requirement, can be adjusted from approximately 2:1 to 3:1. The US capacity of PO in 1999 is about 1715 million lb. (The capacity for PO/styrene and chlorohydrin routes are 1120 and 2150 million lb respectively). At this time, only one domestic PO producer is not using this process.

When this process was first introduced, TBA had a low value, so the TBA/PO ratio was kept to a minimum. Since the use for MTBE as oxygen enhancer in gasoline has became important, a higher TBA/PO value is used. Note, however, that changes in the MTBE picture also apply hereas discussed earlier. Dehydration of TBA yields high-purity isobutylene, which can be converted into MTBE with methanol. Today, almost 900 million lb of isobutylene is produced from TBA.

Other Isobutane-Based Chemicals

Isobutane can be directly dehydrogenated to isobutylene by a modification of the Houdry process. This can then be converted to MTBE. The estimated use is over 1 billion lb of isobutane. Because of their inertness and higher vapor pressures, high-purity propane and butanes have become the important substitutes for fluorocarbons as aerosol propellants. Isobutane can also be used as a solvent in polymer processing, and as a blowing agent for foamed polystyrene.

Butylenes

Butylenes are four-carbon monoolefins that are produced by various hydrocarbon processes, principally catalytic cracking at refineries and steam cracking at olefins plants. These processes yield isomeric mixtures of 1-butene, *cis*- and *trans*-butene-2, and isobutylene. Derivatives of butylenes range from polygas chemicals and methyl t-butyl ether, where crude butylenes streams may be used, to polybutene-1 and LLDPE, which require high-purity 1-butene. In 1997, the estimated consumption of butylenes (in billions of pounds) was: alkylation, 32.0; MTBE, 12.0; other, including polygas and fuel uses, 0.5.

The major chemical uses for n-butylenes are sec-butyl alcohol (and MEK), butadiene, butene-1, heptenes, and octenes. In 1978, butadiene accounted for almost 70 percent of the demand. At the present time sec-butanol and 1-butene are the largest chemical end-uses of butylene, consuming about four fifths of the total.

1-Butene

The largest chemical use of n-butenes is 1-butene used in production of LLDPE, which requires alpha-olefin comonomers. Various processes for the production of LLDPE and HDPE incorporate 1-butene as a comonomer. This accounts for about 70 percent of 1-butene use. The alpha-olefin comonomers control the density and physical properties of the polymer. About 20 percent of the 12,500 million lb of HDPE production in the United States in 1997 utilized 1-butene as comonomer, as did 45 percent of the 6900 million lb of LLDPE. Most of the remainder was split between 1-hexene and 1-octene.

The remaining 30 percent of 1-butene is divided among several uses. About 1015 percent of the 1-butene is polymerized in the presence of a Ziegler-type catalyst to produce polybutene-1 resin. The markets for this resin are pipe, specialty films, and polymer alloys.

Approximately the same volume of 1-butene is reacted with synthesis gas in an oxo reaction to produce valeraldehydes. These C5 aldehydes are then hydrogenated to amyl alcohols or oxidized to valeric acid. Amyl alcohols are consumed in the production of lube oil additives and amyl acetate and in solvent uses. Valeric acid goes into lubricant base stocks and specialty chemicals.

Smaller uses of 1-butene are in 1,2-butylene oxide, butyl mercaptan, and butyl phenols. Butylene oxide, produced by the chlorohydrin process, is used as a corrosion inhibitor in chlorinated solvents. Butyl mercaptan is a precursor for organophosphate herbicides, pharmaceutical intermediates, and is used as a gas odorant.

sec-Butanol and Methyl Ethyl Ketone

The next-largest use for n-butenes is in the manufacture of sec-butanol. A refinery butanesbutylenes stream, usually rich in butene-2, is contacted with 80 percent sulfuric acid to produce the sec-butyl hydrogen sulfate. Dilution with water and steam stripping produce the alcohol. MEK is obtained in high yields at catalytic dehydration of the alcohol at 400500°C.

Solvent applications account for almost 95 percent of all MEK consumption; the rest goes to chemical uses such as MEK peroxide and methyl ethyl ketoxime. The solvent applications include surface coatings, adhesives, lube oil dewaxing, magnetic tape manufacture, and printing inks. Production of MEK in 1999 totaled almost 690 million lb.

Heptenes and Octenes

Heptene and octene are oligomers produced by the polymerization of refinery streams containing C3 and C4 hydrocarbons. Originally these polygas units were developed to provide a source of high-octane blending components from refinery gases, but many have since been adapted to produce heptenes and octenes for chemical uses. Heptenes are used primarily to produce isooctyl alcohol which is a precursor for lube oil additives, diisooctyl phthalate (DIOP), other plasticizers, and herbicide esters. Isooctyl alcohol was once the predominant plasticizer alcohol, but development of plasticizers from competing alcohols such as 2-ethylhexanol and linear alcohols has eroded its market. Octenes are precursors of isononyl alcohol, which in turn is the raw material in the manufacture of diisononyl phthalate (DINP) plasticizers. These plasticizers compete directly with dioctyl phthalate in many applications. The total U.S. consumption of butylenes for production of heptenes and octenes was approximately 240 million lb in 1997.

Butene-2

Most butene-2 in the United States goes into production of gasoline alkylate. Some butene-2 is used in solvent applications, and it is also the intermediate in the disproportionation process for producing propylene from ethylene.

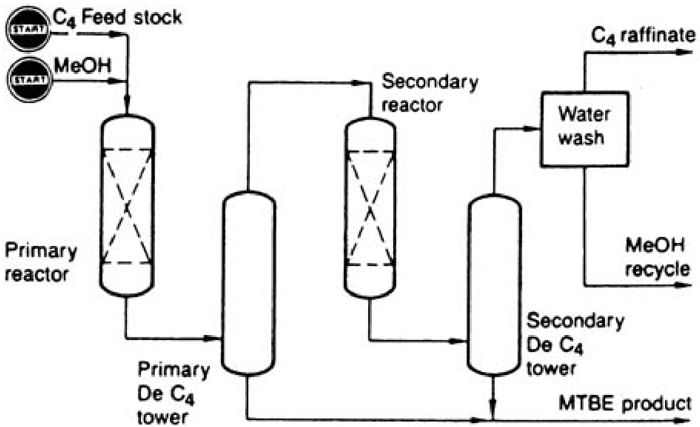
Isobutylene

Methyl t-Butyl Ether

By far the largest use of isobutylene is in the manufacture of methyl t-butyl ether (MTBE). Since its introduction in 1979, the demand for MTBE as an octane improver in gasoline has grown phenomenally. By 2000, production had reached 46 billion lb/year. MTBE use is exclusively as an octane booster/combustion promoter in gasoline. Use of MTBE in gasoline, at least in the United States, is expected to decline in the coming years for the reasons discussed earlier in this chapter. The decline of MTBE production is expected to have a significant impact on C4 uses in the future.

MTBE is produced by reacting methanol and isobutylene under mild conditions in the presence of an acid catalyst. The isobutylene feed is either mixed butylenes, a butylenes stream from catalytic cracking, or a butylenes coproduct from ethylene production. The reaction conditions are mild enough to permit the n-butenes to pass through without ether formation. Figure 22.31 shows a typical process for making MTBE.

Fig. 22.31 Two-stage MTBE process. (*Bitar, L. S., Hazbun, E. A. and Piel, W. J., Hydrocarbons Processing, 63, no. 10, 54. 1984, October. Copyright Gulf Publishing Company and reproduced by permission of the copyright owner.*)



Another approximately 1.5 billion lb of isobutylene goes into other chemical uses. These applications include polybutenes and derivatives of high-purity isobutylene such as butyl rubber, polyisobutylenes, and substituted phenols. Isobutylene is more reactive than the n-butenes, but many of its reactions are readily reversible under relatively mild conditions.

Polybutenes

More than 900 million lb of butylenes are consumed in the production of polybutenes. The process involves the FriedelCrafts polymerization of desulfurized C4 refinery streams. Although the feed is a mixed butylenes stream, the polybutene product is essentially a polymer of isobutylene, with a few n-butene units occurring in the polymer chain. The optimum isobutylene concentration in the feed stream is reported to be in the 2025 percent range. The polymerization reaction yields products with a molecular weight range of 3003000.

The lowest-molecular-weight polymers (mol. wt. 300350) are used for dielectric fluids and specialty lubricants. They comprise a very small part of the market. Polymers in the 700750 range are precursors for caulks, sealants, and gasoline additives. Polymers with molecular weights of 900l500 constitute the segment with by far the largest volume. They are used in the manufacture of lube oil additives via modification with phosphorus, amine, or succinimide groups. The resulting dispersants reduce oil consumption, restore compression, and improve oil viscosity. They are also used as gasoline additives and specialized plasticizers. The higher polymer fraction (mol. wt. 15003000) mainly is formulated into adhesives, caulks, and sealants. The use of polybutenes in gasoline detergents is expected to show continued growth.

Butyl Rubber

Almost two-thirds of the demand for high-purity isobutylene is for production of butyl rubber, which is produced by the cationic polymerization of high-purity isobutylene with isoprene (23% by weight) at low temperatures in the presence of a FriedelCrafts catalyst. The isobutylene must be pure in order to obtain a high molecular weight product. The elastomer's outstanding property of impermeability to air and gas makes it particularly suitable for tire liners and tire bladders and valves. Other important butyl elastomers are the halogenated products, chlorobutyl and bromobutyl. Halogenated butyl rubbers, which are more compatible with other tire elastomers and can be cured faster than butyl rubber, have grown to be of more importance in recent years.

Polyisobutylenes

Polyisobutylenes are produced by the low temperature polymerization of high-purity isobutylene. The main commercial products have two molecular weight ranges: a 40,00050,000 molecular weight polymer, and a polymer with molecular weight two to four times that of butyl rubber. The lower-weight polymer is used in binder systems, and in plasticizers and tackifiers for adhesives and sealants for electrical applications. The largest application for the higher-weight polymer is in the production of lube oil viscosity improvers. In 1997, about 50 million lb of high-purity isobutylene went to make polyisobutylenes.

Other Derivatives of Isobutylene

The production of many smaller-volume chemicals is based on high-purity isobutylene. The major chemicals are para-t-butylphenol, di-t-butyl-*p*-cresol (butylated hydroxytoluene, BHT), 2,6-di-t-butylphenol, t-butylamine, t-butyl mercaptan, and isobutyl aluminum compounds. Isobutylene usage by each of these six ranged from 8 to 18 million lb. The total volume of isobutylene that went for small-volume chemical production was about 140 million lb in 1997.

The substituted phenols and cresols constitute about half the total volume of this group. Para-t-butylphenol is produced by the alkylation of phenol with isobutylene. The principal applications for this derivative are in the manufacture of modified phenolic resins for the rubber industry and in surface coatings. BHT is obtained from isobutylene and *p*-cresol. Technical-grade BHT is an antioxidant for plastics and elastomers, and is a gum inhibitor in gasoline. Food-grade BHT is an antioxidant in edible oils, preserves, and many other foods. 2,6-Di-t-butylphenol is used to produce a wide range of plastics additives, antioxidants, and gasoline additives.

t-Butylamine is formed by the reaction of isobutylene with HCN in the presence of strong sulfuric acid. The intermediate t-butyl formamide is then hydrolyzed to form the amine and formic acid. This amine is used mainly to synthesize sulfonamide rubber accelerator compounds. The major use of n-butyl mercaptan is in odorant formulations for natural gas. The distributors of natural gas inject about one pound of odorant per million cubic feet. Isobutyl aluminum compounds are produced from high-purity isobutylene, hydrogen, and aluminum. Five distinctly different isobutyl aluminum compounds are produced in the United States. The principal end use for these compounds is as polymerization cocatalysts in the manufacture of polybutadiene, polyisoprene, and polypropylene. Other minor uses for high-purity isobutylene are in the manufacture of neopentanoic acid, methallyl chloride, and miscellaneous butylated phenols and cresols.

Butadiene

In the mid-1970s there were several major processes for making butadiene in the United States: steam cracking of naphtha, catalytic dehydrogenation of n-butene, dehydrogenation of n-butane, and oxidative dehydrogenation of n-butene. By 2000, more than 90 percent of all the butadiene was made as a coproduct with ethylene from steam cracking, and the only "on-purpose" production came from the dehydrogenation of butylene.

Thermal cracking of hydrocarbon feedstocks in the presence of steam at 700900°C produces ethylene and several coproducts, including butylenes and butadiene. The yield of butylenes varies widely (to as high as 30%), depending on the feedstock and the severity of the cracking. The yield of butadiene is particularly high with naphthas and heavier feedstocks. Most ethylene producers recover a raw C4 stream that contains butanes, butylenes, and butadiene. The butadiene is recovered by extraction, and the raffinate (containing butanes and butylenes) is used for gasoline blending or the production of chemicals.

The dehydrogenation process feed can be refinery streams from the catalytic cracking processes. This mixed C4 stream typically contains less than 20 percent n-butenes. For use in dehydrogenation, however, it should be concentrated to 8095 percent. The isobutylene generally is removed first by a selective extraction-hydration process. The n-butenes in the raffinate are then separated from the butanes by an extractive distillation. The catalytic dehydrogenation of n-butenes to 1,3-butadiene is carried out in the presence of steam at high temperature (>600°C) and reduced pressure. A typical catalyst could be a chromium-promoted calcium nickel phosphate. The oxidative dehydrogenation process for butadiene reacts a mixture of n-butenes, compressed air, and steam over a fixed catalyst bed of tin, bismuth, and boron.

Of the 17 billion lb of butadiene consumed in 1999, almost two thirds went into the production of elastomers (styrenebutadiene latex rubber (SBR), polybutadiene, nitrile, and polychloroprene). Adiponitrile, ABS resins, styrenebutadiene latex, styrene block copolymers, and other smaller polymer uses accounted for the remainder. The largest single use was for styrenebutadiene copolymers (SBR and latex). Most of it was made by an emulsion process using a free-radical initiator and a styrenebutadiene ratio of about 1:3. More detailed description of the rubber and polymer used can be found in Chapters 18 and 19.

22.5 Higher Aliphatic Hydrocarbons

Cyclopentadiene

Cyclopentadiene is a product of petroleum cracking. It dimerizes exothermically in a DielsAlder reaction to dicyclopentadiene, which is a convenient form for storage and transport. Dicyclopentadiene plus cyclopentadiene demand in the United States amounted to 270 million lb in 1998.

Dicyclopentadiene can be converted back to cyclopentadiene by thermally reversing the DielsAlder reaction. Cyclopentadiene also undergoes the DielsAlder reaction with other olefins, and this chemistry has been used to make highly chlorinated, polycyclic hydrocarbon pesticides. These pesticides are so resistant to degradation in the biosphere, however, that they are now largely banned from use. It is also used as a monomer and a chemical intermediate.

Isoprene

Isoprene is the basic repeating unit in natural rubber and in the naturally occurring materials known as terpenoids. It is a diene like butadiene and is useful as a building block for synthetic polymers. The most frequently used synthetic procedure for making isoprene is acid-catalyzed reaction of formaldehyde with isobutylene, giving a dioxolane intermediate that is thermally cracked to isoprene. Isoprene also can be recovered from petroleum refinery streams. A total of 360 million lb of isoprene was used in the United States in 2000.

Isoprene is converted to elastomers such as poly(*cis*-1,4-isoprene), which is tough, elastic, and resistant to weathering and is used mainly for vehicle tires. Recently, block copolymers of isoprene with styrene have been finding use as thermoplastic elastomers and pressure-sensitive adhesives.

n-Paraffins and Olefins

n-Paraffins are the unbranched fraction of hydrocarbons found in petroleum. They can be separated from the branched and aromatic hydrocarbons by a process using a shape-selective, controlled-pore-size adsorbent. In this process, the small pores exclude branched or aromatic materials and allow them to flow through a column more readily than the linear hydrocarbons, which become adsorbed into the narrow pore structure of the adsorbent. Linear, internal monoolefins are produced by dehydrogenation of n-paraffins.

Primary and Secondary Higher Alcohols

Linear internal monoolefins can be oxidized to linear secondary alcohols. The alpha (terminal) olefins from ethylene oligomerization, described earlier in this chapter, can be converted by oxo chemistry to alcohols having one more carbon atom. The higher alcohols from each of these sources are used for preparation of biodegradable, synthetic detergents. The alcohols provide the hydrophobic hydrocarbon group and are linked to a polar, hydrophilic group by ethoxylation, sulfation, phosphorylation, and so forth.

22.6 Chemicals Derived from Benzene, Toluene, and Xylene

Until World War II, most of the commercial aromatic chemicals in the United States and other countries were derived from the coal industry. The high-temperature carbonization of coal to produce coke from the steel industry also generated a liquid stream that was rich in aromatics. Benzene, toluene, and xylene (BTX) and other aromatics could be recovered from the coal tar by extraction and distillation. Recently, the importance of this source has greatly diminished; now almost all of the BTX in the United States is based on petroleum (17 billion lb of benzene, 12.6 billion lb of toluene, and 12.3 billion lb of xylene).

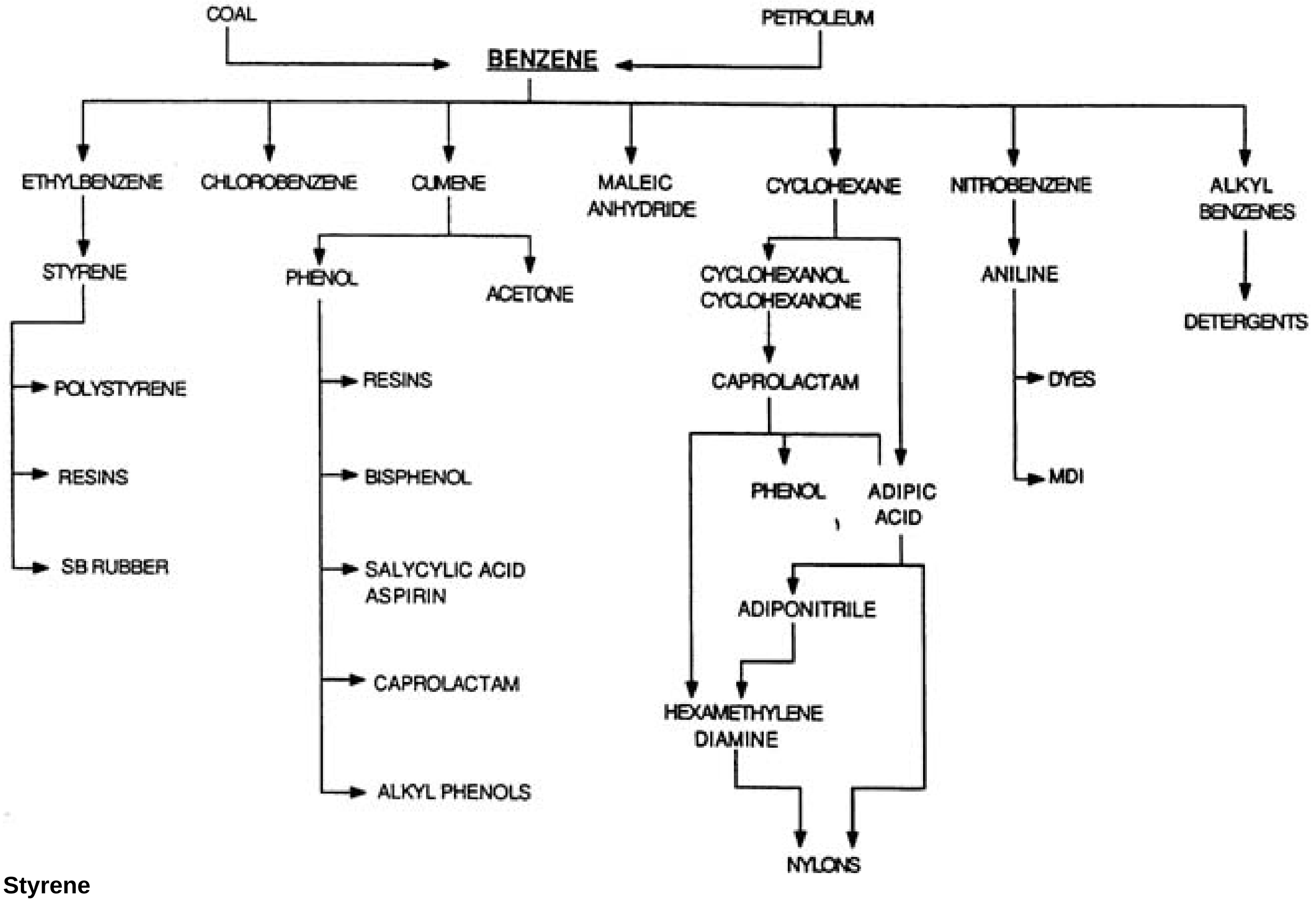
There are two major sources for petroleum-based aromatics, catalytic reformate and pyrolysis gasoline. The catalytic reformate is a refinery product that occurs in the catalytic reforming of naphthenes and paraffins in low-octane naphtha to produce a high-octane product. The pyrolysis gasoline stream that results from the steam cracking of hydrocarbons to produce ethylene and propylene is a very large (and growing) source of benzene and other aromatics. Because the demand for toluene is considerably less than that for benzene, some toluene is converted back to benzene by high-temperature hydrodealkylation (HDA) or by catalytic toluene disproportionation (TDP). The amount of benzene produced by HDA at any given time depends on the relative economics of benzene and toluene. Recently, the ratio of benzene produced from HDA versus TDP process is about 1:3. The specific process details of these production routes can be found in Chapter 15.

Compared to the amount of BTX that goes into fuel, the volume of aromatic chemicals used as chemical building blocks is relatively small. About half of the benzene and more than 90 percent of the toluene and xylenes end up in the gasoline pool.

Chemicals from Benzene

Benzene is by far the most important aromatic petrochemical raw material. During 1999, some 2.8 billion gal were consumed in the United States. This ranks it close to propylene as a chemical building block. Benzene has a broad end-use pattern. Its most important uses are for: ethylbenzene (styrene), 55.6 percent; cumene (phenol), 22.4 percent; cyclohexane (nylon), 13.5 percent; nitrobenzene (aniline), 5 percent; and detergent alkylate, 3 percent. Other significant uses are for chlorobenzenes and maleic anhydride. Benzene is an excellent solvent, but it has been almost entirely replaced by less toxic materials. Some of the major end products are shown in Fig. 22.32.

Fig. 22.32 Chemicals derived from benzene.



Styrene

Styrene is the largest benzene derivative with annual consumption about 11.5 billion lb in the United States. It is produced mainly by catalytic dehydrogenation of high-purity ethylbenzene (EB) in the vapor phase. The manufacture process for EB is based on ethylene alkylation with excess benzene. This can be done in a homogeneous system with aluminum chloride catalyst or a heterogeneous solid acid catalyst in either gas or liquid-phase reaction. In the past decade, the liquid-phase alkylation with zeolite catalyst has won acceptance. Those processes have advantages of easier product separation, reducing waste stream, and less corrosion. In addition, it produces less xylene due to lower process temperature. Since certain xylene isomers have very similar b.p. as styrene, it is difficult to be separated.

The EB dehydrogenation can be done with various commercially available styrene catalysts. The fractionation train separates high-purity styrene, unconverted EB, and minor reaction by-products such as toluene.

Styrene is also produced as coproduct from a PO process. In this route, EB is oxidized to its hydroperoxide and reacted with propylene to yield propylene oxide. The coproduct methyl phenyl carbinol is then dehydrated to styrene. For every pound of PO produced, up to 2.5 lb of styrene can be produced. In 2000, about 25 percent of styrene was produced by this process in the United States.

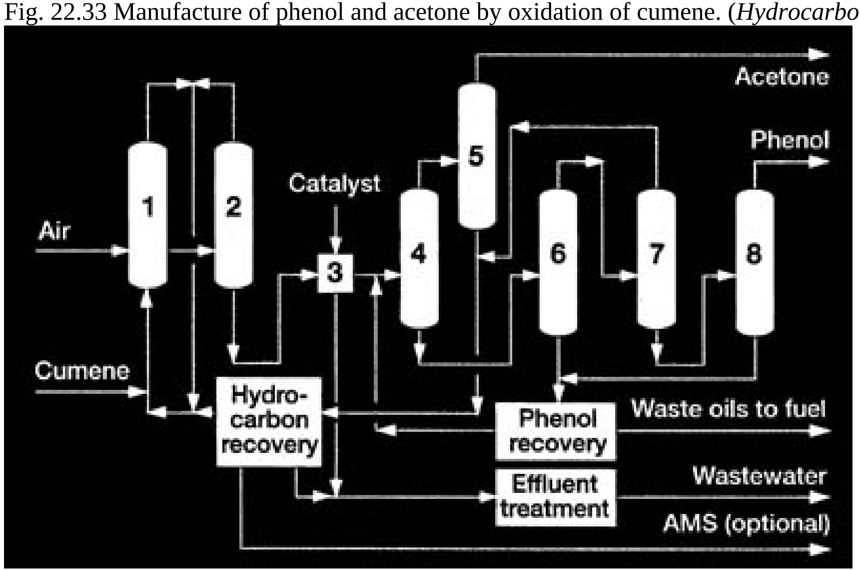
The largest use for styrene (over 70%) is to make homopolymer polystyrene. The U.S. production volume reached 6.3 billion lb in 1998. Other major uses are in plastics, latex, paints, and coatings, synthetic rubbers, polyesters, and styrene-alkyd coatings. In these applications styrene is used in copolymers such as ABS (8%), styrenebutadiene latex (8%) rubber (SBR) (4%), unsaturated polyester resins (6%), and other polymer applications.

Cumene (Phenol)

Cumene has become the second largest chemical use for benzene. It is produced by alkylating benzene with propylene at elevated temperature and pressure in the presence of a solid acid catalyst. The U. S. production was more than 6.9 billion lb in 1999. Of this, about 96 percent then was converted to phenol.

Before 1970, there were five different processes used to make phenol in the United States: the sulfonation route, chlorobenzene hydrolysis, the Raschig process, cumene oxidation, and the benzoic acid route. By 1978, the first three processes had essentially disappeared, and 98 percent of the remaining plant capacity was based on cumene oxidation. The oxidation process is shown in Fig. 22.33.

Fig. 22.33 Manufacture of phenol and acetone by oxidation of cumene. (*Hydrocarbon Processing. p. 117, 2001, March. Copyright 2001 by Gulf Publishing Co.*)



In this process, cumene is oxidized to cumene hydroperoxide by air at about 100°C in an alkaline environment. The oxidation products are separated, and the bottoms are mixed with a small amount of acetone and sulfuric acid and held at 7080°C while the hydroperoxide splits into phenol and acetone. Total domestic phenol capacity with this process is about 4.8 billion lb/year. In the much smaller-volume benzoic acid process, toluene is air-oxidized to benzoic acid with a cobalt catalyst. The benzoic acid then is converted to phenol by an oxidative decarboxylation reaction with air at about 240°C. The three major uses for phenol are in the manufacture of phenolic resins, bisphenol A, and caprolactam.

Phenolic resins

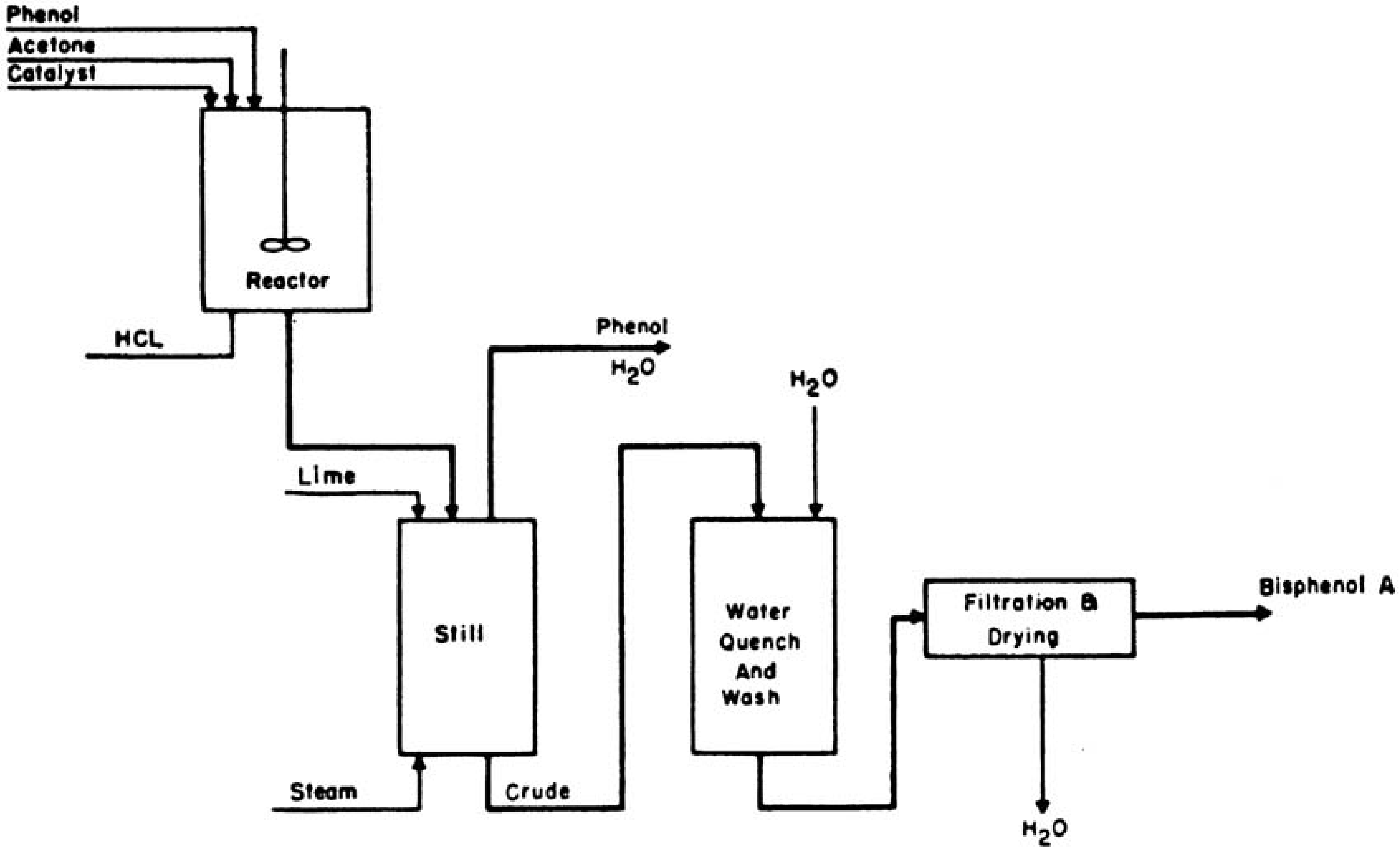
Resins such as those made from phenol and formaldehyde now account for about one third of the phenol consumed in the United States. They are widely used in construction related use such as plywood adhesives, foundry resins, thermoformed plastics, and surface coatings.

Bisphenol

Bisphenol-A (4,4′-isopropylidene-diphenol) accounts for 35 percent of phenol consumption and is used mainly in the production of polycarbonates (55%) and epoxy resins (25%), two of the fastestgrowing families of plastics. Other uses are in the manufacture of flame retardant such as tetrabromobisphenol-A, polysulfone resins, and polyacrylate resins. The consumption of bisphenol in the United States in 1999 topped 2.1 billion lb.

Bisphenol is obtained by the reaction of phenol and acetone with HCl or acid resin as catalyst. In the HCl catalyzed process (Fig. 22.34), phenol and acetone in a molar ratio of about 3:1 are charged to an acid-resistant stirred reactor. A sulfur-containing catalyst is added, and then dry HCl gas is bubbled into the reaction mass. The temperature is maintained at 3040°C for 812 hr. At the end of the reaction, the mixture is washed with water and treated first with enough lime to neutralize the free acid. Vacuum and heat are applied, and water and phenol are distilled separately from the mixture. The batch is finished by blowing the molten product with steam under vacuum at 150°C to remove the odor of the sulfur catalyst. The molten bisphenol is quenched in a large volume of water, filtered, and dried. All new bisphenol plants are based on the acid resin catalyst technology. The total capital investment for a resin plant is about half of a plant using HCl catalyst. The higher investment for the HCl-catalyzed process is due to the need for the corrosion resistant materials in the process equipment handling HCl and removing it after the reaction.

Fig. 22.34 Manufacture of bisphenol A. (*Pet. Ref. 38, no. 11, 225, 1959. Copyright 1959 by Gulf Publishing Co.*)



Other phenolics

Other major derivatives of phenol include caprolactam (14%), salicyclic acid (aspirin), alkylated phenols (3.6%), aniline (3.5%), and xylenols (3.5%).

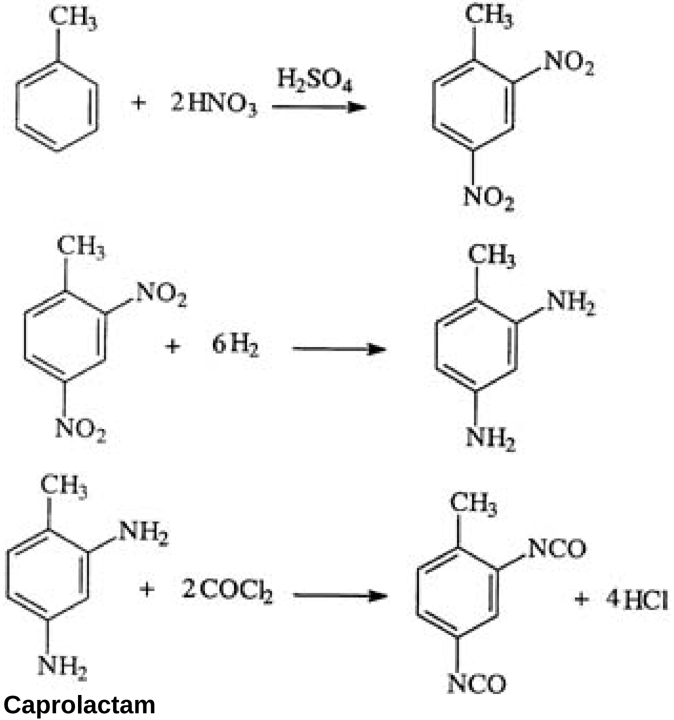
The uses of caprolactam are described in the section "Cyclohexane." Almost 60 percent of the 27 million lb of salicyclic acid went to aspirin, with remainder mainly going to salicylate esters and phenolic resins. Aspirin (acetyl salycylic acid) has long been recognized for its analgesic and antipyretic properties. It must now share this market with the competing pain relievers such as acetaminophen and ibuprofen.

Cyclohexane

Cyclohexane is the basic starting material for nylon fibers and resins via the intermediates adipic acid, caprolactam, and hexamethylenediamine. The world consumption was about 10 billion lb (with 3.5 billion lb in the United States) in 2000. Of these three derivatives, adipic acid and caprolactam account for over 90 percent of cyclohexane consumption. Cyclohexane is also used as a solvent and as a starting material for cyclohexanol and cyclohexanone. Although cyclohexane can be recovered from natural gasoline, most is made by liquid or vapor-phase hydrogenation of benzene. A nickel or platinum catalyst is generally used at elevated temperature and pressure.

Adipic acid

The world consumption for adipic acid was 4.6 billion lb in 1999. Nylon 66, produced from adipic acid and hexamethylenediamine (HMDA), is currently the largestvolume domestic nylon. About 86 percent of all adipic acid goes to make nylon 66 fibers and resins. Although HMDA can be made from adipic acid, a major source is from adiponitrile. The commercial synthesis of adipic acid is a two-step reaction starting with either cyclohexane or phenol. In both cases, a cyclohexanone/cyclohexanol mixture is formed as an intermediate. This mixture is then catalytically oxidized to the adipic acid product with nitric acid. It can also be manufactured as a by-product of the caprolactam process.



Caprolactam

Essentially all caprolactam is used in the manufacture of nylon 6 fibers. In 1998, global demand reached nearly 7.3 billion lb with 1.7 billion lb used in North America. This is a fast-growing nylon with applications in carpets, textiles, and tires. Caprolactam can be produced from cyclohexane, phenol, and toluene via cyclohexanone. It is then reacted with hydroxylamine to give an oxime. The oxime undergoes an acid-catalyzed rearrangement to give caprolactam.

Although these nylon intermediates are derived mainly from cyclohexane, there are various other routes. For example, adipic acid can be made by butylene oxidation as well as by cyclohexane oxidation; HMDA can be made starting with butadiene, adipic acid, or acrylonitrile; and caprolactam is also produced from phenol. See Chapter 21 for a more detailed discussion of nylon fibers and resins.

Maleic Anhydride

Prior to 1975, benzene was the feedstock of choice for maleic anhydride manufacture. By the early 1980s, for economic reasons, many producers had switched to the n-butane process described in the section "n-Butane Derivatives". By 1988, all of the maleic anhydride produced in the United States came from that process. However, about half of the maleic anhydride produced abroad still comes from benzene oxidation, with a small amount being recov-ered as a coproduct in phthalic anhydride manufacture.

Detergent Alkylate

Alkylbenzenes are major intermediates in the manufacture of synthetic detergents. If a straight-chain alkyl group is used, the resulting product is linear alkylbenzene (LAB), a "soft" degradable alkylate. Using a branched olefin (from propylene tetramer) gives branched alkylbenzene (BAB), a "hard" nondegradable alkylate. Approximately 960 million lb (mainly dodecyl- and tridecyl-benzene) of LAB were produced in 1998. The production of BAB was about 188 million lb in 1985; it declined to zero in 1998.

The production of LAB involves the liquid-phase alkylation of benzene with linear monoolefins or alkyl chlorides. Liquid HF is used as catalyst for linear monoolefins. And the AlCl3 is used as the catalyst for alkyl chlorides. Nowadays, acidic zeolite catalyst is used for olefin alkylation which generates less waste and reduces manufacture cost. The alkylate is then sulfonated to produce linear alkylbenzene sulfonate for biodegradable detergents. The manufacture of detergents is described in detail in Chapter 27.

The majority of LAB has a chain length of 1013. This method is also used to make longer-chain (C20C22) alkyl derivatives of benzene. Those alkylates are used as lubricants.

Nitrobenzene (Aniline)

The U.S. nitrobenzene production was about 2 billion lb in 1999. Two types of manufacturing processes were used: the direct nitration and the adiabatic nitration process. In the direct nitration system, benzene is mixed with a mixture of nitric/sulfuric acid. The reaction can be carried out in either a batch or a continuous system. Those reactors require a cooling system to keep it at constant temperature. It also requires a separate system for sulfuric acid reconcentration. In the adiabatic process, water is flashed off under vacuum before the sulfuric acid/nitrobenzene separation. The advantage of the adiabatic process is to eliminate a separated sulfuric acid reconcentration unit. This also will provide a better heat integration. Recently, the disposal of nitrophenols has become a major issue for aniline manufacture. Small amounts of nitrophenols are always made during the benzene nitration. It is more of a problem for the adiabatic process due to higher process temperature. Now an improved adiabatic process has been developed. By using an enhanced mixing and a higher benzene/acid ratio it produces less nitrophenols and it still keeps the advantages of the adiabatic process.

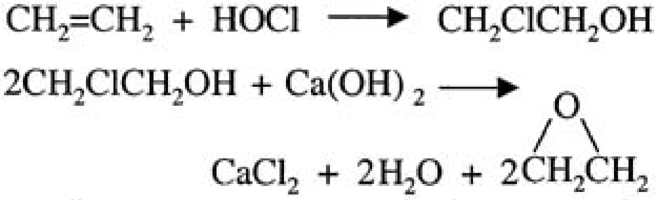
In 2000, the aniline production in the United States was about 1.9 billion lb. Almost 98 percent of nitrobenzene is used for the production of aniline. Consequently many nitrobenzene plants are integrated with facilities for aniline production. The hydrogenation of nitrobenzene can be done in either the vapor over a coppersilica catalyst or in liquid phase over platinumpalladium catalyst. One of the smaller uses for nitrobenzene is the production of the pain reliever, acetaminophen.

Aniline can also be made by two other methods. In the first, nitrobenzene is reduced by reaction with scrap iron in the presence of a hydrochloric acid catalyst. The iron is oxidized to the ferrous state, and the coproduct aniline is separated. This route accounts for less than 5 percent of the current aniline production. The other process avoids nitrobenzene entirely and involves the vapor-phase ammonolysis of phenol, using an alumina catalyst. Aniline is formed with diphenylamine as a by-product. About 20 percent of the aniline is produced by this route.

Aniline is consumed as a raw material in the manufacture of a number of chemicals: *p, p*-methylene diphenyl diisocyanate (MDI), 65 percent; rubber-processing chemicals, 15 percent; herbicides, 5 percent; dyes and pigments, 4 percent; specialty fibers, 2 percent. Other uses are in pharmaceuticals and photo chemicals. Principal growth is occurring from demand for MDI and the small, rapidly growing specialty fibers.

Diisocyanates (MDI)

The first step in the production of MDI (and polymeric "PMDI") is the condensation of aniline and formaldehyde to form diphenylmethylenediamine. The reaction conditions can be varied to change the isomer distribution of the product. This is followed by phosgenation to give an aromatic isocyanate product mix that corresponds to the starting polyaromatic amine.



Typically a mixture containing MDI and its dimer, trimer, and some tetramer is produced. Pure MDI can be separated by distillation. MDI is supplied in several grades, depending on the number of reactive units (NCO groups) per molecule. The most common grade is polymeric MDI with a functionality of 2.33.0. The grades used in rigid foam production typically contain 4060 percent pure MDI, with the balance being dimer and other isomers. Pure MDI is used mainly for RIM (reaction injection molding) systems.

The U.S. production of MDI was about 1.34 billion lb in 1998. Rigid polyurethane foams constitute the largest single use for MDI and its polymers, with total consumption of more than 850 million lb. Typical laminate and board foams contain more than 60 percent MDI, whereas pour systems and spray systems contain somewhat less. The main applications of rigid foam are in construction and in the manufacture of refrigerators and water heaters. Smaller uses are packaging, tank and pipe insulation, and transportation. A minor amount of polymeric MDI is used to make foundry sand binders. Pure MDI finds use in RIM systems, specialty coatings, thermoplastic resins, high-performance casting elastomers, and spandex fibers. A more detailed description of polyurethane polymers can be found in Chapter 19.

Other uses for aniline

Aniline currently is used as a raw material in most of the major groups of rubber-processing chemicals: accelerators, antioxidants and stabilizers, and anti-ozonants. The most important of these are the thiazole derivatives and substituted *para*-phenylenediamines. The demand for aniline in these rubber-processing uses is expected to grow at less than 1 percent/year. In agricultural chemicals, the major use for aniline is as a raw material in the manufacture of amide herbicides for controlling annual grasses and broadleaf weeds in various crops. More than 175 commercial dyes can be made from aniline, and many others are produced from aniline derivatives. This market, however, is not expected to show further growth. Since their introduction in 1979, the use of aniline in polyaramid specialty fibers has shown rapid growth. This trend is expected to continue in the future. Among the important pharmaceutical derivatives of aniline are the sulfonamides, a group of compounds used to combat infections.

Chlorobenzenes

Of the 12 different chlorobenzenes that can result from the chlorination of benzene, three are of most commercial importance: monochlorobenzene (MCB), *o*-dichlorobenzene (ODCB), and *p*-dichlorobenzene (PDCB). Chlorination of benzene can be done either batchwise or continuously in the presence of a catalyst such as ferric chloride, aluminum chloride, or stannic chloride. It is usually run as a three-product process; the current product distribution is about 52 percent to MCB, 17 percent ODCB and 31 percent PDCB. The pure compounds are separated from the crude by distillation and crystallization.

Production of monochlorobenzenes peaked in the 1960s with production volume at about 600 million lb. It was down to 152 million lb in 1998. The most significant cause for the decline is the replacement of monochlorobenzene by cumene as the preferred raw material for phenol manufacture. Other reasons include the elimination of the herbicide DDT, the change of diphenyl oxide process from chlorobenzene to phenol and a significant drop in solvent use. The production volume for ODCB and PDCB were 50 and 91 million lb, respectively, in 1998.

Monochlorobenzene

The largest use for monochlorobenzene, accounting for about 59 percent of the consumption, is in the production of chloronitrobenzenes. *p*-Nitrochlorobenzene (NCB) is converted into *p*-phenylenediamine for use as antioxidants in rubber processing. A smaller use for NCB is in the synthesis of the pain reliever, acetaminophen. *Ortho*-nitrochlorobenzene is a raw material for producing insecticides and several azo pigments. A large number of dyes also can be derived from either chlorobenzene or nitrochlorobenzene. About 13 percent of the monochlorobenzene is used as a solvent for pesticide formulation and in MDI processing. About 18 percent is used to make dichlorodiphenylsulfone, an intermediate in the manufacture of sulfone polymers.

Dichlorobenzenes

In 1998, approximately 16 million lb of the *o*-dichlorobenzene was converted into 3,4-dichloroaniline, the raw material for several major herbicides. Also a small amount (3 million lb) goes to various solvent applications. The major demands for *p*-dichlorobenzene come from uses in polyphenylene sulfide resins (50 million lb), room deodorants (16 million lb), and moth-control agents (11 million lb). Any future growth will have to come from the phenylene sulfide resins.

Trichlorobenzenes

A mixture containing trichlorobenzene is always obtained when chlorinating benzene. It can also be made through further chlorination of dichlorobenzene. Most trichlorobenzenes are produced as a mixture of 1,2,3-/1,2,4-trichlorobenzene. The 1,2,3-trichlorobenzene is sold as a raw material for pesticides. The 1,2,4-trichlorobenzene is used for the manufacture of the herbicide Banvel (dicamba). The estimated U.S. consumption was about 15 million lb in 1997.

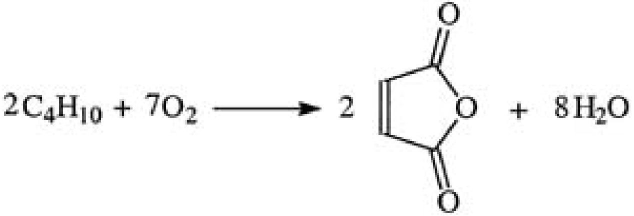
Derivatives of Toluene

Although the bulk of the toluene is never isolated from the gasoline pool, approximately 12.6 billion lb of toluene was produced for nonfuel consumption in the United States in 2000. Approximately 80 percent of this toluene is used as feed stock for benzene and xylene. The second largest end-use of toluene is as a solvent for coatings, paints, and lacquers. Also of importance is the use of toluene as an intermediate in the manufacture of other chemicals, mainly toluene diisocyanate, and also benzoic acid and benzyl chloride.

Toluene is converted into benzene by a catalytic hydrodealkylation (HDA) process at elevated temperature and pressure. The importance of this process is influenced by the relative value and demand for benzene, as benzene from this source is normally more costly than that isolated directly from refinery reformate streams. Benzene (along with xylenes) can also be obtained by the catalytic TDP. It has became favorable in recent years. Toluene consumption for toluene disproportionation versus HDA has changed from about 1/5 in 1990 to 2/1 in 2000. The volume of toluene that finds use as a solvent is expected to show a continued decline because of regulations controlling the emission of VOCs.

Toluene Diisocyanate (TDI)

TDI is manufactured from toluene by the route indicated in the following equations:



The synthesis of TDI begins with the nitration of toluene, using a nitric acidsulfuric acid mixture. The nitration product typically contains at least 75 percent 2,4-dinitrotoluene with the balance mostly 2,6-dinitrotoluene, which is catalytically reduced to toluene diamine. Lastly, the diamine mixture is dissolved in chlorobenzenes and reacted with phosgene to produce the TDI. After phosgenation, the mixture is stripped of the solvent and separated by distillation. The final product is an 80:20 isomer mixture.

The annual U.S. production of TDI was 960 million lb in 1999. Most of the TDI is reacted with polyols to produce flexible polyurethane foams. These foams are widely used as cushioning materials in furniture, automobiles, carpets, and bedding. A small amount of TDI is used to make polyurethane coatings. Polyurethanes are discussed in detail in Chapter 19.

Benzoic Acid

Benzoic acid can be produced by the LPO of toluene using a catalyst such as cobalt or manganese. Domestic production of benzoic acid was about 130 million lb in 2000. Of this amount, about one half went to make phenol or phenolic derivatives. Other uses are in the synthesis of caprolactam and terephthalic acid, and as food additive, and as a plasticizer and resin intermediate.

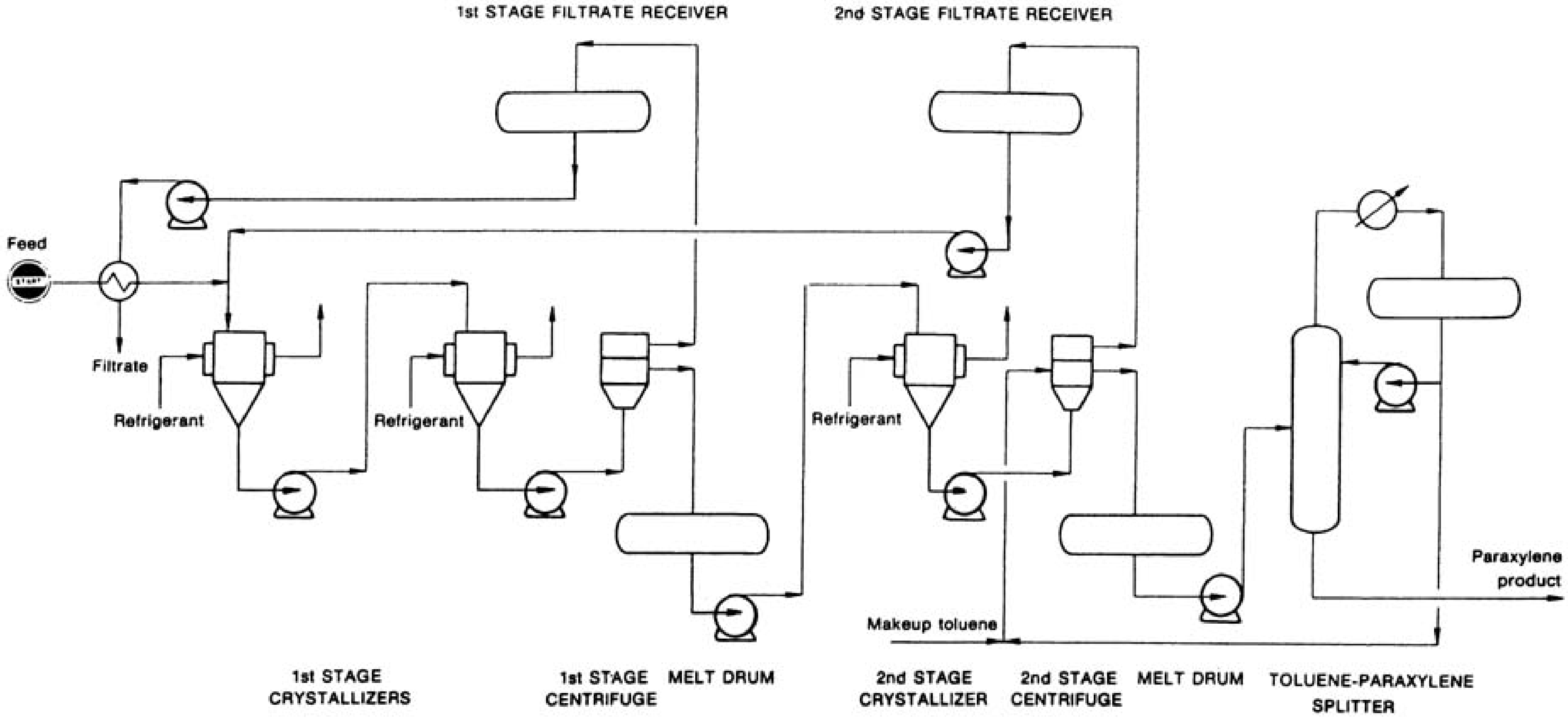
Benzyl Chloride

The principal method for producing benzyl chloride involves the photochlorination of toluene, followed by neutralization and distillation. In 1999, 75 million lb of benzyl chloride was produced in the United States. About two-thirds was used to manufacture benzyl phthalates (mainly butyl benzyl phthalate), which are widely used as plasticizers. The other use was to make benzyl quarts. Benzyl chloride can also be used as raw material in the manufacture of benzyl alcohol, for use in photography, perfumes, and cosmetics. The production has increased considerably in Western Europe because of the greater use in solvents such as benzyl esters. But the U.S. production was stopped in 1999.

Chemicals from Xylene

Xylenes are obtained mainly (80%) from petroleum reformate streams in the form of "mixed xylenes." A typical composition of this stream is about 18 percent *p*-xylene, 40 percent *m*-xylene, 22 percent *o*-xylene, and 20 percent ethylbenzene. The major chemical uses of xylene, however, require the pure isomers. The purification process involves a number of steps. First the *o*-xylene is separated from the other aromatics by distillation, with the *meta* and *para* isomers going overhead along with the ethylbenzene. *p*-Xylene can be recovered by either adsorption or crystallization processes. The flow diagram in Fig. 22.35 depicts a two-stage crystallization process for recovery of high-purity *p*-xylene from mixed xylenes. In the adsorption process, the stream is charged to a fixed bed of molecular sieves, and the selectively adsorbed *p*-xylene is recovered by washing the bed with solvent. Because the demand for *p*-xylene is far greater than that for m-xylene, the raffinate usually is isomerized to form more of the *para* isomer. Toluene may also be disproportionated to form equivalent amounts of benzene and xylenes without any ethylbenzene. Recently, this TDP process became more favorable. Today, it contributes to 18.5 percent of xylene production.

Fig. 22.35 Two-stage crystallization process for recovery of high-purity *para*-xylene from mixed xylenes. (*Reproduced from Hydrocarbon Processing, p. 175, Nov. 1985. Copyright 1985 by Gulf Publishing Co.*)



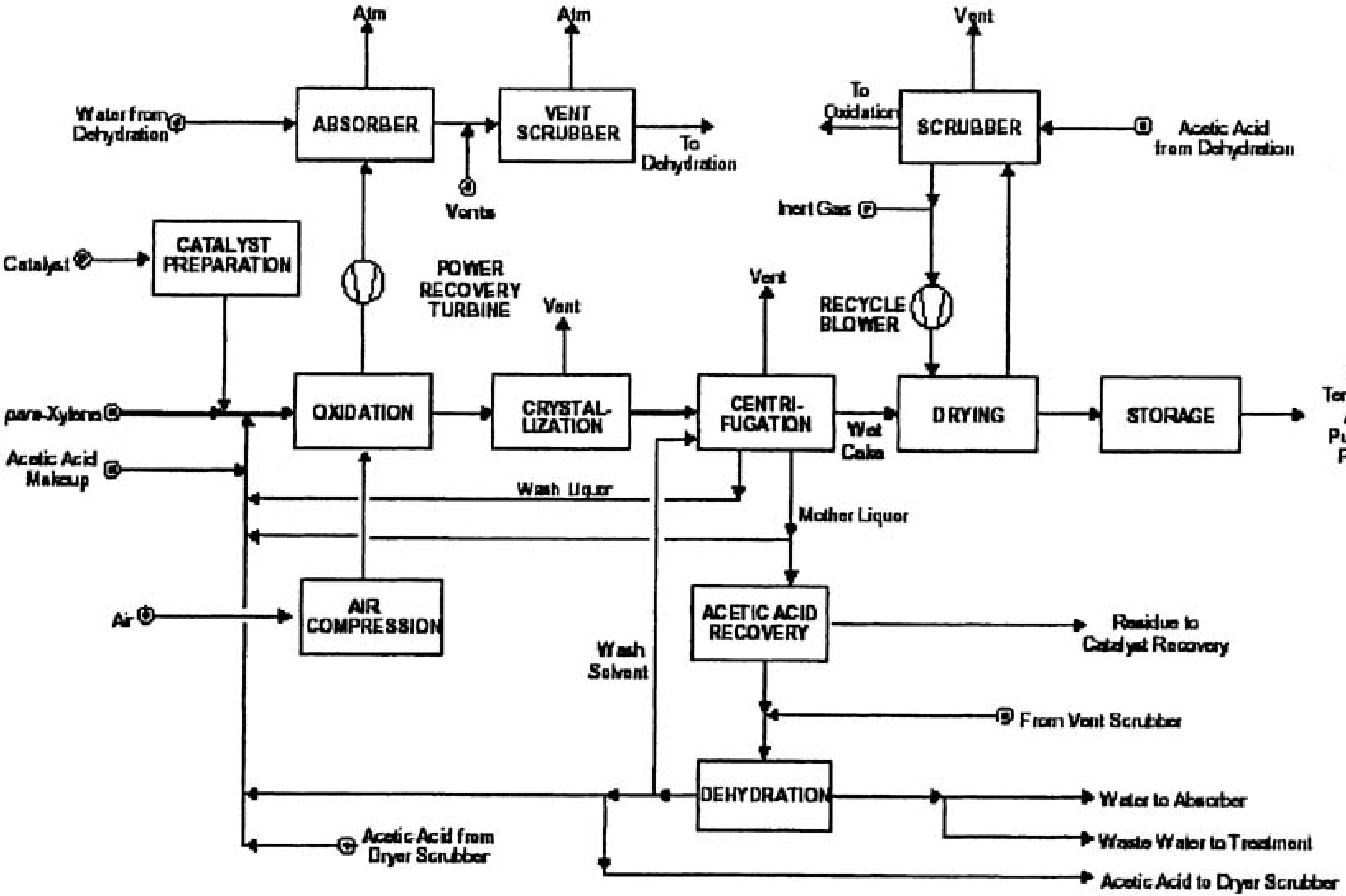
In 1999, the total demand for the xylenes (12.3 billion lb) is roughly comparable to that for toluene. The volume of *o*-, *m*- and *p*-xylene were approximately 1.1, 0.27, and 9.9 billion lb, respectively. The principal uses of the three xylene isomers are the production of terephthalic acid (or di-methyl terephthalate), phthalic anhydride, and isophthalic acid, respectively.

Terephthalic Acid (Dimethyl Terephthalate)

Terephthalic acid (TPA) and dimethyl terephthalate (DMT) are precursors for polyethylene terephthalate (PET), which in turn is used in the production of polyester fibers and film polyester thermoplastic PET bottles, and other resins. In 1999 the total U.S. production was more than 9 billion lb. In the past, the relative ease of producing high-quality DMT gave it the largest share of the terephthalate market. The trend is now toward TPA, as the result of technological advances that permit better purification of TPA and the use of the acid directly in polymer formation. The capacity is about 3 to 1 split in favor of TPA process.

One process for making TPA involves the air oxidation of a solution of *p*-xylene in acetic acid in the presence of a catalyst containing cobalt, manganese, and bromide (Fig. 22.36). The liquid-phase reaction is conducted at about 200°C and 20 atm. pressure. The mixture is cooled to recover TPA by crystallization, dissolved in hot water, and hydrogenated to remove aldehydic by-products. Fiber-grade TPA is recovered by recrystallization.

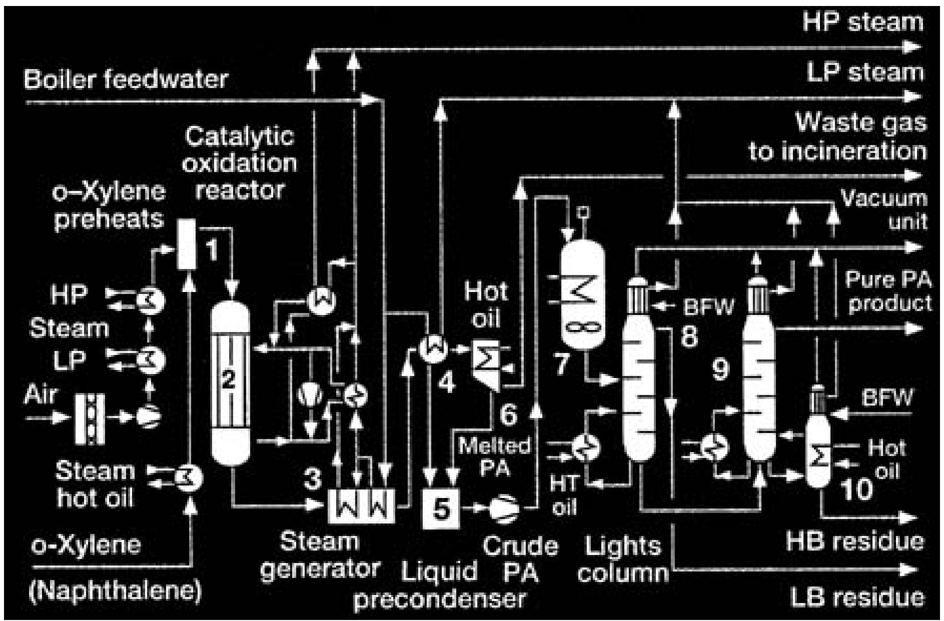
Fig. 22.36 Block flow diagram of generic crude TA process. (*Chem Systems Report No. 97/98-5. Copyright Nexant Chem Systems, Inc. and used by permission of the copyright owner*.)



Phthalic Anhydride

Although phthalic anhydride first was made commercially from the oxidation of naphthalene, by 1999, nearly 90 percent of the production had been converted to *o*-xylene as the feedstock. This is now essentially the only major use for the *o*-xylene. The flowsheet in Fig. 22.37 shows a typical process for making phthalic anhydride. The *o*-xylene is vaporized by injection into the hot gas stream and then passes through a catalyst-filled multitube reactor. The crude phthalic anhydride is desublimated, and any acid present is dehydrated in the predecomposer vessel. The crude is finally purified in two distillations. Although the fixed-bed process currently is important, there are a number of plants in which a fluidized-bed reactor is used.

Fig. 22.37 Production of phthalic anhydride. (*Hydrocarbons Processing. p. 118, March, 2001*. Copyright 2001 by Gulf Publishing Co.)



About 1.1 billion lb of phthalic anhydride are produced annually in the United States. The major uses are in plasticizers, alkyd resins, and unsaturated polyester resins. The plasticizers are esters made by reacting two moles of an alcohol, such as 2-ethylhexanol, with one mole of phthalic anhydride. These plasticizers find major use in vinyl chloride polymers and copolymers. Alkyd resins are a type of polyester resin used in surface coatings. The most rapidly growing end use is in unsaturated polyester resins for reinforced plastics.

Isophthalic Acid

Although *m*-xylene is an abundant material, it has limited demand as a chemical raw material. The only major outlet is in the manufacture of isophthalic acid. More than 220 million pounds were produced in the United States in 1992, primarily for use in preparing alkyd resins and unsaturated polyester resins. Small amounts also are used in PET bottle applications.

Naphthalene Derivatives

The use of naphthalene for production of chemicals has been declining steadily, as it is being replaced by other petroleum-based materials. Current domestic consumption is only about 235 million lb. There are two commercial processes for producing naphthalene. One is the recovery of naphthalene from coal tar, and the other involves its recovery from certain aromatized petroleum fractions. Until the end of the 1950s, the only commercial source of naphthalene in the United States was coal tar. At that time petroleum-derived naphthalene became a commercial product and was quickly established as a desirable source of raw material for phthalic anhydride because of its quality, low sulfur content, and stable supply.

The oxidation of naphthalene to phthalic anhydride has long been its principal end use. In the last 30 years, however, *o*-xylene has supplanted naphthalene as the preferred raw material for phthalic anhydride, and it now accounts for about 90 percent of the PA production. The manufacture of phthalic anhydride is described above in the section "Chemicals from Xylene."

Presently the major end uses for naphthalene are phthalic anhydride (62%), carbaryl insecticide (15%), surfactants and dispersants (20%), and synthetic tanning agents (3%). Carbaryl (1-naphthyl, n-methylcarbamate) is a broad-spectrum insecticide. The major surfactants and dispersants are derivatives of naphthalene sulfonates; their principal applications are as commercial wetting agents, concrete additives, and rubber dispersants, and in agricultural formulations. The synthetic tanning agent are derivatives of naphthalene sulfonic acid and formaldehyde. Other smaller uses of naphthalene are in beta-naphthol and as a moth-repelling agent.

22.7 References

Chem System Inc, Process Evaluation/Research Planning: Extending the Methane Value Chain 99-S9 (2000, October), Acetic Acid/Acetic Anhydride 97/98-1 (1999, January), Acetic Acid from Ethane Oxidation 99-S5 (2001, January), Methanol 98/99-4 (2000, May), Developments in Methanol Production Technology 96/97-S14 (1998, August), Methyl Methacrylate 99-2 (2001, September), Formaldehyde 00-8 (2001, April), Ammonia 97/98-6 (l998, August), Impact of MTBE Phaseout on Chemical Market 00/01 S2 (2001, June), Ethanol 99/00-8 (2001, August), Ethylbenzene/Styrene 99/00-6 (2000, August), Maleic Anhydride 99/00-5 (2000, May), Caprolactam 99/00-4 (2001, March), Adipic Acid 98/99-3 (1999, March), Nylon 6/Nylon 6,6 99-S1 (2000, March), Acrylic Acid/Acrylates 00/01-7 (2001, May), Development in Propylene Oxide Technology 00-S12 (2001, November), Glycerin 00-S4 (2001, November), Ethylene/Propylene 00-4 (2001, January), Acrylic Acid/Acrylates (2001, May), Epichlorohydrin 99-S11 (2000, August), Vinyl Chloride/Ethylene Dichloride 99-3 (2000, May), TDI/MDI 98/99-S8 (1999, October), Benzene/Toluene 98/99-6 (1999, May), Terephthalic Acid 97/98-5 (1999, February), Bisphenol A 97/98-4 (1998, May), Naphthalene and Derivatives 96/97-S9 (1998, March), Phenol/Acetone/Cumene 96/97-2 (1997, November), Development in p-Xylene Technology 96/97-S7 (1997, July), meta-Xylene/Isophthalic Acid 94/95 S-14 (1997, February).

Chemical Economic Handbook: SRI International. Menlo Park, CA, Ethylene Oxide (2000, August), Acetaldehyde (2001, September), Linear Alpha-Olefins (2000, September), Ethylene-Propylene Elastomers (2000, February), Ethylene (1999, April), Isopropyl Alcohol (IPA) (2000, May), Propylene (2000, March), Detergent Alcohols (2000, May), Plasticizer Alcohols (C4-C13) (1998, June), Propylene (2000, March), Butylenes (1998, August), 1,4-Butanediol (2001, March), Nonene (Propylene Trimer) and Tetramer (2001, August), Butadiene (2000, November), Methyl Ethyl Ketone (2000, April), Butanes (1999, July), Isoprene (2001, October), Cyclopentadiene/Dicyclopentadiene (1999, August), Benzene (2000, October), Toluene (2000, December), Xylene (2001, March), Bisphenol A (2001, February), Benzoic Acid (2001, September), Benzyl Chloride (2001, July), Aniline (2001, May), Chlorinated Methanes (2001, December), C2 Chlorinated Solvents (1999, January), Chlorobenzene (1999, December), Cyclohexane (2000, April), Cyclohexanol and CyclohexanoneUnited States (1998, May), Dimethyl Terephthalate (DMT) and Terephthalic Acid (TPA) (2001, February), Epichlorohydrin (2000, December), Epoxy Resins (2001, January), Ethylbenzene (1999, September), Styrene (1999, August), Glycerin (1999, August), Isophthalic Acid (1999, March), Naphthalene (2000, December), Petrochemical Industry Overview (2001, April), Phenol (1999, March), Phthalic Anhydride (2001, June), Propylene Oxide (1999, December), Propylene Glycols (2000, December), Ethylene Dichloride (2001, January), Vinyl Chloride Monomer (2001, December), Natural Gas (1999, January), Ammonia (2001, March), Acetylene (2001, August), Acetic Acid (2000, February), Adipic Acid (2000, April), Caprolactam (2000, April), Carbon Disulfide (2001, December), Ethanol (1999, April), Formaldehyde (2000, July), Hydrogen Cyanide (2000, June), Ketene (1999, January), Methanol (1999, August), Methyl Methacrylate (2000, March), Maleic Anhydride (1999, August), Urea (2001, January), Vinyl Acetate (2001, March).

Process Economics Program Report: SRI International. Menlo Park, CA, Isocyanates IE, Propylene Oxide 2E, Vinyl Chloride 5D, Terephthalic Acid and Dimethyl Terephthalate 9E, Phenol 22C, Xylene Separation 25C, BTX, Aromatics 30A, o-Xylene 34 A, m-Xylene 25 A, p-Xylene 93-3-4, Ethylbenzene/ Styrene 33C, Phthalic Anhydride 34B, Glycerine and Intermediates 58, Aniline and Derivatives 76C, Bisphenol A and Phosgene 81, C1 Chlorinated Hydrocarbons 126, Chlorinated Solvent 48, Chlorofluorocarbon Alternatives 201, Reforming for BTX 129, Aromatics Processes 182 A, Propylene Oxide Derivatives 198, Acetaldehyde 24 A2, 91-1-3, Acetic Acid 37 B, Acetylene 16A, Adipic Acid 3 B, Ammonia 44 A, Caprolactam 7 C, Carbon Disulfide 171 A, Cumene 92-3-4, 22 B, 219, MDA 1 D, Ethanol 53 A, 85-2-4, Ethylene Dichloride/Vinyl Chloride 5 C, Formaldehyde 23 A, Hexamethylenediamine (HMDA) 31 B, Hydrogen Cyanide 76-3-4, Maleic Anhydride 46 C, Methane (Natural Gas) 191, Synthesis Gas 146, 148, 191 A, Methanol 148, 43 B, 93-2-2, Methyl Methacrylate 11 D, Nylon 6-41 B, Nylon 6,6-54 B, Ethylene/Propylene 29 A, Urea 56 A, Vinyl Acetate 15 A.

Chemical Profile of the Chemical Marketing Reporter, Schnell Publishing. (http://www.chemexpo.com), Nonylphenol (2001, July 9), Ethanolamines (1998, May 4), Ethylene Oxide (2001, August 27), Vinyl Acetate (2000, August 21), Propylene Glycol (1998, July 20), Isopropanol (1998, December 28), Cumene (1999, March 22), Acetone (1999, April 5), Acrylic Acid (1999, May 24), Butene-1 (1999, August 16), Butadiene (2000, March 6), Dicyclopentadiene (2001, July 23), Aniline (1999, February 8), Benzene (1999, December 6), Benzoic Acid (1996, November 4), Bisphenol A (2001, November 5), Cyclohexane (2001, May 28), Dipropylene Glycol (1998, July 27), Ethylbenzene (2001, April 30), Ethylene Dichloride (2001, March 19), Glycerine (2001, December 16), Linear Alkyl Benzene (2001, July 30), Methyl Chloride (2000, October 2), Methylene Chloride (2000, October 9), Chloroform (2000, October 16), Monochlorobenzene (1999, May 17), Naphthalene (1999, May 31), Nitrobenzene (1999, February 15), o-Xylene (1998, May 18), p-Xylene (1998, May 11), o-Dichlorobenzene (1996, September 9), p-Dichlorobenzene (1999, June 7), Perchloroethylene (2000, October 30), Phenol (1999, March 29), Phthalic Anhydride (2001, February 12), Propylene Oxide (1998, July 13), Propylene Glycol (1998, July 20), PTA/DTM (1998, July 6), Styrene (2001, May 14), Toluene (2000, October 23), Toluene Diisocyanate (TDI) (1999, April 12), Methyl Diphenyl Diisocyanate (MDI) (2000, January 3), Trichloroethylene (2000, September 25), Vinyl Chloride (1998, February 9), Formaldehyde (2001, May 21), Hydrogen Cyanide (2001, November 26), Methanol (2000, July 31), Methyl Methacrylate (1999, June 28), Maleic Anhydride (2001, February 5), Ammonia (1999, November 29), Acetylene (2001, December), Acetic Acid (2001, February 26), Adipic Acid (1998, June 15), Caprolactam (2001, April 16), Carbon Disulfide (2000, November 20), Cyclohexane (2001, May 28), Ethanol (2000, March 13).

23.1 Introduction

Dyeing

It is difficult if not impossible to determine when mankind first systematically applied color to a textile substrate. The first colored fabrics were probably nonwoven felts painted in imitation of animal skins. The first dyeings were probably actually little more than stains from the juice of berries. Ancient Greek writers described painted fabrics worn by the tribes of Asia Minor. But just where did the ancient craft have its origins? Was there one original birthplace or were there a number of simultaneous beginnings around the world?

While it is difficult to determine just when each respective civilization began to use dyes, it is possible to date textile fragments and temple paintings, which have survived the ensuing centuries. The ancient Egyptians wove linen as early as 5000 B.C., and paintings on tomb walls infer that colored wall hangings were in use by 3000 B.C. By 2500 B.C., dyer's thistle and safflower were used to produce red and yellow shades. Egyptian dyers developed a full range of colors by 1450 B.C.

Another cradle of civilization was the Indian subcontinent where religious and social records dating to 2500 B.C. refer to dyed silk and woven brocades of dyed yarn. Cotton, first cultivated in the Indus valley of Pakistan was woven as early as 2000 B.C. A book written about 300 B.C. included a chapter on dyes. It is believed that systematic dyeing occurred in China as early as 3000 B.C. near the city of Xian in the Hoang (Yellow) River Valley, although there is no conclusive proof. Empress Si-Ling-Chi is credited with the discovery of silk about 2640 B.C. Kermes and indigo were used as dyes as early as 2000 B. C. Fragments of silk have been found in the corrosive patina of bronze swords of the Shang dynasty (15231027 B.C.), but most assuredly these samples are not the oldest.1

The New World was similarly active in developing the textile art. With help from the desert climate in the high Andes of Peru, dyed samples of wool have been preserved and recovered from burial sites. These fragments have been dated to the millennium before the Christian era. The western and southwestern regions of the United States provided homes for the Anasazi, or ancient ones, who dwelt in the region of Mesa Verde National Park in southwestern Colorado, northern Arizona and New Mexico, and eastern Utah. Again the dry climate has helped to preserve samples from these early civilizations.

Very little in the nature, of large, intact textile samples have survived in Europe. Remains of a large woolen robe, the Thorsberg Robe, found in northern Germany and dated prior to 750 B.C. indicate a highly developed dyeing and weaving technology.2 Indirect evidence is more plentiful, for example, a tombstone of a purpurarius, a Roman purple dyer, was found near Parma, northern Italy, and a dyer's workshop excavated in Pompeii. This great center of the Roman Empire was destroyed by the eruption of Mount Vesuvius in 79 A.D. Similar stone vats for dyeing have been excavated in the tells of Israel and in present day Turkey.

The Dark Ages following the fall of the Roman Empire were dark indeed, with little development of the dyer's art. The robes of a number of the monastic orders were brown and blacksurely a dark age. By the end of the 1300s, however, civilization began making the swift and certain strides, which have led to our present level of development. In 1371, the dyers of Florence, a city famous for its Renaissance art, formed a guild, or association of like merchants and craftsmen, which lasted for eleven years. Other guilds were being formed in other centers across Europe. Some of these guilds exist to this day. The Worshipful Company of Dyers was formed in 1471 in London. One of the legacies of this guild is a Publications Trust, which has underwritten the publication of a number of books on dyeing in cooperation with The Society of Dyers and Colourists.

The art and craft of dyeing was largely passed down from father to son or from craftsman to apprentice by word of mouth and example until the early 1500s. The *Plictho* of Gioaventura Rosetti, a Venetian armory superintendent, is believed to be the first published book on dyeing. It certainly is the oldest surviving European text to have come down to us in the twenty-first century. Five known Italian editions were published between 1548 and 1672. A French edition appeared in 1716. It is interesting that no known English translation was made until 19683 when Sidney Edelstein of Dexter Chemical Company and Hector Borghetty collaborated to reproduce a facsimile of the original 1548 edition along with a complete translation into English. During his extensive travels, Rosetti collected dyeing recipes and processes used in the flourishing city states of Venice, Genoa, and Florence. He published: *Plictho de L'arte de Tentori che insegna tenger pani telle banbasi et sede si per larthe magiore come per la comune* or *Instructions in the Art of the Dyers, which Teaches the Dyeing of Woolen Cloths, Linens, Cottons, and Silk by the Great Art as Well as by the Common* or simply *Instruction in the Art of Dyeing*. The book was divided into four sections: the first and second sections were devoted to the dyeing of wool, cotton and linen; the third to the dyeing of silk and the use of fugitive colors; and the fourth to the dyeing of leather and skins. Approximately 160 complete recipes were preserved in the first three sections. Edelstein and Borghetty labored diligently in determining the meanings of terms in recipes written in the dialect of 16th century Italy.

Synthetic Dyes

The father of modern synthetic dyes was William Henry Perkin (18381907), who synthesized Mauve, or Aniline Purple, in 1856. The story behind this great story bears telling. William's father was a builder who wanted him to become an architect, but like many others, Perkin did not follow his father's chosen profession. Perkin studied at the City of London School where he became interested in chemistry at the age of 12. A teacher, Mr. Hall, gave him work in the laboratory, which in turn, inspired Perkin to follow his natural curiosity. At age 15, Perkin entered the Royal College of Science and listened to the lectures of the great German chemist, August Wilhelm von Hofmann (18181892). He was granted an assistantship under von Hofmann at age 17. Since his work did not allow time for his own research, he set up a separate laboratory at home and it was there that he discovered aniline purple, the first dyestuff to be commercially produced. Another dye, based on naphthalene, and prepared in collaboration with Arthur H. Church, actually preceded aniline purple, but was not commercially produced before aniline purple. Aniline purple was discovered at this home during Easter vacation while looking for quinine, an antimalarial drug. After oxidizing aniline with potassium dichromate and getting a black precipitate, extraction with ethanol gave a brilliant purple solution. Almost immediately, he sent a sample of this dye to a dyer in Perth with a request to dye silk fabric. The dyer's report read: "If your discovery does not make the goods too expensive, it is decidedly one of the most valuable that has come out for a long time." Trials on cotton were not as successful because the need for a mordant was not realized. Perkin later reported, "The value of mauve was first realized in France, in 1859. English and Scotch calico printers did not show any interest in it until it appeared on French patterns, although some of them had printed cloth for me with that colour."4

Since that beginning, thousands of dyes have been synthesized; some 15002000 are commercially successful today. Until 1884, however, all synthetic dyes required a mordant to give acceptable wash fastness on the textile substrate. In 1884, Böttiger produced Congo Red, which could dye cotton directly without a mordant. These dyes were commonly called direct dyes. In order to improve washfastness, the path taken in some synthetic dye chemistry was to build the dye from two or more components, directly in the fiber, or "in situ." By building a large molecule without solubilizing groups within the fiber, washfastness was markedly improved. The first practical development along these lines was by A. G. Green who synthesized primuline; a dye, which because of poor light fastness was not commercially important, but later, led the way to many important commercial dyes. Table 23.1 lists a number of classes of dyes along with the date of discovery.

TABLE 23.1 Dyesin Order of Discovery

Basic	Mauve or Aniline Purple, Perkin, 1856
	Fuchsin, Verguin, 1859
Acid	Alkali Blue, Nicholson, 1862
Vat	Alizarin, Gräbe & Liebermann, 1868
	Indigo, von Bayer, 1880 (discovered structure and synthesized indigo) Indanthrene, Böhn, 1901
Direct	Congo Red, Böttiger, 1884
Direct developed	Primuline coupled with beta-naphthol, Green, 1887
Sulfur	Vidal, 1893
Azoics	Zitscher & Laske, 1911
Disperse	A simple azo dye, 1920s
Phthalocyanine	Linstead & Diesbach, 1928/29
Reactive	commercialized, I.C.I., Rattee and Stevens, 1956

World wide, 80 percent of all dyes go into textiles, 20 percent into paper, leather, food, and the like.

The Development of the U.S. Dyestuff Industry

The natural dyes industry was more than just a cottage industry in Colonial America. Indigo was a very important cash crop in South Carolina among the coastal islands and for some distance inland. Plantations existed well into the early 1900s despite the growth of the synthetic dyestuff industry.

The modern synthetic dye industry in the United States dates from World War I. However, in 1864, Thomas Holliday of Great Britain, and in 1868, the Albany (NY) Aniline Company with participation of Bayer of Germany began coal-tar dye manufacture. In the early 1900s, most synthetic dyes used in the United States were imported from Germany and Switzerland. With the outbreak of World War I, the British naval blockade of Germany prevented export of dyes from Europe. In spite of the blockade, the German submarine, *Deutschland*, ran the British blockade and sailed into American ports twice with dyestuffs and drugs. The Germans needed critical war material and export moneys; the United States and others needed dyes. Ironically, in 1914, German dyes were used by French dyers to dye the official French Army uniforms. The outbreak of war and ensuing blockade showed the United States how important dyes were to the American economy. Several companies began investigative work, which would lead to dye synthesis; they found that dyestuffs were very difficult to makethe chemistry was much more complex than imagined. A real boost to the U.S. industry came after World War I, when the German patents were given over to the Allies via the Alien Property Custodian. According to Lehner, DuPont reportedly spend $43 million, a tremendous sum of money in the early 20th century, before ever showing a profit.5 Obviously, only financially strong companies could afford to enter the business. The early pioneers included Allied (formed by merging five companies), American Cyanamid, and DuPont, to name only a few who survived to become major factors later in the 20th century. In 1938, others included Dow, German-owned General Aniline and Film (GAF), and Swiss-owned Cincinnati Chemical Company (Ciba, Geigy, and Sandoz). In the 1960s, 5060 percent of all U.S. manufacturing was in the hands of four principal U.S. companies:

1. Allied Chemical (later sold to Bayer of Germany and to independent investors as Buffalo Color, 1977)
2. American Cyanamid
3. GAF, the result of the breakup of the German cartel, I. G. Farben, which was nationalized during World War I, sold to BASF in 1978
4. DuPont (sold in 19801981 with various lines going to Crompton & Knowles (C & K), Ciba, and Blackman-Uhler).

Today, there are at least 42 dyestuff manufacturers, distributors and repackaging agents in the United States.6 Of the major companies, Swiss and German-based companies tend to dominate the U.S. market. Those companies include:

1. Ciba, formerly Ciba-GeigySwitzerland
2. Clariant (split off from Sandoz in 1995)Switzerland
3. DyStarformed by the merger of Hoechst and Bayer (1995), acquisition of BASF textile colors which include the former Zeneca, and Mitsubishi of JapanGermany
4. C & K sold to Yorkshire Group plc and is now Yorkshire Americas

There is no major surviving U.S.-based company. Raghavan7 and Mock8 give an interesting description of these mergers and the reasons behind them. Most of the international companies have limited manufacturing facilities in the United States and major facilities in other countries where environmental laws are not as stringent or where the parent companies have a modern integrated low pollution facility. These facilities in the United States minimize the tariffs paid and also allow quicker response to the market place. Ciba has a manufacturing facility in St. Gabriel, LA; Clariant in Martin, SC; DyStar near Charleston, SC; and Yorkshire Americas at Lowell, NC.

Today well over 1500 dyes are produced in commercial quantities, although only a select handful in each class are the true "workhorse" colors found in virtually every dyehouse dyeing a particular substrate for a particular end-use. Approximately two thirds of the dyes and pigments consumed in the United States are used by the textile industry. One sixth of the dyes and pigments are used for coloring paper; and the rest are used chiefly in the production of organic pigments and in the dyeing of leather and plastics.

Dyes are cataloged and grouped under a set of rules established by the Colour Index committee, consisting of representatives from the Society of Dyers and Colourists (SDC), Bradford, England, and the American Association of Textile Chemists and Colorists (AATCC), Research Triangle Park, NC. Table 23.2 shows how over 9000 dyes are enumerated in the current Index.9 The *Colour Index*, now in its fourth edition, is updated periodically with newly released information and is available in book form and on CD-ROM. Volumes 13, published in 1971, contain the C. I. name and number, chemical class, fastness properties, hue indication, application and usage. Volume 4, also published in 1971, contains the structures of all disclosed structures. Volume 5, last published as part of Volume 9 in 1993, contains the commercial names of all known dyes and pigments. Volumes 6, 7, and 8 are supplements with updates to information in Volumes 14 up to 1976, 1981, and 1987, respectively. An online version of the *Colour Index*, Fourth Edition, containing a significant number of new entries is now available.

TABLE 23.2 Dyes Listed in the Colour Index

Shade	Yellow	Orange	Red	Violet	Blue	Green	Brown	Black	Sum
Acid	255	178	439	131	356	121	453	236	2169
Direct	171	121	263	104	303	100	242	189	1493
Disperse	243	155	371	102	371	9	27	31	1309
Reactive	188	123	252	44	250	27	48	46	978
Pigment	194	69	262	50	74	52	42	32	775
Basic	108	68	115	50	164	6	23	11	545
Vat	48	29	61	17	74	44	84	65	405
Sulfur	23	5	14		20	37	96	18	213
Mordant	65	47	95	60	82	36	92	96	573
Natural	26	4	34			5	13		82
Solvent	174	105	232	49	134	33	58	52	837
Food	15	8	17		5	4	3	3	55
Total									9475

In addition to the *Colour Index*, AATCC publishes a *Buyer's Guide* annually in July.6 Part A lists dyes, pigments, and resin-bonded pigment colors available from companies who choose to list this information.

The textile industry uses a large number of dyestuffs from each of the dye categories, the choice depending on the shade, fiber and dyeing process, end use of the textile product, requirements for fastness, and economic considerations. To provide an understanding of the interrelationships that exist among the various dye classes and fiber types, a brief survey of the major fibers follows.

23.2 Textile Fibers

In this survey, commercially important textile fibers are grouped by their origin. First there are the natural fibers from plant sources, cotton and flax, and those from animal sources, wool and silk. A second group consists of those fibers that are regenerated or chemically modified natural materialsthe rayon and acetate fibers. The final group consists of synthetic fibers, which include polyester, nylon, acrylics, polyolefins, and elastane.

Natural Fibers

Cotton

Cotton fibers are comprised mainly of cellulose; a long-chain polymer of anhydroglucose units connected by ether linkages. The polymer has primary and secondary alcohol groups uniformly distributed throughout the length of the polymer chain. These hydroxyl groups impart high water absorption characteristics to the fiber and can act as reactive sites. The morphology of the cotton fiber is a complex series of reversing spiral fibrils. The fiber in total is a convoluted collapsed tube with a high degree of twist occurring along the length of the fiber. This staple fiber occurs in nature in lengths of 1/22 in., depending on the variety and growing conditions. The diameter ranges from 1621 µm (1 µm is 1 × 10-6 m).

Flax

Flax is also a cellulosic fiber that has a greater degree of crystallinity than cotton. The morphology of flax is quite different from that of cotton. Flax fibers have a long cylindrical shape with a hollow core. The fibers range in length from 1/22 to 1/22 inches, with a diameter of 1216 µm. Flax staple is comprised of bundles of individual fibers. Historians believe that flax was among the first fibers to be used as a textile fiber. In recent years, its commercial importance as a textile fiber has decreased significantly.

Wool

Wool fibers are comprised mainly of proteins: the polypeptide polymers in wool are produced from some 20 alpha-amino acids. The major chemical features of the polypeptide polymer are the amide links, which occur between the amino acids along the polymer chain, and the cystine (sulfursulfur) cross-links, which occur in a random spacing between the polymer chains. The polymer contains many amine, carboxylic acid, and amide groups, which contribute in part to the water-absorbent nature of the fiber.

The morphology of wool is complex. There is an outer covering over the fiber, the cortical. There are also overlapping scales having a ratchet configuration that causes shrinkage and felting. The coefficient of friction in wool fibers is vastly different between the tip and the root, depending on which way the scales point. Wool can be made washable by chemically abrading the scales or coating the fibers with another polymer.

Wool fibers are not round but are oval in cross section. The cortical cells constitute the major component of the fiber, and are aligned along the axis of the fiber. There is a medulla section at the center region of the fiber. Each fiber has a bicomponent longitudinal crystalline arrangement. One side of the fiber contains alpha-keratin crystalline regions, and the other contains beta-keratin crystalline regions. Alpha-keratin and beta-keratin have different moisture absorption characteristics, and this difference is what gives wool fibers crimp and springiness. It is also the reason why wool fibers kink in conditions of, high humidity.

Wool fibers are sheared from about 30 major sheep breeds. The length of the wool fibers varies from 1 to 14 in. and depends on the breed, the climate, and the location on the sheep's anatomy. The fibers can be very fine to very coarse, ranging from 1050 µm in diameter. The longer, coarser fibers normally are used for woolen fabrics, whereas the shorter, finer ones are used for worsted fabrics.

Silk

Silk, like wool, is a protein fiber, but of much simpler chemical and morphological makeup. It is comprised of six alpha-amino acids, and is the only continuous-filament natural fiber. Historians claim that silk was discovered in China in 2640 B.C. Silk fiber is spun by the silkworm as a smooth double-strand, each part having a trilobal cross section. This configuration helps give silk its lustrous appearance. The fiber is unwound from the cocoon the silkworm spins as it prepares its chrysalis. The filaments are smooth and have no twists in their length, which can vary from 300 to 1800 yards. The diameter of silk is very fine, ranging from 2 to 5 µm. Because of the labor-intensiveness of sericulture and subsequent preparation of the fiber, silk remains a luxury fiber.

Regenerated Fibers

Rayon

Viscose rayon, like cotton, is comprised of cellulose. In the manufacturing process, wood pulp is treated with alkali and carbon disulfide to form cellulose xanthate. Subsequently, the reaction mass is forced through a spinneret and precipitated in an acid coagulation bath as it is formed into a continuous filament. The fiber has a round striated cross section. Rayon staple is made by "breaking" the continuous strands into staple-length fibers. Viscose rayon is conventionally produced in diameters varying from 9 to 43 µm.

Acetate

Triacetate and diacetate fibers are manufactured by the chemical treatment of cellulose obtained from refined wood pulp or purified cotton lint. Most of the hydroxyl groups are acetylated (esterified) by treating the cellulose with acetic acid. This determines the chemical configuration of triacetate. Acetate or diacetate is made by the saponification of one of the acetylated groups, thus restoring a hydroxyl to each cellulosic monomer unit. Theoretically, then, diacetate has two acetylated groups in each glycoside unit. The conversion of the hydroxyl groups causes these fibers to be hydrophobic and changes the dyeing characteristics drastically from those of the normal cellulosic fibers. Triacetate fibers are spun by mixing the isolated reaction product (flake) with methylene chloride and alcohol. The spinning solution (dope) is forced through a spinneret and dry spun into continuous filaments.

An alternate way of wet spinning is also possible. Acetate fibers are spun by mixing the isolated reaction product with acetone and water. The spinning solution is formed into filaments by evaporating the solvent and coagulating the acetate in a manner similar to that for triacetate (i.e., by the dry-spinning method).

Synthetic Fibers

Nylon

In 1939 the DuPont Company introduced the first truly synthetic textile fiber. Wallace Carothers invented nylon as a result of his basic research into polymer science. Chemically, nylon is a polyamide fiber. The two major types of nylon polymer are used in textiles: type 66, which is made by using hexamethylene glycol and adipic acid, and type 6, which is made by polymerizing epsilon-caprolactam. Nylon fibers are made by melt spinning the molten polymer. The result is a continuous filament fiber of indeterminate length. It is spun in many deniers, with its diameter varying from 10 to 50 µm. The cross section usually is round, trilobal, or square with hollow channels when used as carpet fiber.

Polyester

Polyester is made by the polymerization reaction of a diol and a diester. The main commercial polymer is formed by a condensation reaction using ethylene glycol and terephthalic acid. Fibers are formed by melt spinning. Commercially introduced in 1953 by the DuPont Company as Dacron, polyester fibers have high strength, and very low moisture absorbance. The fiber is usually spun with a round cross section. Polyester is the most used synthetic fiber around the world.

Acrylics

The DuPont Company introduced the first commercial acrylic fiber, Orlon, in 1950. Acrylics are made from the polymerization of acrylonitrile and other comonomers to allow dyeability and to open the internal structure. The fibers are produced by either solvent spinning (Orlon), or wet spinning (Acrilan). In the solvent-spinning process, the polymer is dissolved in a low-boiling liquid solvent such as dimethyl formamide and extruded in a warm air chamber. In wet spinning, the polymer is dissolved in a suitable solvent, extruded into a coagulation bath, dried, crimped, and collected. Although the acrylic fibers are extruded as continuous filaments, they are subsequently cut into staple-length fibers. Acrylics have found a niche market as a substitute for wool or in wool blends (blankets, sweaters, etc.) and in awnings and boat covers. The cross section of the filament varied among manufacturers, Orlon having a dog-bone configuration and Acrilan having a lima-bean shape. Acrylic fibers are quick drying and wrinkle resistant.

Polyolefins

Polyolefin fibers are produced from the polymerization of ethylene or propylene gas. The catalysis research of Ziegler and Natta led to the development of these polymers to form crystalline polymers of high molecular weight. Hercules Inc. produced the first commercial fibers in 1961. The fibers made from these polymers are melt-spun. The cross sections are round, and the fibers are smooth. They have extremely low dye affinity and moisture absorbance. Colored fiber is normally produced by mixing pigments in the melt polymer prior to extrusion.

Elastane

The DuPont Company commercialized the first manufactured elastic fiber, Lycra, in 1958. Originally categorized as a spandex fiber, the name "elastane" is becoming more common around the world. This specialty fiber is described as a segmented polyurethane that contains "hard" and "soft" segments; their ratio determines the amount of stretch built into the fiber. Elastane fibers are formed by dry spinning or solvent spinning. The continuous filaments can be coalesced multifilaments or monofilaments, depending on the manufacturer.

Because most dyeings are applied from water solutions or dispersions, the effect of water absorption by the fiber is an important criterion. Table 23.3 shows the hydrophobic/hydrophilic characteristics of the important fibers. The cellulosic and natural fibers are the most hydrophilic, and polyolefin is the most hydrophobic.

TABLE 23.3 Hydrophobic/Hydrophilic Characteristics of Various Fibers

Fiber	Moisture Content, % (at 65% RH -70°F)	Water Retention, % (Weight Change in Water)	Swelling, % (Volume Change in Water)
Acetate	6.5	2025	NA
Acrylic	1.01.5	4.56	2
Cotton	7	45	45
Elastane	0.31.5	low	low
Nylon	3.05.0	912	13
Polyester	0.30.5	35	0.5
Olefin	0.010.1	Very low	Very low
Triacetate	3.2	1218	NA
Viscose	13	90100	95
Wool	1315	42	42

Microdenier Fibers

The first commercial production of microfiber in the United States was in 1989 by the DuPont Company. Today microfibers are produced in a variety of synthetic fibers (i.e., polyester, nylon, acrylic, etc.) A microfiber is a fiber that is less than one denier per filament. Yarns made from microdenier filaments are able to give silk-like hand to fabrics.

< previous page

next page >

23.3 Classification of Dyes

In this section, the various classes of dyes are described, with emphasis placed on structural features that serve as the basis for their classification and determine the substrates for which they have affinity. Specifically, acid, azoic, basic, direct, disperse, reactive, sulfur, and vat dyes will be covered. With regard to the application of organic dyes to textiles, it is well known that dyeing of fibers with textile dyes is accomplished in four steps, namely exhaustion, diffusion, migration, and fixation. In step 1, dissolved dye molecules move from the dyebath to the fiber surface and in step 2, the surface-adsorbed dye molecules move into the amorphous regions of the fiber. In step 3, dye molecules are uniformly distributed within the polymer matrix, and in step 4, dye molecules interact with groups in the polymer chain via primary or secondary valency forces. Dyepolymer interactions can involve ionic bonding (e.g., acid dyes on nylon or wool), covalent bonding (e.g., reactive dyes on cotton), mechanical entrapment (e.g., vat dyes on cotton), secondary valency forces (direct dyes on cotton), or solidsolid solution (e.g., disperse dyes on polyester).

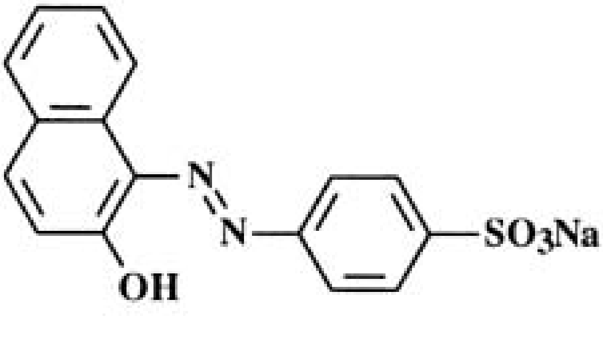
Acid Dyes

Acid dyes owe their name to the fact that they are generally applied to textile fibers from dyebaths containing acid. Most have one or two sodium sulfonate (SO3Na) groups and, consequently, are water soluble and capable of bonding with fibers having cationic sites (cf. Fig. 23.1). They give a wide range of bright colors on textiles, especially when monoazo and anthraquinone structures are used.

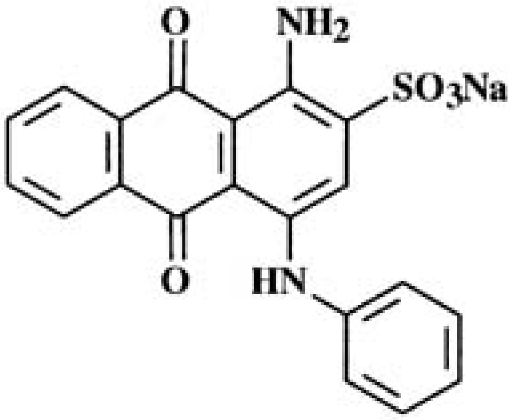
Fig. 23.1 Ionic bond formation between nylon and an acid dye.

$$\text{Nylon–NH}_3^{+}\ {}^{-}\text{O}_3\text{S–Dye}$$

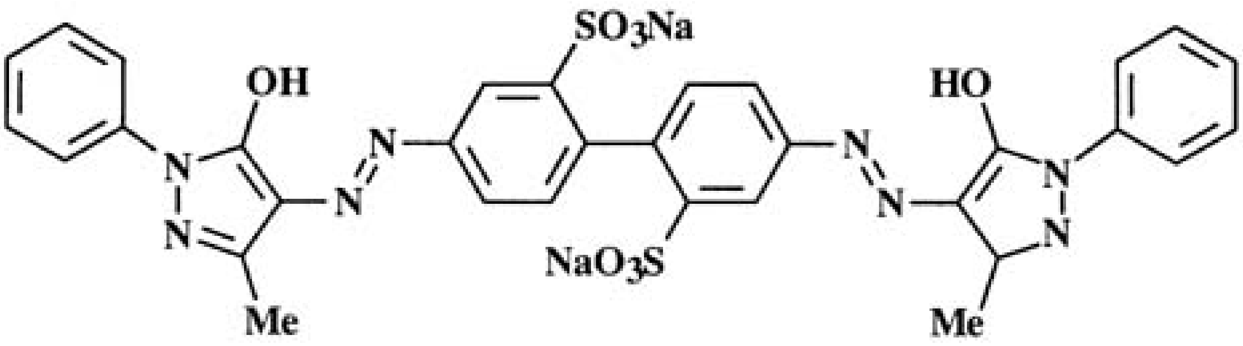
Acid dyes vary widely in molecular structure and in the amount of acid required during the dye application step. They include relatively low molecular weight dyes such as C.I. Acid Orange 7 and C.I. Acid Blue 25, both of which are readily applied to polyamide and protein fibers and are known as *level dyeing acid dyes*. As the name suggests, these dyes are characterized by good migration and, therefore, level well with time. In addition they give reasonably good lightfastness and barré coverage. The application of these dyes to nylon and wool employs weak acid and strong acid, respectively. For applications requiring good washfastness, *milling and dyes* or *supermilling acid dyes* are employed. Both of these types of dyes have poor barré coverage properties, however. Dyes of the former type are applied from weakly acidic dyebaths while the latter are applied at neutral pH, with molecular size increasing as acid strength decreases. Examples of milling acid dyes are C.I. Acid Yellow 42 and C.I. Acid Red 151, while supermilling acid dyes include C.I. Acid Blue 138 and C.I. Acid Red 138.



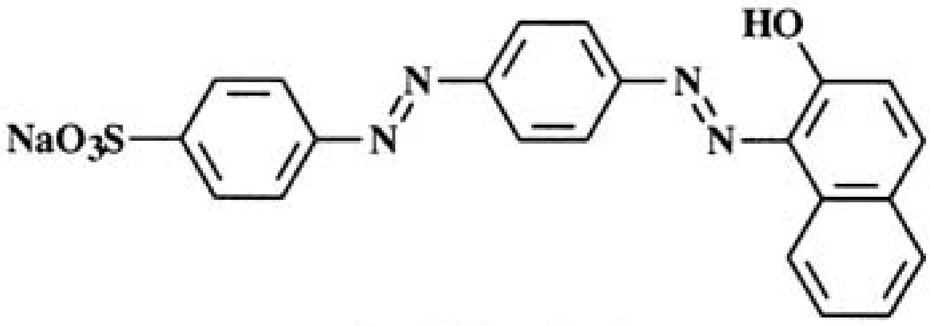
Acid Orange 7



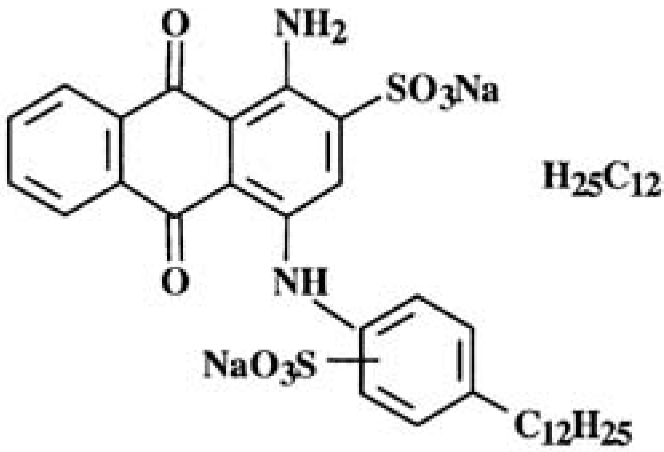
Acid Blue 25



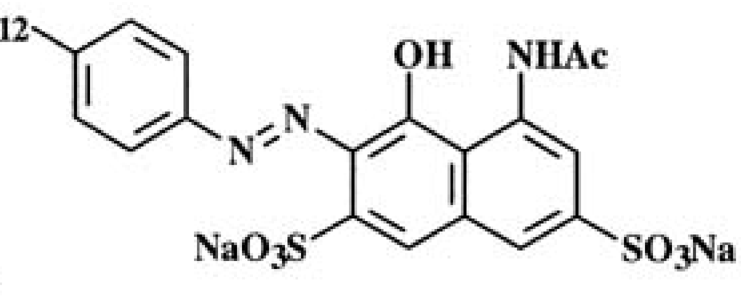
Acid Yellow 42



Acid Red 151



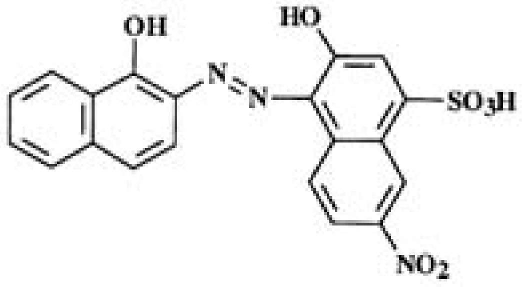
Acid Blue 138



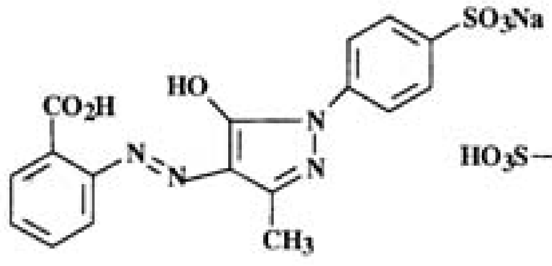
Acid Red 138

Acid dyes include metal-complexed azo structures, where the metals used are cobalt, chromium, and iron.[10a, b] Examples are 1:1 and 1:2 chromium complexes and 1:2 cobalt complexes, where the numbers employed represent the ratio of metal atoms to dye molecules. Metal-complexed dyes can be formed inside textile fibers by treating suitably dyed fibers with a solution containing chromium ions.11 In this case, the metal-free forms of these azo dyes are known as *mordant dyes* and mainly contain *ortho*, *ortho′*-bis-hydroxy or *ortho*-carboxy, *ortho′*-hydroxy groups (e.g., C.I. Mordant Black 11, Mordant Yellow 8, and Mordant Orange 6). When the metal complexes are formed prior to the dye application process, the resultant dyes are known as *pre-metallized acid dyes* and vary in the acid strength required in the application step.12 The 1:1 chromium complexes (e.g., C.I. Acid Blue 158) are stable only in very strong acid, making them suitable for wool but not nylon. Neutral dyeing pre-metallized acid dyes contain SO2NH2 or SO2CH3 groups in lieu of SO3Na groups (see Acid Black 172 vs. Acid Red 182). In this case, dye-fiber fixation occurs because the combination of trivalent metal ion (e.g., Cr3+) and four attached negatively charged ligands gives the complex a net negative charge.

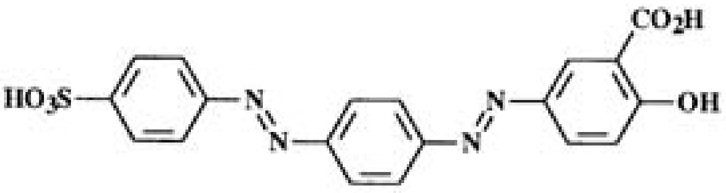
Metallization of azo dyes enhances lightfastness, reduces water solubility, causes a bathochromic shift in color, and dulls the color. Iron complexes typically have brown colors (e.g., C.I. Acid Brown 98) and are most often used to dye leather.



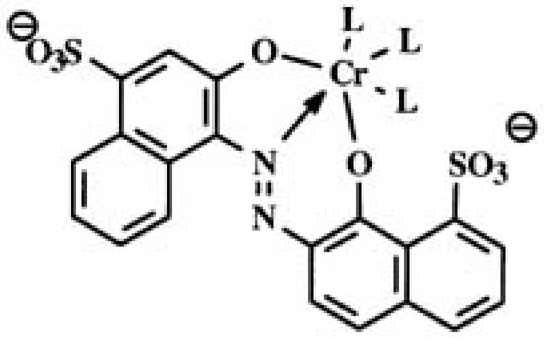
Mordant Black 11



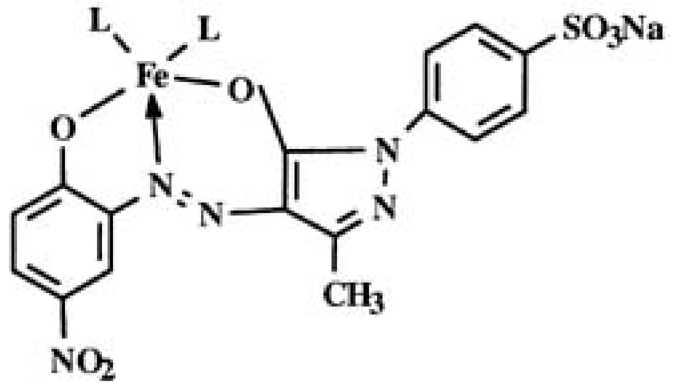
Mordant Yellow 8



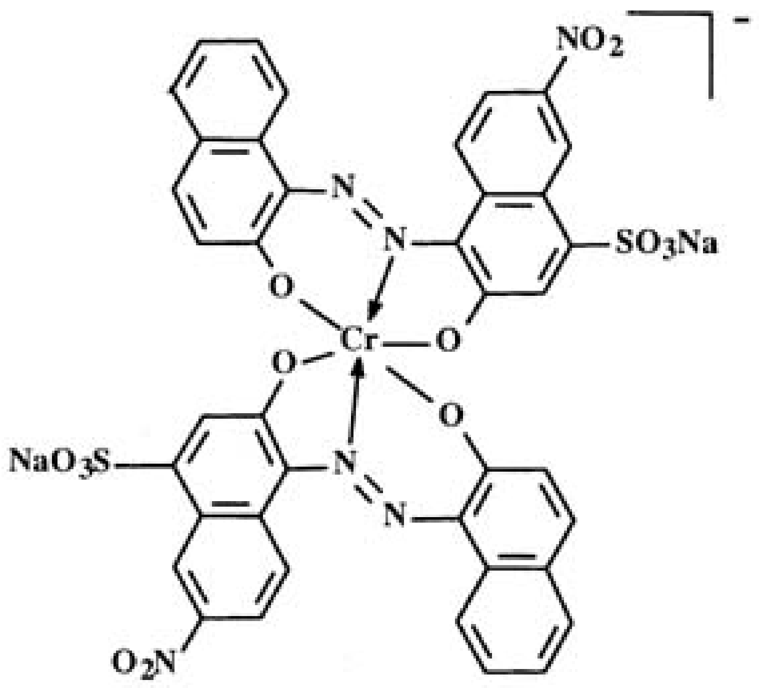
Mordant Orange 6



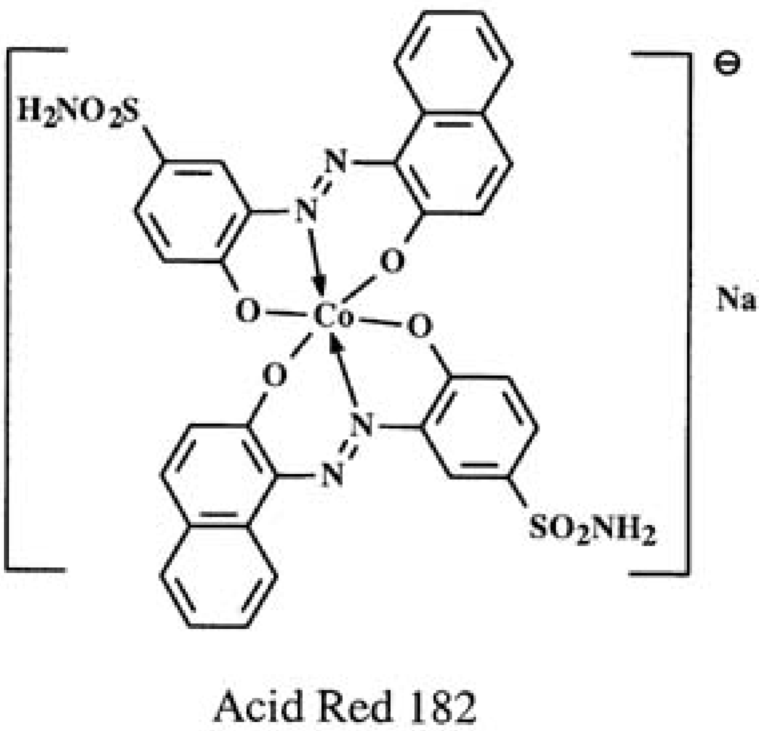
Acid Blue 158



Acid Brown 98



Acid Black 172



Acid Red 182

Azoic Dyes

Azoic dyes are mainly bright orange and red monoazo dyes for cotton, with dull violet and blue colors also possible.13 They are water insoluble and consequently give high washfastness. They are also referred to as azoic combinations rather than "dyes" because they do not exist as colorants until their constituent parts are combined inside the pores of cotton fibers.14 They are quite useful for printing on cotton and often give good lightfastness in heavy depths. Their bleachfastness is better than direct and sulfur dyes and crockfastness requires efficient soaping after dye formation. The formation of these dyes requires two constituentsan azoic coupling component and an azoic diazo component, examples of which are shown in Figs 23.2 and 23.3. The azoic coupling components are beta-naphthol and β-oxynaphtholic acid (BON acid) derivatives and the azoic diazo components are substituted anilines.

Fig. 23.2 Structures of C.I. Azoic Coupling Components 18 (a), 12 (b), 15 (c), and 25 (d).

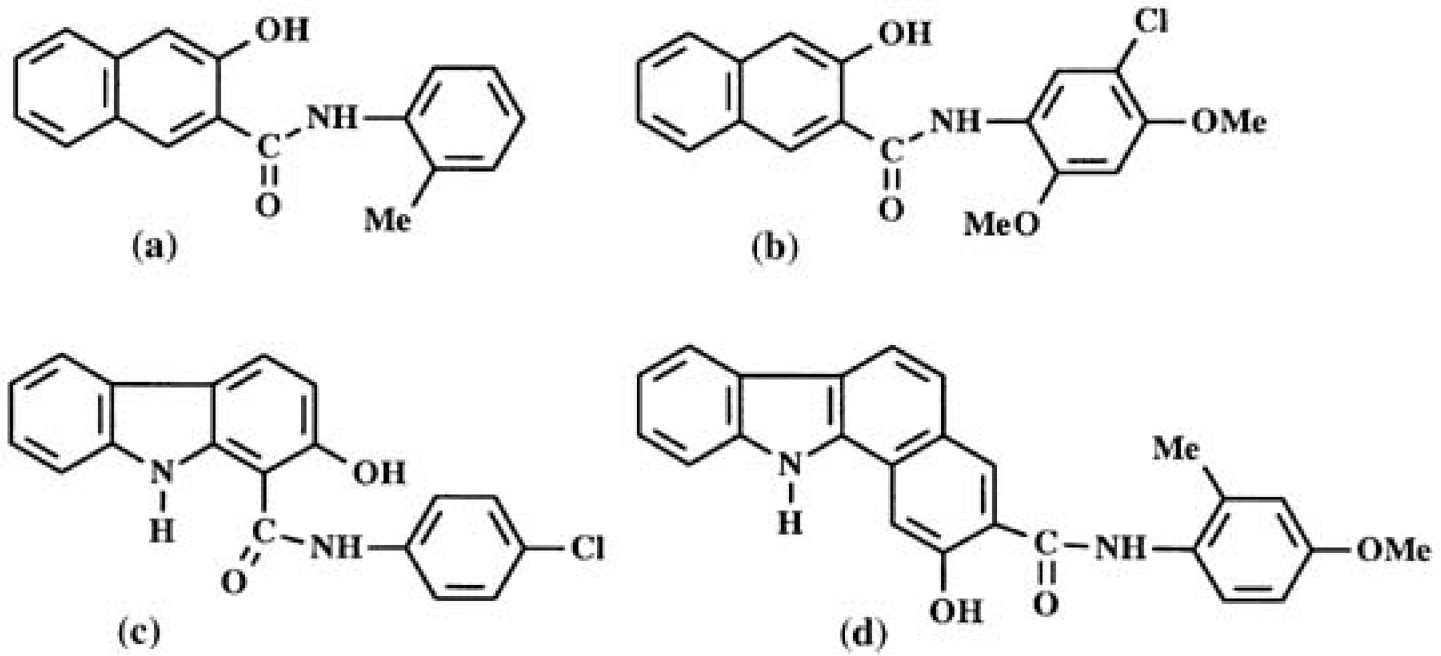
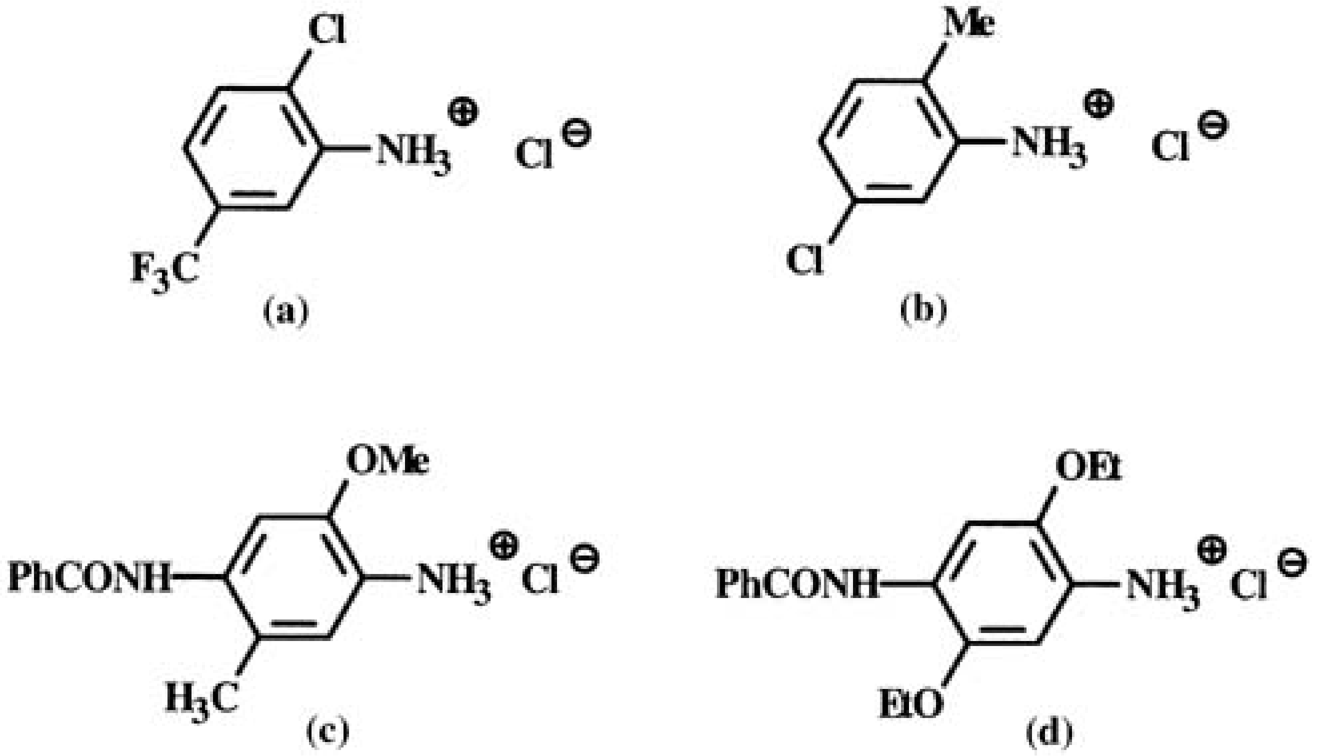
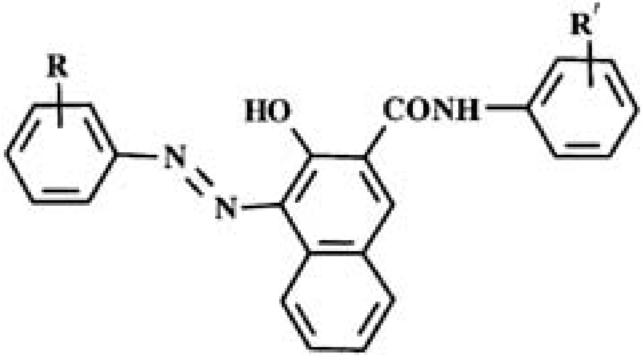


Fig. 23.3 Structures of C.I. Azoic Diazo Components 49 (a), 32 (b), 41 (c), and 20 (d).



Azoic dyes are also known as naphthol dyes, since all employ a naphthol component in their formation, and can be produced in batch or continuous processes. Since they have a limited shade range, they are best known for their ability to provide economical wetfast orange and red shades on cotton. A generic azoic dye structure is shown in Fig. 23.4.

Fig. 23.4 Generic structure for azoic dyes, where R and R′ = alkyl, alkoxy, halo, and nitro groups.

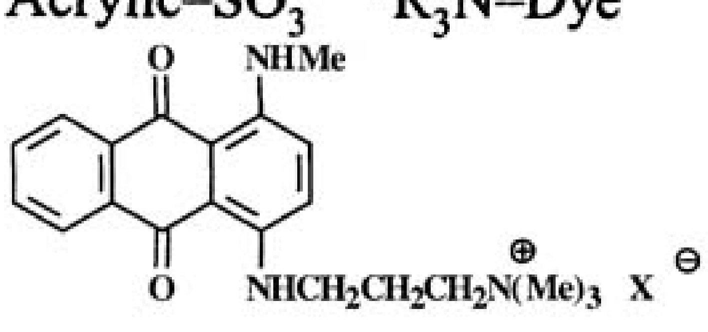


Basic or Cationic Dyes

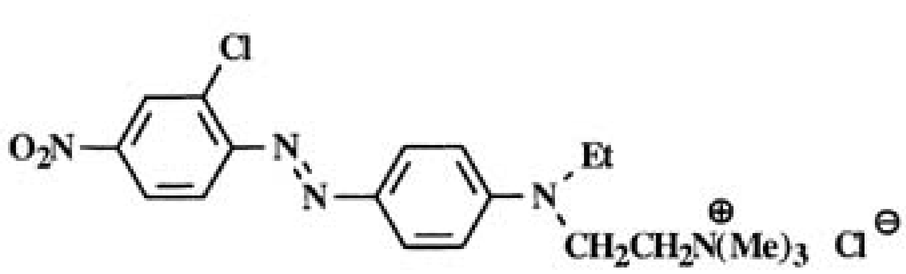
Basic dyes were developed to dye negatively charged acrylic fibers, forming ionic bonds (Fig. 23.5).15 They owe their name to the presence of aromatic amino (basic) groups, and in this case a cationic amino group is present. Generally, they have excellent brightness and color strength, especially among the triarylmethane types; however their lightfastness is often low, when they are applied to fibers other than acrylics. Basic dyes include those containing mobile and fixed cations, examples of which are C.I. Basic Blue 22, C.I. Basic Red 18 and C.I. Basic Green 4. The triarylmethane dye (Basic Green 4) has a mobile cation that produces a pair of extreme forms and resonance stabilization.

Fig. 23.5 Ionic bond formation between an acrylic fiber and a cationic dye.

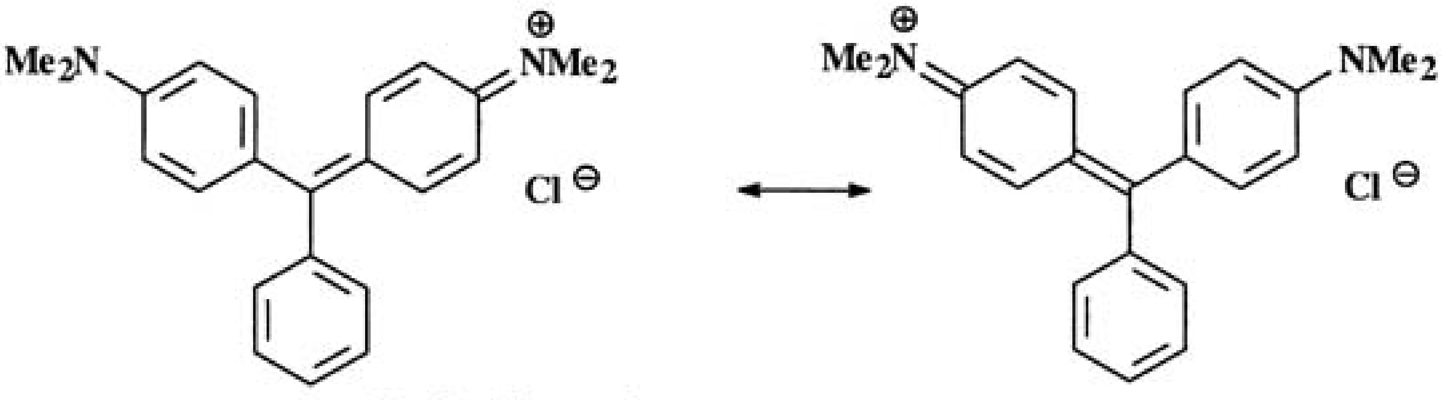
$\text{Acrylic–SO}_3^{-}\ {}^{+}\text{R}_3\text{N–Dye}$



Basic Blue 22



Basic Red 18



Basic Green 4—resonance structures

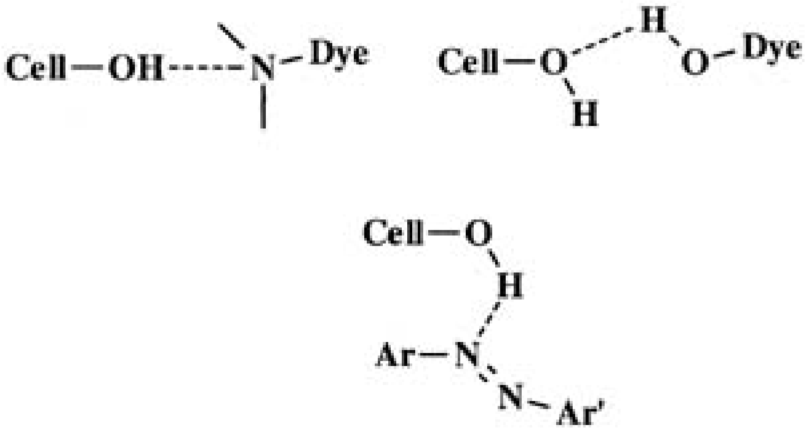
Basic dyes are applied from weakly acidic dyebaths (pH = 4.55.5) and often require the use of anionic or cationic retarding agents, to control the rate of dye strike and give level dyeings. Suitable retarding agents either form a weak bond with dyesites along the polymer chain or interact with the dye in the dyebath. In the former case, the cationic retarder employed is displaced by the dye as dyeing progresses because the dye has higher affinity for the fiber. When anionic retarding agents are used, the dye retarder bond is broken by increasing the dyebath temperature, giving controlled release of dye molecules to facilitate levelling.

To help determine which basic dyes can be combined for shade matching, key dyebath parameters have been developed.16 The first pertains to the dyes themselves and is known as the *combinability constant* (*k*). This value provides a measure of how fast a basic dye will dye the fiber, and the dyes are rated on a scale of 1 (fast) to 5 (slow). The second parameter pertains to the fiber type involved and is known as the *fiber saturation value* (*S*F). This value provides an indication of how much dye the fiber will hold at the saturation point. In this regard, the dye used is C.I. Basic Green 4 and typical saturation levels are 1.04.0 based on the weight of the fibers (owf) for light to deep dyeings. The third parameter is the *dye saturation factor* (*f*), which is a measure of the capacity of a basic dye for saturating a fiber. This factor is influenced by the molecular size and purity of the dye. In this case, the goal is to avoid placing more dye on the fiber than the number of dye sites, and the standard is C.I. Basic Green 4 (*f* = 1 percent).

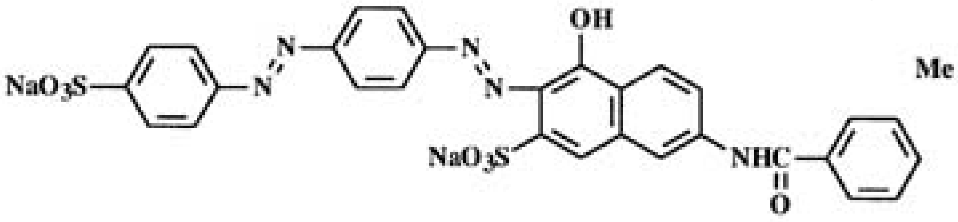
Direct Dyes

Direct dyes are anionic colorants that have affinity for cellulosic fibers.17 They were the first dyes with the ability to dye cotton in the absence of a mordanting agent, giving rise to the term *direct-cotton dyes*. Like acid dyes, direct dyes contain one or more SO_3Na groups, making them water soluble. Unlike acid dyes, they interact with cellulose chains via secondary valency forces (e.g., H-bonding and dipoledipole interactions), as illustrated in Fig. 23.6. The combined effects of these rather weak forces and sulfonated structures cause direct dyes to have low intrinsic washfastness.

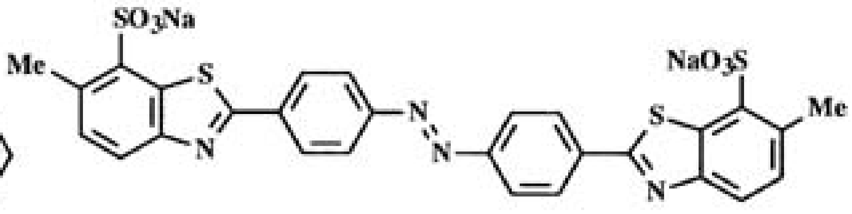
Fig. 23.6 H-bonding interactions between cellulose and polar groups in direct dyes.



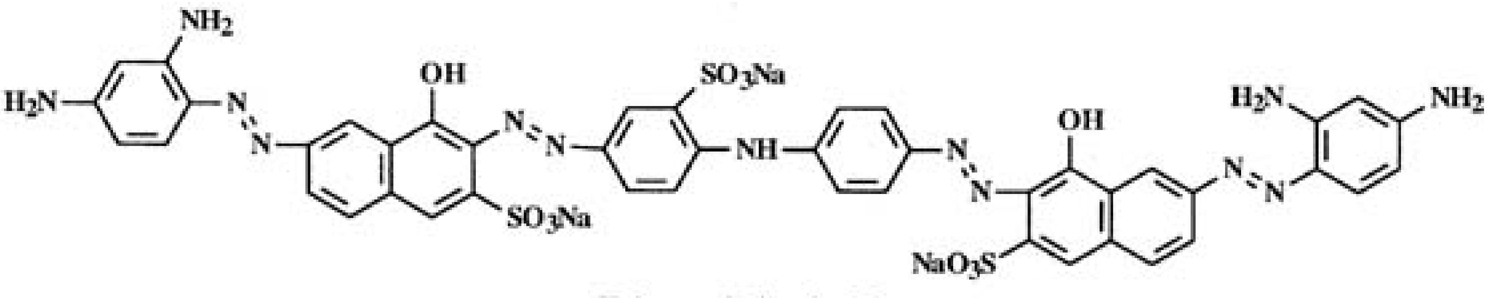
Direct dye structures are based on four main chromophoresazo (e.g., C.I. Direct Red 81, C.I. Direct Yellow 28, and C.I. Direct Black 22), stilbene (e.g., C.I. Direct Yellow 12 and C.I. Direct Yellow 11), oxazine (e.g., C.I. Direct Blue 106 and C.I. Direct Blue 108), and phthalocyanine (e.g., C.I. Direct Blue 86 and C.I. Direct Blue 199). About 82 percent of all direct dyes have disazo or polyazo structures, with stilbene and monoazo structures occupying about 5 percent each and thiazole, phthalocyanine, and dioxazine structures covering the remaining few percent.18



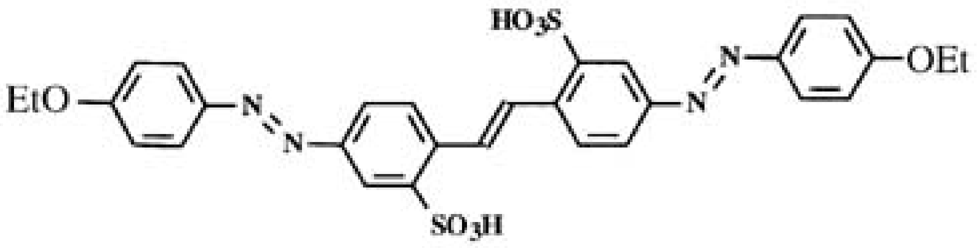
Direct Red 81

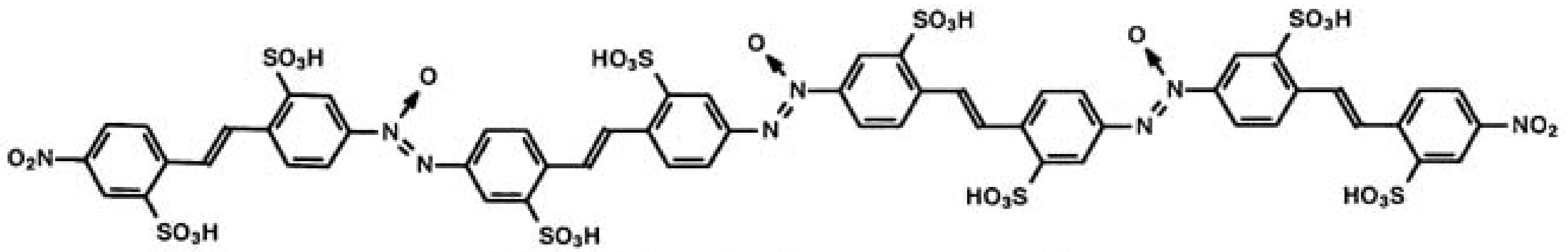


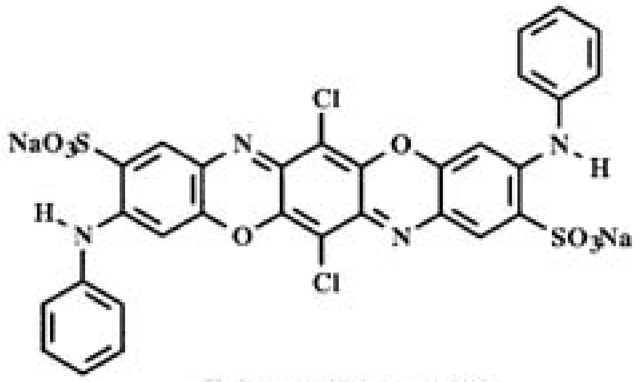
Direct Yellow 28

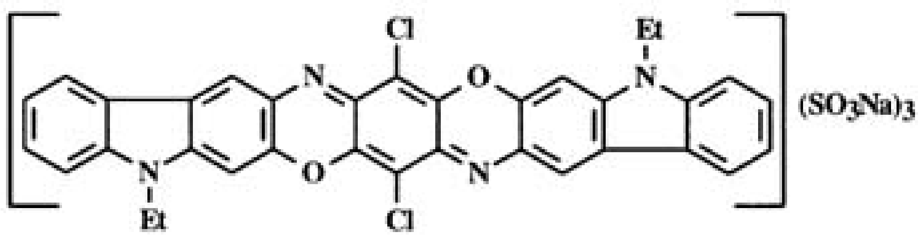


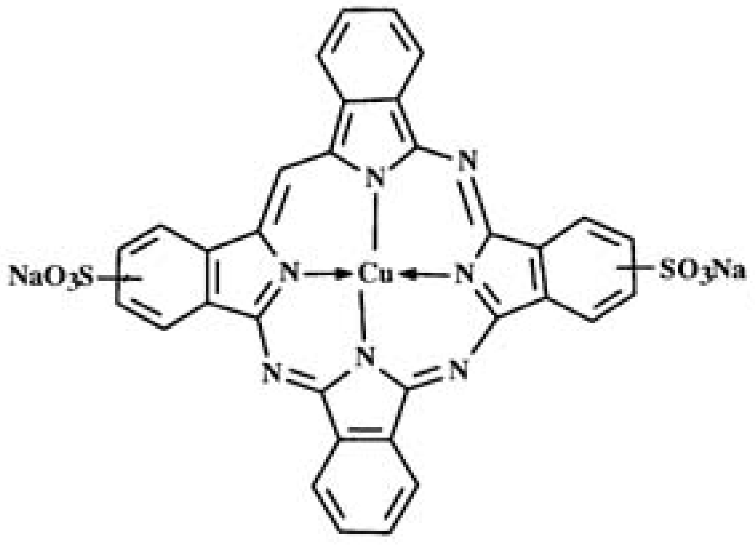
Direct Black 22

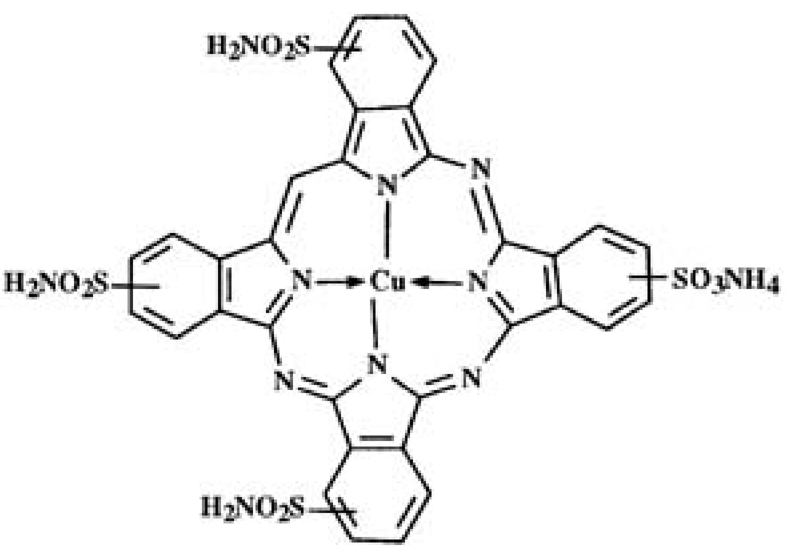

Direct Yellow 12


Direct Yellow 11 (major component)

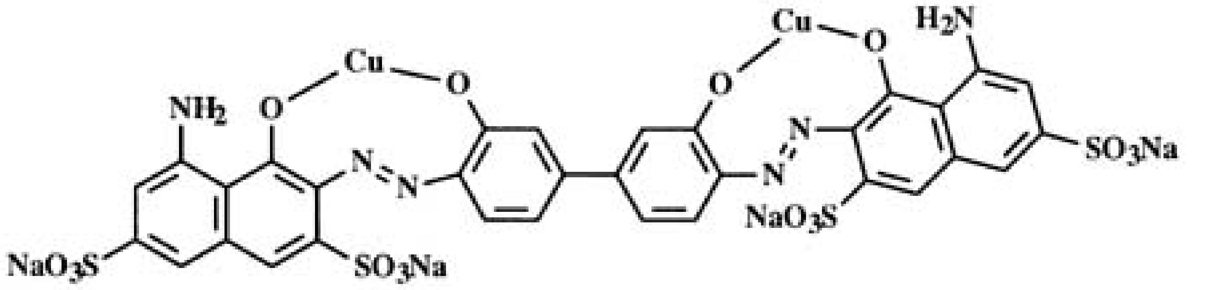

Direct Blue 106


Direct Blue 108

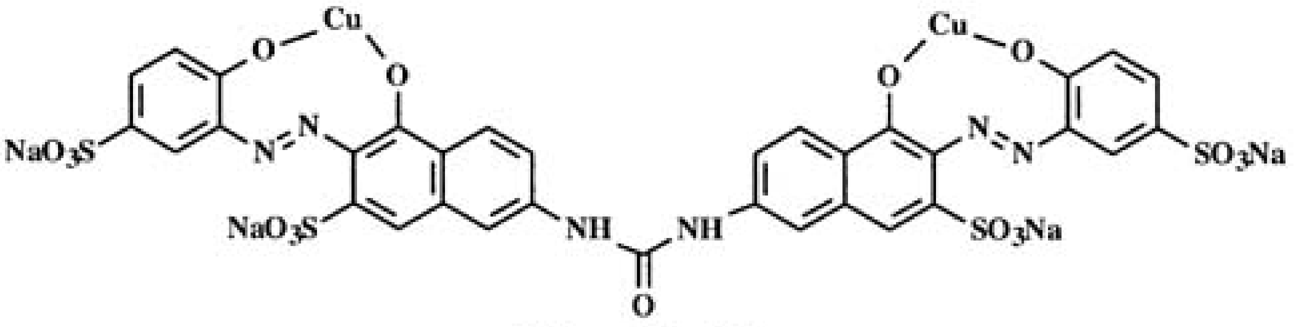

Direct Blue 86


Direct Blue 199

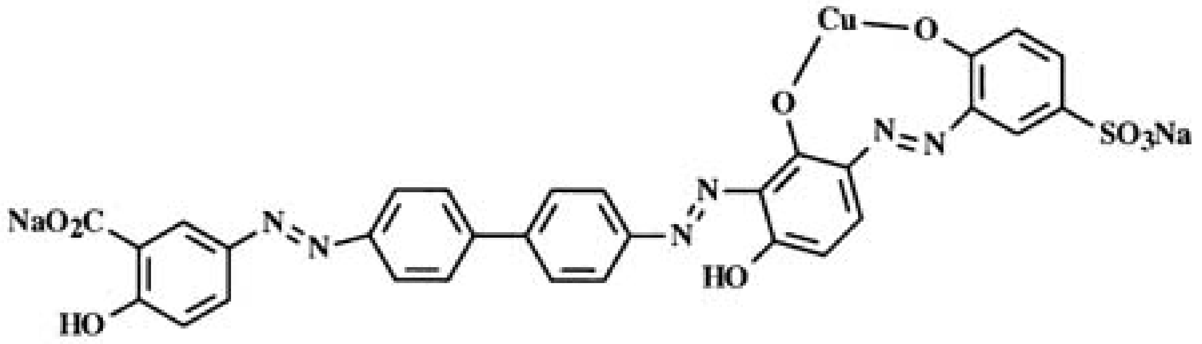
Suitably substituted direct dyes can be converted to metal complexes. In this regard, Cu is the metal of choice and examples are C.I. Direct Blue 218, C.I. Direct Red 83, and C.I. Direct Brown 95. About 5 percent of all azo direct dyes are metal complexes and unlike most direct dyes, these dyes have good lightfastness, as would be anticipated.



Direct Blue 218

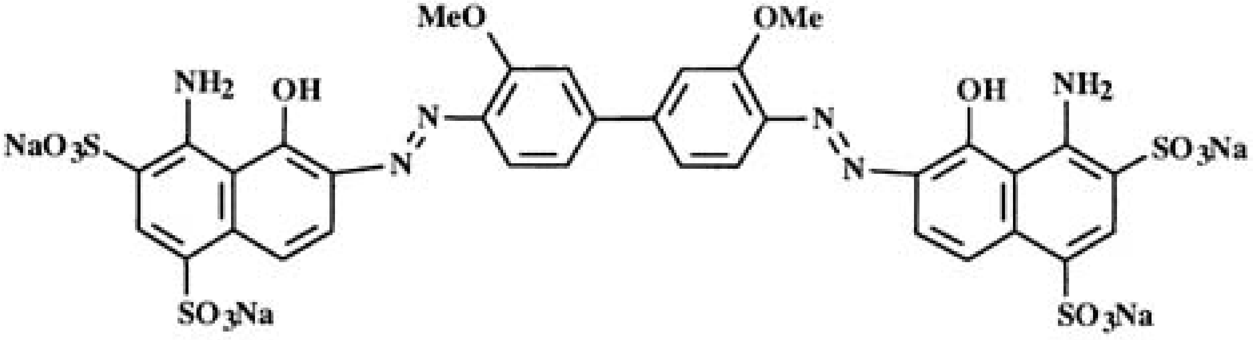


Direct Red 83



Direct Brown 95

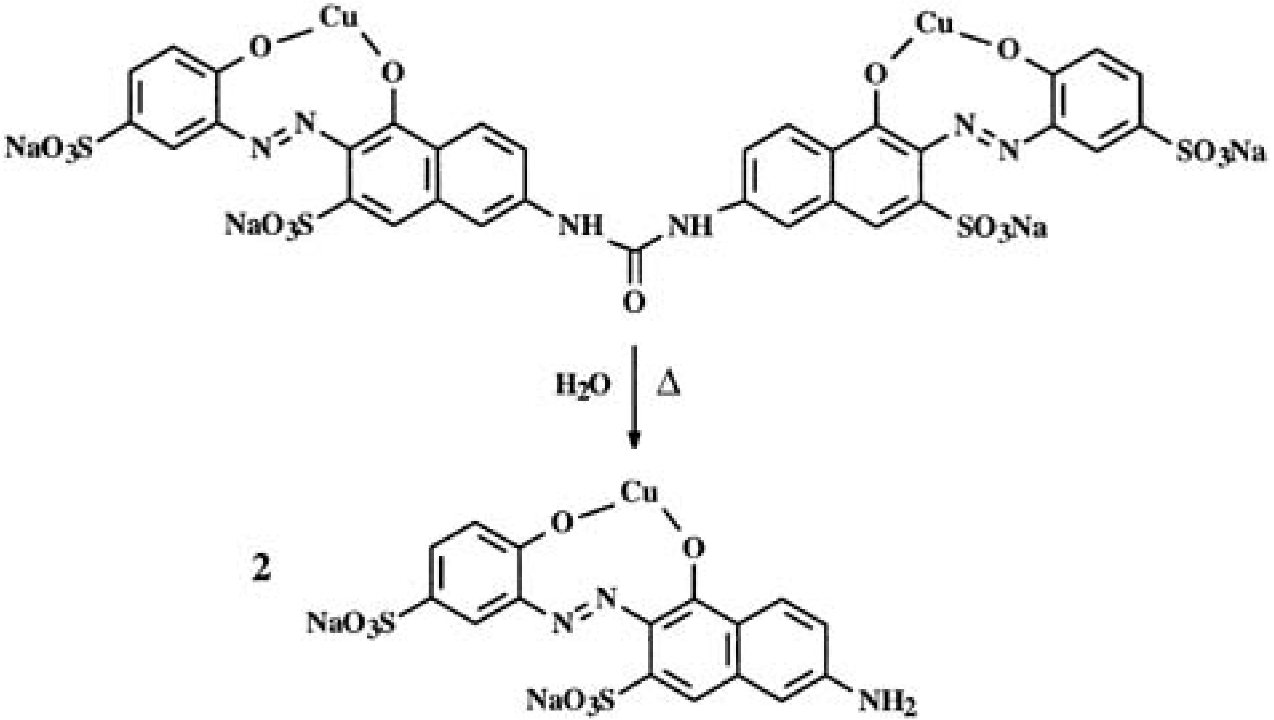
Direct dyes are subdivided into three classes (A, B, and C), to assist the dyer in selecting appropriate combinations for color matching.19 Class A direct dyes give good migration and leveling with time. The dyer employs 520 percent salt for their application and in this case all of the salt may be added at the beginning of the dyeing cycle. An example of this direct dye class is C.I. Direct Yellow 12. Class B direct dyes have poor migration and leveling properties and require the controlled addition of salt to afford level dyeings. They are larger than the former types and have better washfastness. An example is C.I. Direct Blue 1. Class C dyes are the largest of the direct dyes and, consequently, have the best washfastness but poorest leveling properties. Leveling requires careful control of the rate of temperature rise during the dyeing process. Some salt may be added but less than the amount used with classes A and B. An example of this dye class is C.I. Direct Black 22.



Direct Blue 1

The high temperature stability of direct dyes is an important consideration if one wishes to use these dyes as the colorant for cotton when dyeing a polyester/ cotton blend at 130°C.20 The key to success is to choose dyes that are resistant to hydrolysis. Suitable dyes include C.I. Direct Yellow 105, C.I. Direct Orange 39, and C.I. Direct Blue 80, while unsuitable dyes include C.I. Direct Yellow 44, C.I. Direct Red 80, and C.I. Direct Red 83. A quick examination of the structures of the unsuitable dyes reveals that they have groups that are subject to hydrolysis. In the case of Direct Red 83, hydrolysis essentially cuts the molecule in half, eliminating fiber affinity (Fig. 23.7).

Fig. 23.7 High temperature hydrolysis of Direct Red 83.



Since many direct dyes do not have good washfastness and lightfastness, their dyeings on cotton are often treated with a chemical agent, in what is commonly known as an aftertreatment process. The most widely used aftertreatment methods involve (1) cationic fixatives, (2) copper sulfate, or (3) diazotization and coupling. The first and third methods are designed to enhance washfastness and are illustrated in Figs 23.8 and 23.9. The use of cationic fixatives ties up sodium sulfonate groups, reducing the water solubility of the treated dye. Diazotization and coupling enlarges the size of the dye, making desorption more difficult, and simultaneously makes the dye less hydrophilic. This process requires the presence of at least one diazotizable primary arylamino (ArNH2) group in the dye structure. In this 2-step process, the amino group, is diazotized by treatment with nitrous acid (HNO2) and the resultant diazonium groups are coupled with a naphthoxide to give new azo groups. It should be pointed out that the addition of new azo groups can also affect dye color. Therefore, this process is most often used for navy and black shades, where the differences in shade variations from batch to batch are less objectionable.

Fig. 23.8 Use of a cationic fixative to enhance direct dye washfastness.

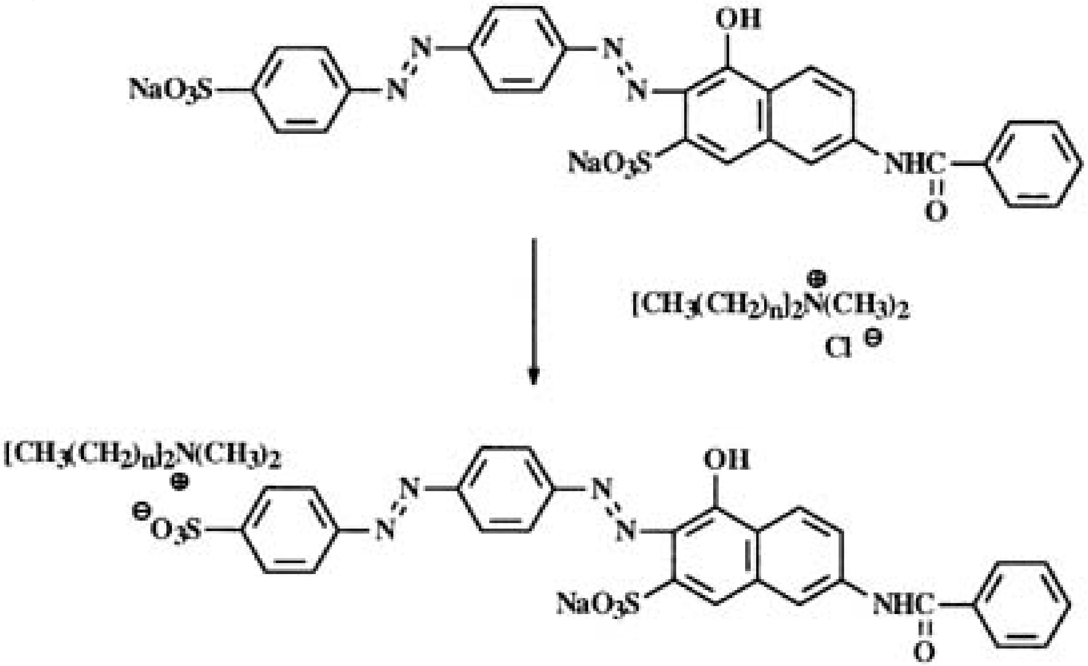
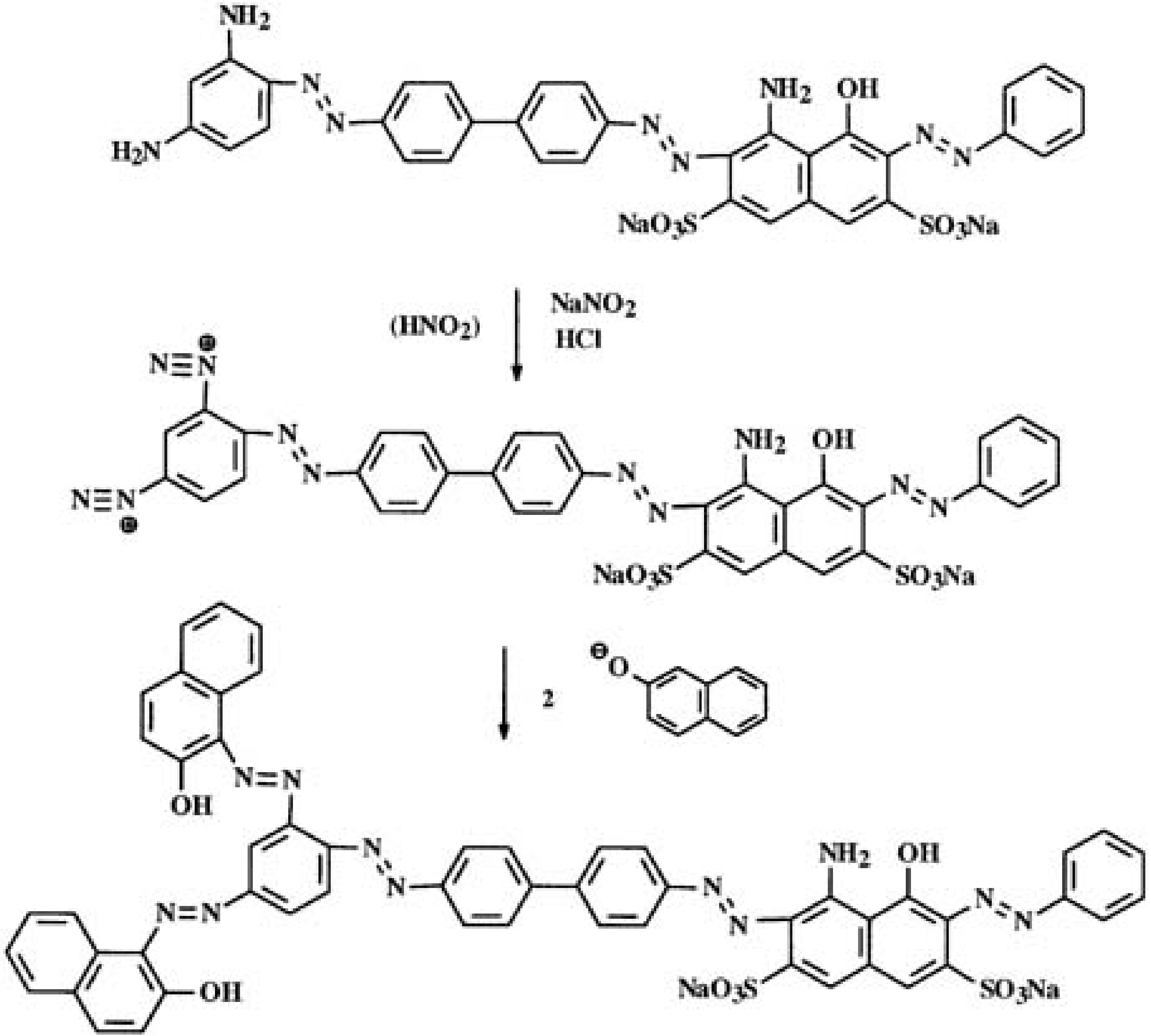


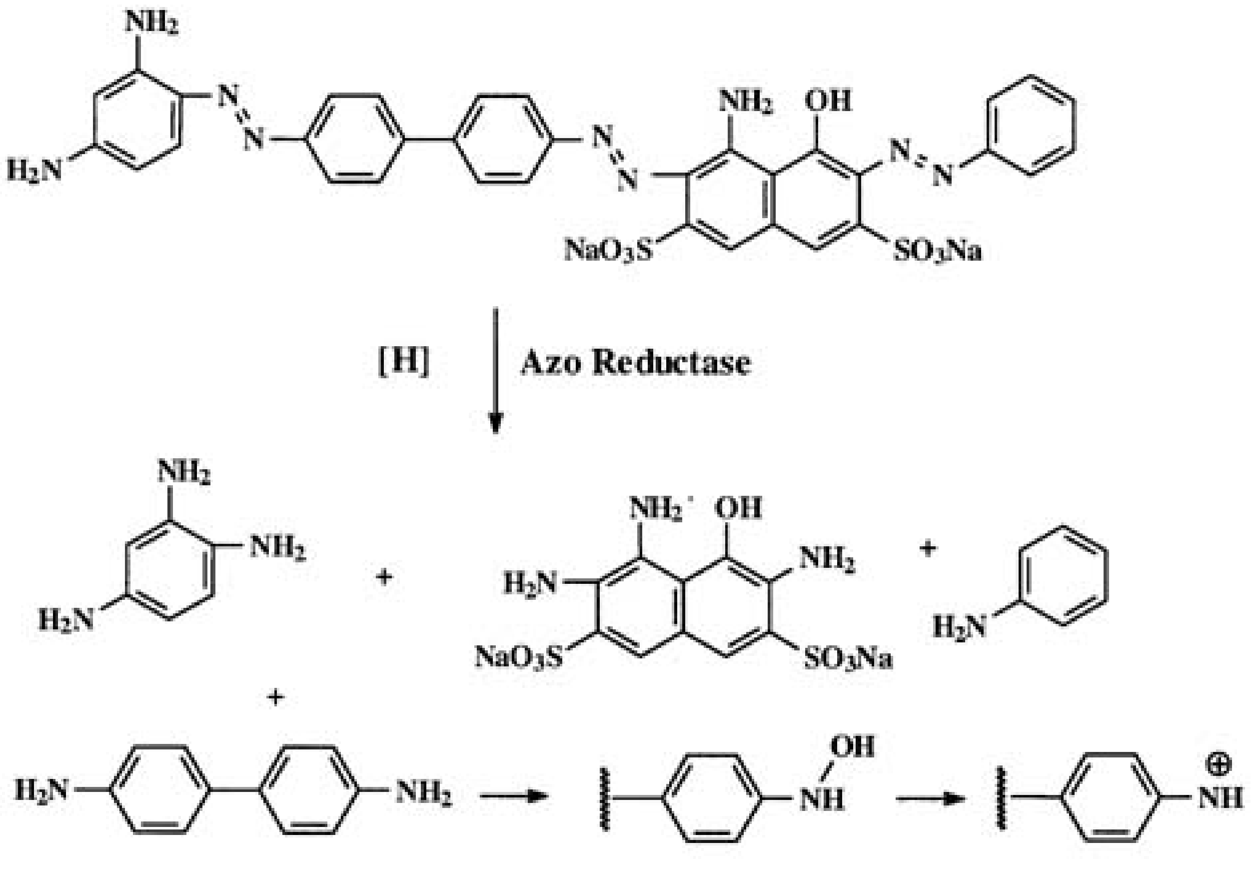
Fig. 23.9 Use of the diazotization and coupling process to enhance direct dye washfastness.



While copper sulfate aftertreatments are designed mainly to enhance lightfastness, the reduction in water solubility that accompanies Cu-complex formation can have a beneficial effect on washfastness. This treatment also dulls the fabric shade and causes a shift in dye color, so that the resultant color must be the one the dyer is seeking.

The dye used as an example in Fig. 23.9 is C.I. Direct Black 38. It is worthwhile to note that this dye is one of many that were synthesized from benzidine, an established human carcinogen.21 Nowadays, such dyes are regarded as cancer-suspect agents because of their potential to generate free benzidine upon metabolic breakdown (see Fig. 23.10).22 With this point in mind, regulations preventing the use of azo dyes derived from benzidine and 20 other aromatic amines in textiles have appeared.23 This requires dye chemists to consider the genotoxicity of potential metabolites in the design of new azo dyes.24

Fig. 23.10 Metabolic breakdown of a direct dye by azo reductase enzymes.



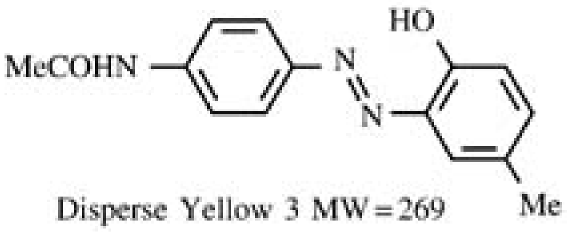
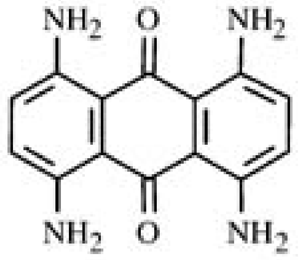
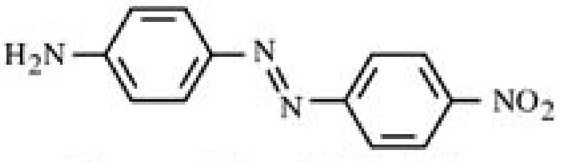
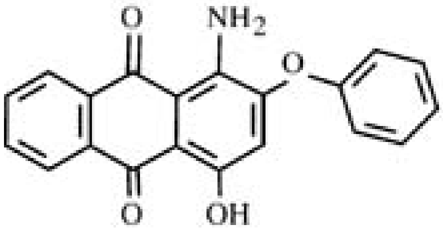
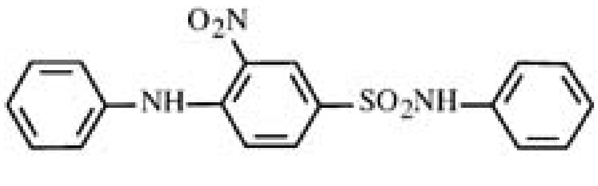
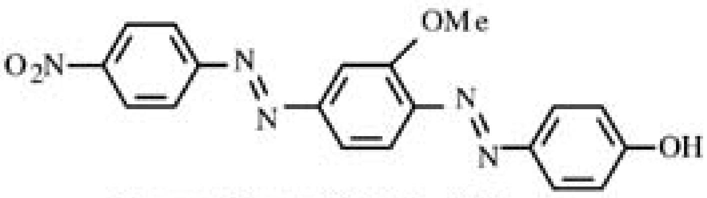
Benzidine

Disperse Dyes

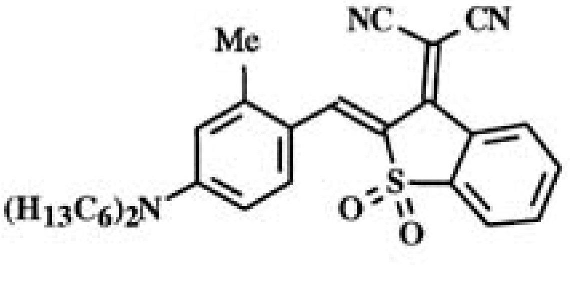
Disperse dyes were invented to dye the first hydrophobic fiber developed, namely cellulose acetate, and were initially called acetate dyes.25 The term disperse dyes is more appropriate, since (1) these dyes are suitable for a variety of hydrophobic fibers and (2) it is descriptive of their physical state in the dyebath. Disperse dyes have extremely low water solubility and to be applied from this medium they must be (1) milled to a very low particle size (13 µm) and (2) dispersed in water using a surfactant (dispersing agent). These nonionic, hydrophobic dyes can be used on acetate, triacetate, polyester, nylon, acrylic, and polyolefin fibers, and their mechanism of fixation involves solidsolid solution formation.

Disperse dyes provide a wide range of bright colors on textiles and many have excellent build up and barré coverage properties. In addition, they have good washfastness properties but their lightfastness varies with structure. They are suitable for continuous dyeing, a process that takes advantage of their sublimation properties. Disperse dye end-use applications are often based on their classification. The classification system employed is shown in Table 23.4. Low energy disperse dyes are the easiest to exhaust under atmospheric dyeing conditions but have the lowest thermostability, with the latter property making them unsuitable for automotive applications. They are used to dye acetate, triacetate, and nylon fibers. On the other hand, the high energy dyes are best applied under pressure (*T* = 130°C) and are most appropriate for polyester body cloth for automobile interiors. Medium energy dyes are also used to dye polyester and can be applied at atmospheric pressure using a carrier.

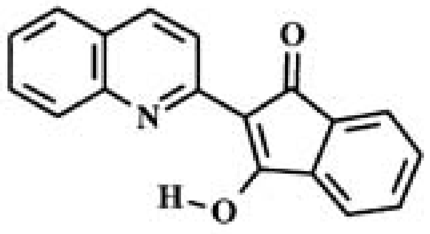
TABLE 23.4 Disperse Dyes: Classification and Examples

Class	Molecular Mass	Sublimation Temperature
Low Energy	<300	<150°C
 Disperse Yellow 3 MW = 269	 Disperse Blue 1 MW = 269	 Disperse Yellow 3 MW = 242
Medium Energy	300400	150210°C
 Disperse Red 60 MW = 331	 Disperse Yellow 42 MW = 369	 Disperse Orange 29 MW = 377
High Energy	>400	>210°C
 Disperse Red 167 MW 507	 Disperse Blue 79 MW 639	

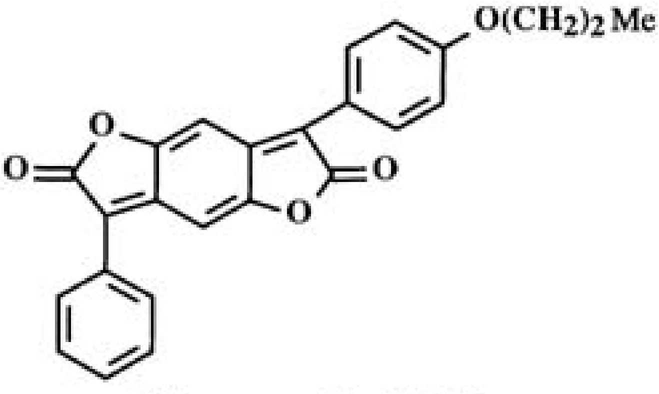
Disperse dyes vary in the type of chromophore present and include azo, anthraquinone, nitro, methine, benzodifuranone, and quinoline-based structures. Examples of the first three types are given in Table 23.4, and representative of the latter three types are C.I. Disperse Blue 354, C.I. Disperse Yellow 54, and C.I. Disperse Red 356. Most disperse dyes have azo (•59%) or anthraquinone (•32%) structures. Azo disperse dyes cover the entire color spectrum, while the important anthraquinone disperse dyes are mainly red, violet, and blue. The azo types offer the advantages of higher extinction coefficients (εmax = 30,00060,000) and ease of synthesis, while the anthraquinones are generally brighter and have better photostability (lightfastness). The key weakness associated with the anthraquinone dyes are their low extinction coefficients (εmax = 10,00015,000) and less environmentally friendly synthesis.



Disperse Blue 354

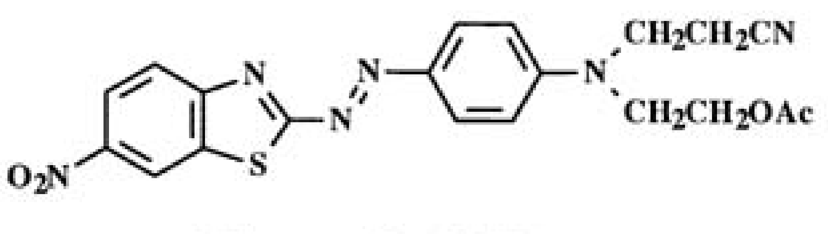


Disperse Yellow 54

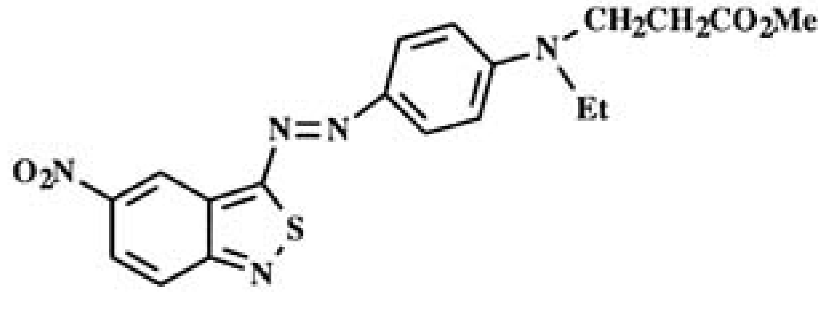


Disperse Red 356

To produce disperse dyes having the brightness of the anthraquinone system and the color strength of the azo system, azo dyes based on heteroaromatic amines were developed.2628 Examples are C.I. Disperse Red 177, Disperse Blue 148, Disperse Red 156, and C.I. Disperse Blue 102. These dyes employ aminated thiazoles, benzothiazoles, benzisothiazoles, and thiadiazoles in their synthesis. They have εmax = 50,00080,000, good brightness, and good washfastness. Another key feature of disperse dyes with heretoaromatic systems is their less complex structures. Compare, for example, the fewer number of substituent groups in the diazo component (left side of the azo bond) of Disperse Red 167 versus Disperse Red 156 and Disperse Blue 165 versus Disperse Blue 102. However, these dyes are more expensive than disperse dyes derived from benzeneamines, owing to their low reaction yields and have lower lightfastness than the anthraquinone dyes.

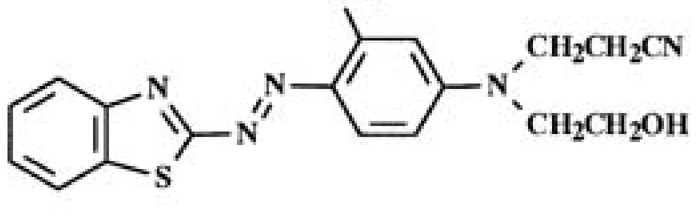


Disperse Red 177

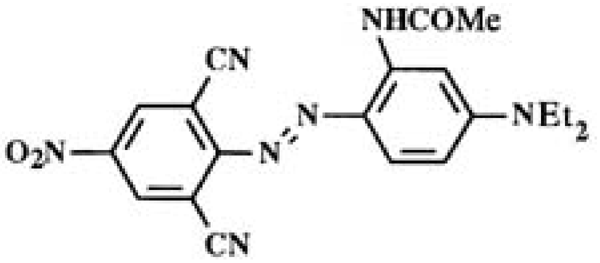


Disperse Blue 148

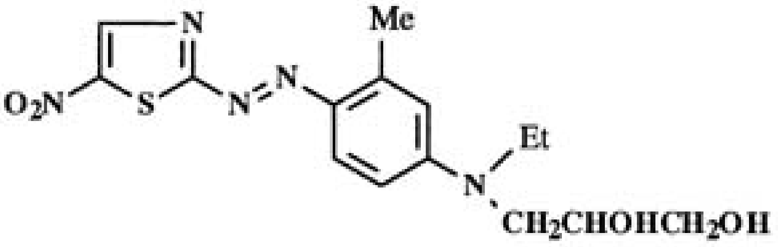
Disperse Red 167



Disperse Red 156



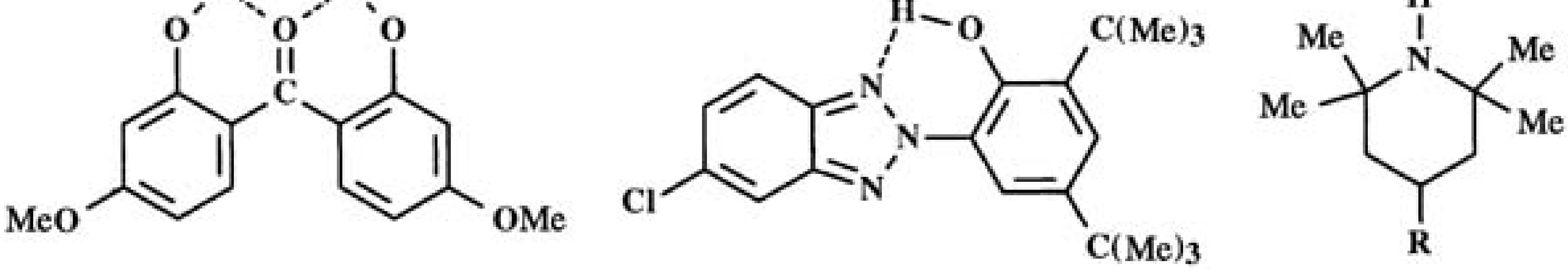
Disperse Blue 165



Disperse Blue 102

The use of disperse dyes in applications requiring high lightfastness involves the coapplication of photostabilizers. These agents enhance dye stability by quenching the excited states of disperse dyes, probably via energy transfer, or by preferential absorption (screening) of UV radiation. They are also known as UV absorbers and exhaust from the dyebath like disperse dyes. They encompass benzophenone, benzotriazole, oxalanilide, and hindered amine/phenol structures (see Fig. 23.11).29

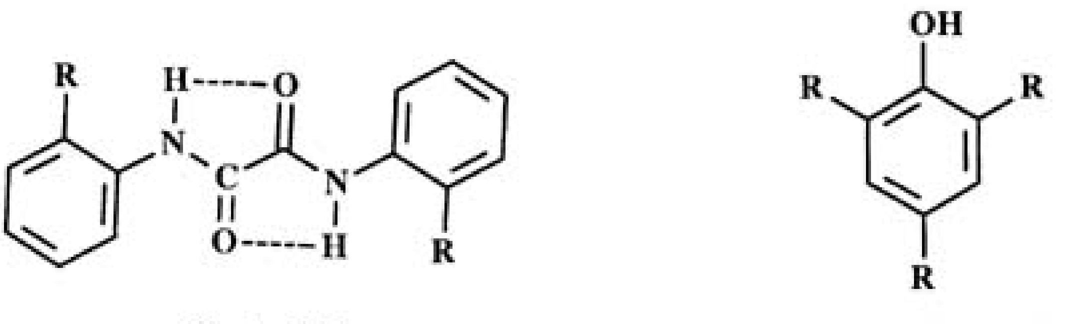
Fig. 23.11 Structures of disperse dye photostabilizers.



Benzophenone

Benzotriazole

Hindered amine



Oxalanilide

Hindered phenol

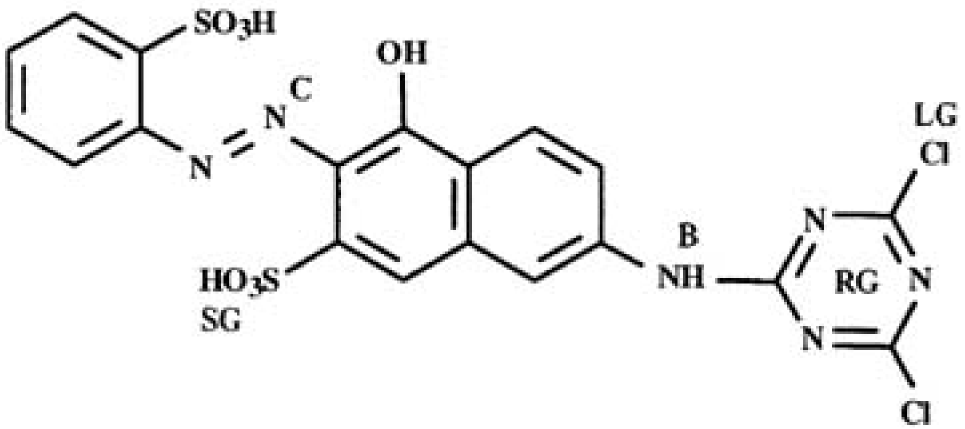
Reactive Dyes

Reactive dyes are used mainly as colorants for cotton, although they are also suitable for nylon and wool.30 They are water soluble, due to the presence of one or more SO_3Na groups, and undergo fixation to polymer chains via covalent bond formation. Reactive dyes have very high washfastness and are used for leisure wear and other applications requiring stability to repeating laundering. Each dye is composed of 5 basic parts:

SG–C–B–RG–LG

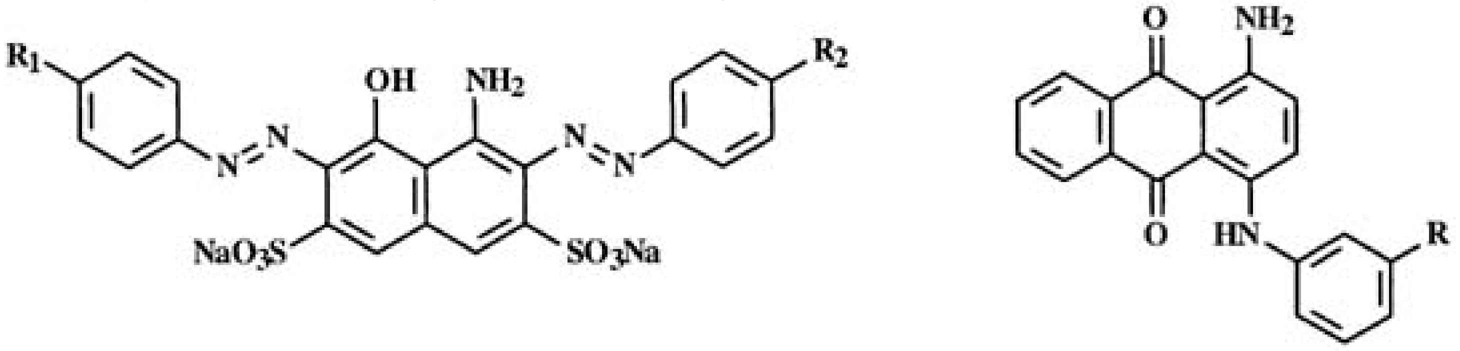
In this regard, SG = water solubilizing group (SO_3Na), C = chromogen (e.g., azo, anthraquinone), B = bridging or linking group (e.g., NH), RG = reactive group (e.g., chlorotriazine, vinylsulfone), and LG = leaving group (e.g., Cl, F, HSO_4). These parts are illustrated for the structure in Fig. 23.12. This structure also shows that reactive dye structures can be quite small, much smaller in fact than those characterizing direct dyes. As a consequence, reactive dyes have significantly lower inherent affinity for cotton and can require high levels of salt (200300 g/L) in their dyebaths to promote exhaustion.31

Fig. 23.12 Structure showing the basic parts of fiber-reactive dye.



In addition to giving high washfastness on cotton, reactive dyes usually give bright shades. The latter property arises from the fact that reactive dyes are often acid dye structures linked to reactive groups, as shown in Fig. 23.13. Reactive dyes have moderate to good lightfastness and fair to poor chlorine fastness.

Fig. 23.13 Comparison of some acid dye and fiber-reactive dye structures.



Acid Black 1: $R_1 = H$; $R_2 = NO_2$ Acid Blue 25: $R = H$

Reactive Black 5: $R_1 = R_2 = SO_2(CH_2)_2OSO_3Na$ Reactive Blue 19: $R_1 = SO_2(CH_2)_2OSO_3Na$

While the most commonly used reactive systems involve the halotriazine and sulfatoethyl sulfone (vinyl sulfone) groups, halogenated pyrimidines, phthalazines, and quinoxalines are also available (Fig. 23.14). In all of these systems, alkali is used to facilitate dyefiber fixation, and fixation occurs either by nucleophilic substitution or addition (Figs 23.15 and 23.16).

Fig. 23.14 Examples of reactive groups found in fiber-reactive dye chemistry (a) dichlorotriazine, (b) monochlorotriazine, (c) trichloropyrimidine, (d) sulfatoethyl sulfone, (e) dichlorophthalazine, (f) monofluorotriazine, (g) dichloroquinoxaline, (h) difluropyrimidine, and (i) difluorochloropyrimidine.

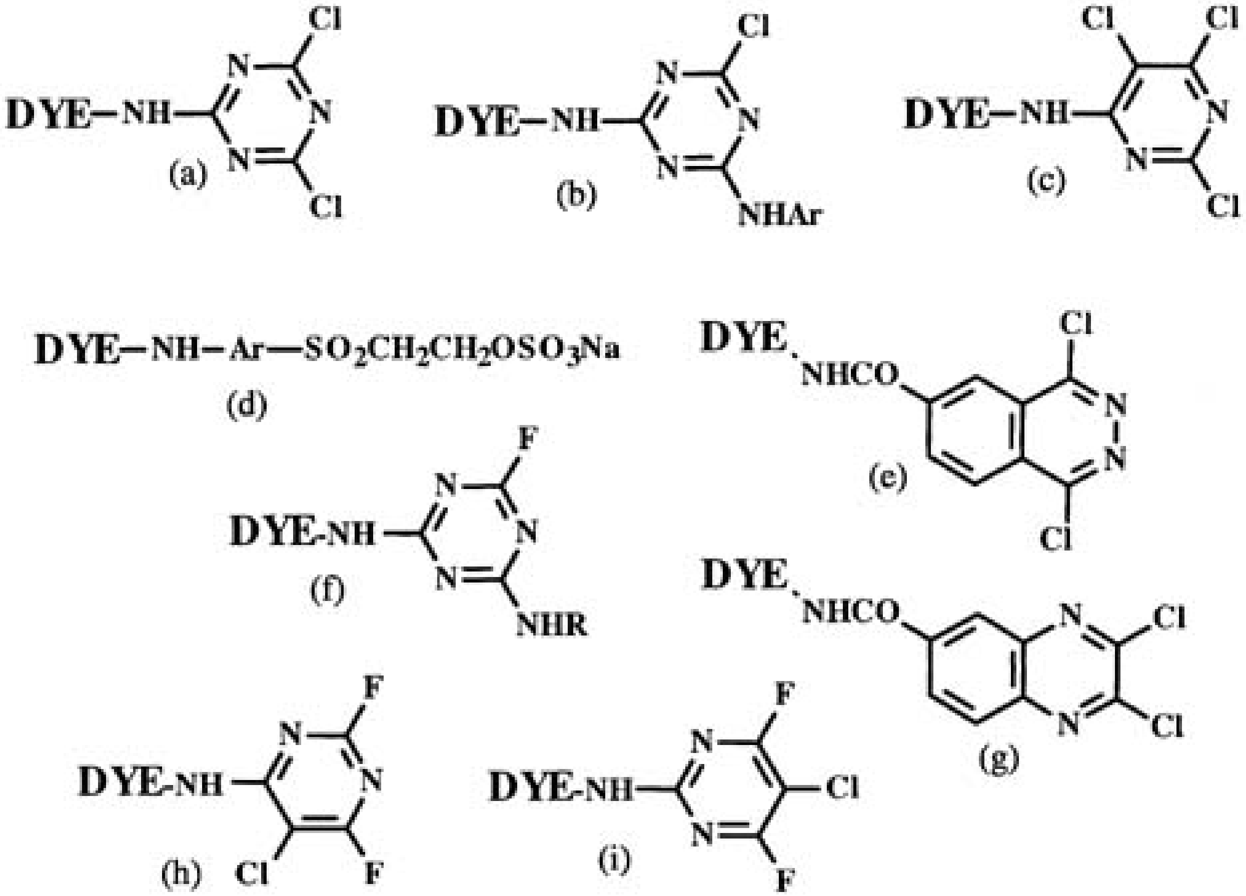


Fig. 23.15 Reactive dye fixation to cellulose via nucleophilic substitution.

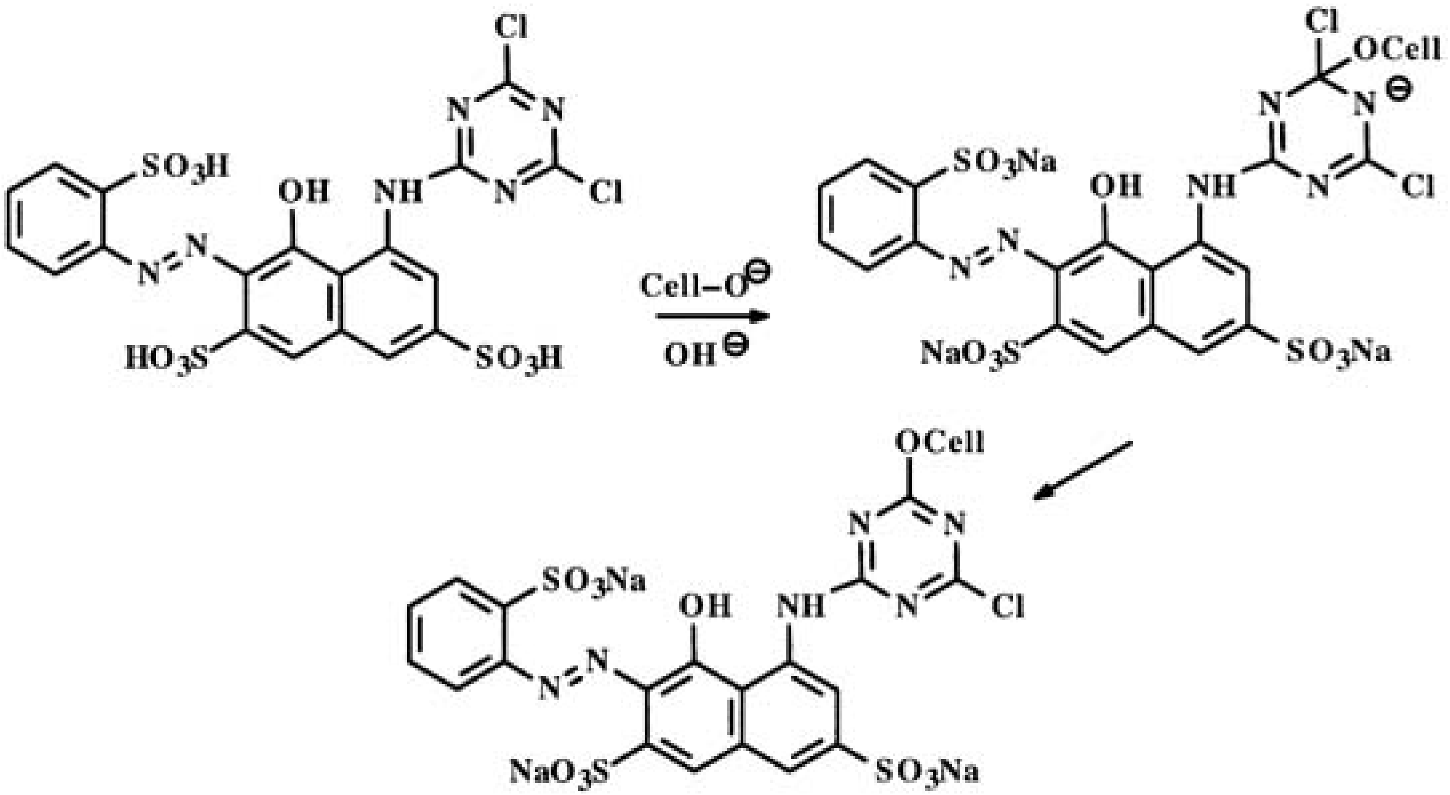
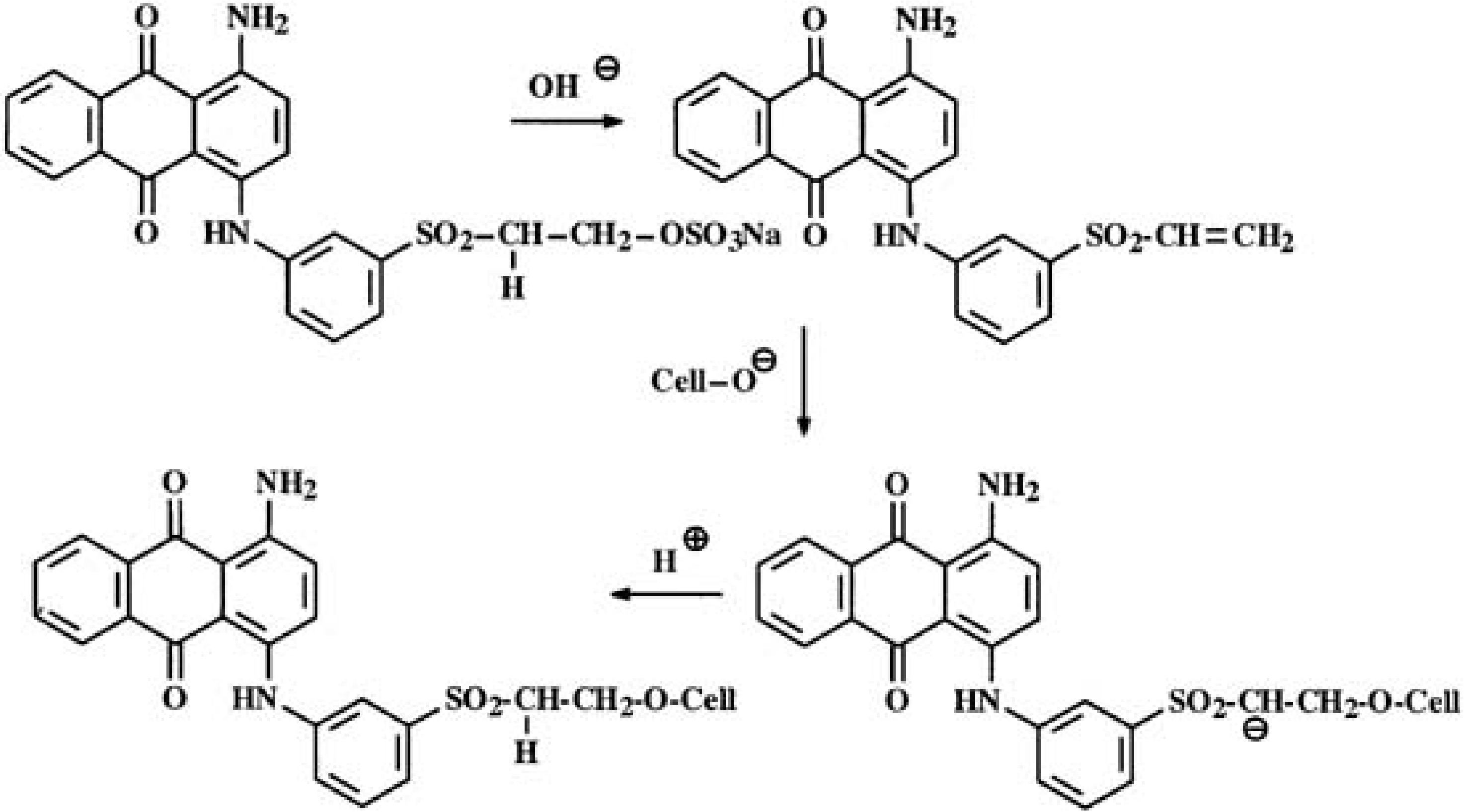


Fig. 23.16 Reactive dye fixation to cellulose via nucleophilic addition.



The requirement for alkali in the application of reactive dyes to cotton leads to an undesirable side reaction, namely hydrolysis of the reactive groups before dyefiber fixation can occur (Fig. 23.17). Since the hydrolyzed dye cannot react with the fiber, this leads to wasted dye and the need to treat the residual color in the wastewater prior to dyehouse discharges. To improve the percentage of fixation, dyes with two or more reactive groups were developed (Fig. 23.18). This makes it possible for dyefiber fixation to occur even when one reactive group undergoes hydrolysis.32

Fig. 23.17 Competing reactions when reactive dyes are applied to cotton.

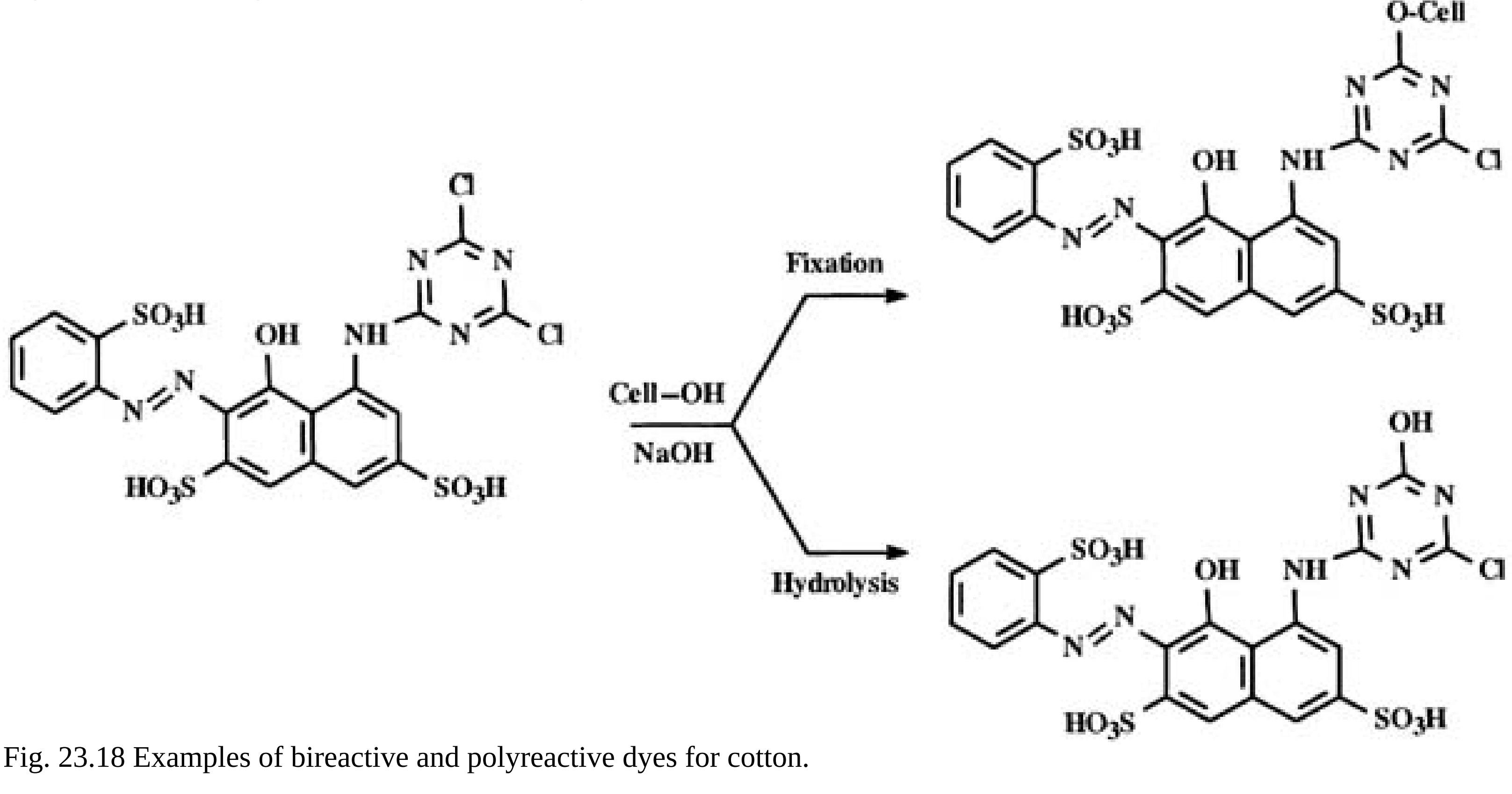
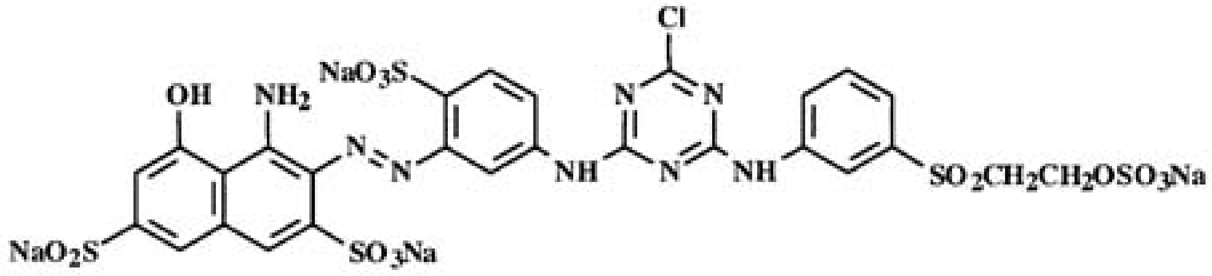
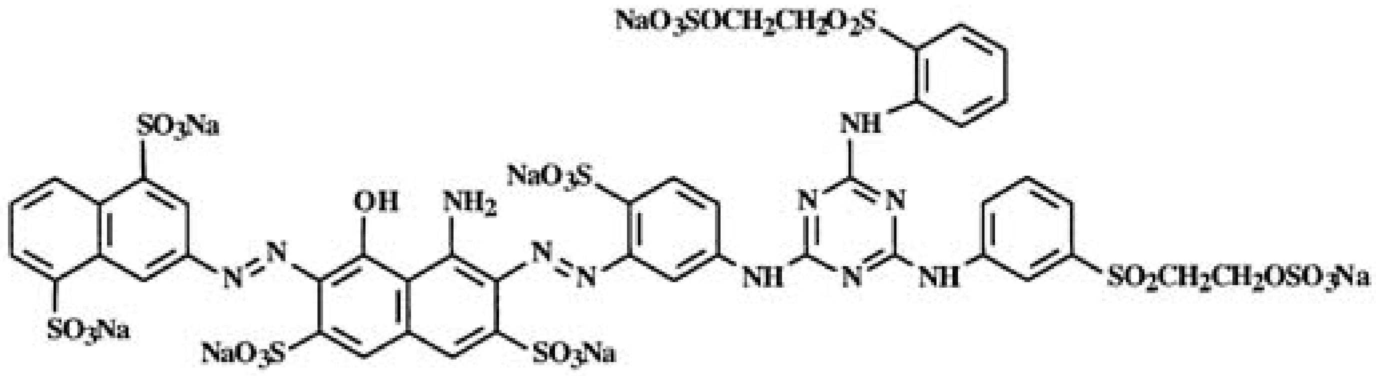
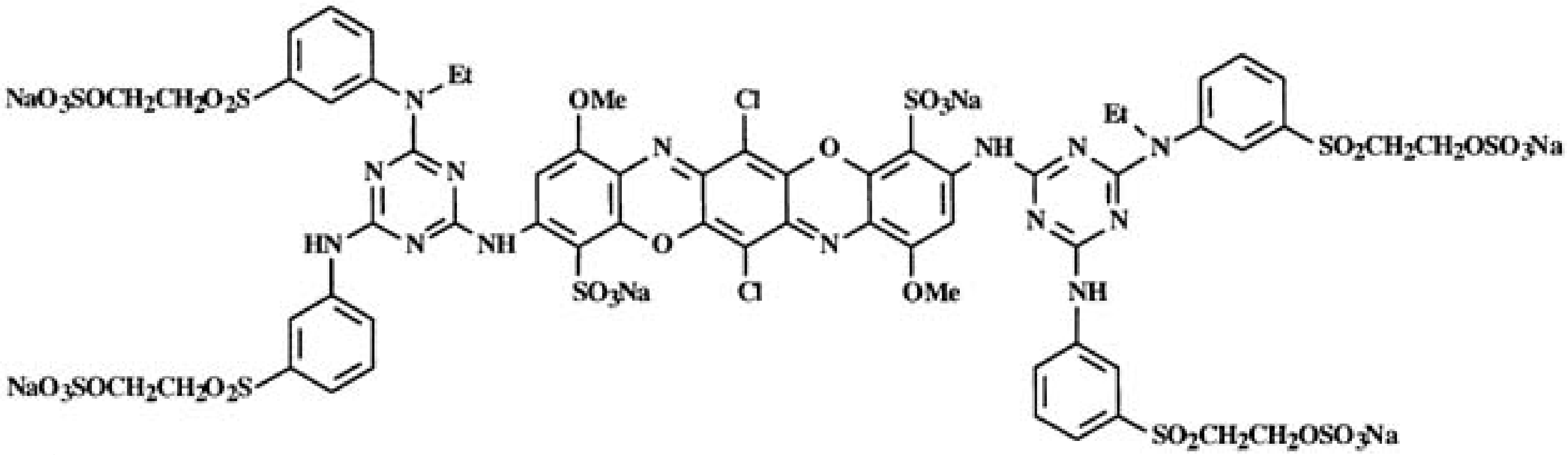


Fig. 23.18 Examples of bireactive and polyreactive dyes for cotton.





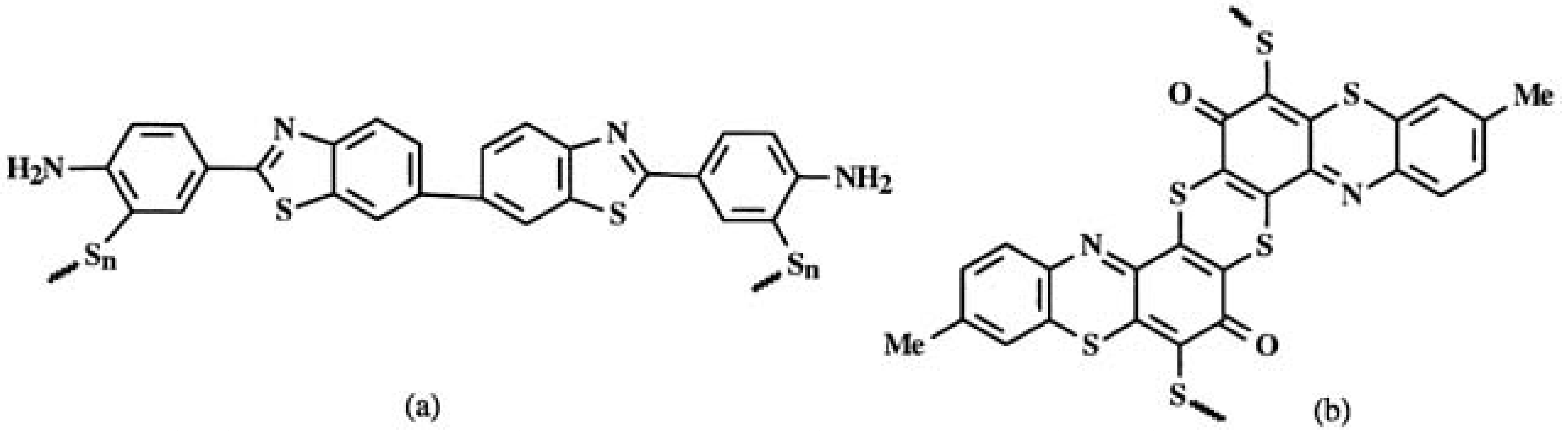


Sulfur Dyes

Sulfur dyes are water-insoluble dyes that are applied to cotton.33 They are used primarily for their economy and high washfastness, are easy to apply, and mainly give dull shades. While yellow, red, brown, olive, and blue colors can be produced, sulfur dyes are most important for their ability to deliver washfast black shades on cotton. In this regard, C.I. Sulfur Black 1 is the main dye used commercially. Sulfur dyes have acceptable lightfastness but poor bleachfastness.

Due to extremely low solubility, the precise structures of most sulfur dyes remain unknown. Much of what we know about sulfur dye structures arises from the characterization of certain degradation products or reaction precursors.34 Based on such work, it has been possible to determine that structures of the type shown in Fig. 23.19 are covered in this dye class. A key common feature of sulfur dyes is the presence of sulfide (S*n*) bonds, and it is this feature that makes dye application from an aqueous medium possible.

Fig. 23.19 Examples of (a) sulfur yellow and (b) red dye structures.


(a) (b)

The reaction of sulfur dyes with sodium sulfide (Na2S) at pH > 10 effects the reduction of the sulfide bonds, giving their water-soluble (leuco) forms. The reduced forms behave like direct dyes, in that they exhaust onto cotton in the presence of salt. Once exhausted, the reduced dyes are reoxidized to their water-insoluble forms, giving dyeings with good washfastness. This chemistry is illustrated in Fig. 23.20. While the oxygen in air can be used for the oxidation step, an agent such as hydrogen peroxide is used because it works faster. Sulfur dyes have also been marketed in their prereduced form (DyeS-Na+), as ready to use C.I. Leuco Sulfur dyes. Dye exhaustion in the presence of salt is followed by oxidation. Similarly, water-soluble sulfur dyes containing thiosulfate groups are sold as C.I. Solubilized Sulfur dyes. They are known as "Bunte salts,"35 have better levelling properties than the C.I. Sulfur dyes, and are attractive for package dyeing. The C.I. solubilized sulfur dyes are applied with Na2S and the chemistry associated with their 2-step application is summarized in Fig. 23.21, along with a representative dye structure.

Fig. 23.20 Steps involved in the application of sulfur dyes to cotton.

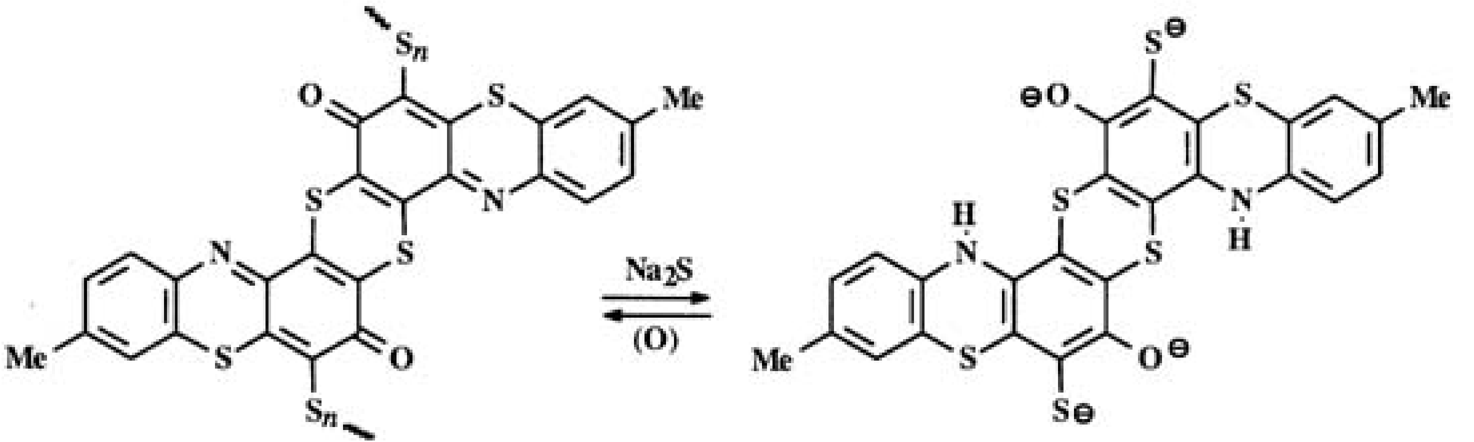
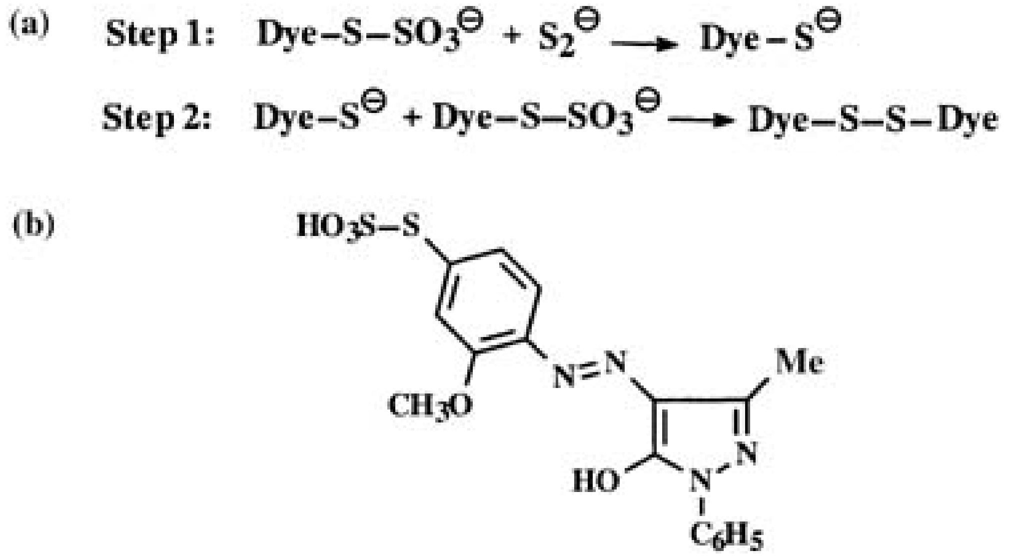


Fig. 23.21 Chemistry employed in the application of (a) C.I. solubilized sulfur dyes and (b) a representative dye.



Vat Dyes

Like sulfur dyes, vat dyes are water-insoluble colorants for cotton that must be reduced to their soluble "leuco" forms to be applied from an aqueous dyebath.36a, b Their name originates from their early application from wooden vessels known as vats. The term "vatting" is used to refer to the application of these dyes via chemical reduction followed by oxidation. Vat dyes are easier to reoxidize than sulfur dyes and the oxygen in air is often the agent used. As would be anticipated, most vat dyes display high washfastness. As a class, they have the best lightfastness and bleach fastness among the dye families suitable for cotton. Some cause catalytic fading or phototendering on cotton.37

Vat dyes have mainly anthraquinone (82%) or indigoid/thioindigoid (9%) structures, with the former having much better fastness properties. The anthraquinone vat dyes exhibit a bathochromic color shift (λmax of higher wavelength) upon reduction to their leuco forms, while the indigoids exhibit a hypsochromic shift. Examples of the two structural types are shown in Figs 23.22 and 23.23. While anthraquinone vat dyes having a single anthraquinone unit exist, those with the best fastness properties seem to have the equivalent of two anthraquinone units.

Fig. 23.22 Representative anthraquinone vat dye structures: (a) C.I. Vat Red 13, (b) Vat Black 27, (c) Vat Orange 2, (d) Vat Blue 4, and (e) Vat Green 1.

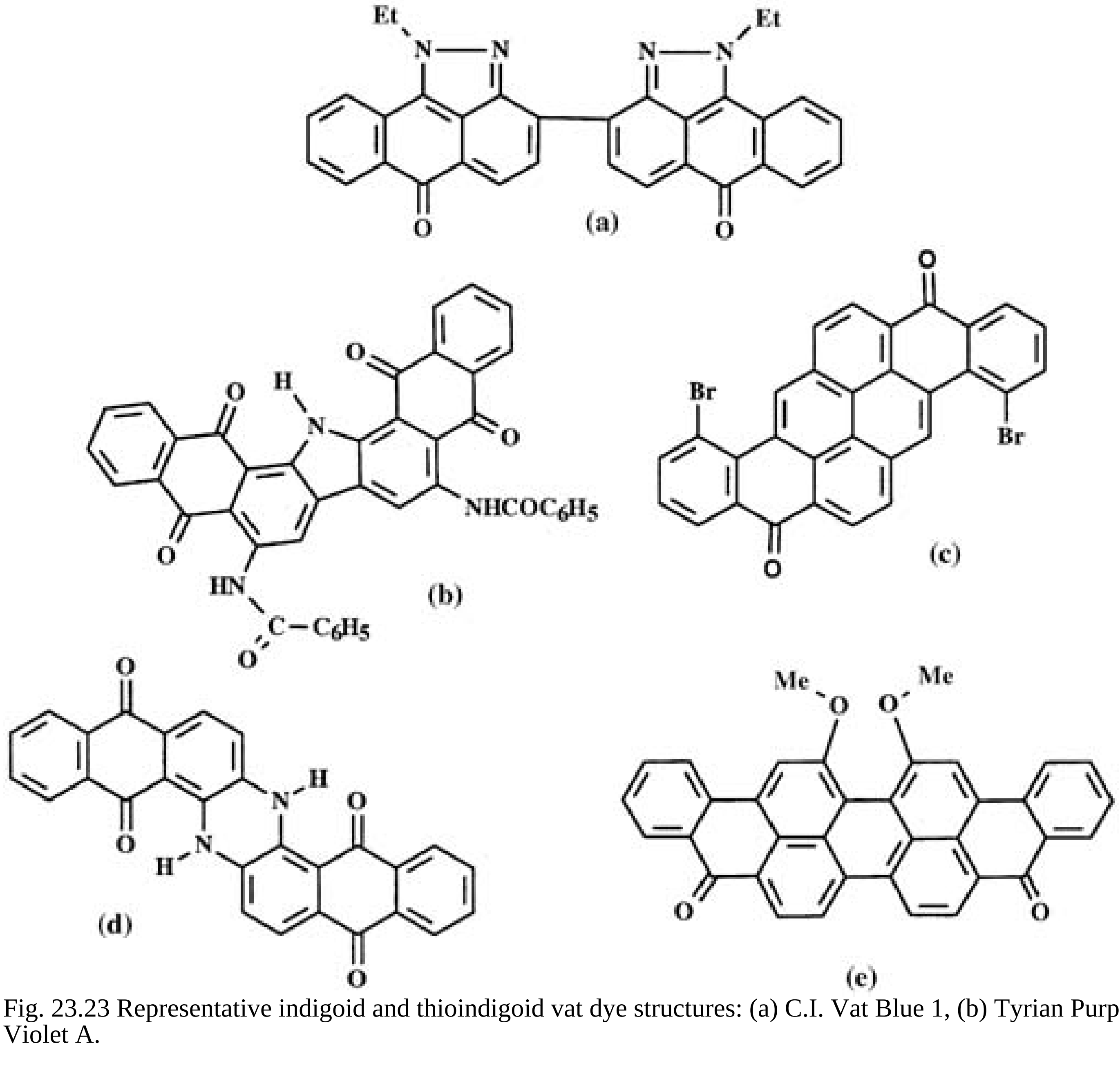
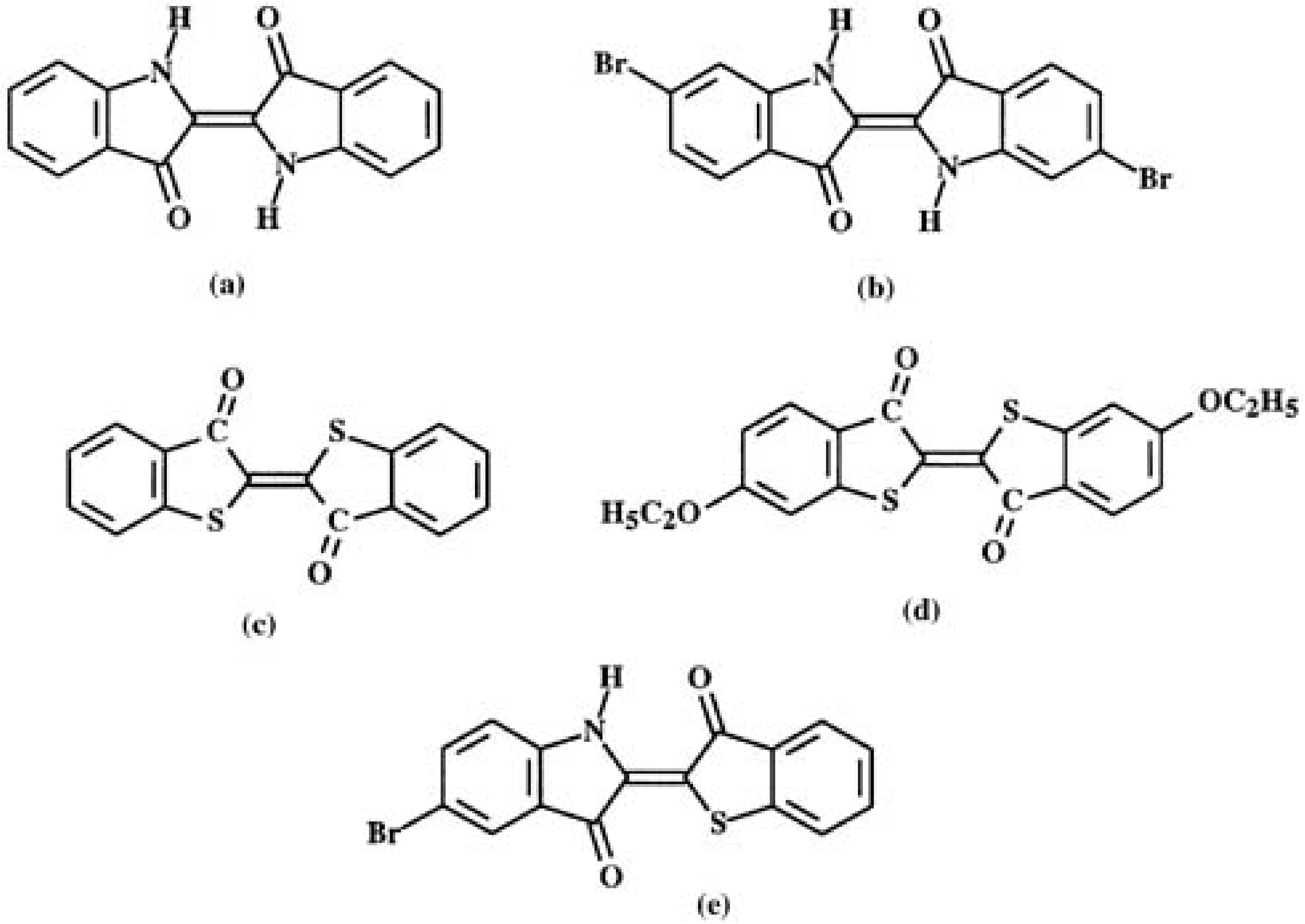


Fig. 23.23 Representative indigoid and thioindigoid vat dye structures: (a) C.I. Vat Blue 1, (b) Tyrian Purple, (c) Vat Red 41, (d) Vat Orange 5, and (e) Ciba Violet A.

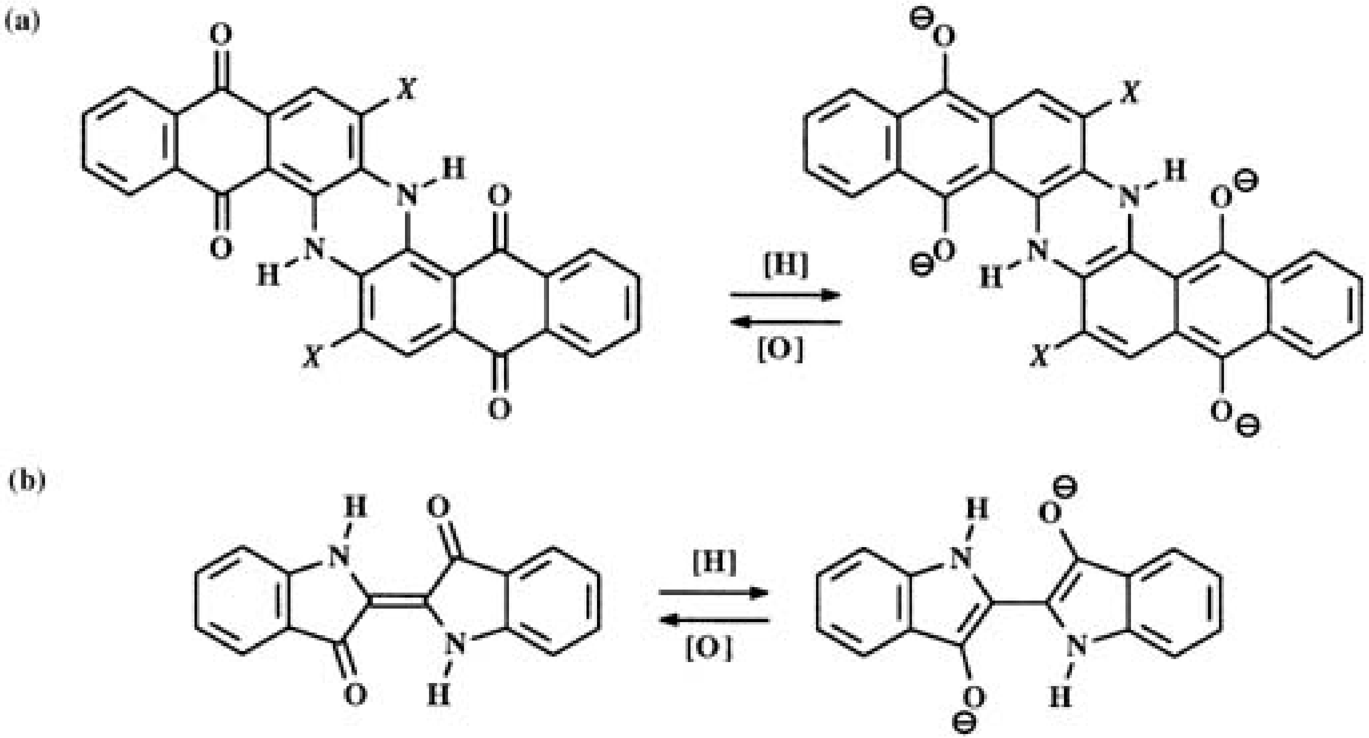


(a) (b) (c) (d) (e)

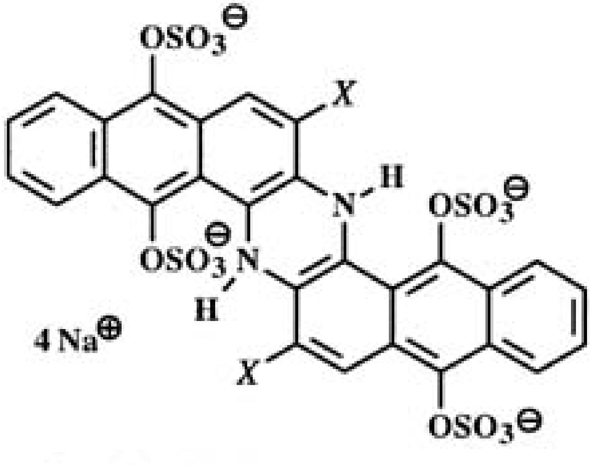
No doubt the best known and biggest volume vat dye is C.I. Vat Blue 1, indigo, the denim blue dye. Closely related structures are the thioindigoids (4%), which have a sulfur atom in lieu of the NH group (Fig. 23.23). The thioindigoids are mainly used as colorants for printing and give orange and red hues. A few dyes having the features of both indigoid types are also known (e.g., Ciba Violet A).

The chemistry associated with the vatting process is illustrated in Fig. 23.24. For the reduction step, a mixture of sodium hydroxide (caustic) and sodium hydrosulfite (hydro, $Na_2S_2O_4$) is used. Depending upon the amount of caustic and hydro employed, one or both of the anthraquinone rings may undergo reduction.

Fig. 23.24 Chemistry involved in the vatting of the C.I. Vat dyes (a) Blue 4 and (b) Blue 1.



Vat dyes are also available in prereduced forms (3%), an example of which is the leuco sulfuric acid ester C.I. Solubilized Vat Blue 4. These water-soluble forms have affinity for cellulose and exhaust like direct dyes. They are oxidized to the insoluble form using hydrogen peroxide.



C.I. Solubilized Vat Blue 4

Vat dyes are brighter than direct and sulfur dyes but less so than reactive dyes. They are the colorants of choice when dye bleachfastness on cotton is important. They span the entire color spectrum and can be applied to cotton using a variety of methods. With regard to the latter point, they can be further classified based on the temperatures involved in their application. Accordingly, there are hot (5060°C), warm (4050°C), and cold (2530°C) dyeing vat dyes. The hot dyeing types are large planar leuco forms having high affinity and no salt is required for their application. The cold types are small molecules with low affinity and require repeated application to get good build up. Indigo falls into the cold dyeing category.

23.4 The Application of Dyes

The process of dyeing may be carried out in batches or on a continuous basis. The fiber may be dyed as stock, yarn, or fabric. However, no matter how the dyeing is done, the process always is fundamentally the same: dye must be transferred from a bathusually aqueousto the fiber itself. The basic operations of dyeing include: (1) preparation of the fiber; (2) preparation of the dye bath; (3) application of the dye; and (4) finishing. There are many variations of these operations, depending on the kind of dye. The dyeing process is complicated by the fact that single dyes seldom are used. The matching of a specified shade may require from two to a dozen dyes.

Fiber Preparation

Fiber preparation ordinarily involves scouring to remove foreign materials and to ensure even access to dye liquor. Some natural fibers are contaminated with fatty materials and dirt, and synthetic fibers may have been treated with spinning lubricants or sizing that must be removed. Some fibers may also require bleaching before they are ready for use.

Dye-Bath Preparation

Preparation of the dye bath may involve simply dissolving the dye in water, or it may be necessary to carry out more involved operations such as reducing the vat dyes. Wetting agents, salts, "carriers," retarders, and other dyeing assistants may also be added. "Carriers" are swelling agents that improve the dyeing rate of very hydrophobic fibers such as the polyesters. Examples are *o*-phenylphenol and biphenyl. Retarders are colorless substances that compete with dyes for dye sites or form a complex with the dye in the bath and act to slow the dyeing rate. Their use is necessary when too rapid dyeing tends to cause unevenness in the dyeings.

Finishing

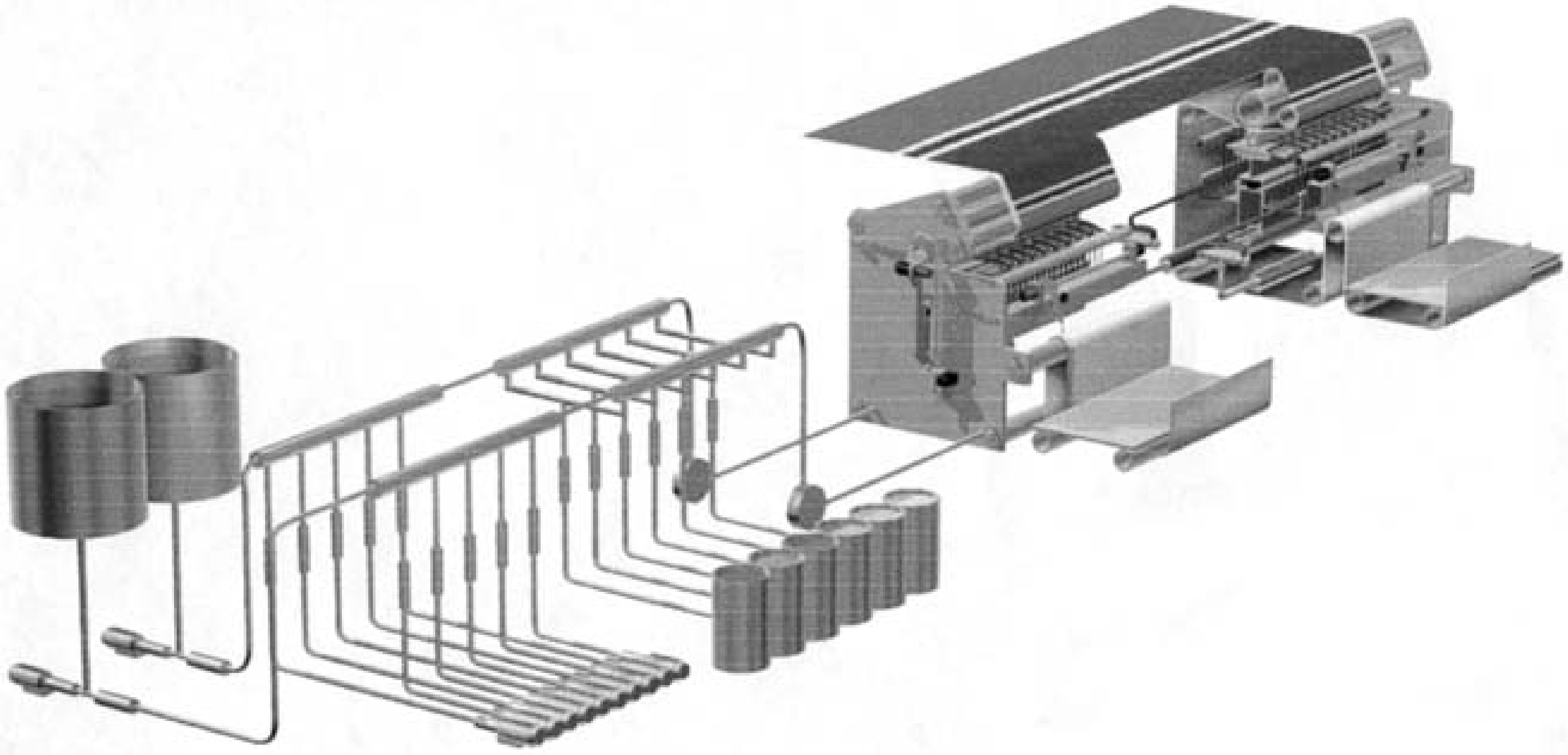
The finishing steps for many dyes, such as the direct dyes, are very simple: the dyed material merely is rinsed and dried. Vat-dyed materials, on the other hand, must be rinsed to remove reducing agent, oxidized, rinsed again, and soaped before the final rinsing and drying steps are carried out. Generally, the finishing steps must fix the color (if fixation has not occurred during application) and remove any loose dye from the surface of the colored substrate. Residual dyeing assistants such as "carriers" also must be removed.

The types of textile structures that lend themselves to continuous dyeing methods are wovens and tufted carpets. Continuous dyeing is designed for long runs of similar product; it is a high-output method of dye application.

The first volume-yardage continuous process was the continuous pad-steam process for vat dyes on cotton. The vat dye dispersion was padded onto the cloth and dried; this was followed by passage through a reducing bath, steaming for 30 sec, passage through an oxidizing bath and, finally, washing. When it was discovered that disperse dyes could be thermosoled into polyesters by treatment with dry heat for 60 sec and 400°F, this procedure was readily adapted to continuous processing. The advent of large volumes of dyed polyestercotton-blend fabrics in the late 1960s made it possible to combine these two processes into one thermosol pad-steam system.

Tufted nylon carpet grew to be the number one floor covering in the United States in recent decades. Continuous open-width ranges were developed but not without a great deal of ingenuity to deliver the precise loading of liquid to the tufted surface. This was accomplished by a dye applicator that flooded the dye solution onto the carpet surface. The advancing technology in continuous, metered dyeing systems has created a need for dyes in liquid form, both dispersions and solution. The dyes used in carpet dyeing, for the most part, are supplied by the dye manufacturers as liquids (Fig. 23.25).

Figure 23.25 Fluidyer Carpet Range Dye Applicator. (Courtesy of Zima Corporation.)



Dyeing Methods/Batch

Virtually all types of textile structures are dyed by batch (or exhaust) methods of dyeing, such as stock, yarn, circular knits, warp knits, woven fabrics, garments, carpets, and so on. Batch methods include beck dyeing, jig dyeing, pad-batch, beam dyeing, and others. These methods are dictated primarily by the physical structure of the textile product and the type of fiber(s) it contains. Each of these batch methods employs a different type of machine. As an example, a circular knit fabric comprised of cotton could be dyed in a beck, whereas the same structure comprised of polyester most likely would be dyed in a high pressure jet machine, and a garment constructed from the circular knit cotton likely would be dyed in a garment machine.

Stock dyeing often is carried out in large heated kettles made of stainless steel or other corrosion-resistant metal. These kettles can be sealed and used for dyeing at temperatures somewhat above the boiling point of water at atmospheric pressure.

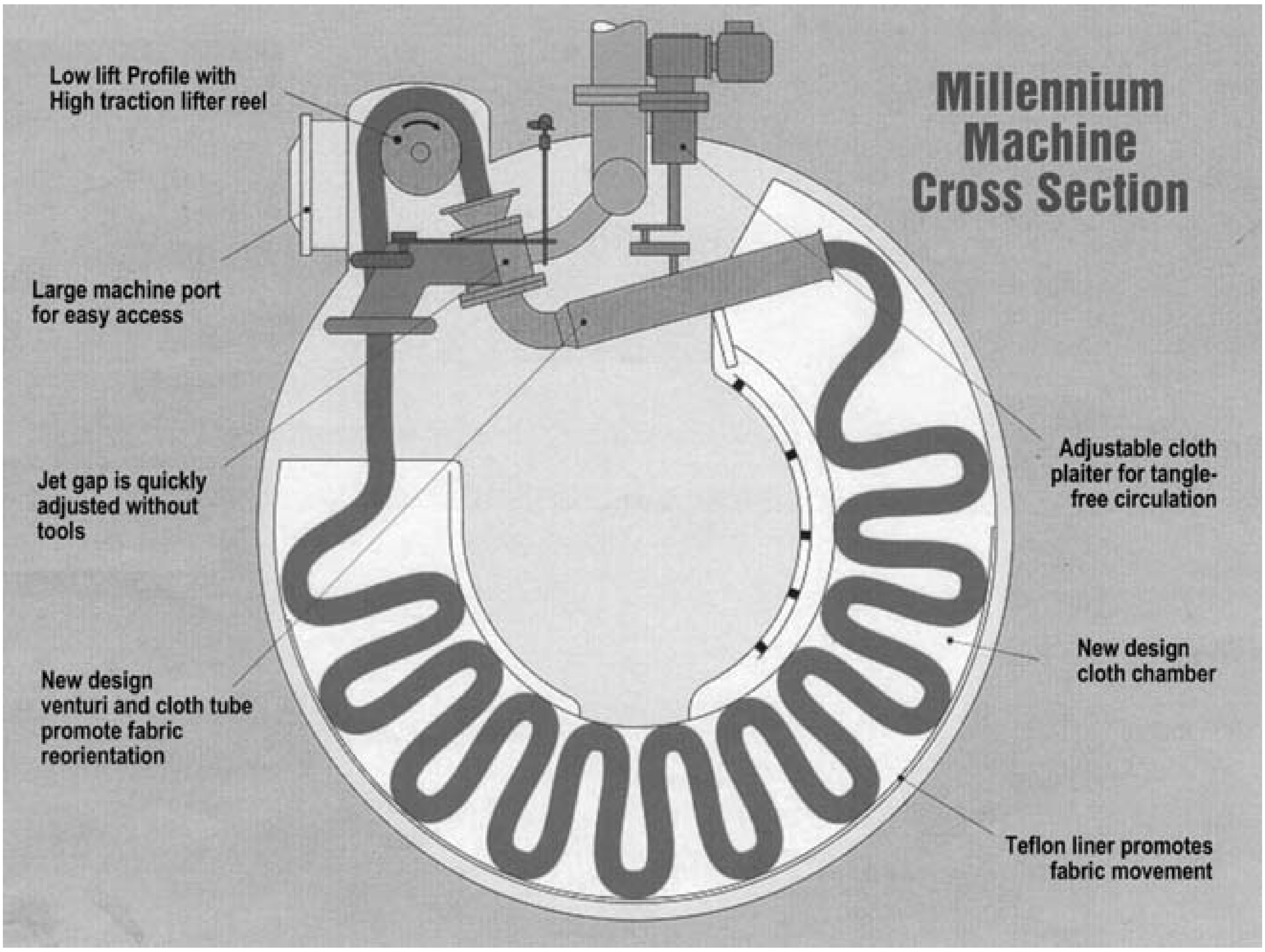
Yarns are dyed in package machines. In this arrangement the yarn is wound onto perforated dye tubes and placed on spindles that are fitted into a closed kettle. The dye solution is heated and pumped through the spindle and yarn package. A cycle of insideoutside flow usually is used to provide level dyeing by equal exposure of the dye to yarns. Although the basis of package dyeing has not changed, a number of refinements have been introduced in recent years. Precision winding of the yarn has improved quality by giving a more uniform package density. Horizontal machines and valving between chambers to allow reconfiguration of the dye machine to control the size of the dyeing have changed the way package dye houses are built. Robotization has been widely utilized to load and unload machines. Also lower-ratio dyebaths with higher flow rates have improved the energy efficiency of the newer machines.

Fabrics are dyed in machines that move them through the dye liquor either under tension (jig) or relaxed (beck). Fabrics can also be dyed in full width by winding them on a perforated beam through which hot dye liquor is pumped. This is the principle of the beam dyeing machine.

The pressure-jet dyeing machine is unique in that it has no moving parts. The cloth, in rope form, is introduced into a unidirectional liquid stream enclosed in a pipe. Liquor is pumped through a specially designed venturi jet imparting a driving force that moves the fabric. The two fabric ends are sewn together to form a continuous loop.

The first jet machine was introduced in 1965. There are two major types of jet dyeing machines: the vertical kier and the elongated horizontal kier (see Fig. 23.26). In general, the kier uses small water volumes, whereas the elongated types use larger-volume ratios in dyeing. The kier types are normally used for more substantial fabrics, and the elongated types are suited for fine or delicate fabric styles. Important features in today's machines are improved corrosion-resistant alloys and the ability to operate at higher efficiencies with minimum energy consumption. The control systems have been refined; there is simultaneous loading and unloading. Larger-capacity machines are also being built; a jet dye machine has been developed for carpet dyeing.

Figure 23.26 Millennium Jet Dyeing Machine. (Courtesy of Gaston County Dyeing Machine Co.)


Millennium Machine Cross Section
Low lift Profile with High traction lifter reel
Large machine port for easy access
Jet gap is quickly adjusted without tools
New design venturi and cloth tube promote fabric reorientation
Adjustable cloth plaiter for tangle-free circulation
New design cloth chamber
Teflon liner promotes fabric movement

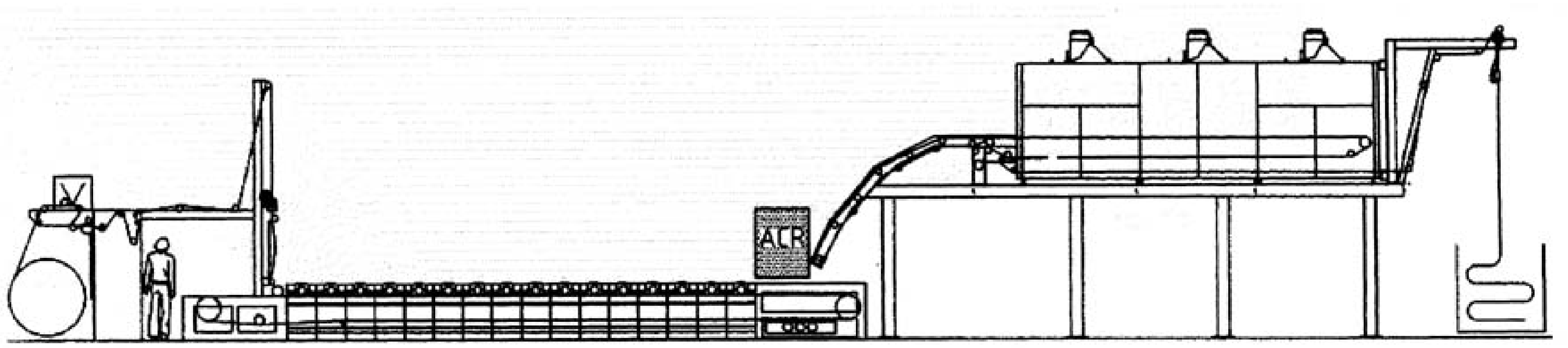
23.5 Printing

Printing is a special kind of localized dyeing that produces patterns. Four kinds of printing have long been recognized: (1) direct, (2) dyed, (3) discharge, and (4) resist. In direct printing, a thickened paste of the dye is printed on the fabric to produce a pattern. The fabric is then steamed to fix the dye and is finished by washing and drying. Dyed printing requires that the pattern be printed on the fabric with a mordant. The entire piece is then placed in a dye bath containing a mordant dye, but only the mordanted areas are dyeable. Washing then clears the dye from the unmordanted areas, leaving the pattern in color.

In discharge printing, the cloth is dyed all over and then printed with a substance that can destroy the dye by oxidation or reduction, leaving the pattern in white. When a reducing agent such as sodium hydrosulfite is used to destroy the dye, the paste may contain a reduced vat dye. Finishing the goods by oxidation and soaping then produces the pattern in color. In resist printing, certain colorless substances are printed on the fabric. The whole piece then is dyed, but the dye is repelled from the printed areas, thus producing a colored ground with the pattern in white.

Printing is most often done with rotary screens etched in the design to be printed. Printing paste is fed constantly to the center of the rotating screen from a nearby supply, and a squeegee pushes the colored paste through the holes in the screen, leaving the dye paste only in the intended areas; a separate screen is required for each color in the pattern (see Fig. 23.27).

Fig. 23.27 Rotascreen V Rotary Screen Print Machine. (Courtesy of Zimmer Machinery Co.)



An important recent advance in the pattern-coloring of textiles is ink-jet or digital printing. Milliken's Millitron and Zimmer's ChromoJet have been successfully used for carpet and upholstery markets for over two decades. Finer resolution machines began to emerge in the late 1980s when Stork introduced a prototype machine. During the past five years a number of manufacturers have introduced digital ink-jet printers which use either CYMK (Cyan, Yellow, Magenta, and Black) to make a composite color or true-color machines which use mixed pigment systems. The current machines are very successful at furnishing one of a kind and for use in rapid prototyping.

23.6 Pigment Dyeing and Printing

Pigment dyeing and printing are processes that compete with the more conventional means of dyeing and printing described above. These processes use water-insoluble dyes or pigments that are bound to the surfaces of fabrics with resins. A paste or an emulsion, containing pigment and resin or a resin-former, is applied to the fabric. The goods are then dried and cured by heat to produce the finished dyeing or print. During the heating or curing, fabric, resin, and pigment become firmly bonded together. This method of color application is economical and produces good results. It should be noted that the pigment is confined to the surface of the fabric and can be selected without regard for fiber affinity.

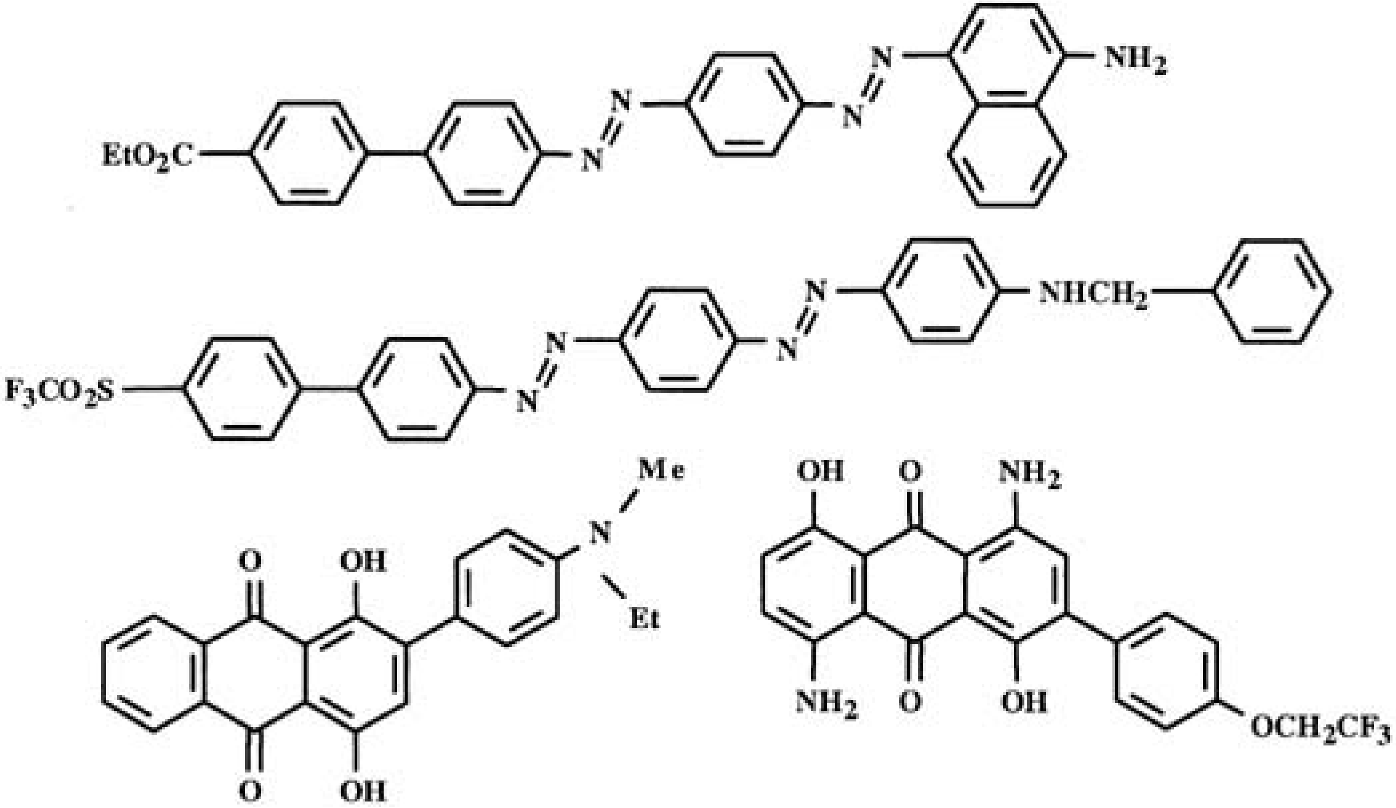
23.7 Nontextile Uses of Dyes

Colorants for nontextile use have been developed mainly for use in hair dyeing, photography, biomedical applications, and electronics and reprographics.3339 Dye application areas involving the latter types include ink-jet printing, thermal or pressure dye transfer, laser printing, liquid crystal displays, optical data storage, and nonlinear optics. In several nontextile applications, dyes are not used for their ability to deliver color. Instead, they are used because of their potential electrical properties, such as photoconduction and electrostatic charging of toners, and in some cases they are used because they absorb IR radiation, which induces heating effects. The latter property is important in optical data recording.

Liquid Crystal Dyes

Dyes for liquid crystalline media must have (1) nonionic structures, (2) high purity, (3) solubility and compatibility with the medium, (4) a transition dipole that is parallel with the alignment axis of the molecular structure, and (5) good alignment with the liquid crystal molecule.40 Examples include the disazo and anthraquinone dyes in Fig. 23.28.

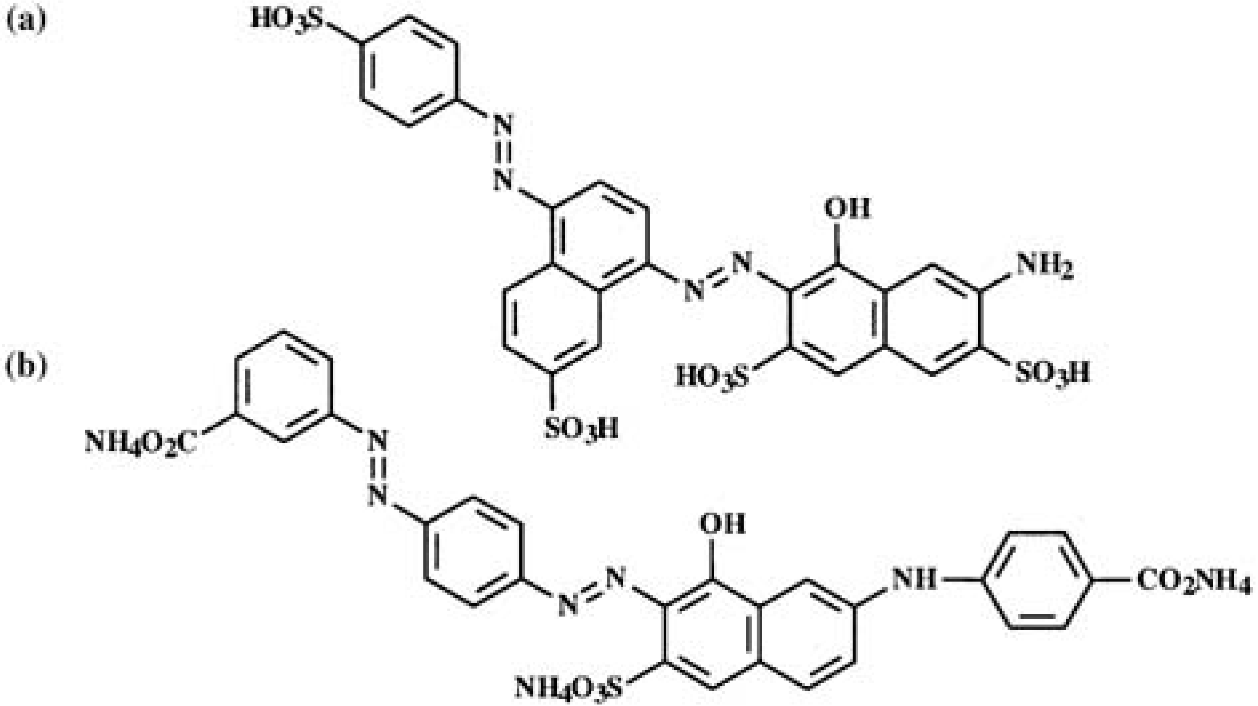
Fig. 23.28 Azo and anthraquinone dyes used in liquid crystalline media.



Ink-jet Dyes

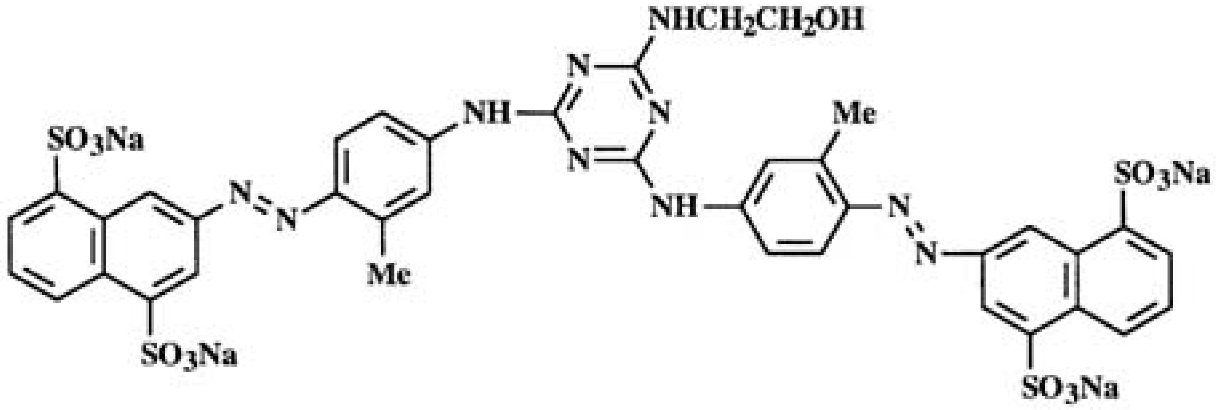
Dyes used in this area must have the following properties: (1) water solubility at high concentrations, (2) low toxicity, (3) good stability to color loss due to exposures to UV light, heat, and moisture, (4) quick fixation to paper following application (deposition), and (5) good color strength.41 To achieve high resistance to removal by water (wetfastness), ink-jet dyes often contain fewer sulfonate groups and one or more carboxylate groups. This change in structural features allows the dyes to have good solubility in alkaline ink formulations but high wetfastness following deposition. This change in solubility behavior is known as *differential solubility*.42 Structures in Fig. 23.29 illustrate the type changes made to C.I. Food Black 2, an early ink-jet dye, to provide good wetfastness.

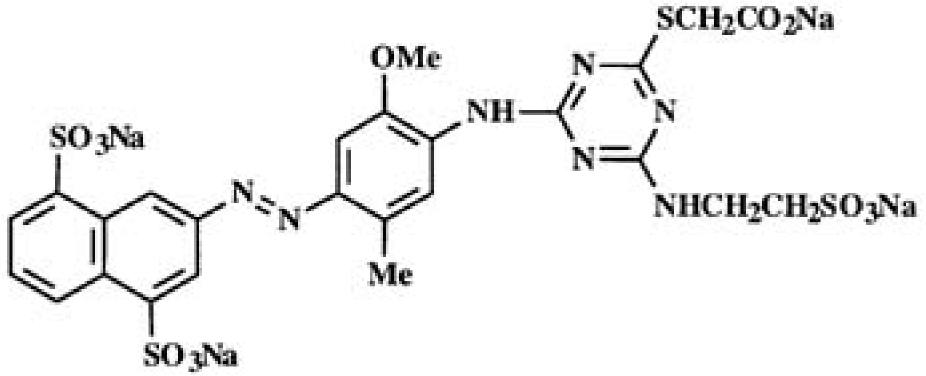
Fig. 23.29 (a) Food Black 2 and (b) its wetfast analog.



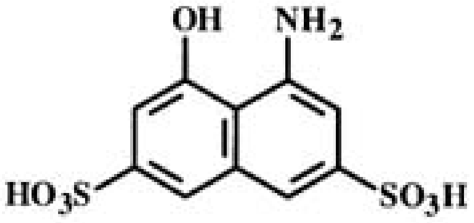
New water-soluble yellow dyes for ink-jet printing are similar to the initially used dye C.I. Direct Yellow 86, except that they are smaller (Fig. 23.30). The size change permits the solubility needed for high-throughput ink cartridge systems, without clogging the ink-jet nozzles.

Fig. 23.30 Examples of yellow ink-jet dyes.

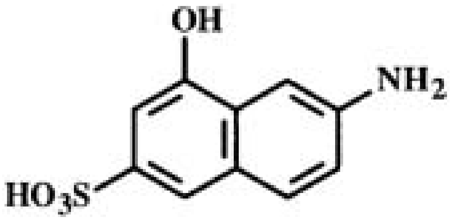




Most of the new water-soluble magenta dyes are based on H-acid. Examples include dyes that contain a fluorocyanophenyl group (Fig. 23.31). To improve the lightfastness of magenta dyes, gamma acid can be used as the coupling component. For very bright magenta prints, dyes based on xanthene structures can be used. Examples include C.I. Acid Red 52, which has low photostability, and a carboxylated analog, which has better photostability and wetfastness (Fig. 23.32).



H-acid



Gamma acid

Fig. 23.31 Examples of magenta ink-jet dyes based on (a) H-acid and (b) gamma acid.

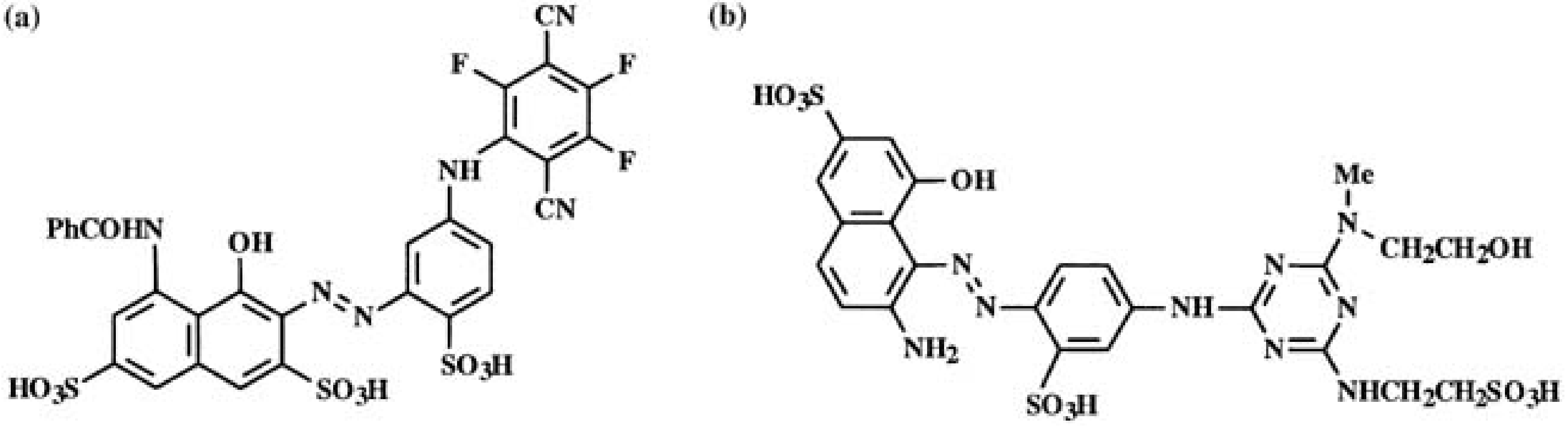
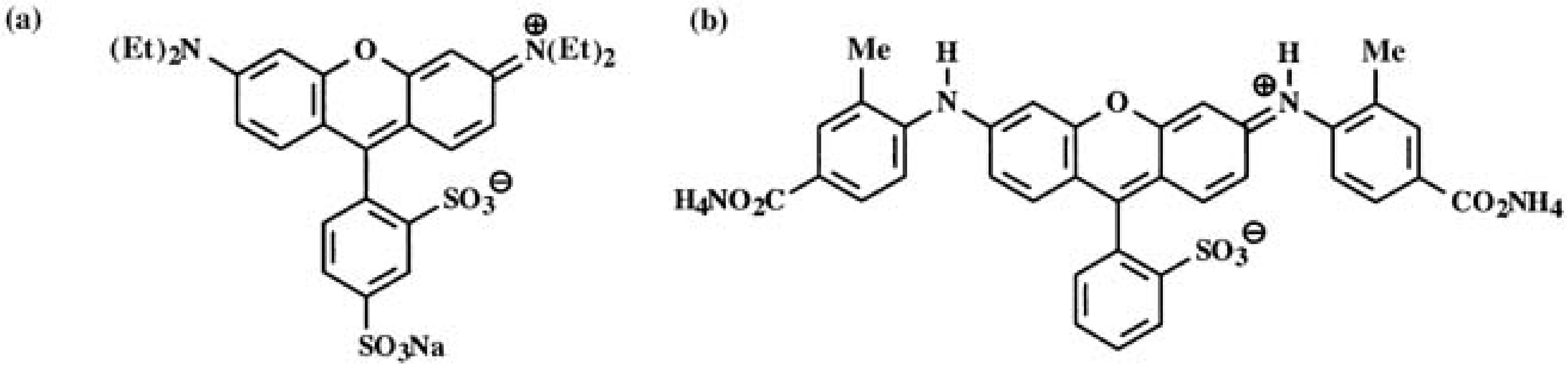
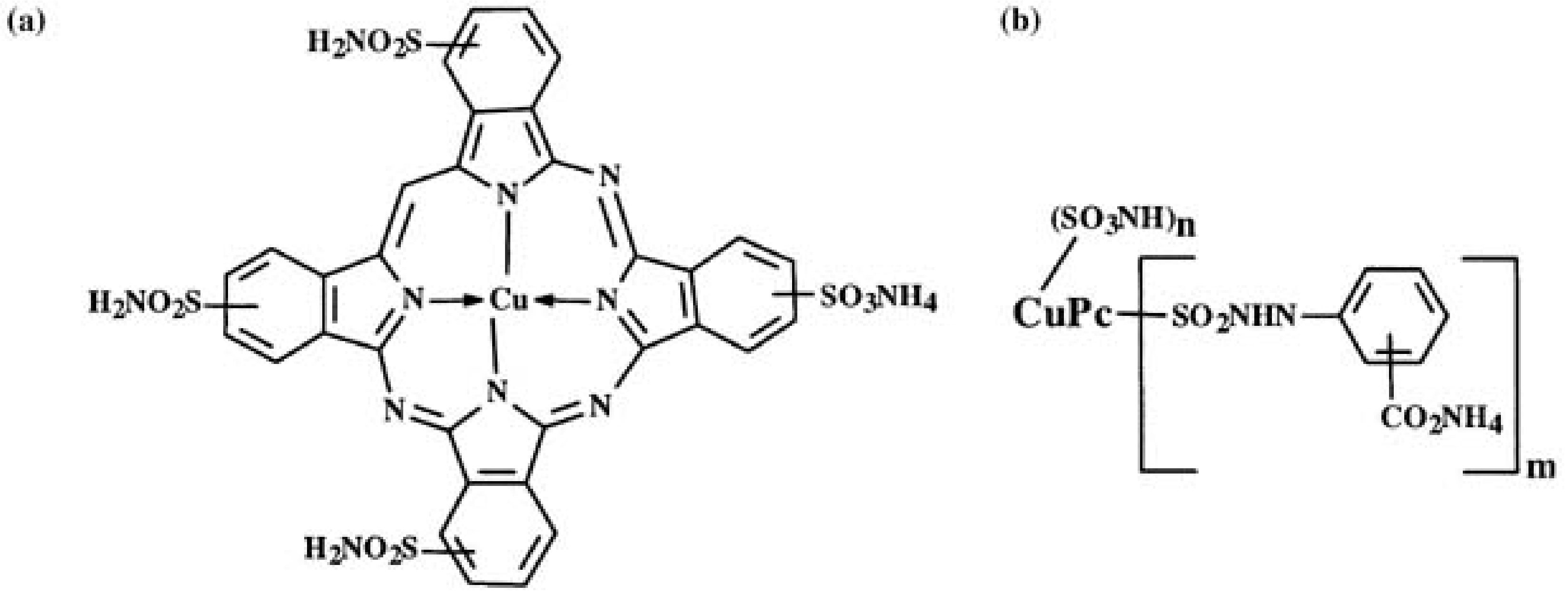


Fig. 23.32 (a) Acid Red 52 and (b) a more lightfast analog.



The water-soluble cyan dyes continue to be based on the copper phthalocyanine system. In this regard, C.I. Direct Blue 199 has proved effective, due to its good color strength and photostability. Carboxylated analogs of this type of dye have also been developed to enhance wetfastness (Fig. 23.33).

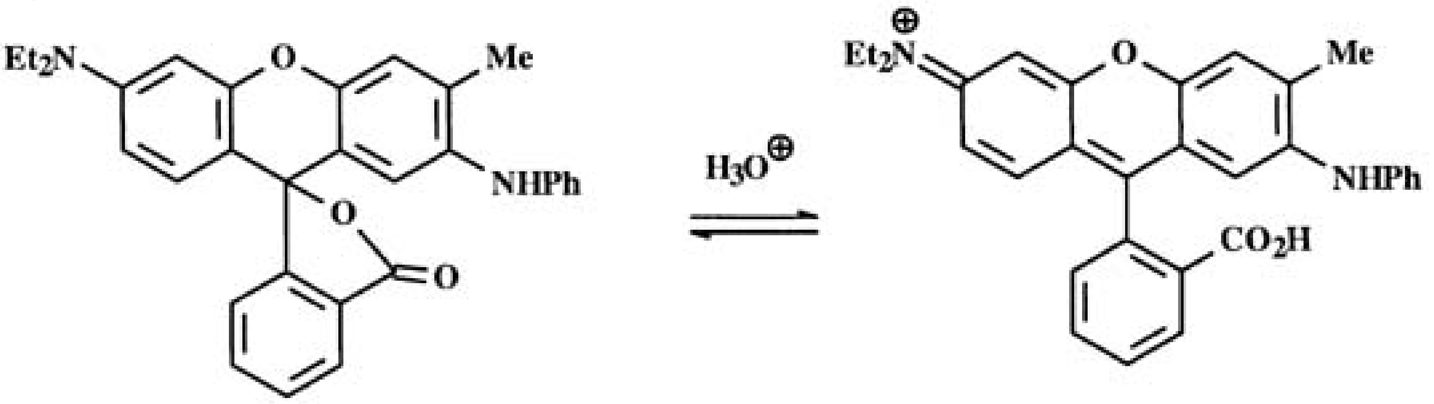
Fig. 23.33 (a) Direct Blue 199 and (b) a wetfast analog, where Pc = phthalocyanine.



Thermal and Pressure Sensitive Printing

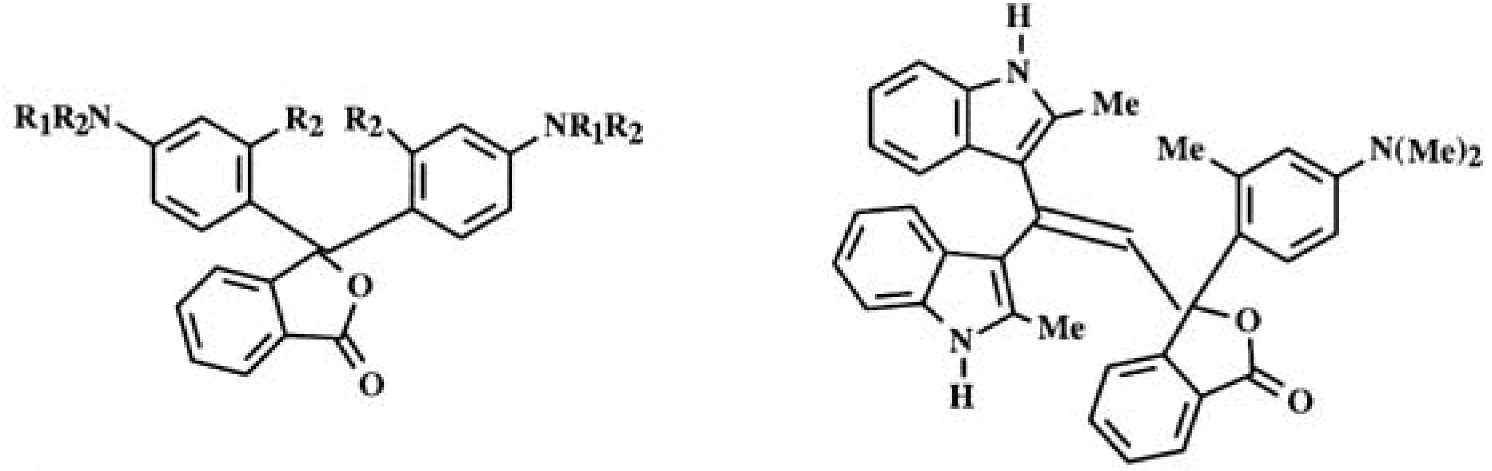
In direct thermal printing, a color former (colorless) and a developer (acidic) are brought into contact in the presence of heat, to produce color on paper.43 The most important color in thermal printing is black and the majority of the color formers are fluorans (Fig. 23.34). The most important application of direct thermal printing is in facsimile machines.

Fig. 23.34 Acid-induced conversion of a color former to its colored form.



In pressure sensitive printing technology the color former is dissolved in a solvent and encapsulated.44 The use of pressure (pen, typewriter key) ruptures microcapsules containing the color former, which generates color upon contacting a developer. Black prints are usually obtained either from fluorans or from color former mixtures. Compounds of the type shown in Fig. 23.35 can be used in two- and three-component mixtures.

Fig. 23.35 Color former combinations suitable for pressure-sensitive printing.

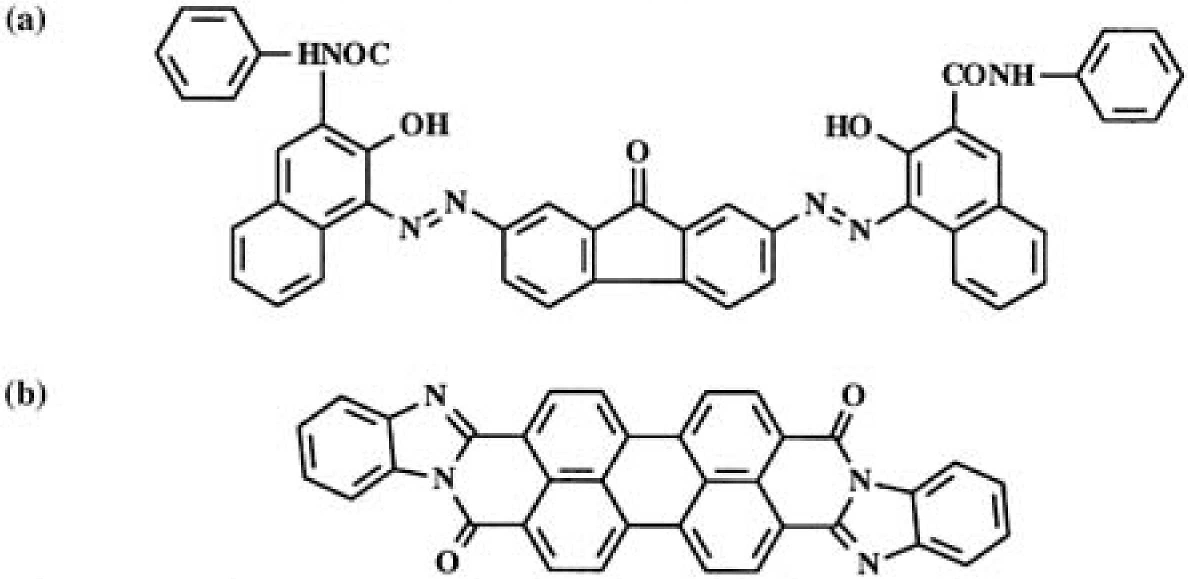


R_1: $C_{1\text{-}6}$ alkyl, R_2: $C_{1\text{-}6}$ alkyl, alkoxyalkyl

Organic Photoconductors and Toners

Photoconductors and toners are used in photocopiers and laser printers to produce images.45 Organic photoconductors consist of a charge-generating layer and a charge-transporting layer. The former is comprised of pigments and the latter is comprised of electron-rich organic compounds that are usually colorless. Suitable organic pigments for charge generation include azo pigments, tetracarboxydiimides, polycyclic quinones, phthalocyanines, perylenes and squarylium compounds (e.g., Fig. 23.36).

Fig. 23.36 (a) Azo and perylene pigments for charge generation.

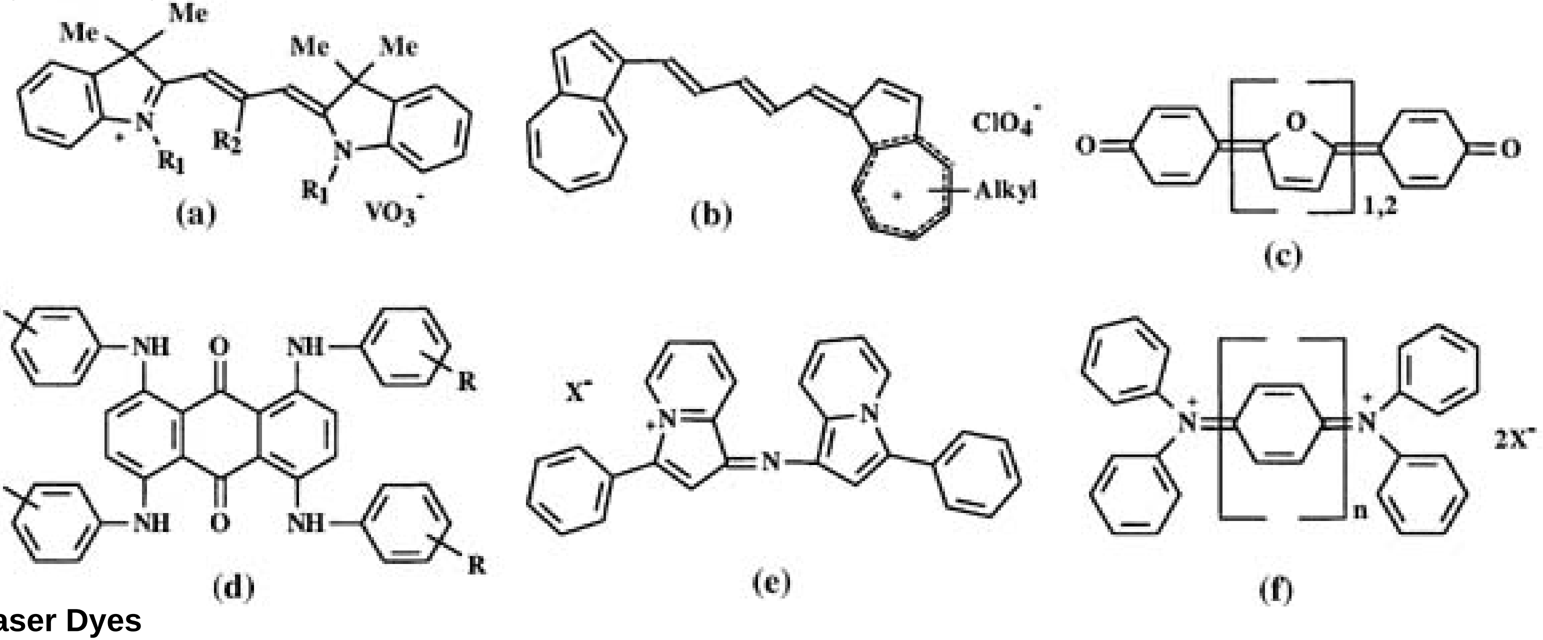


Colorants are used in toners to provide color and control the electrostatic charge on toner particles. Diarylides and monoarylides have been used as the yellow pigments in colored toners. The magenta pigments are often quinacridones and the cyan pigments are copper phthalocyanines.

Infrared Absorbing Dyes

Infrared dyes include indoleninecyanines, azulenium compounds, both of which are used in optical recording materials.46 Other examples are metal (Mn, Fe, Co, Cd, Al, Cu, Pd)-complexed phthalocyanines, quinones, quinonoids, and iminium and diiminium compounds (Fig. 23.37).

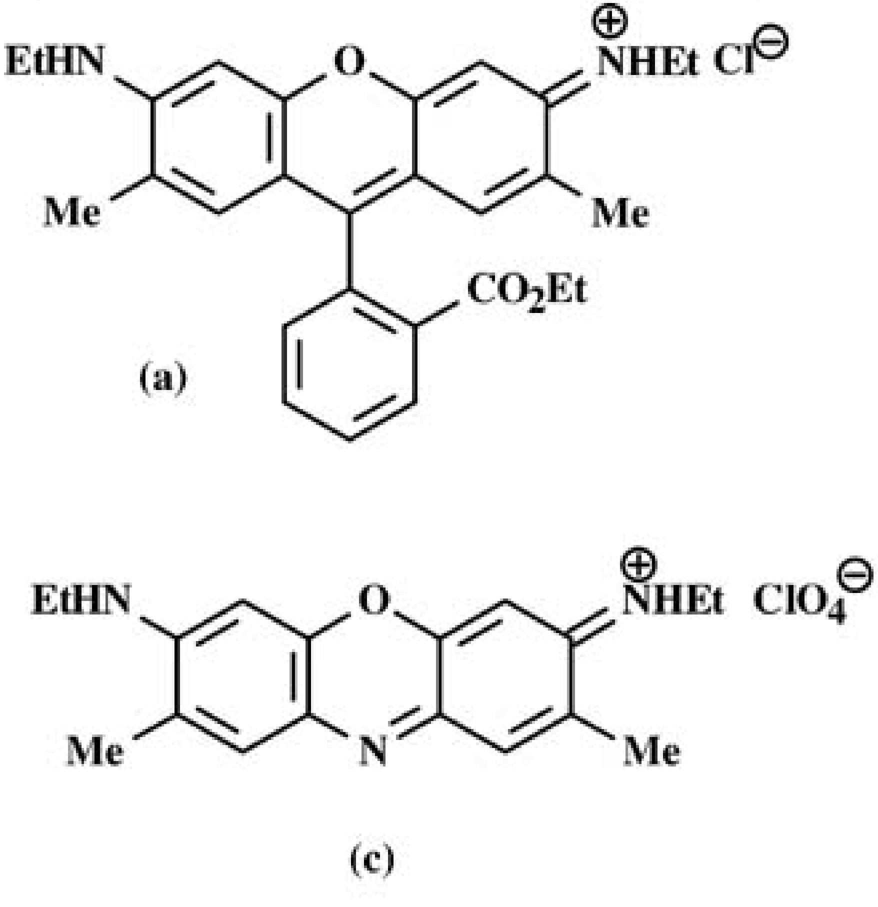
Fig. 23.37 Structures of (a) indoleninecyanine, (b) azulenium, (c) quinonoid, (d) anthraquinone, (e) iminium and (f) diiminium IR-absorbing dyes, where R = alkyl or alkoxy.



Laser Dyes

Lasers in which dyes comprise the active medium have become one of the most widely used types.47 The key virtue of these systems is their ability to cover virtually the entire fluorescence spectral region. Accordingly, the most commonly used dyes are highly fluorescent and include coumarin, rhodamine, oxazine, and *syn*-bimane structures (Fig. 23.38). Dye lasers are employed in liquid form, which allows them to dispel excessive heat by recirculating the dye solution. Good photostability and efficient laser action under flashlamp excitation are important properties.

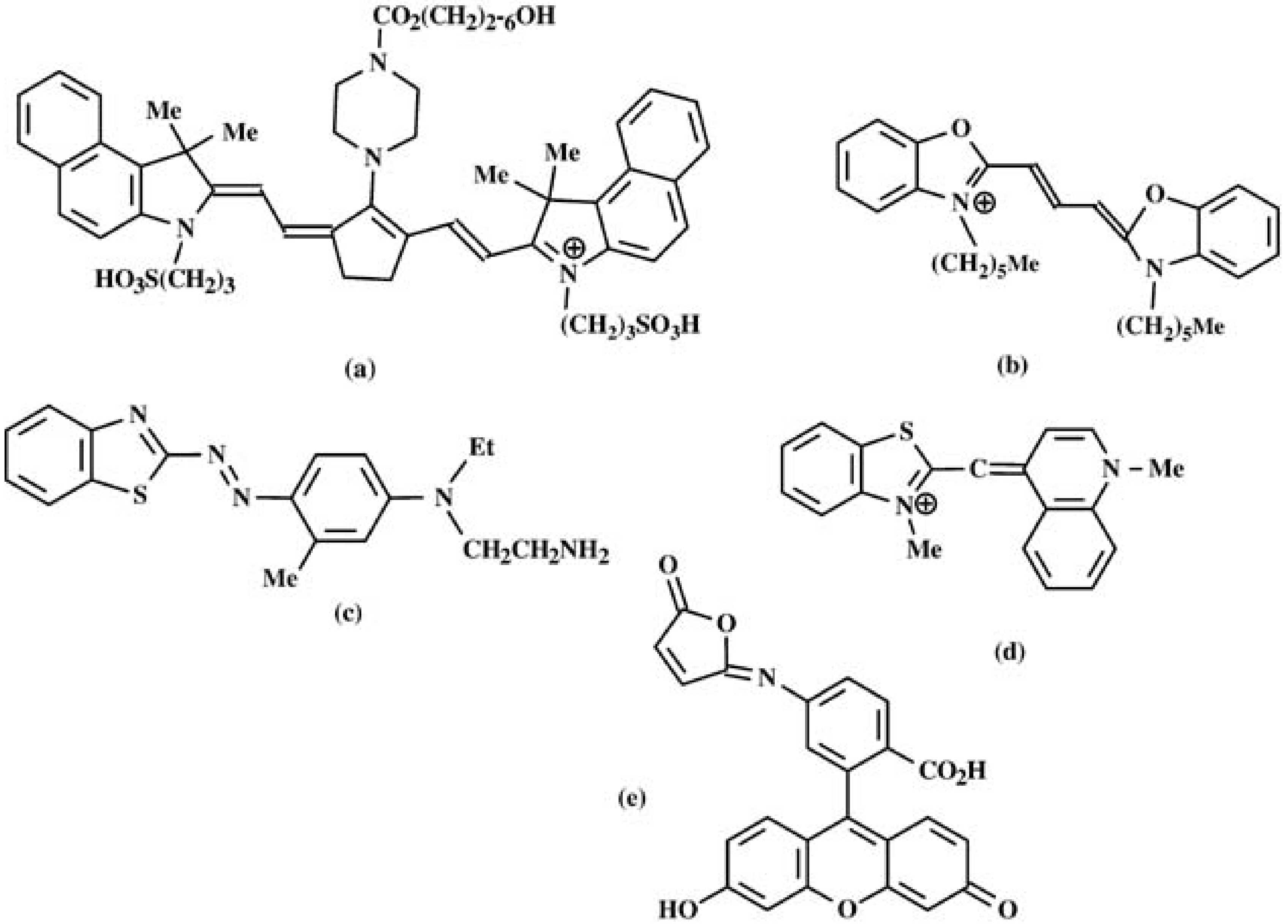
Fig. 23.38 Rhodamine (a), coumarin (b), oxazine (c), and *syn*-bimane (d) laser dye structures.



Biomedical Dyes

Dyes can be used clinically in bioanalysis and medical diagnostics and in the treatment of certain diseases (cf. Fig. 23.39). For instance, near infrared (NIR) and IR-absorbing dyes can be used in automated DNA sequencing (dye (a)), fluorescent dyes can be used in cancer detection (dye (b)), and certain azo and heterocyclic dyes can be used in virus (dye (c)), cell (dye (d)), and bacteria (dye (e)) detection. An in-depth summary of dyes in this area has been published recently.48

Fig. 23.39 Examples of experimental dyes used in biomedical applications.



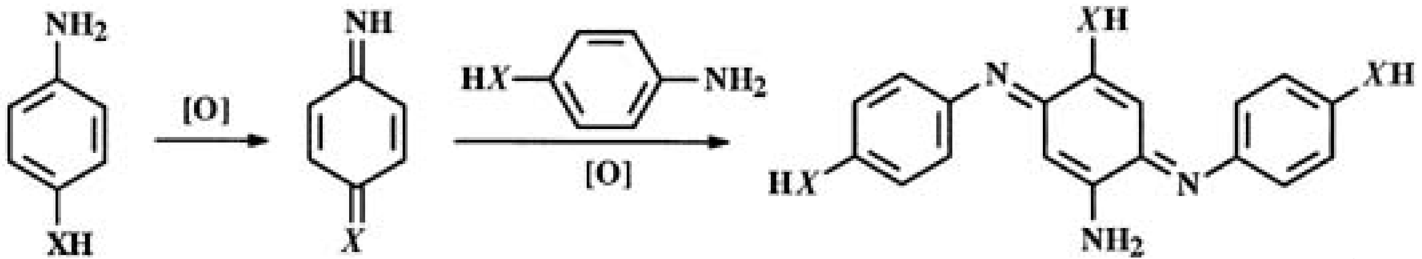
Hair Dyes

About 80 percent of the dyes used in hair coloring are known as oxidation hair dyes.49,50 The remaining 20 percent of the available hair dyes are mainly synthetic dyes that have affinity for protein substrates. Oxidation dyes are produced directly on hair by oxidizing aromatic diamines (e.g., *para*-phenylenediamine or 2,5-diaminotoluene) with a suitable oxidizing agent. In this regard, the diamines have been referred to as "primary intermediates" and the oxidizing agents (e.g., hydrogen peroxide) as "developers." Other suitable primary intermediates are aminodiphenylamines, aminomethylphenols, and *para*-aminophenol.

When used alone, the primary intermediates give a quite limited shade range following oxidation on hair. To enhance the range of available hair colors, the primary intermediates are oxidized in the presence of suitable "couplers." Whereas most couplers do not produce colors when exposed to developers alone, they give a wide array of hair shades in combination with primary intermediates. Suitable couplers include 3-aminophenol, resorcinol, and α-naphthol.

The chemistry associated with the oxidation of primary intermediates is now reasonably well known. For *para*-phenylenediamine and *para*-aminophenol, this involves the process outlined in Fig. 23.40. It can be seen that dye formation is a two-step process involving oxidation and self-coupling.

Fig. 23.40 Oxidation of primary intermediates, where X = O, NH.



C.I. Basic dyes such as Yellow 57, Red 76, Blue 99, Brown 16, and Brown 17 have been used in color refreshener shampoos and conditioners. Similarly, C. I. Acid dyes such as Yellow 3, Orange 7, Red 33, Violet 43, and Blue 9 have been used in shampoos, in this case to deliver highlighting effects.51

Photographic Dyes

Color photography is still one of the most important and interesting nontextile uses for synthetic dyes. The chemistry employed is comparable to that described above for oxidation hair dyes, in that an oxidizable substrate (e.g., phenylenediamine) is combined with a coupler to produce the target colorant. In this case the diamine is referred to as the "developer," and it is oxidized by silver halide in the photographic film. The oxidized developer then reacts with the coupler to form the dye. This process produces a negative dye image consisting of yellow, magenta, and cyan dyes in proportion to the amount of red, blue, and green light absorbed by the film.52

Some widely used developers are shown in Fig. 23.41. They can be used to produce the yellow, magenta, and cyan dyes shown in Fig. 23.42. These dye structures demonstrate that acetoacetanilide, pyrazolone, and indoaniline intermediates are useful for producing yellow, magenta, and cyan colors, respectively.

Fig. 23.41 Structures of some developers used in color photography.

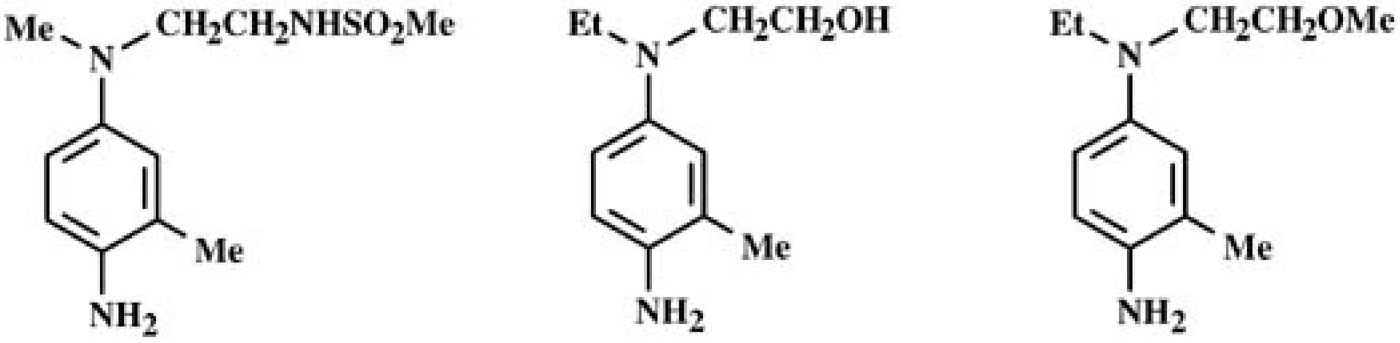
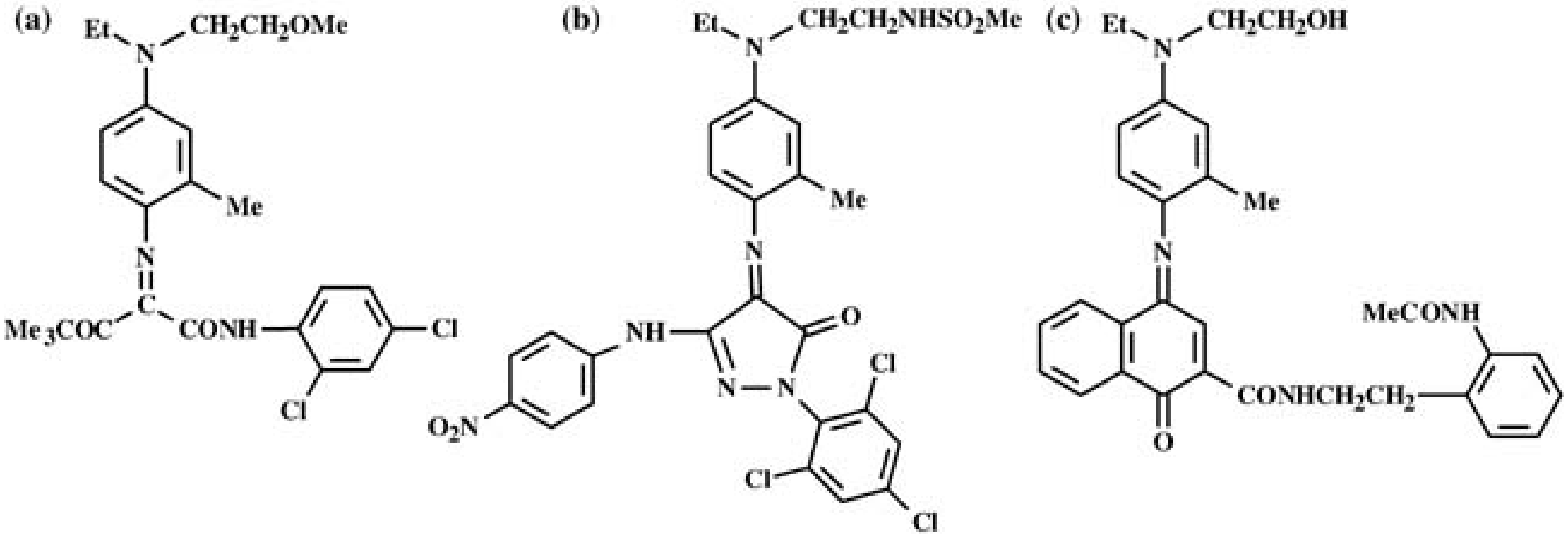


Fig. 23.42 Structures of (a) yellow, (b) magenta, and (c) cyan photographic dyes.

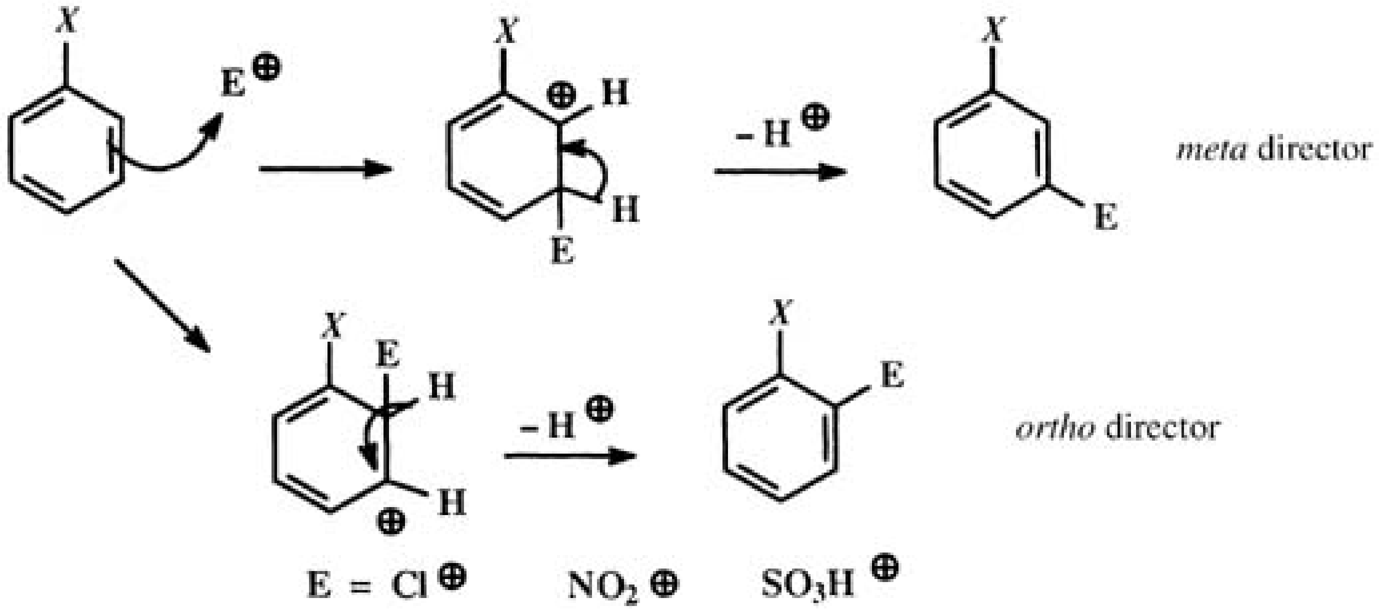


23.8 Dye Intermediates

The raw materials used to synthesize organic dyes are commonly referred to as dye intermediates. Largely, they are derivatives of aromatic compounds obtained from coal tar mixtures. The majority of these derivatives are benzene, naphthalene, and anthracene based compounds. This section provides an overview of the chemical reactions used to prepare the key intermediates employed in dye synthesis. In this regard, emphasis is placed on halogenated, aminated, hydroxylated, sulfonated, and alkylated derivatives of benzene, naphthalene, and anthraquinone.

Most dye intermediates are prepared by reactions involving electrophilic or nucleophilic substitution processes. The electrophilic processes include nitration, sulfonation, and halogenation reactions, while the nucleophilic processes include, hydroxylation and amination reactions. Electrophilic substitution reactions are of the form shown in Fig. 23.43. In this regard, the incoming electrophile (electron-seeking species) reacts with the more electron-rich positions. When the aromatic ring contains ring-activating groups (e.g., hydroxy, alkoxy, amino, alkyl), the incoming group will attack *ortho/para* positions. If ring-deactivating groups (e.g., nitro, sulfonic acid, carboxylic) are present, the positions *meta* to the deactivating groups will be attacked.

Fig. 23.43 Electrophilic attack of an aromatic ring containing deactivating and activating groups.



Other key dye intermediates are prepared by oxidation and reduction processes. Examples of each of these processes are covered in the sections that follow.

Nitration

For dye intermediates, this process involves the introduction of one or more nitro (NO2) groups into aromatic ring systems. Nitro groups serve as chromophores (color bearers), precursors for amino groups, and as auxochromes (color aiders). Since they are *meta*-directing groups they are also useful in the strategic placement of another incoming group.

Nitric acid (HNO3) is the chemical agent commonly used in nitration reactions. Depending, upon the degree of ring activation, HNO3 may be used in combination with other acids. In fact, nitrations are often conducted by using a mixture of HNO3 and sulfuric acid (H2SO4). This combination is known as "nitrating mixture" or "mixed acid", and it is especially effective when deactivated ring systems are to be nitrated. Dilute HNO3 or a HNO3/acetic acid (CH3CO2H) mixture can be used for nitrating very reactive ring systems. When the former is used there is also the potential for ring oxidation to occur rather than the desired nitration, depending upon the actual compound undergoing nitration. Examples of nitration reactions are shown in Figs 23.4423.46.

The nitration of toluene (Fig. 23.44) is selected because it illustrates what can happen when monosubstituted benzenes having a ring-activating group are used. In this case, the principal products reflect a statistical mixture of *ortho* and *para* isomers, with only a small amount of the *meta* isomer obtained. Nitration is conducted near 20°C and the products are separated by distillation.

Fig. 23.44 Nitration of toluene using mixed acid.

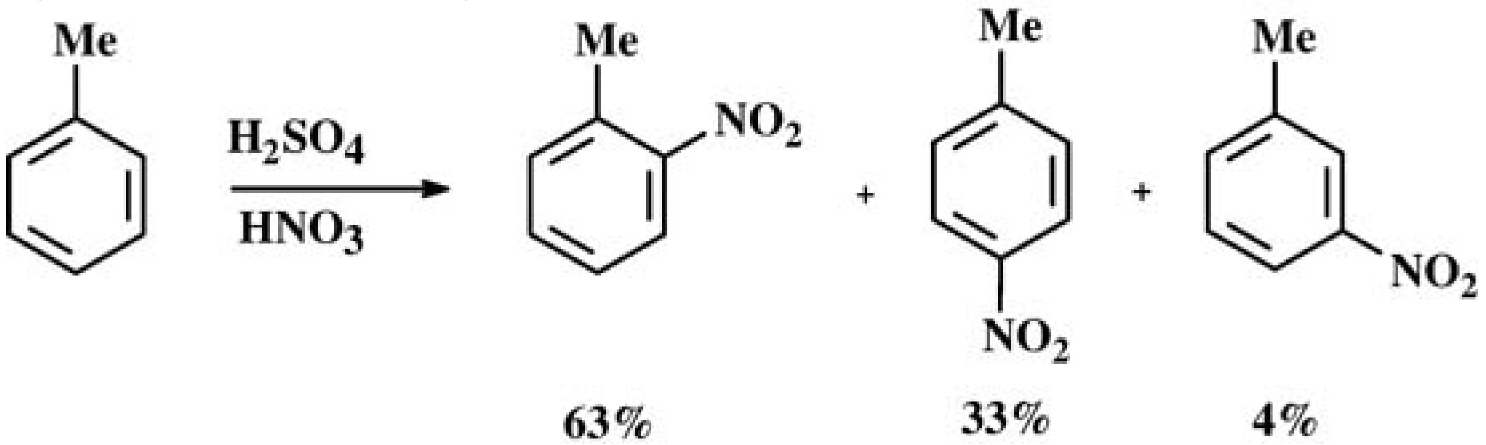


Fig. 23.45 Nitration of naphthalene using mixed acid.

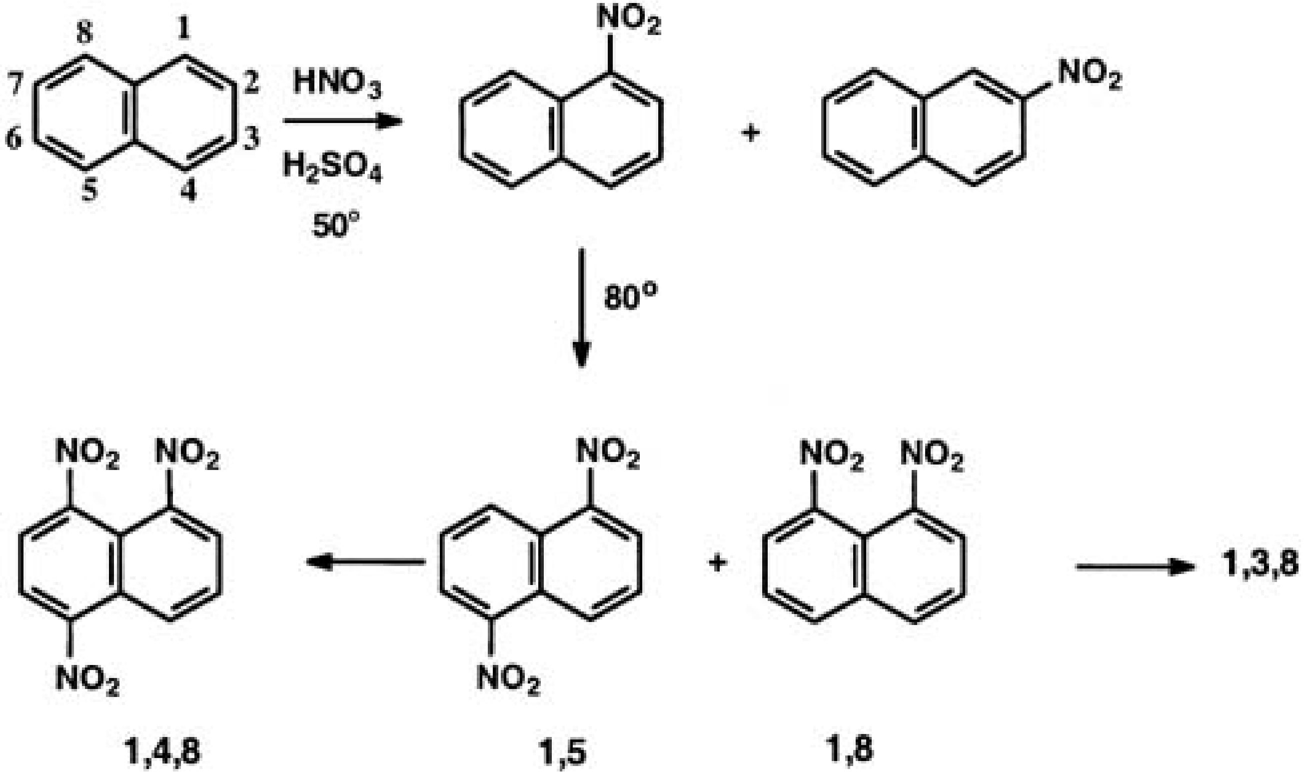
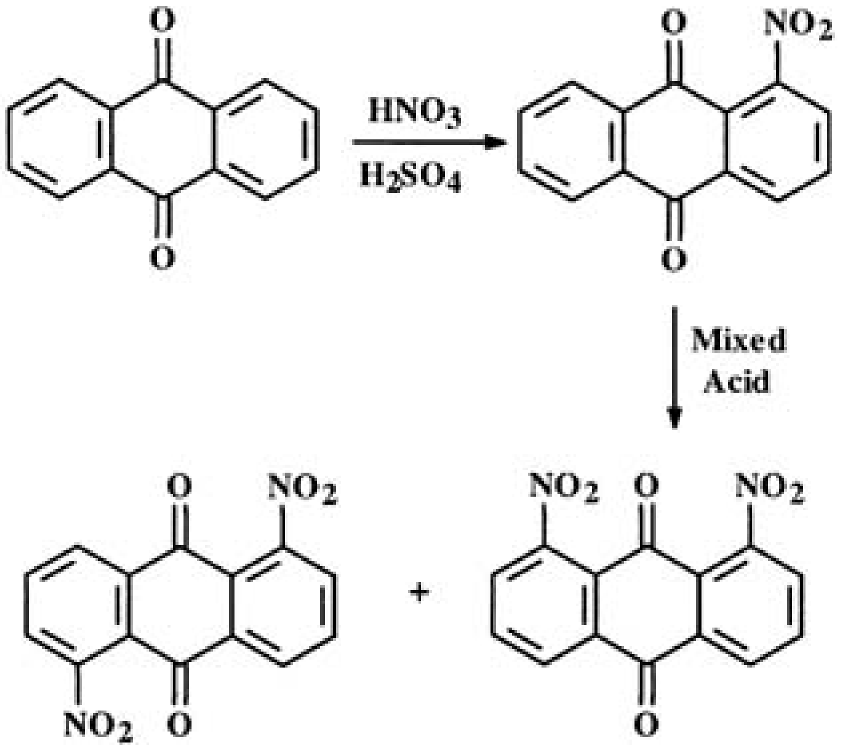


Fig. 23.46 Nitration of anthraquinone using mixed acid.

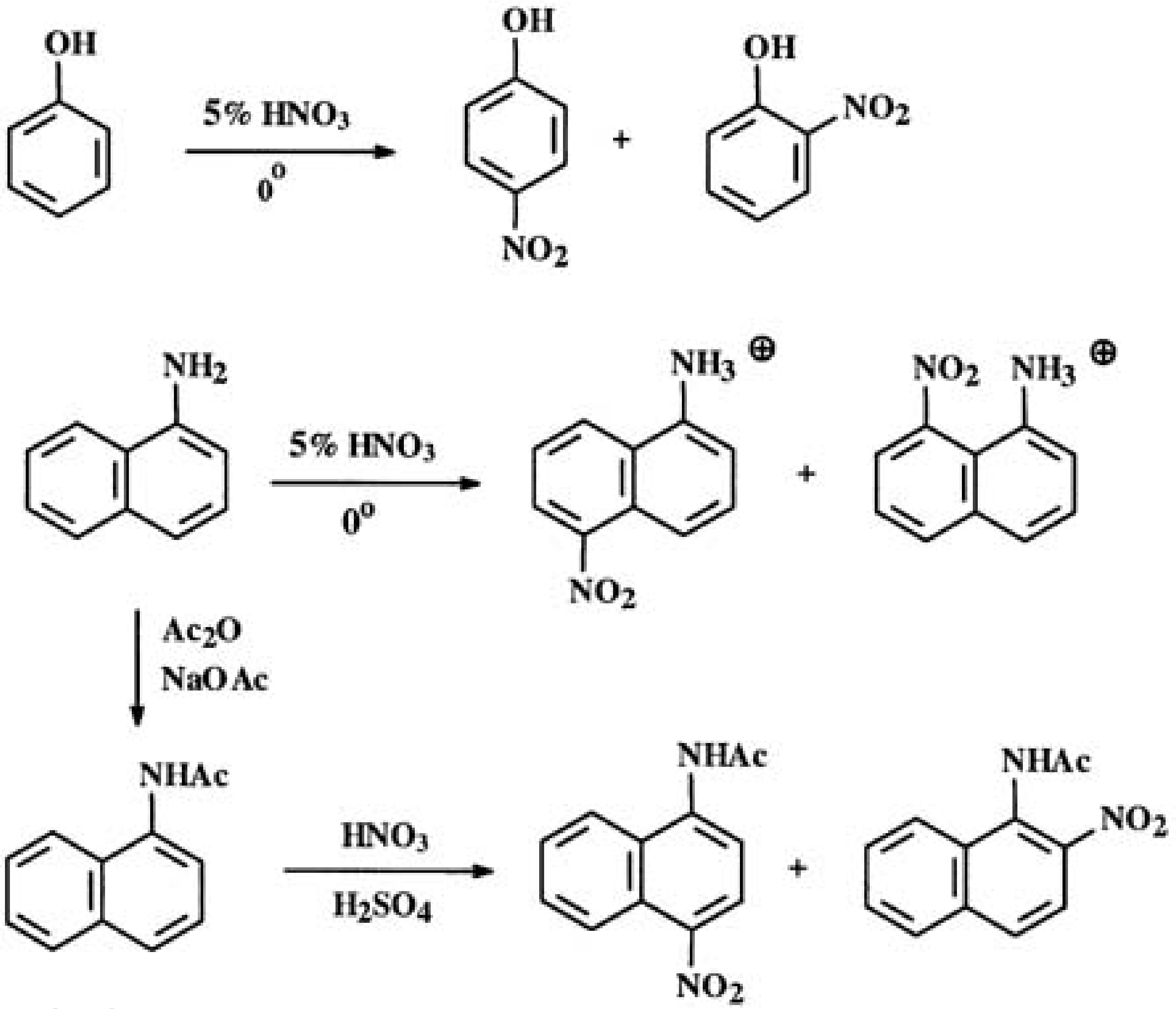


Nitration of naphthalene gives mostly the 1-nitro isomer (•90%), initially. Introduction of a second nitro group takes place in the opposite ring since the existing nitro group reduces the reactivity of the ring to which it is attached. Although, nitro groups are *meta* directors, in this case they can also direct the incoming second (or third) nitro group to a *peri* position. In the naphthalene ring system, the *peri* positions are those that are 1,8 and 4,5 to each other (Fig. 23.45).

The nitration of anthraquinone at 50°C gives, initially, the 1-nitro isomer, and if nitration continues at 8090°, the 1,5 and 1,8 isomers are obtained. Further nitration is impractical and serves to point out that the anthraquinone ring is appreciably less reactive than the naphthalene system. This will be more evident as the chemistry reported in this section continues to unfold.

The nitration of phenols and amines must be conducted with care, as these systems are subject to ring oxidation if the temperature gets too high. For instance, the nitration of phenol itself is conducted near 0°C using 5 percent HNO_3. This gives a mixture of *ortho* and *para* isomers that can be separated by steam distillation (Fig. 23.47). Aromatic amines are often protected by *N*-acetylation prior to nitration. This reduces both the potential for ring oxidation and the amount of *meta* isomer that forms when the amino group undergoes protonation. The protonated amino group (NH_3^+) is a *meta* director, unlike the free amino (NH_2) and the acetylated amino (NHAc) groups. This chemistry is illustrated in Fig. 23.47 for 1-naphthylamine. Following nitration, the acetyl group can be removed by hydrolysis.

Fig. 23.47 Nitration of phenol and 1-naphthylamine.



Reduction

The most important reduction reactions are those leading to aromatic amines that are suitable for azo dye formation. While this usually involves the reduction of a nitro group to an amino (NH_2) group, the reduction of azo groups to amino groups is also an important process. Agents that are commonly used to effect chemical reductions include: (1) Fe + HCl or H_2SO_4, (2) Na_2S, (3) NaSH, (4) Zn + NaOH, (5) H_2 + transition metal catalysts, and (5) $Na_2S_2O_4$. Examples of these reductions are given in Figs 23.4823.51. While the reduction of nitrobenzene can be conducted in a number of ways, a key commercial process involves the method in Fig. 23.48, where high temperature hydrogenation is used.

Fig. 23.48 Commercial process for the reduction of nitrobenzene.

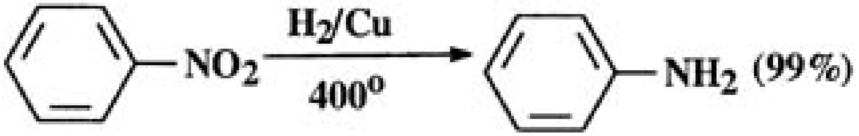


Fig. 23.49 Reduction of (a) azo and (b) nitroanthraquinone compounds.

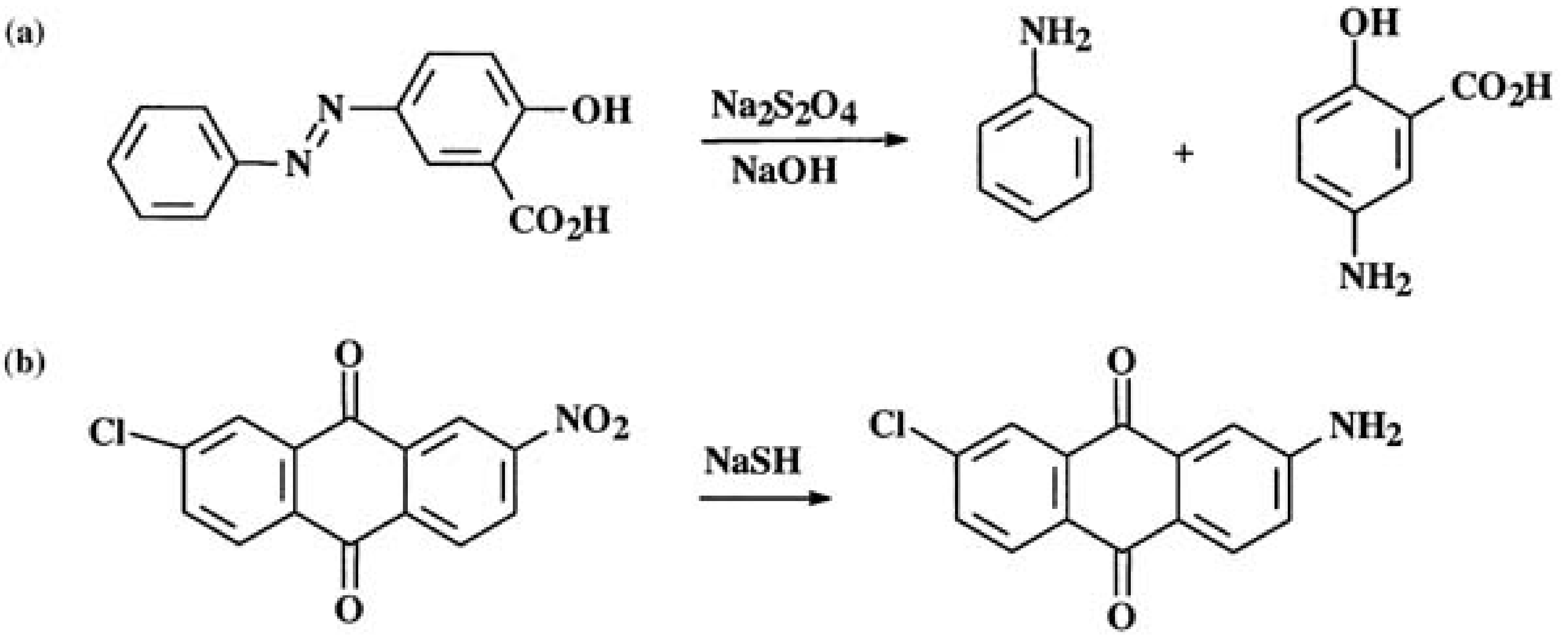


Fig. 23.50 Formation of Koch acid via nitration and reduction steps.

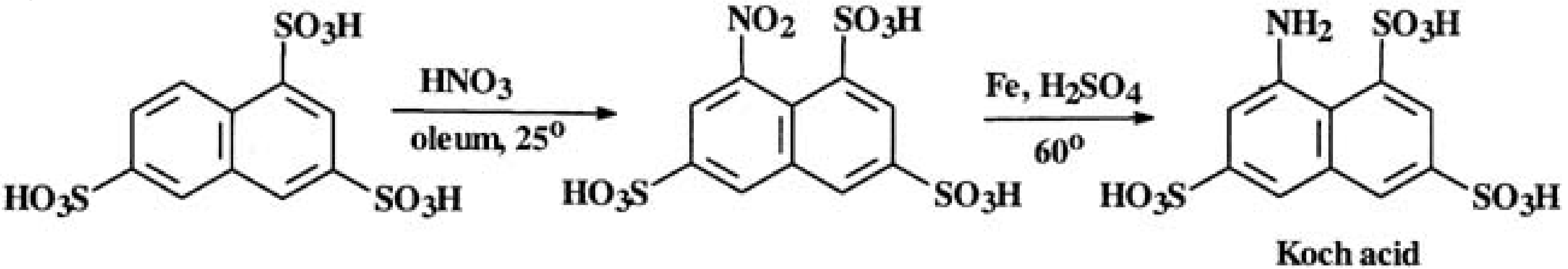
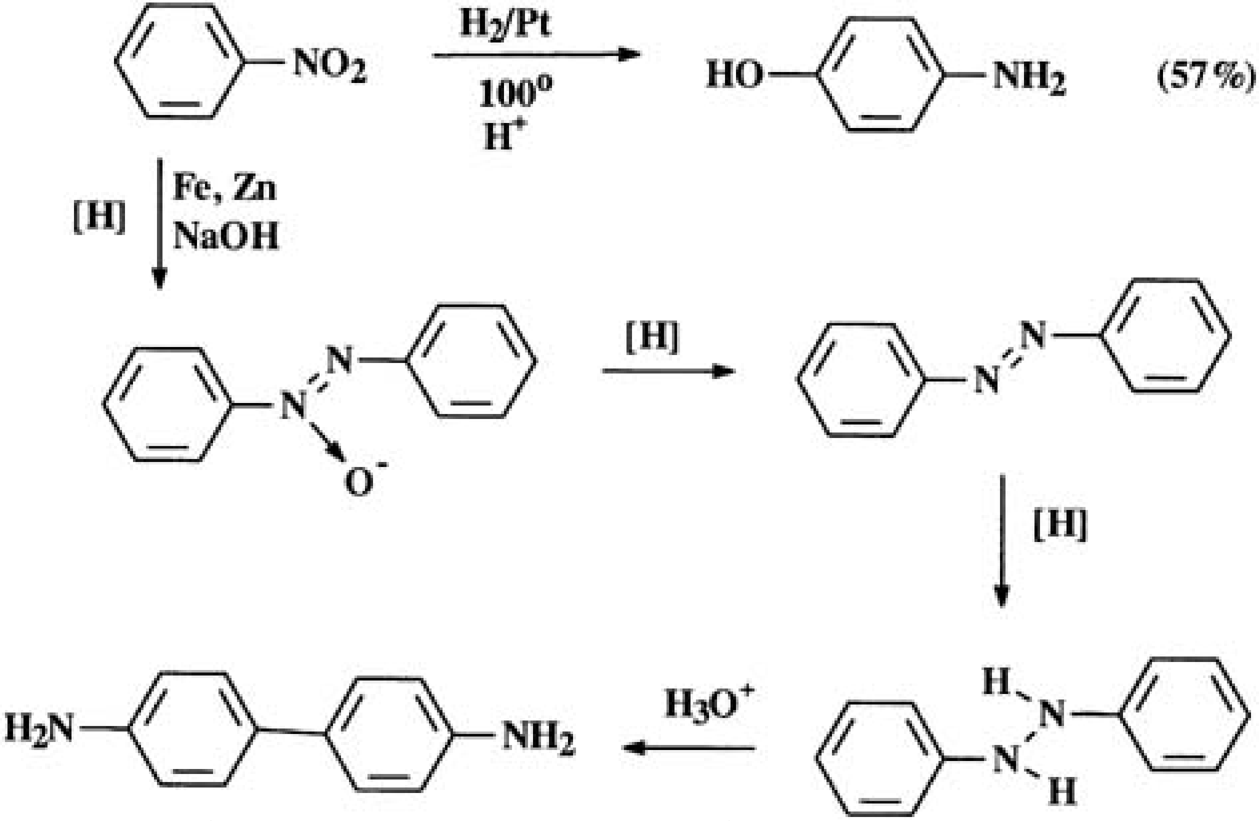


Fig. 23.51 Reduction of nitrobenzene in acidic and alkaline media.



The reduction of azo compounds using sodium hydrosulfite ($Na_2S_2O_4$) and NaOH is an important reaction, as it provides an indirect method for the amination of phenols and naphthols (Fig. 23.49). The reduction of nitro groups in anthraquinone compounds works best when a mild reducing agent (e.g., sodium hydrosulfide, NaSH) is used. In this way one avoids reducing the quinoid system.

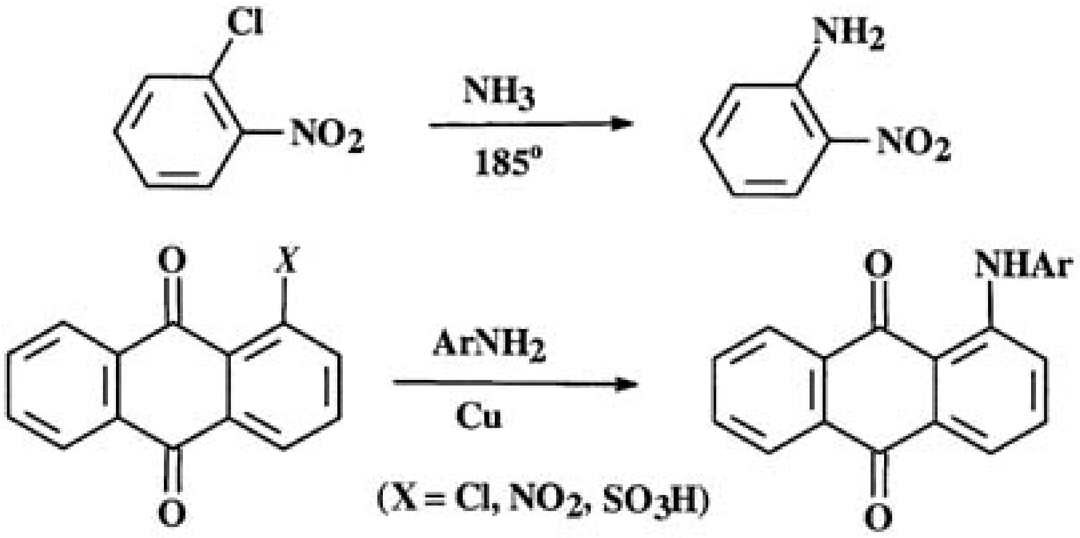
An example of an important reduction reaction involving $Fe + H^+$ is shown in Fig. 23.50. In this case the sequential use of nitration and reduction is illustrated.

It must also be pointed out that the medium employed in the reduction process can play a major role in the outcome of the reaction. A good example is the reduction of nitrobenzene in the presence of acid or alkali. While one should expect the reduction to follow the course shown in Fig. 23.48 under normal conditions, in acidic media the product obtained is mainly *para*-aminophenol. In fact, this has long been the key step in the commercial route to acetaminophen,52 which is obtained by *N*-acetylation of the reduction product. When the reduction is conducted in the presence of alkali and Zn, the nitro compound is converted to a hydrazo compound via azoxy and azo intermediates. The hydrazo compound is important because it can be treated with acid to form diaminobiphenyls known as benzidines. These reactions are shown in Fig. 23.51. Since benzidine (4,4′-diaminobiphenyl) itself is known to be a human carcinogen, its use as a dye intermediate is substantially curtailed in the western world.

Amination

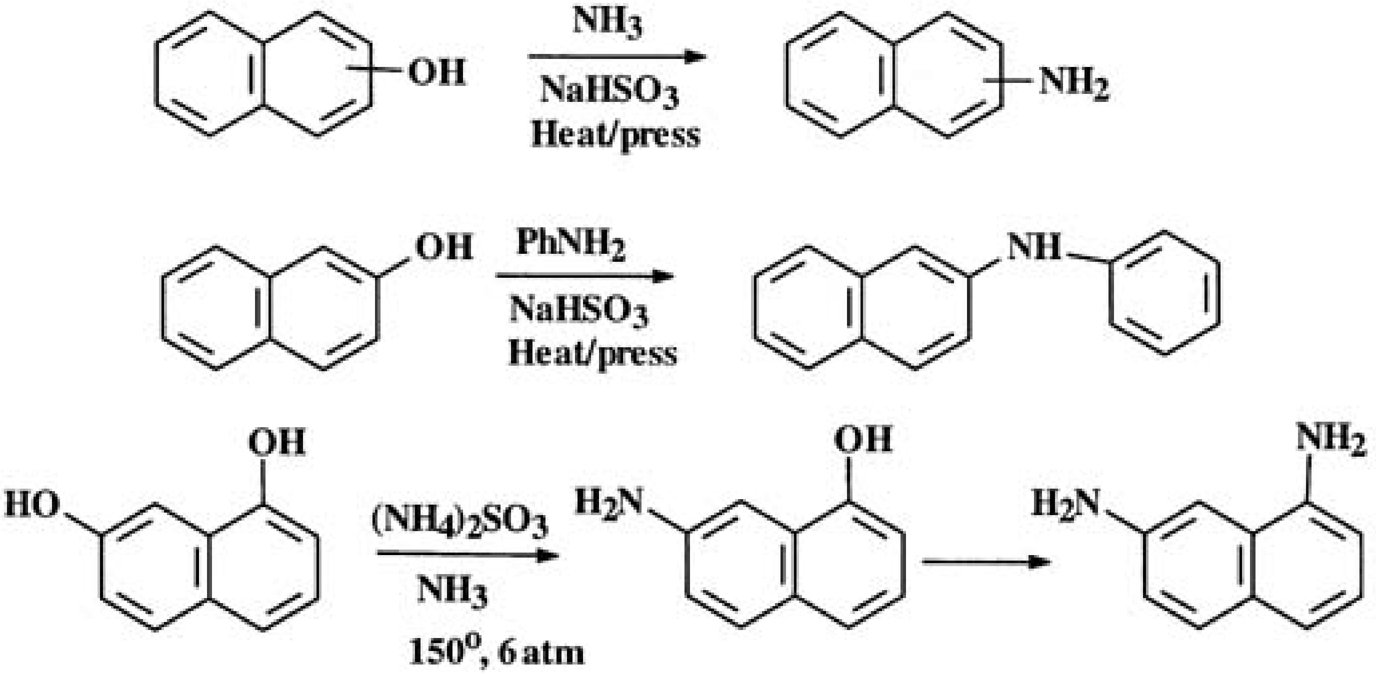
Since the previous section covers the reduction of nitro and azo compounds as a method for introducing amino groups, the focus of this section will be direct aminations involving replacement reactions and examples of indirect amination. In the former case, amination via the replacement of activated halogens using an alkyl or arylamine is widely used. The examples given in Fig. 23.52 show that halogens positioned *ortho* to a nitro group or in an α-position on the anthraquinone ring can be replaced by amino groups. The former reaction also works well when the groups are *para*. However, the reaction is difficult and usually impractical when electron-donating rather than electron-attracting groups are situated *ortho* and/or *para* to the halogen. In the case of the anthraquinone system, α-sulfonic acid and α-nitro groups can also be replaced.

Fig. 23.52 Amination reactions involving benzene and anthraquinone compounds.



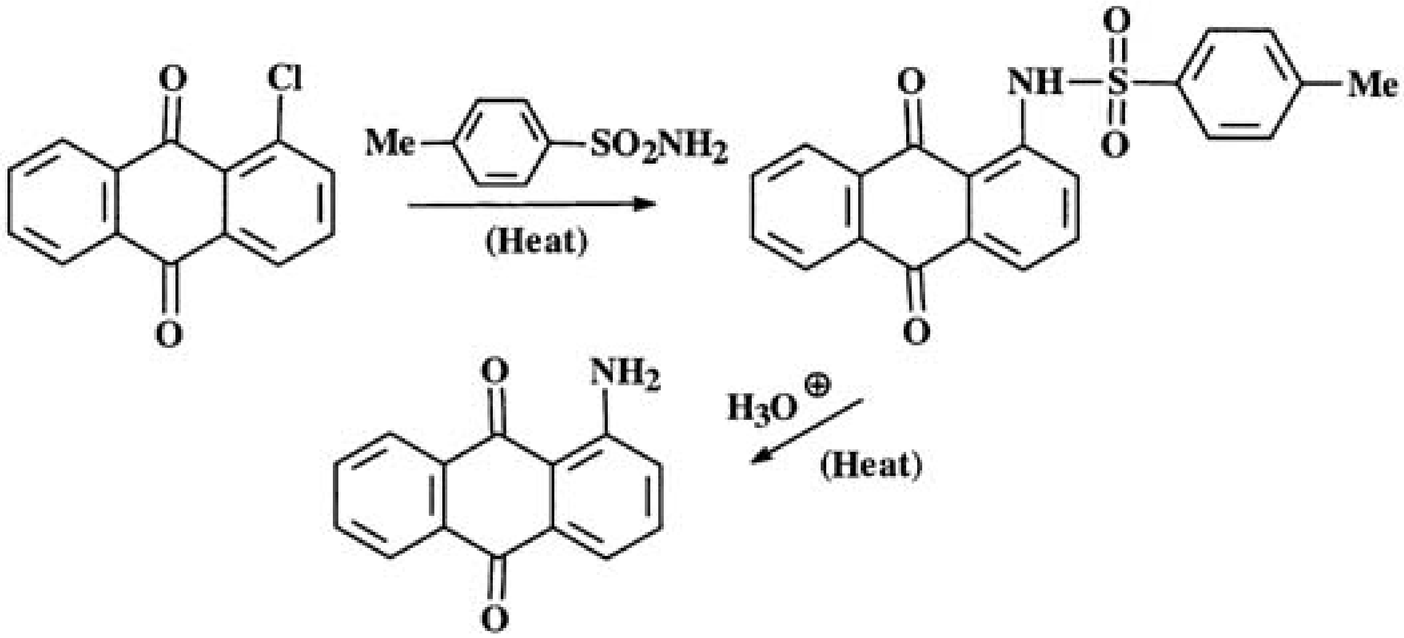
An important amination reaction involves hydroxy-substituted naphthalenes (Fig. 23.53). In a process known as the Bucherer reaction, naphthols are heated under pressure with a mixture of ammonia and sodium bisulfite. As the second and third examples indicate, the reaction works with aromatic amines and is selective. Note that the β-hydroxy group reacts preferentially when an α-hydroxy group is also present, and that two hydroxy groups in the same compound can be replaced.

Fig. 23.53 Amination of naphthalene compounds via the Bucherer reaction.



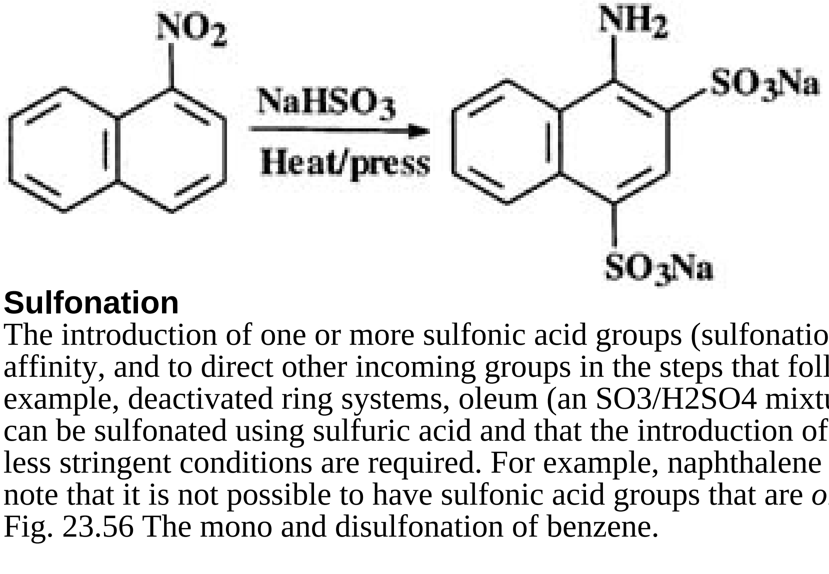
An alternative route to the synthesis of aminoanthraquinones is the 2-step sequence shown in Fig. 23.54. In this case, amination occurs via the condensation of *para*-toluenesulfonamide with chloroanthraquinone followed by hydrolysis of the sulfonamide bond. This method provides a way to introduce an NH_2 group without the use of ammonia gas and the associated high temperatures and pressures.

Fig. 23.54 An indirect amination of the anthraquinone ring system.



Another interesting reaction is shown in Fig. 23.55. In this example, amination and sulfonation occur when α-nitronaphthalene is reduced by heating it under pressure with $NaHSO_3$.

Fig. 23.55 A one-step amination and sulfonation of naphthalene.



Sulfonation

The introduction of one or more sulfonic acid groups (sulfonation) into dye intermediates is often conducted to confer water solubility, to provide fiber affinity, and to direct other incoming groups in the steps that follow sulfonation. In most cases this process employs sulfuric acid but in difficult cases, for example, deactivated ring systems, oleum (an SO_3/H_2SO_4 mixture) is used. This chemistry is illustrated in Fig. 23.56 for benzene. Here we see that benzene can be sulfonated using sulfuric acid and that the introduction of a second sulfonic acid group requires oleum. When a more reactive system is sulfonated, less stringent conditions are required. For example, naphthalene (Fig. 23.57) is readily sulfonated up to four times without using oleum. It is important to note that it is not possible to have sulfonic acid groups that are *ortho, para*, or *peri* to each other in the naphthalene system.

Fig. 23.56 The mono and disulfonation of benzene.

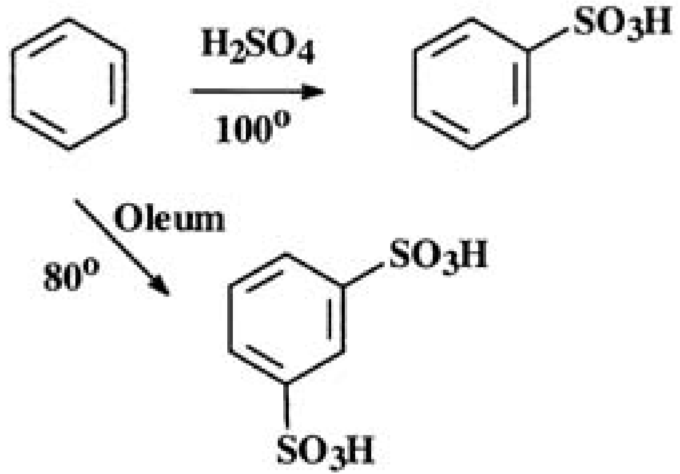
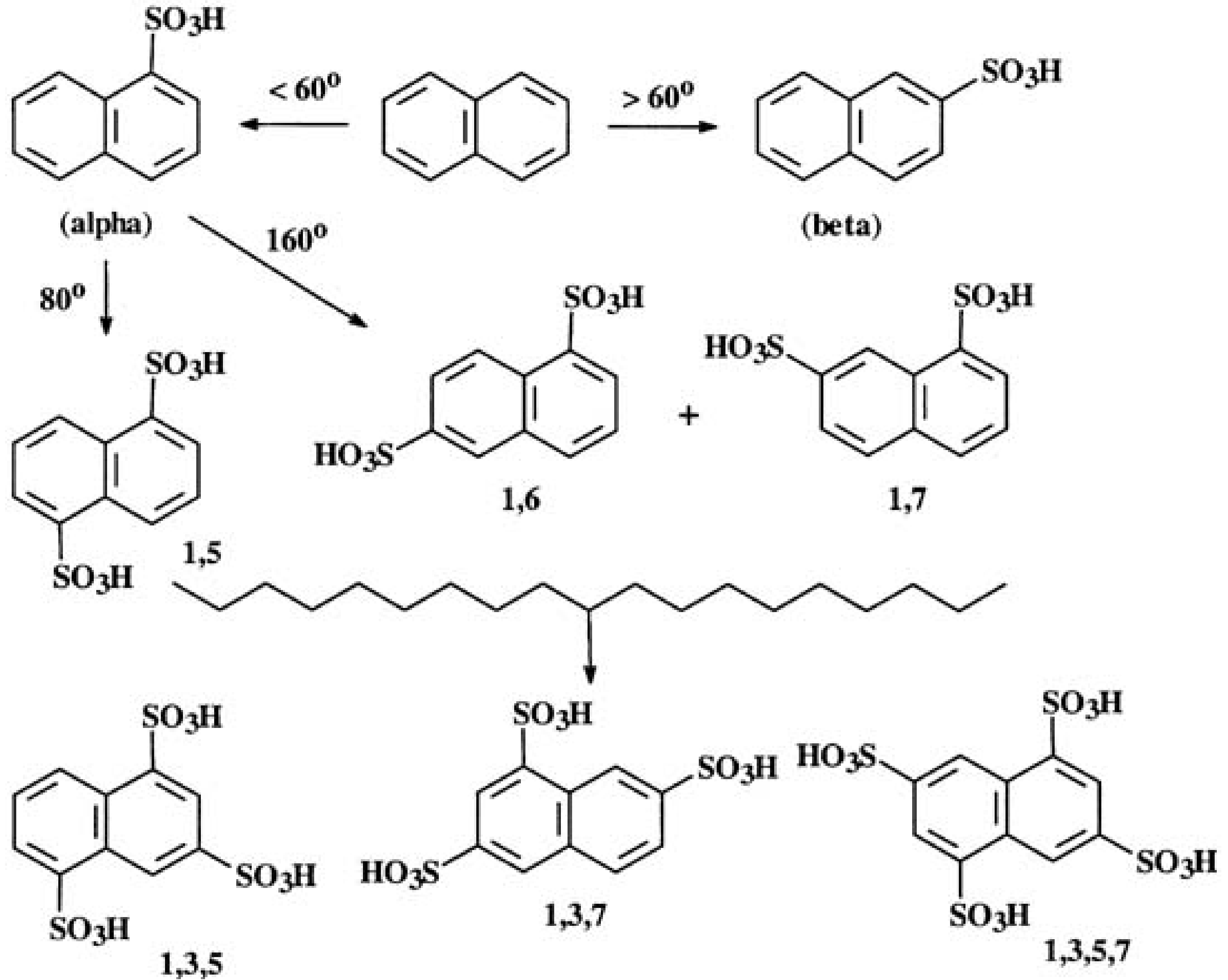
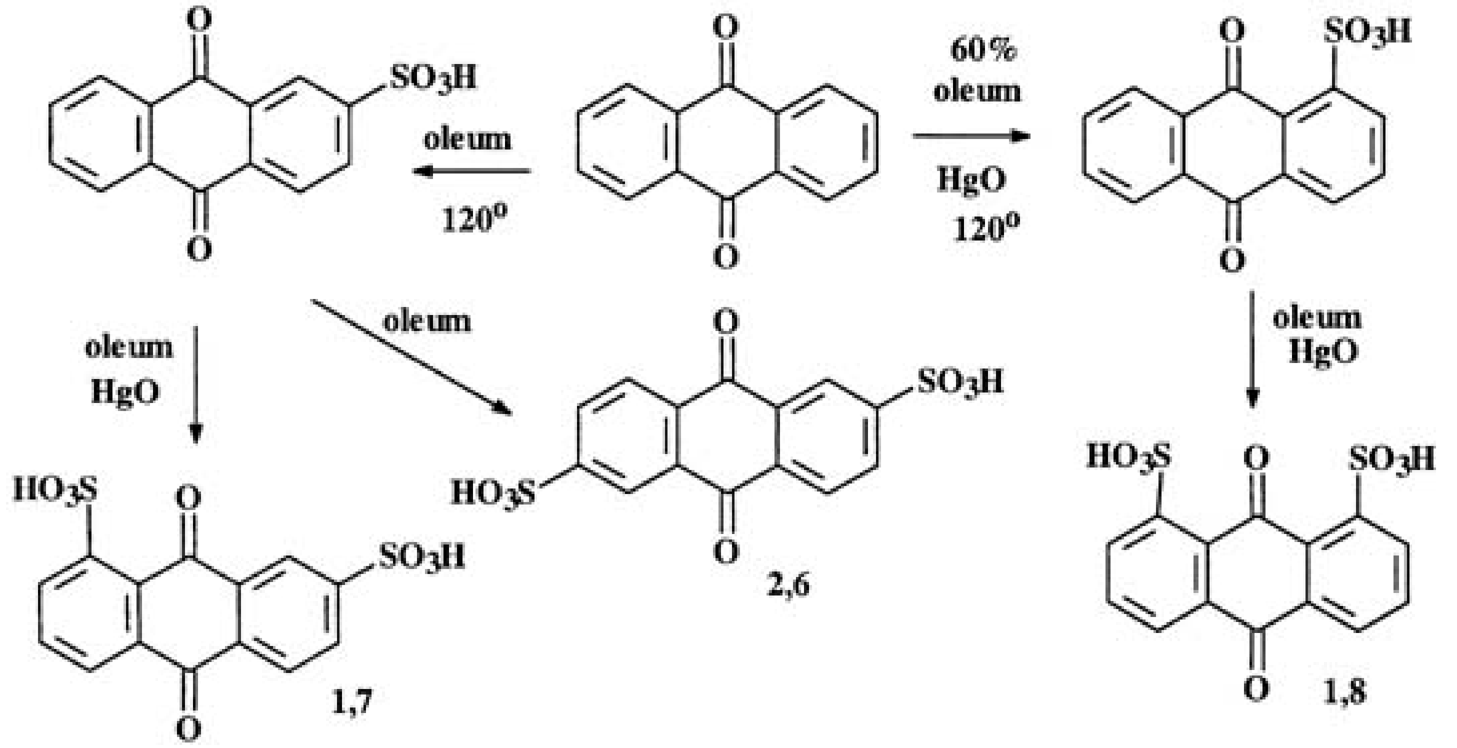


Fig. 23.57 Examples of products obtained from the sulfonation of naphthalene.



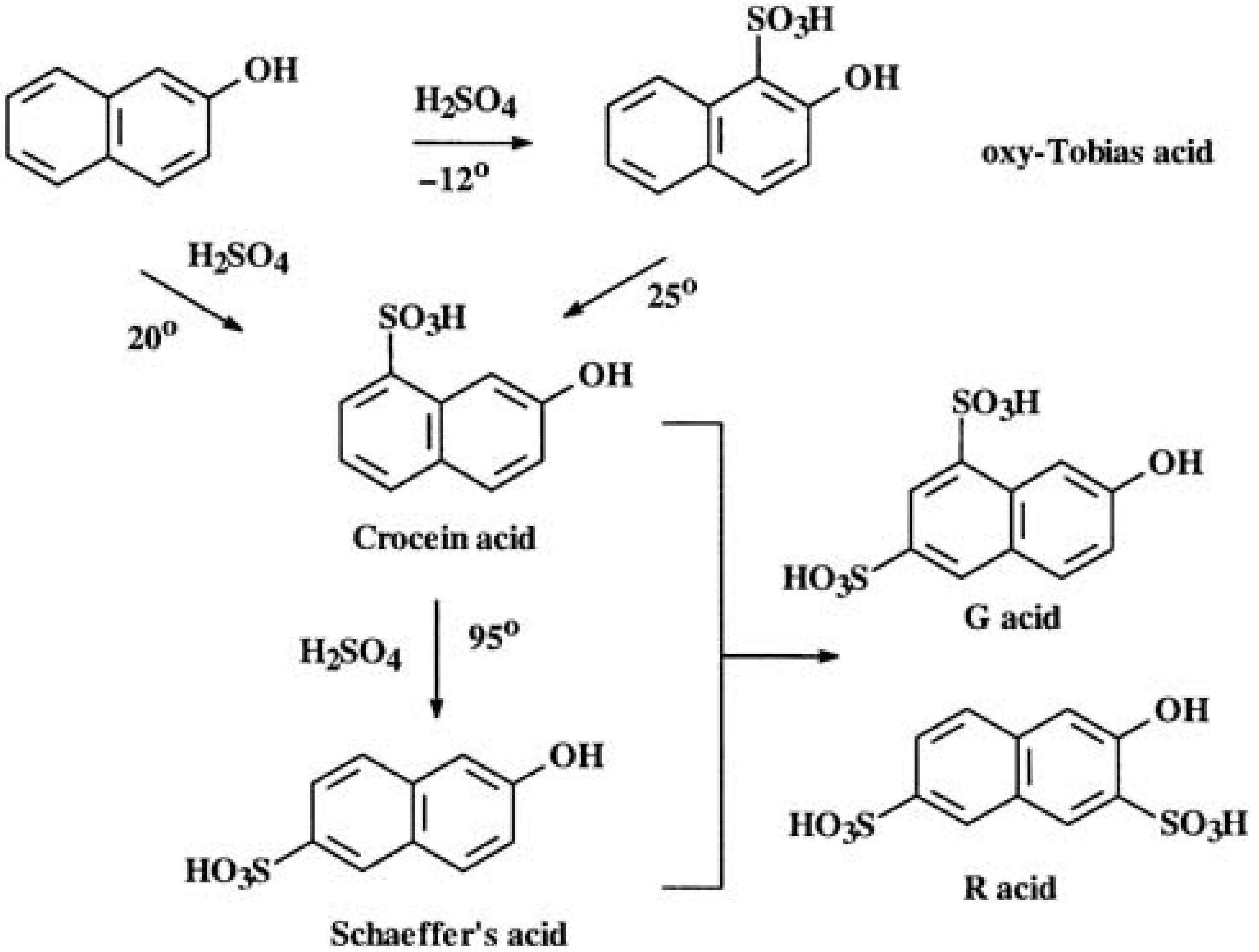
By contrast, the sulfonation of anthraquinone requires oleum and no more than two sulfonic acid groups can be introduced. In this system, sulfonation in the α-position requires the use of HgO as a catalyst. Examples of the possible products are shown in the scheme in Fig. 23.58.

Fig. 23.58 Typical products produced from the sulfonation of anthraquinone.



The sulfonation of β-naphthol produces several important dye intermediates, the nature of which depends upon the conditions employed (Fig. 23.59). At low temperatures, sulfonation occurs in the α-position to give oxy-Tobias acid. Under ambient conditions Crocein acid is produced and at elevated temperatures three other products are obtained, including two that are disulfonated.

Fig. 23.59 Dye intermediates prepared from the sulfonation of β-naphthol.



The sulfonation of aromatic amines such as aniline can give a mixture of products that must be separated prior to dye synthesis. When a single product is sought, the "Baking" reaction is often employed (Fig. 23.60). In this process, the sulfate salt of aniline is prepared, dried, and then "baked" in an oven under vacuum. The product in this case is the important dye intermediatesulfanilic acid. Similarly, naphthylamine sulfonic acid can be produced, and if the *para*-position is occupied, sulfonation of an *ortho*-position occurs (Fig. 23.61).

Fig. 23.60 Direct sulfonation of (a) aniline versus (b) the baking reaction.

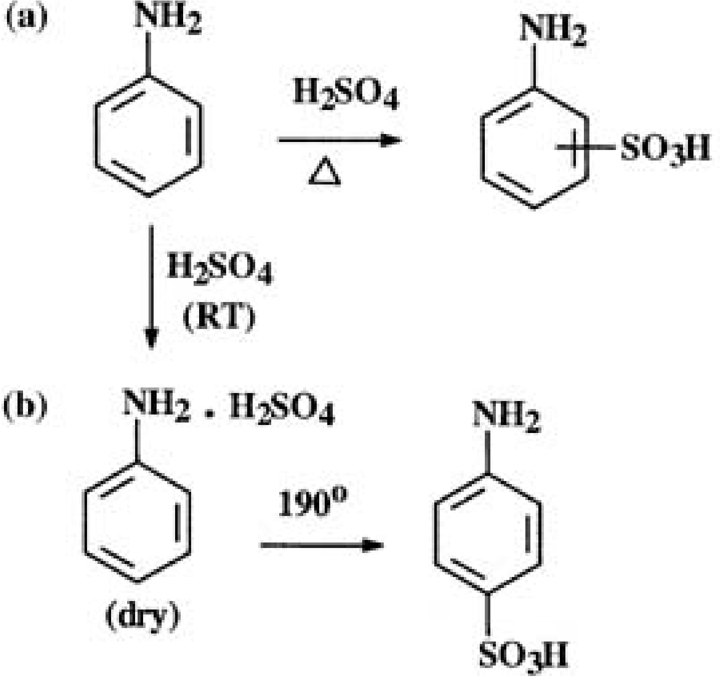
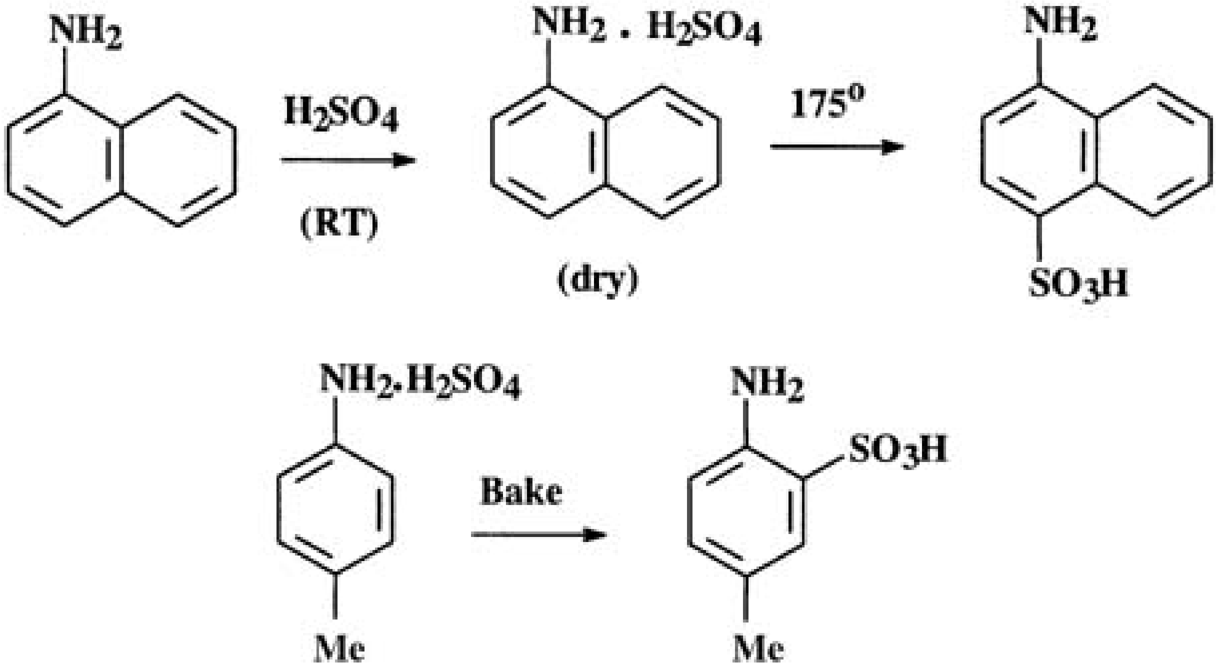


Fig. 23.61 Additional examples of the baking reaction.



Halogenation

For dye intermediates, halogenation most often involves the incorporation of chloro groups. As pointed out earlier, halogens are important as leaving groups in the amination process, but they can also be used to enhance brightness and influence color. Later, we will see that halogens are important as leaving groups in reactive dye chemistry, and in this regard chloro and fluoro groups are used.

Figs 23.6223.65 provide examples of chlorination reactions. In the first example, the commonly used agent FeCl3/Cl2 is employed for the chlorination of benzene and naphthalene rings. This method is not practical for the chlorination of anthraquinone. In this case the most important reaction is the tetrachlorination process shown in Fig. 23.63.

Fig. 23.62 Halogenation of benzene and naphthalene rings.

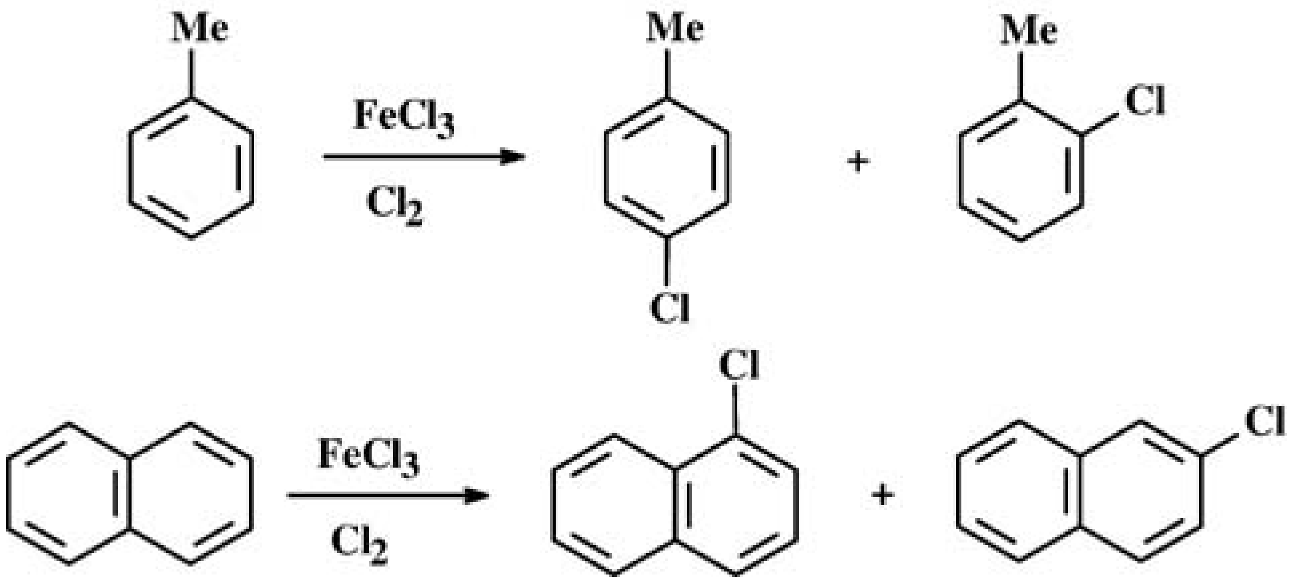


Fig. 23.63 Synthesis of the dye intermediate 1,4,5,8-tetrachloroanthraquinone.

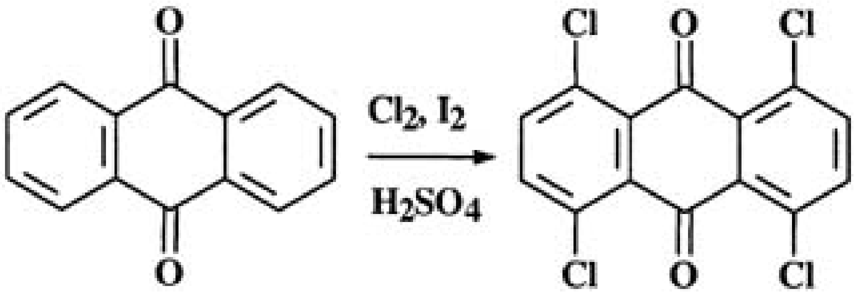


Fig. 23.64 Chlorination of phenol and anilines.

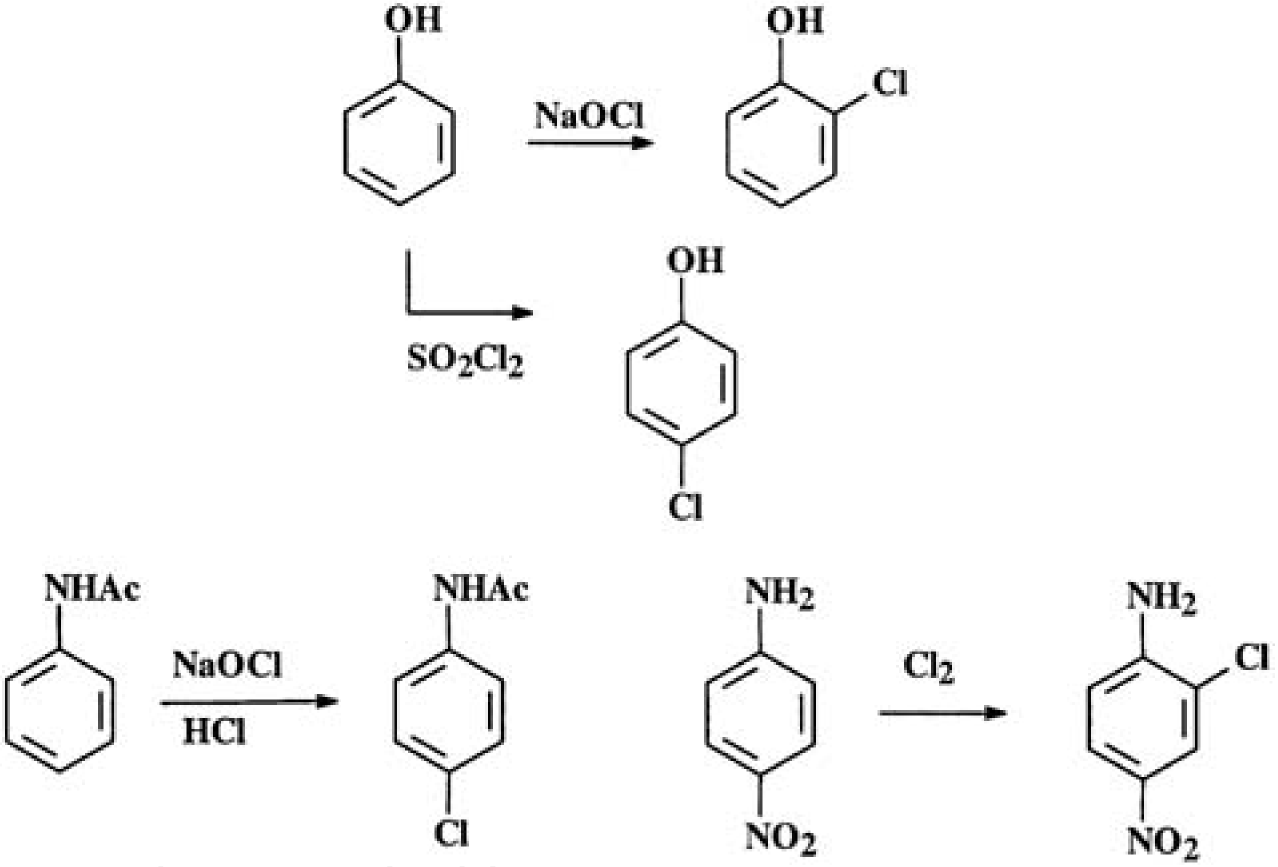
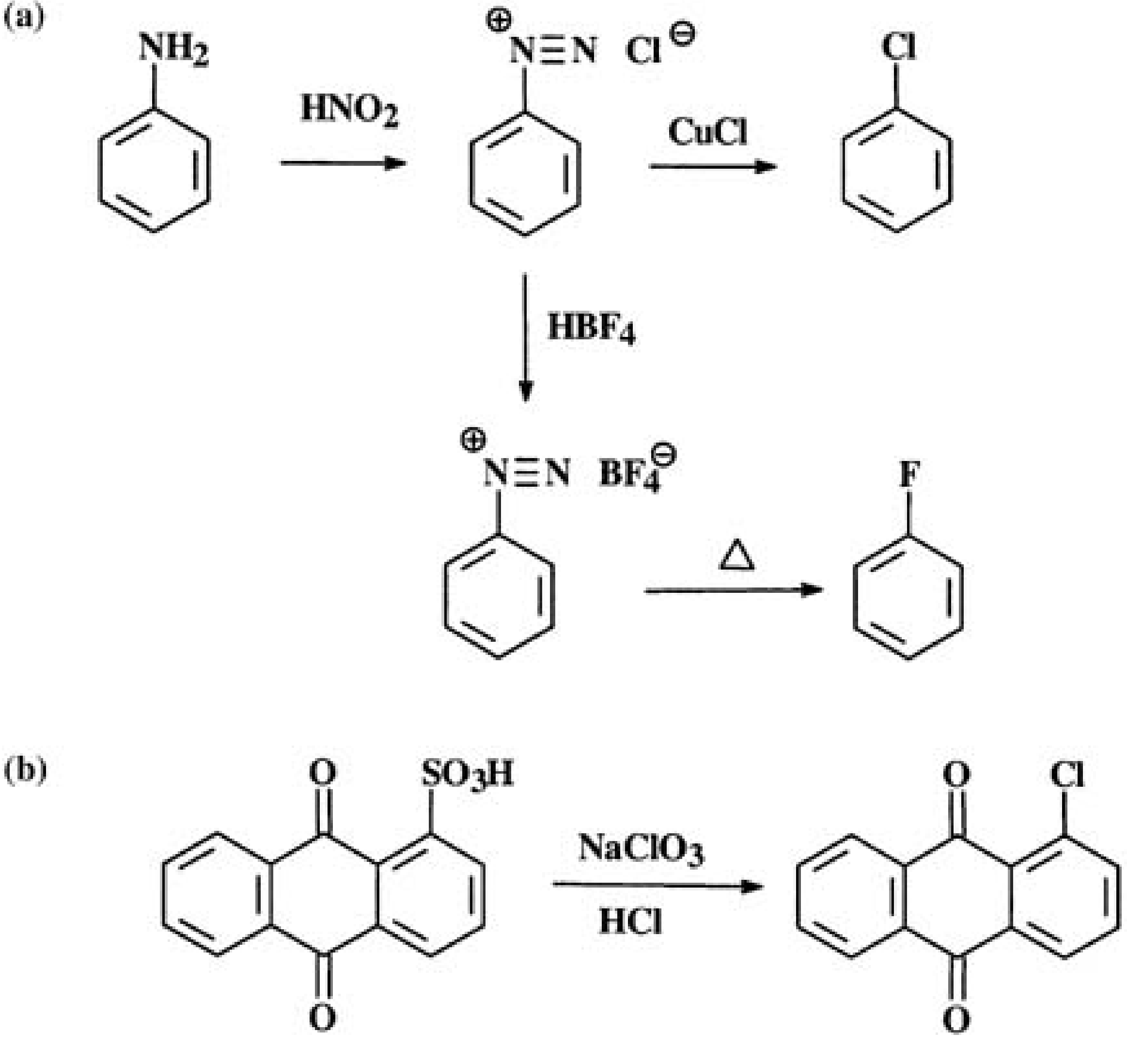


Fig. 23.65 Three important indirect halogenation reactions.

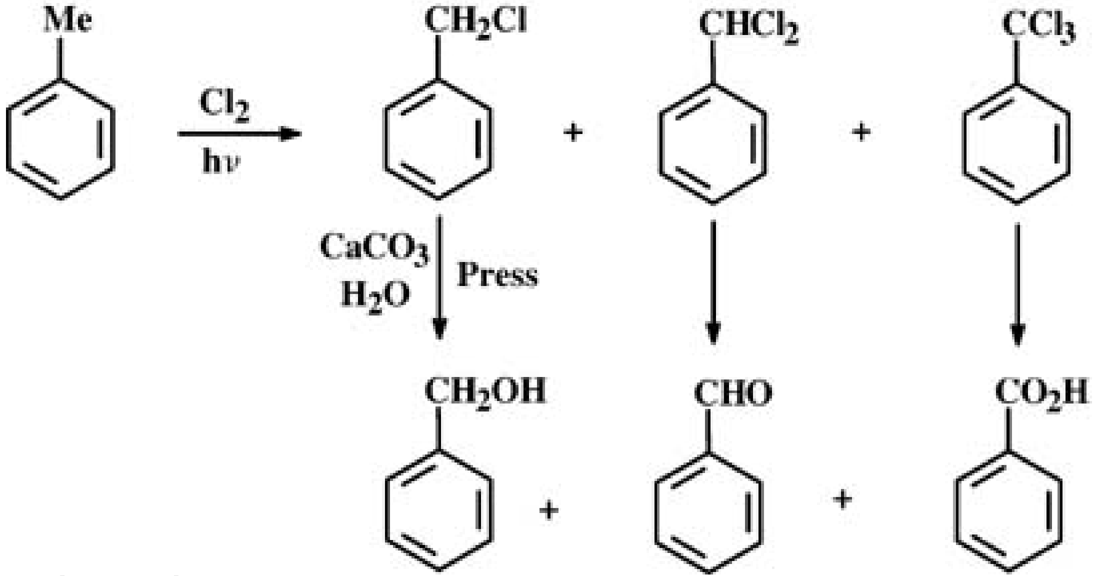


Since the chlorination of phenols and aromatic amines can be difficult to control, chlorination of these systems usually employs agents that will give a single chloro group when this is the desired outcome. In this regard, NaOCl and SO2Cl2 are quite useful chlorinating agents (see Fig. 23.64). In cases involving amines, often such compounds are protected by acetylation prior to chlorination. If the reactivity of the ring has been reduced by the presence of a deactivating group (e.g., NO2), acetylation may not be needed.

Halogens are also introduced via indirect methods, three examples of which are shown in Fig. 23.65. In the first case (sequence (a)), aniline is diazotized and the resultant diazonium compound is heated with cuprous chloride to give chlorobenzene, in a process known as the Sandmeyer reaction.53 Alternatively, the diazonium compound can be converted to the tetrafluoroborate salt, which in turn is heated to give fluorobenzene. In sequence (b), anthraquinone-2-sulfonic acid is converted to the corresponding chloro by treatment with NaClO3/HCl. All three reactions can be used to prepare a wide array of halogenated aromatics.

Chlorination is also important step in the synthesis of oxygenated aromatic compounds. In this case, chlorination takes place at alkyl groups attached to the rings and is conducted in the absence of iron. The use of UV light speeds up this reaction, which is illustrated for toluene in Fig. 23.66. This free radical chlorination of toluene gives a mixture of benzyl chloride, benzal chloride, and benzotrichloride, which in turn can be hydrolyzed to benzyl alcohol, benzaldehyde, and benzoic acid.

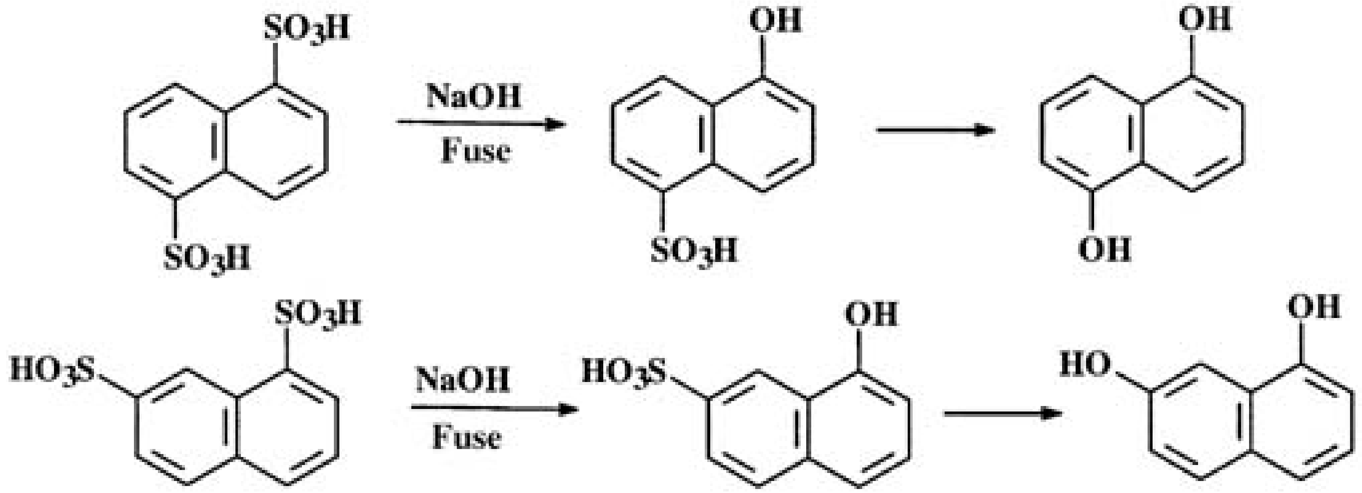
Fig. 23.66 Free radical chlorination of toluene and hydrolysis of the products.



Hydroxylation

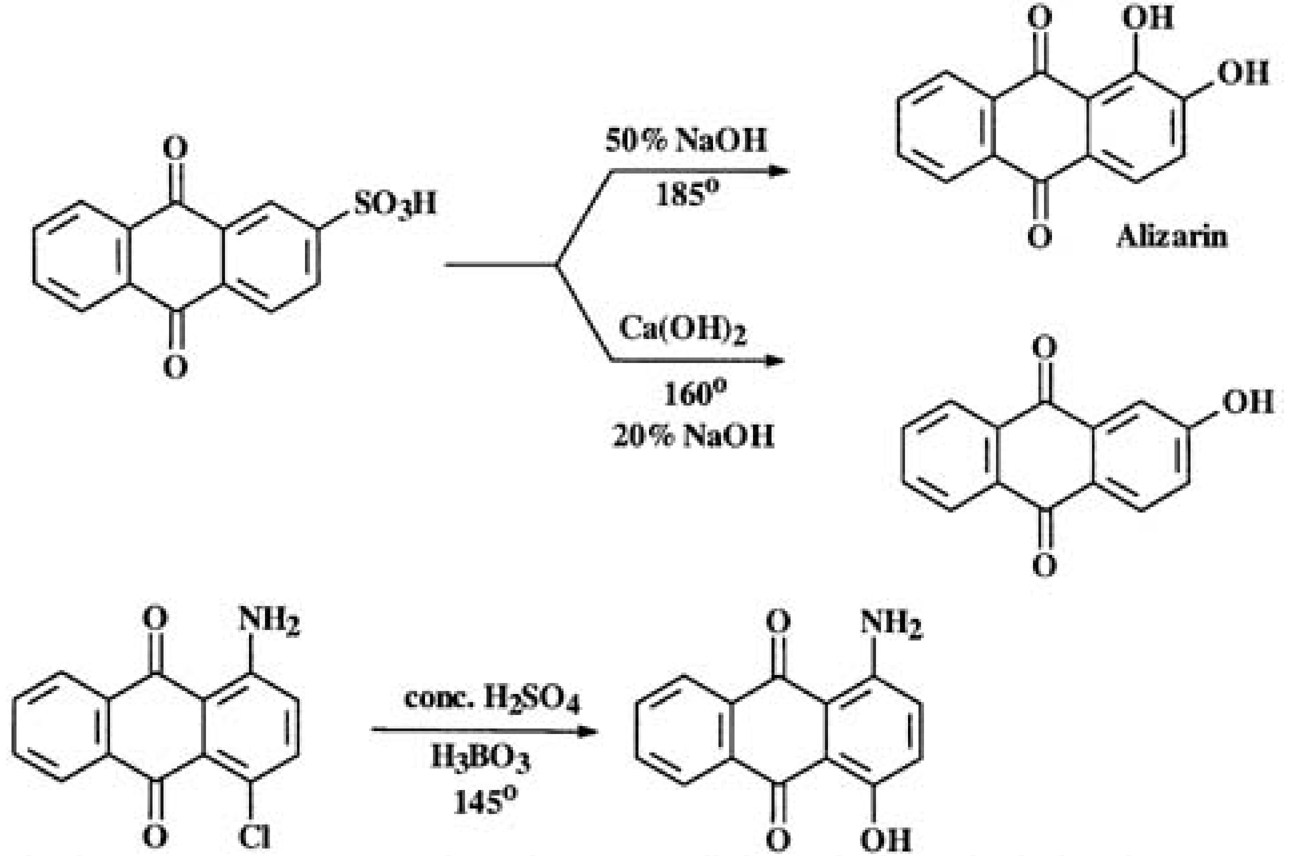
The introduction of hydroxy groups is important in dye chemistry because it opens the door to azo dye formation, using phenols and naphthols, and provides an important auxochrome. Hydroxylation methods include alkali fusion, replacement of labile groups, and the reverse Bucherer reaction. In the alkali fusion reaction, naphthalene sulfonic acids are reacted with molten NaOH, KOH, or combinations of the two, as illustrated in Fig. 23.67. When disulfonated naphthalenes are used, the reaction can be stopped at the monohydroxylation stage if this is the desired outcome. The second example shows that the α-sulfonic acid group reacts faster.

Fig. 23.67 Sodium hydroxide fusion as a hydroxylation process.



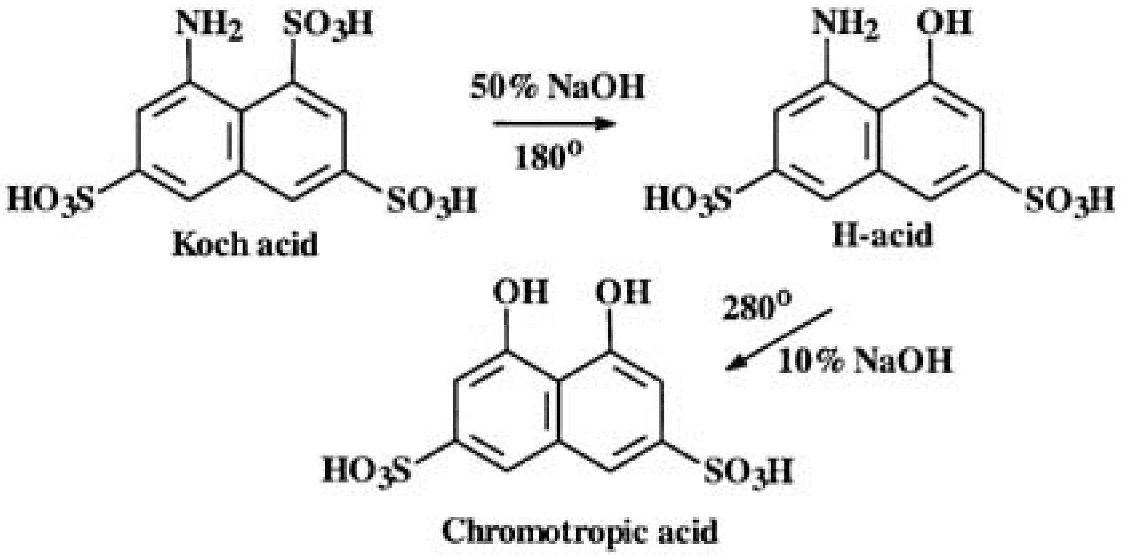
When sulfonated anthraquinones are used, hydroxylation is conducted with Ca(OH)2 to avoid over oxidation that occurs when hot NaOH is used. Example reactions are shown in Fig. 23.68.

Fig. 23.68 Hydroxylation of the anthraquinone system.



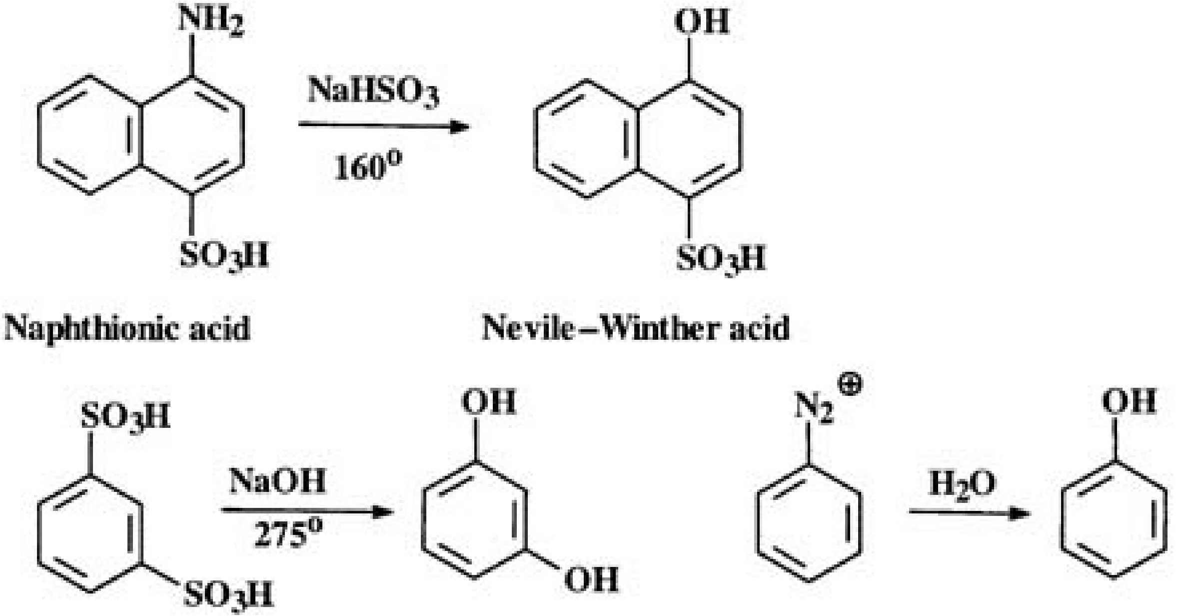
The short sequence in Fig. 23.69 shows that aqueous alkali can also be used in hydroxylation reactions. In both cases, however, elevated temperatures are required.

Fig. 23.69 Synthesis of two key dye intermediates using hydroxylation steps.



Other important hydroxylation reactions are shown in Fig. 23.70. Here it can be seen that the Bucherer reaction is reversible, that the fusion reaction works for sulfonated benzene compounds, and that diazonium compounds undergo hydrolysis to produce phenols/naphthols.

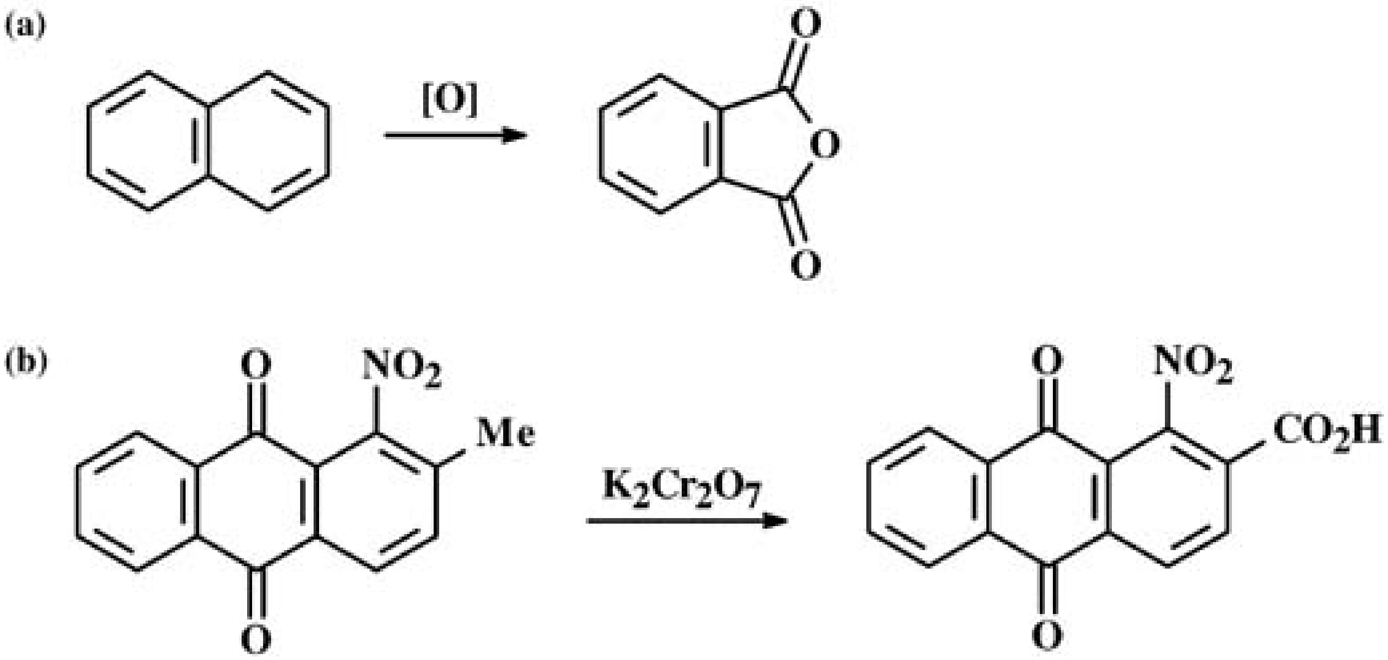
Fig. 23.70 Other examples of hydroxylation reactions.



Oxidation

Although the oxidation of aromatic methyl groups can be conducted via the 2-step sequence shown in Fig. 23.66, a convenient alternative involves potassium dichromate. In this case, the ring system employed must be stable to the conditions of the reactions. Another important oxidation involves the conversion of naphthalene to phthalic anhydride, which can be accomplished using hot $KMnO_4$ or Va_2O_5. These two reactions are illustrated in Fig. 23.71. Later we will see that the oxygen in air can be used as the oxidant for certain organic dyes.

Fig. 23.71 Oxidation of (a) naphthalene and (b) a methylated anthraquinone.

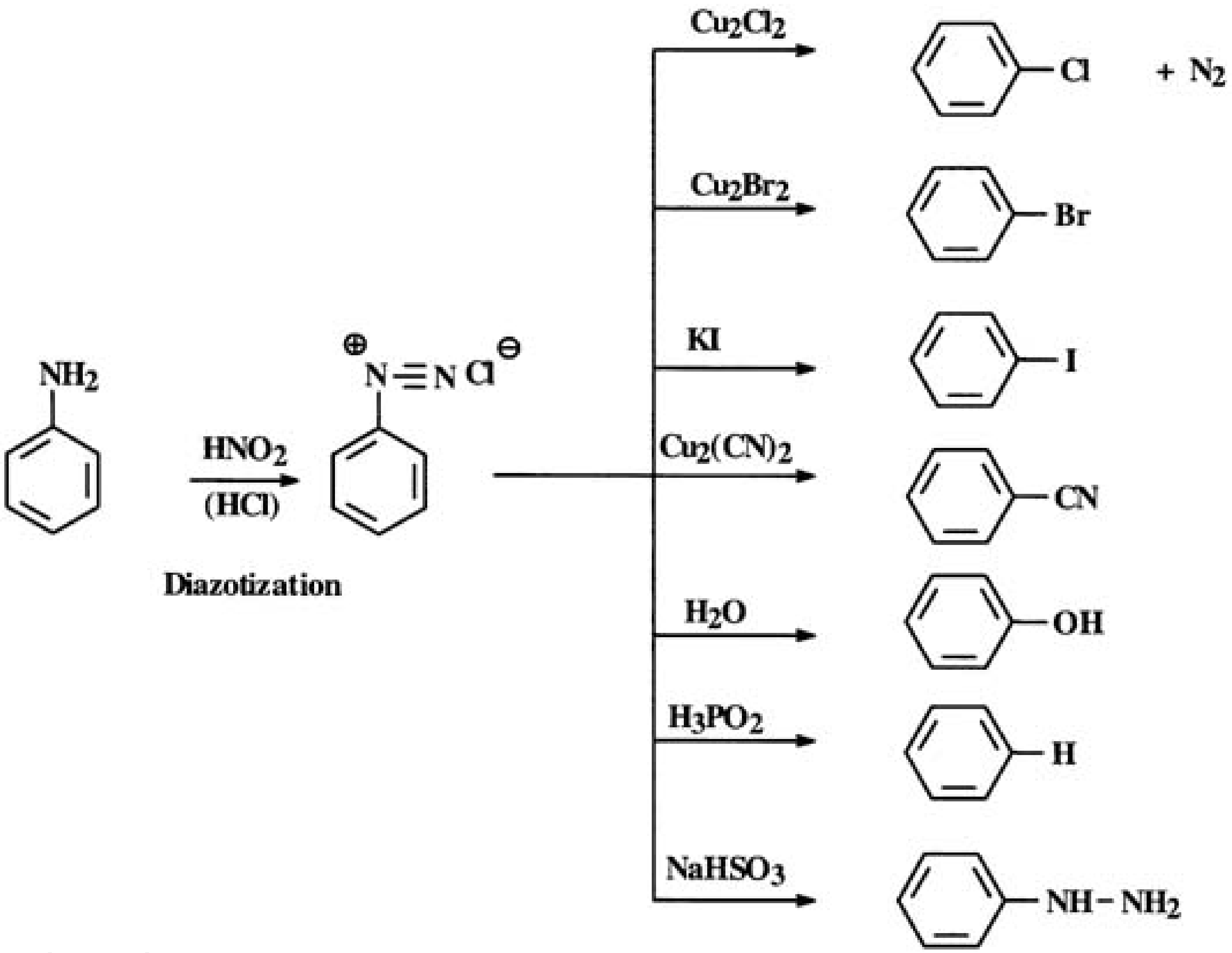


Other Important Reactions

Diazotization

The conversion of a primary aromatic amine to a diazonium compound is known as diazotization. While this process will be covered in more detail in our discussion of azo dye synthesis, it is worthwhile to point out that the diazonium group (N_2^+) is used to produce a wide range of intermediates. As indicated in Fig. 23.72 diazotization is often achieved through the action of nitrous acid (HNO_2) and the resultant diazonium group can be replaced by various groups or reduced to give arylhydrazines.

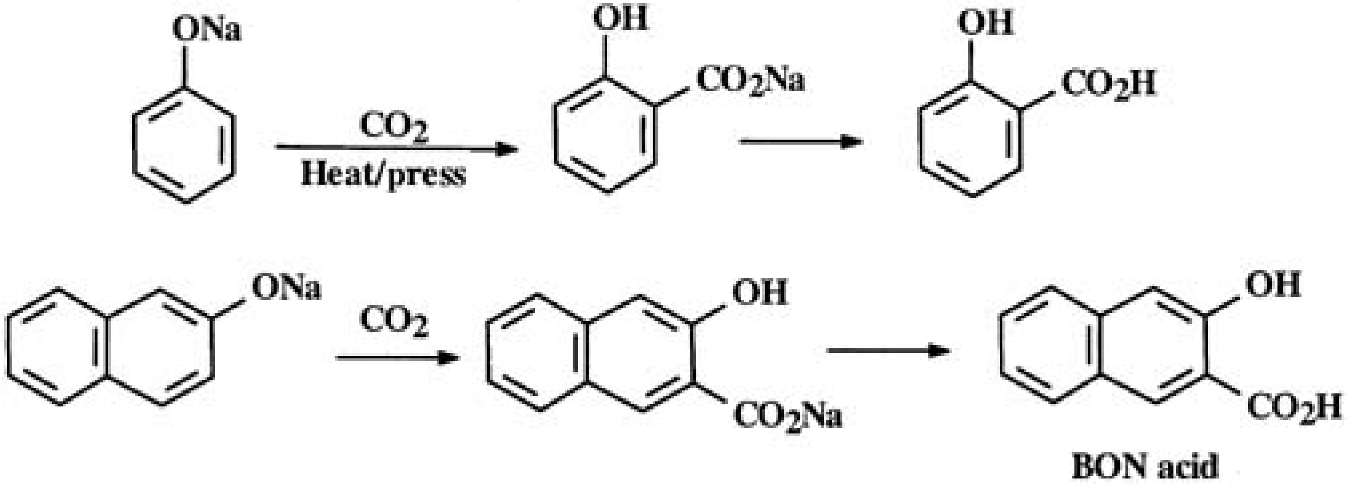
Fig. 23.72 Diazotization of aniline and its conversion to other types of intermediates.



Carboxylation

The introduction of carboxyl groups into the structures of phenols and naphthols produces some important dye intermediates, including salicylic acid and BON acid. This process is conducted under pressure at elevated temperatures using the sodium salts of phenols/naphthols and in the case of β-naphthol, the carboxyl group enters the 3-position (cf. Fig. 23.73). The free acid (CO2H) group is produced by acid treatment in the final step.

Fig. 23.73 Carboxylation of phenol and β-naphthol.



23.9 Manufacture of Dyes

In this section, we summarize the principal methods of synthesis for different classes of dyes. Emphasis is placed on dyes presently in commerce and the industrial methods suitable for making them. Before doing so, we will review the important principles that set dyes apart from other classes of organic compounds.

Unlike other organic compounds dyes possess color because they (l) absorb light in the visible spectrum (400700 nm), (2) have at least one chromophore (color bearing group), (3) have a conjugated system (system of alternating double and single bonds), and (4) exhibit resonance (a stabilizing force in organic compounds). Table 23.5 shows the relationships between wavelength of visible and color absorbed/seen and the other three factors are illustrated in Figs 23.7423.76.

Fig. 23.74 Chromophores commonly found in organic dyes.

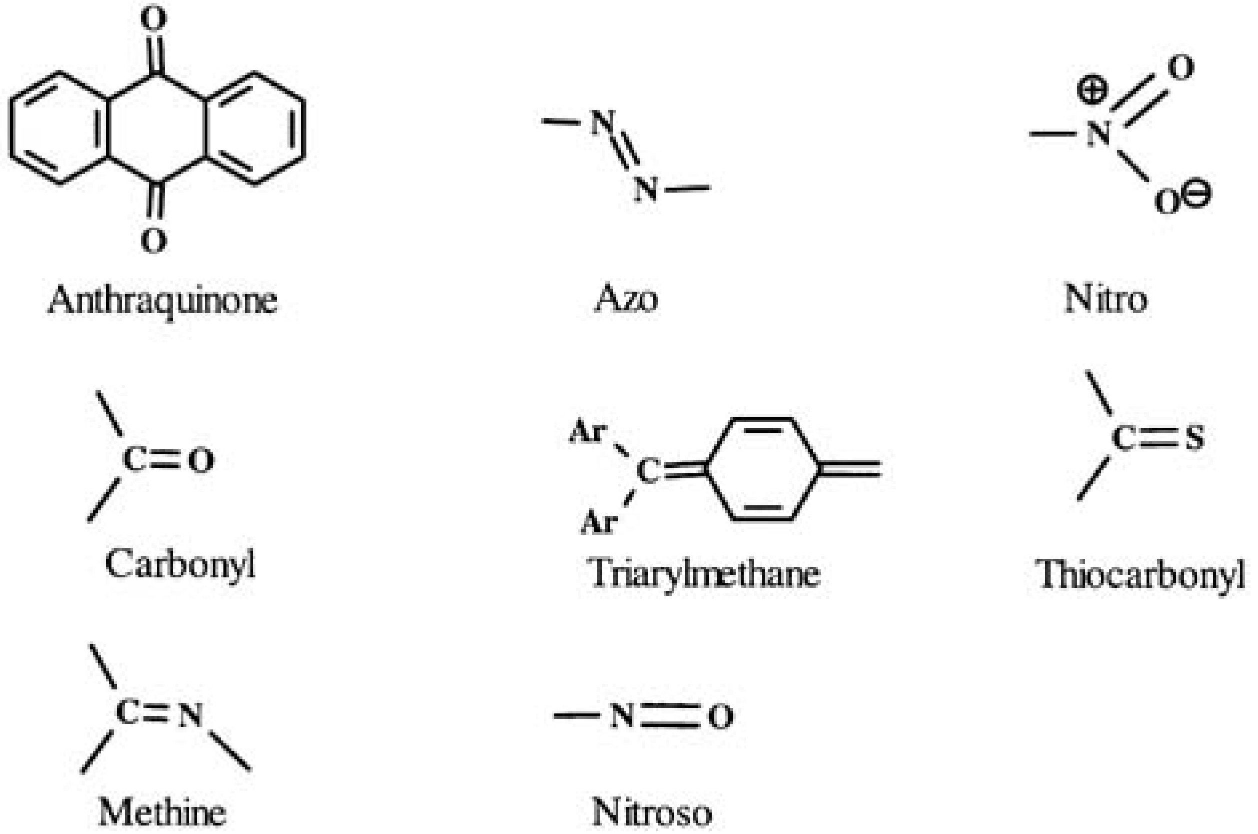


Fig. 23.75 Comparison of the conjugated systems in (a) vitamin A and (b) β-carotene.

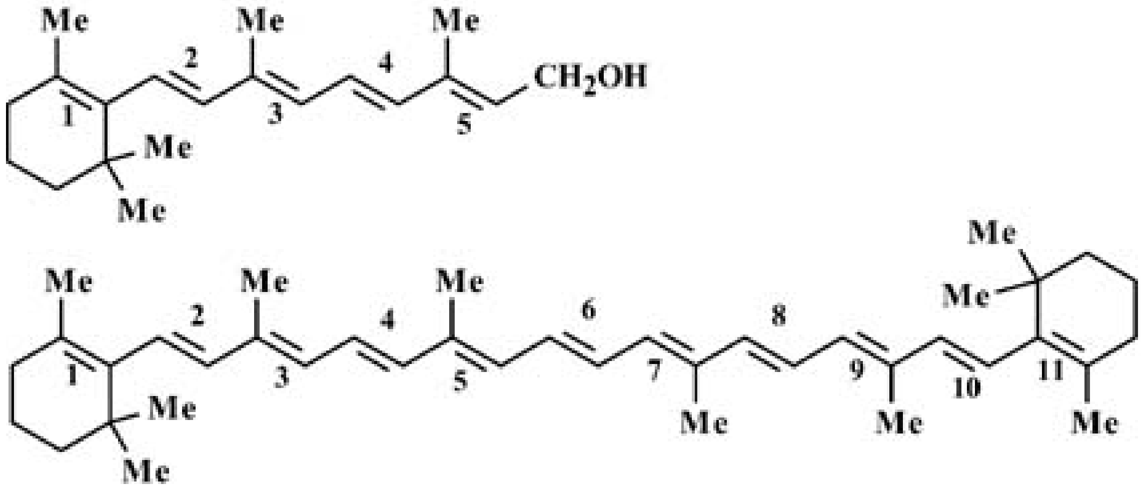


Fig. 23.76 Resonance structures for Malachite Green (C.I. Basic Green 4).

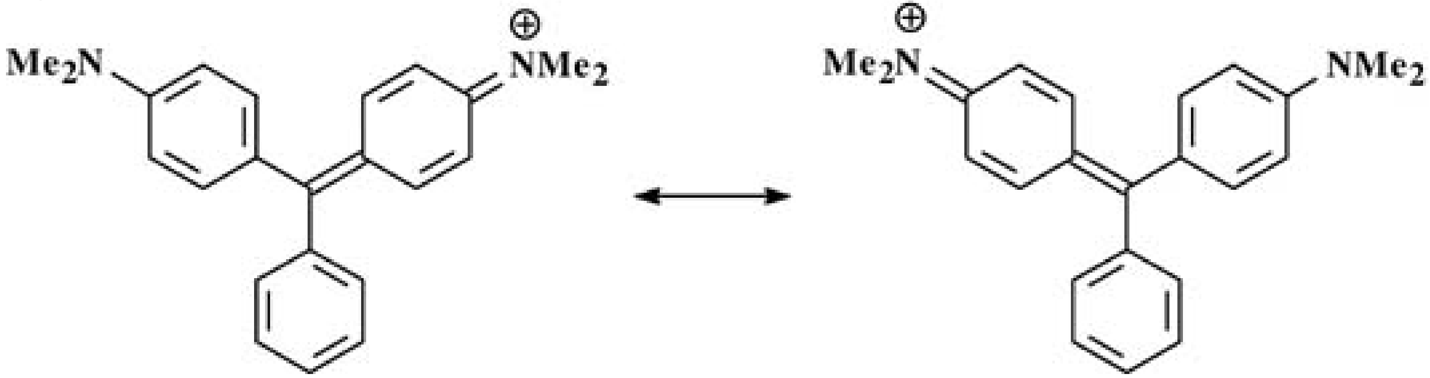
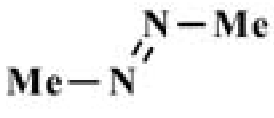


TABLE 23.5 Wavelength of Light versus Color

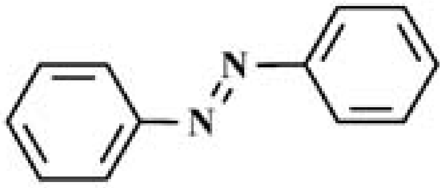
Wavelength Absorbed	Color Absorbed	Color Seen
400435	Violet	Yellow-green
435480	Blue	Yellow
480490	Green-blue	Orange
490500	Blue-green	Red
500560	Green	Purple
560580	Yellow-green	Violet
580595	Yellow	Blue
595605	Orange	Green-blue
605700	Red	Blue-green

Concerning the various factors responsible for color in organic compounds, it is worthwhile to point out that the chromophore must be part of a conjugated system. This is illustrated through the examples in Fig. 23.77. When the azo group is connected to methyl groups the resultant compound is colorless. When it is attached to aromatic rings, the compound possesses color. Similarly, the structures in Fig. 23.75 illustrate the importance of having an extended conjugated system. In this case, doubling the length of the conjugated system for vitamin A to give β-carotene causes the λmax value to shift from 325 to 466 and 497 nm.

Fig. 23.77 Impact of having a chromophore (a) apart from or (b) part of a conjugated system.



Colorless

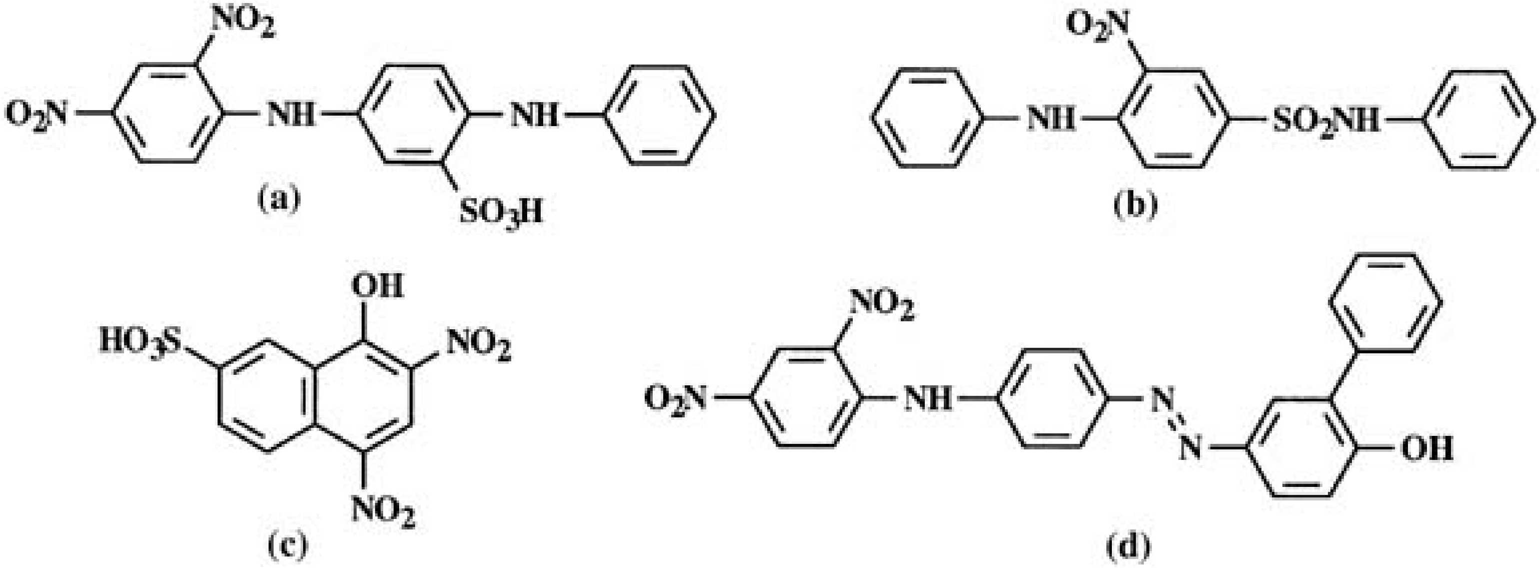


Orange

Nitro Dyes

As the name suggests, this very small class of organic dyes has at least one nitro group as the chromophore. Nitro dyes invariably are yellow or orange and are important for their economical cost and good lightfastness. Examples include the dyes shown in Fig. 23.78C.I. Acid Orange 3 (a), C.I. Disperse Yellow 42 (b), C.I. Acid Yellow 1 (c), and C.I. Disperse Yellow 70 (d). A key disadvantage of nitro dyes is their low color strength (εmax = 50007000). Improvements in color strength have been achieved by incorporating an azo group, as illustrated in dye (d).

Fig. 23.78 Examples of nitro dye structures.



Representative syntheses are shown in Figs. 23.79 and 23.80. In the first example, C.I. Disperse Yellow 42 is prepared by condensing two molecules of aniline with one molecule of 4-chloro-3-nitrobenzenesulfonyl chloride, using ethanol as the solvent. In the second example, C.I. Acid Orange 3 is prepared in a 3-step synthesis, starting from 2-chloro-5-nitrobenzenesulfonic acid.

Fig. 23.79 Synthesis of Disperse Yellow 42.

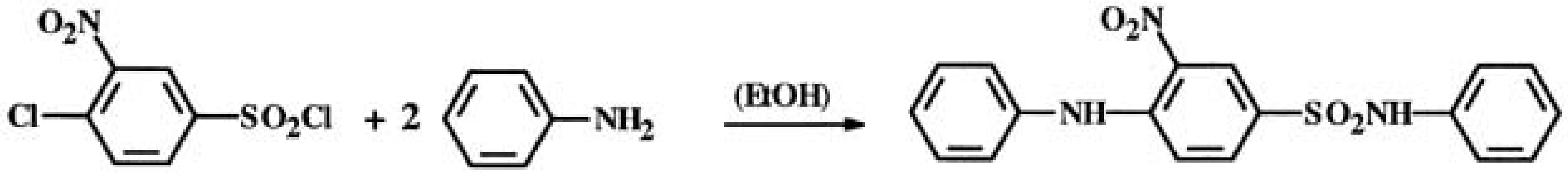
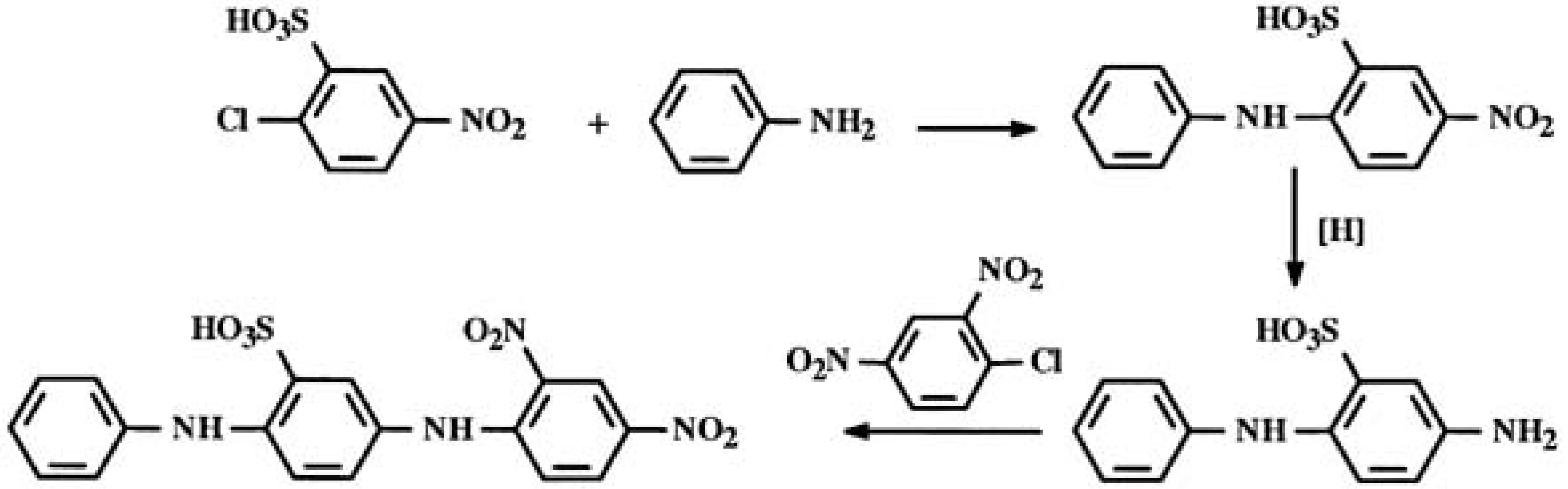


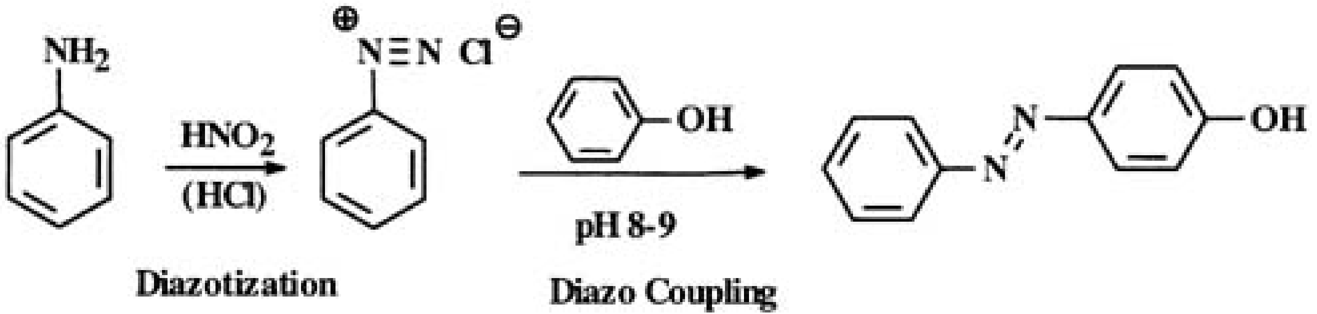
Fig. 23.80 Synthesis of Acid Orange 3.



Azo Dyes

Azo dyes are by far the largest family of organic dyes. They play a prominent role in acid, direct, reactive, azoic, and disperse dye structures, as shown in the section "Classification of Dyes", and include structures that cover the full color spectrum. Generally, the synthesis of azo dyes involves a process known as diazo coupling. In this process, a diazotized aromatic amine is coupled to a phenol, naphthol, aromatic amine, or a compound that has an active methylene group, as illustrated in the 2-step synthesis in Fig. 23.81. Step 1 is the conversion of aniline to benzenediazonium chloride, a process known as diazotization, and step 2 is the reaction of the diazo compound with phenol to produce the corresponding azo dye, a process known as diazo coupling.

Fig. 23.81 Two-step synthesis of an azo dye from aniline and phenol.



Diazotizations are normally conducted in an aqueous medium containing nitrous acid, generated in situ from HCl + NaNO2, and a primary aromatic amine. When weakly basic or heteroaromatic amines are used in the synthesis of azo dyes, H2SO4 is often used as the reaction medium, forming H(NO)SO4 (nitrosylsulfuric acid) as the diazotizing agent.54 The stoichiometry associated with this reaction is given in Fig. 23.82, and although only 2 moles of acid per mole amine are required, in practice 2.22.5 moles are used. Diazotizations are most often conducted at 010°C because the resultant diazo compounds are usually unstable at higher temperatures.

Fig. 23.82 Summary of the diazotization process.

$$\mathrm{Ar{-}NH_2 + 2\ HCl + NaNO_2 \longrightarrow Ar{-}\overset{\oplus}{N}{\equiv}N\ \ Cl^{\ominus} + 2\ H_2O + NaCl}$$

Examples of aromatic amines that can be diazotized are shown in Fig. 23.83. This extremely abbreviated list is designed to show that a wide variety of amines can be used, including hydrophobic, weakly basic, hydrophilic, and heterocyclic compounds. *ortho*-Diamines are not typically used because of their propensity to undergo triazole formation (Fig. 23.84).

Fig. 23.83 Representative aromatic amines used in azo dye synthesis.

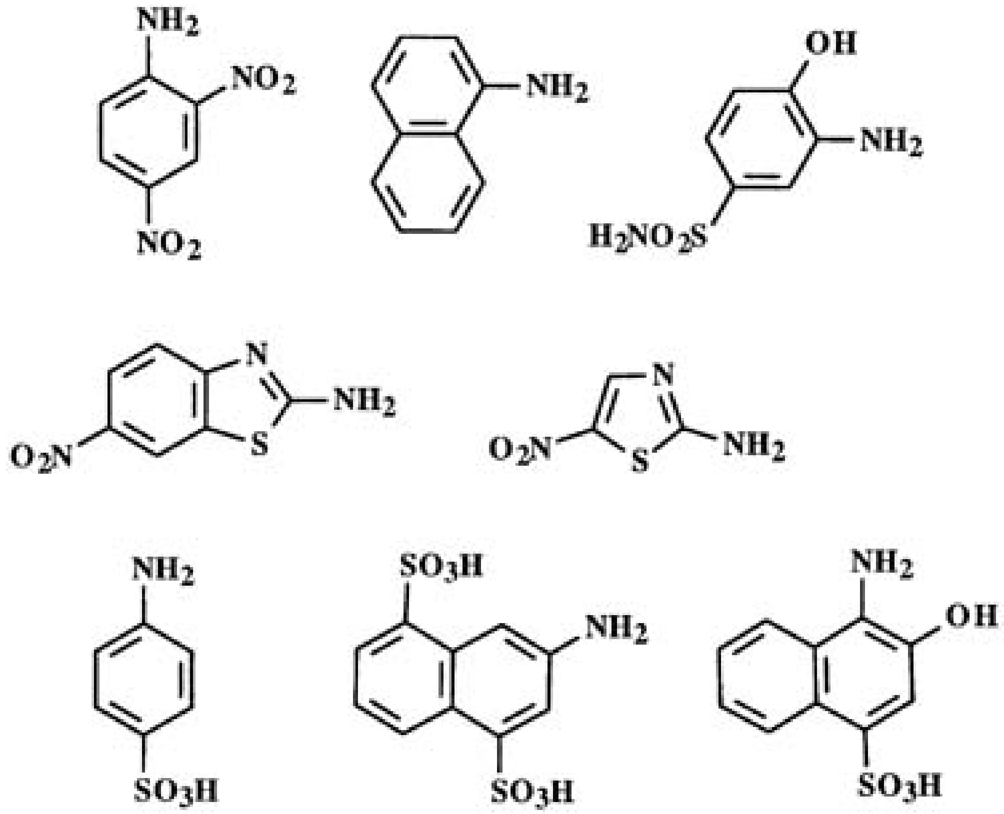
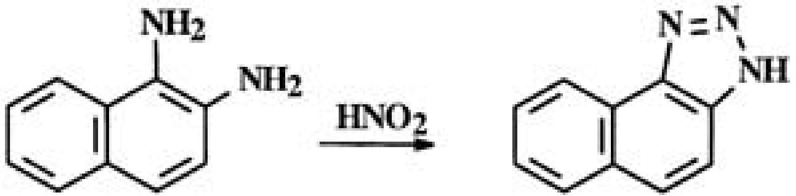
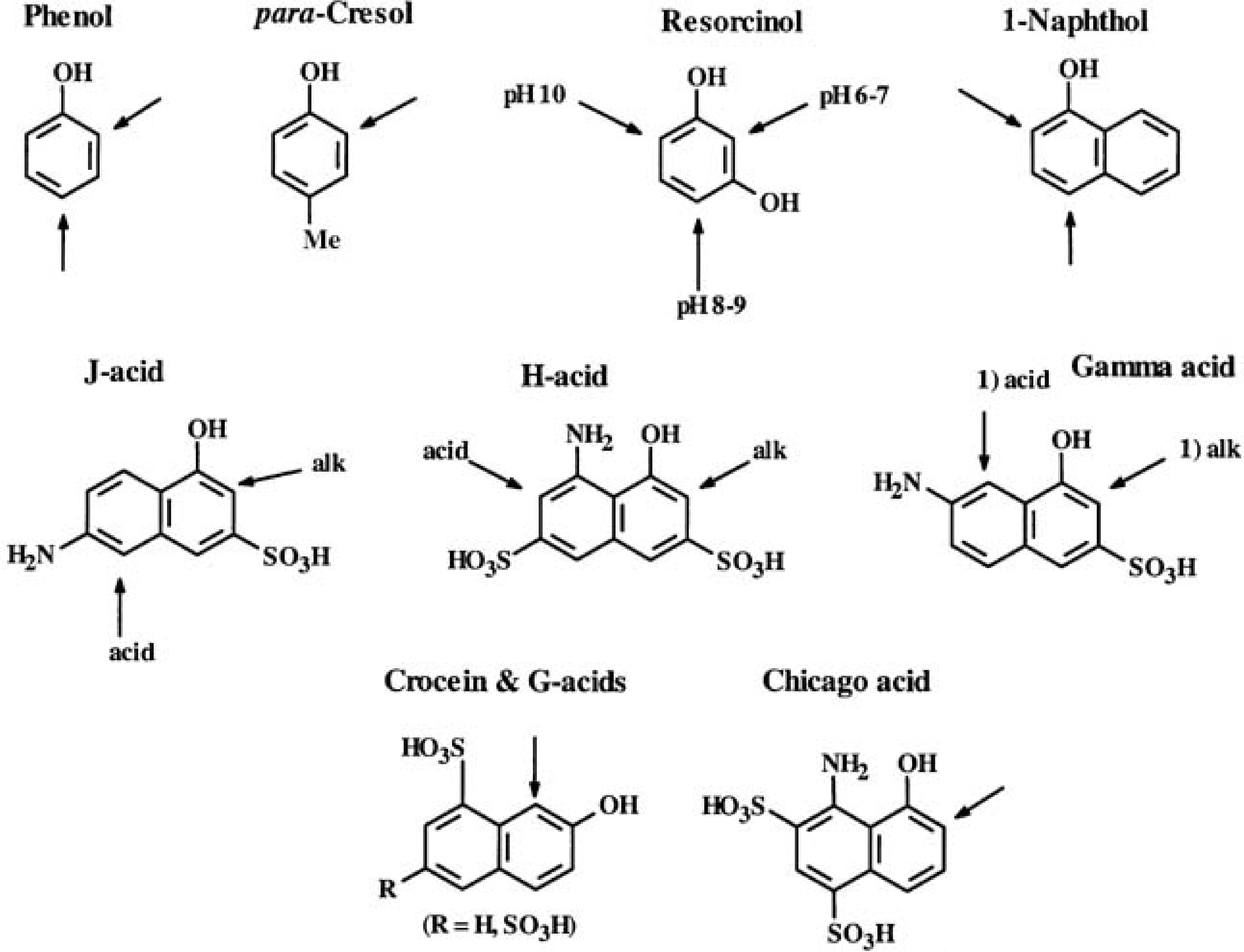


Fig. 23.84 Triazole formation from the diazotization of an *ortho*-diamine.



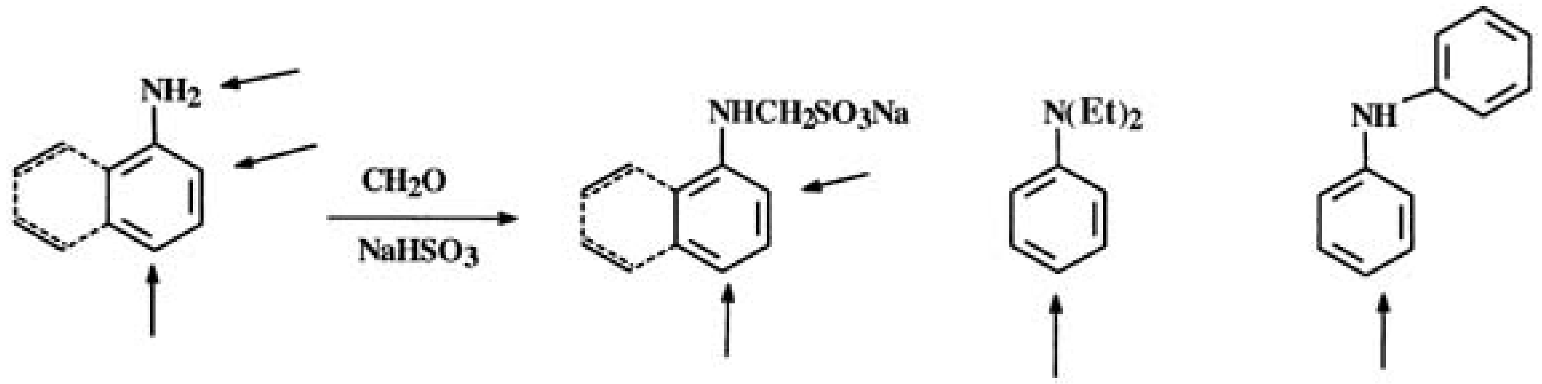
Examples of compounds that can be used as coupling components in azo dye synthesis are shown in Figs 23.8523.87. The first group is comprised of phenols and naphthols, the second group is comprised of amines that couple, and the third contains couplers that have an active methylene group. Compounds in the first and third groups require ionization using alkali, to give sufficient reactivity for diazo coupling, and the pH employed is usually 89. Since aromatic amines are appreciably more reactive, they couple at pH 56. Arrows have been used to indicate the coupling positions for the various couplers. Compounds such as 1-naphthol or 1-naphthylamine give a mixture of monoazo dyes by coupling in the 2-position or the 4-position. When couplers containing OH and NH2 groups are employed (see Fig. 23.85), coupling may occur twice, giving disazo dyes. In such cases, coupling is first conducted in acid, *ortho* to the NH2 group, and then in alkali. This is important because the introduction of the first azo group decreases the reactivity of the coupler. The ability to ionize the OH group provides sufficient ring activation for the second coupling. In the case of gamma acid, one has the lone option of coupling under acidic or alkaline conditions.

Fig. 23.85 Examples of phenol and naphthol-based couplers.



When primary amines are used as couplers, coupling can occur on the ring or at the amino group itself unless the amino group is blocked. One good way to block this group is by converting it to the *N*-sulfomethyl group, as illustrated in Fig. 23.86. The products formed are also known as omega salts.55 The blocking group can be removed following the coupling step, by treating the resultant azo dye with an alkaline solution.

Fig. 23.86 Examples of aromatic amine couplers.



There are also important examples of phenolic compounds that do not couple (see Fig. 23.88). In these examples, the required coupling positions are blocked, the ring is too deactivated, or the compounds undergo oxidation in the presence of the diazo compound.

Fig. 23.87 Examples of couplers having active methylene groups.

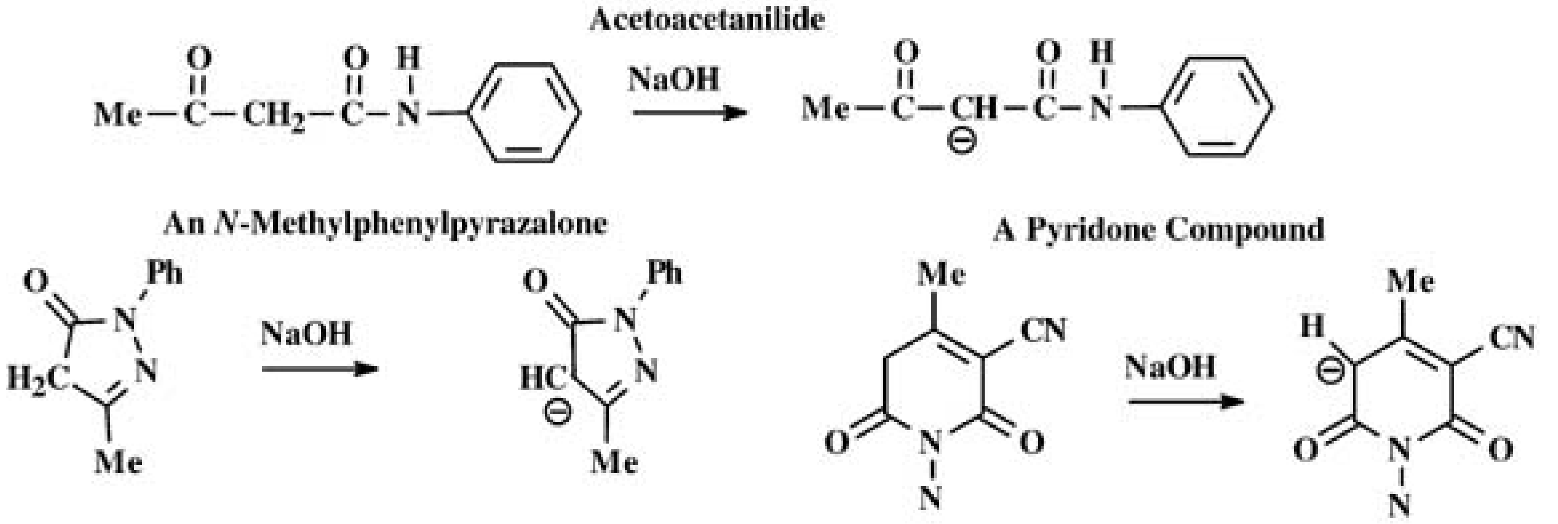
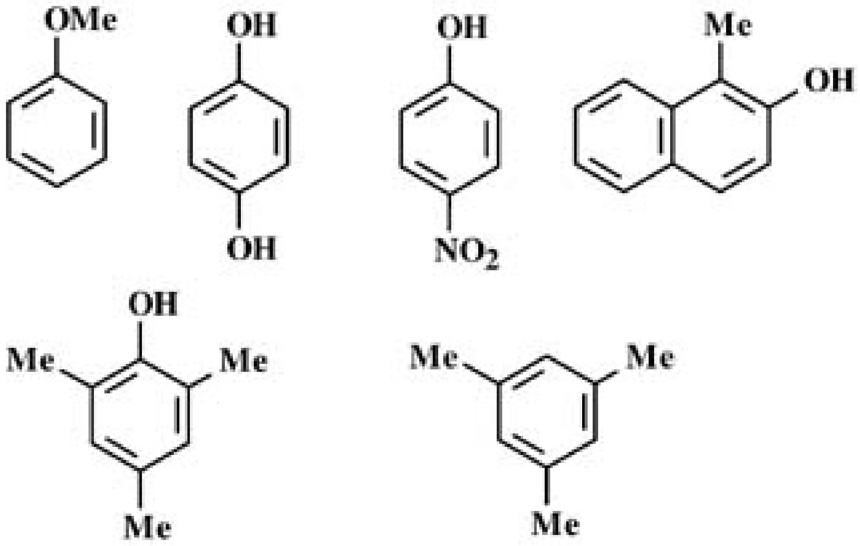


Fig. 23.88 Examples of compounds that do not couple.



The synthesis of azo dyes can be illustrated using the following letter designations:
A = diazotizable amine
D = tetrazotizable diamine
E = coupler that couples once
M = 1° amine that couples once and is diazotized and coupled again
Z = coupler that couples twice
Z·X·Z = binuclear coupler that couples twice

These designations are used to provide an indication of how a given dye has been assembled, and will be used in describing the syntheses of azo dyes covered in the subsections that follow.

Monoazo Dyes

Azo dyes of this type are manufactured predominantly by the reaction between a diazotized amine ("A") and a type "E" or "Z" coupler. The synthesis can be as simple as coupling diazotized aniline to H-acid, in an A → E process, to produce C.I. Acid Red 33. An example of a reactive dye that is manufactured via an A → E process is C.I. Reactive Red 1. In this case, the target dye is manufactured as shown in Fig. 23.89, which shows that the reactive group can be introduced prior to (sequence 1) or after (sequence 2) the coupling step. Similarly, monoazo bireactive dyes are made by this process (Fig. 23.90). This illustrates that a quite complex arylamine can be used as the diazo compound.

Fig. 23.89 Two approaches to the synthesis of Reactive Red 1.

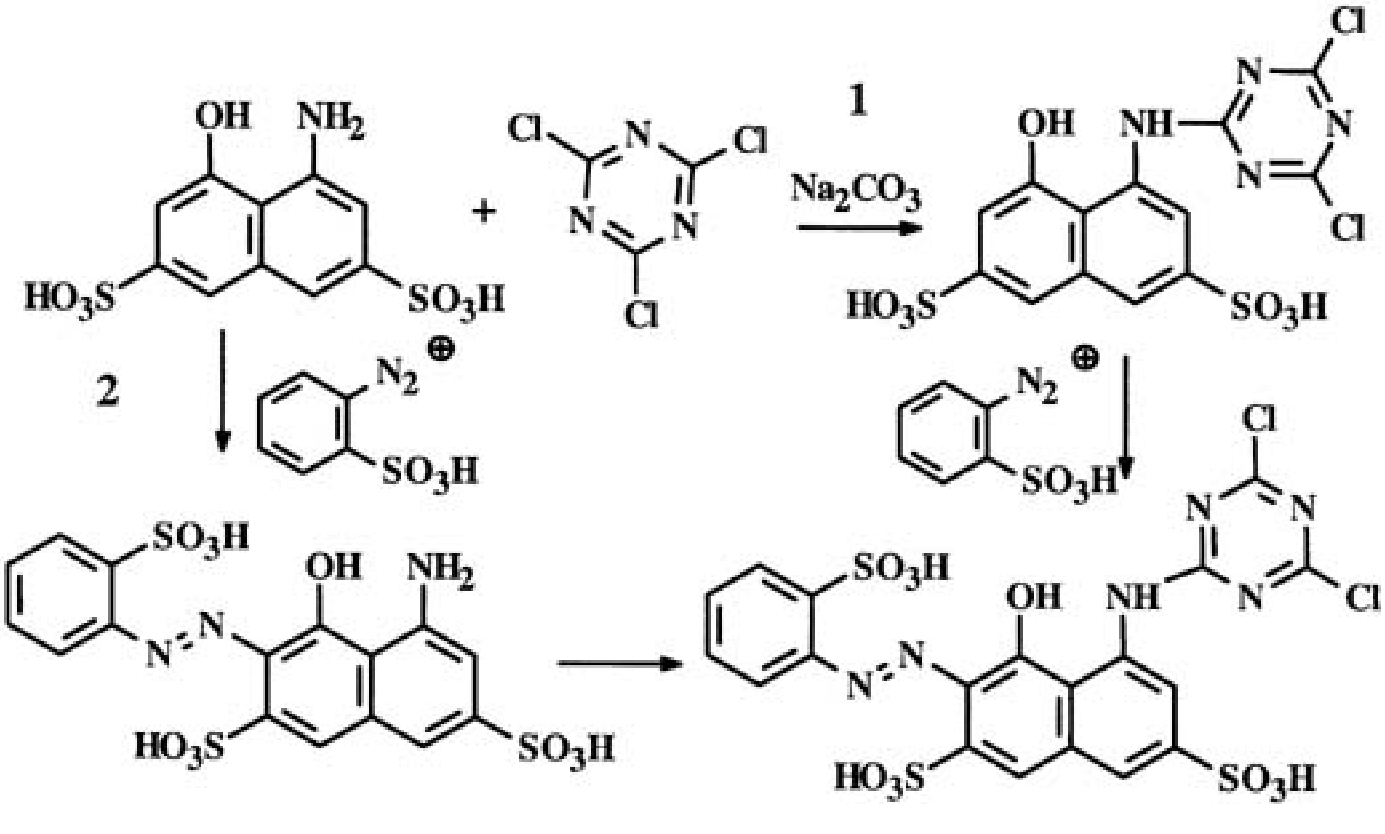
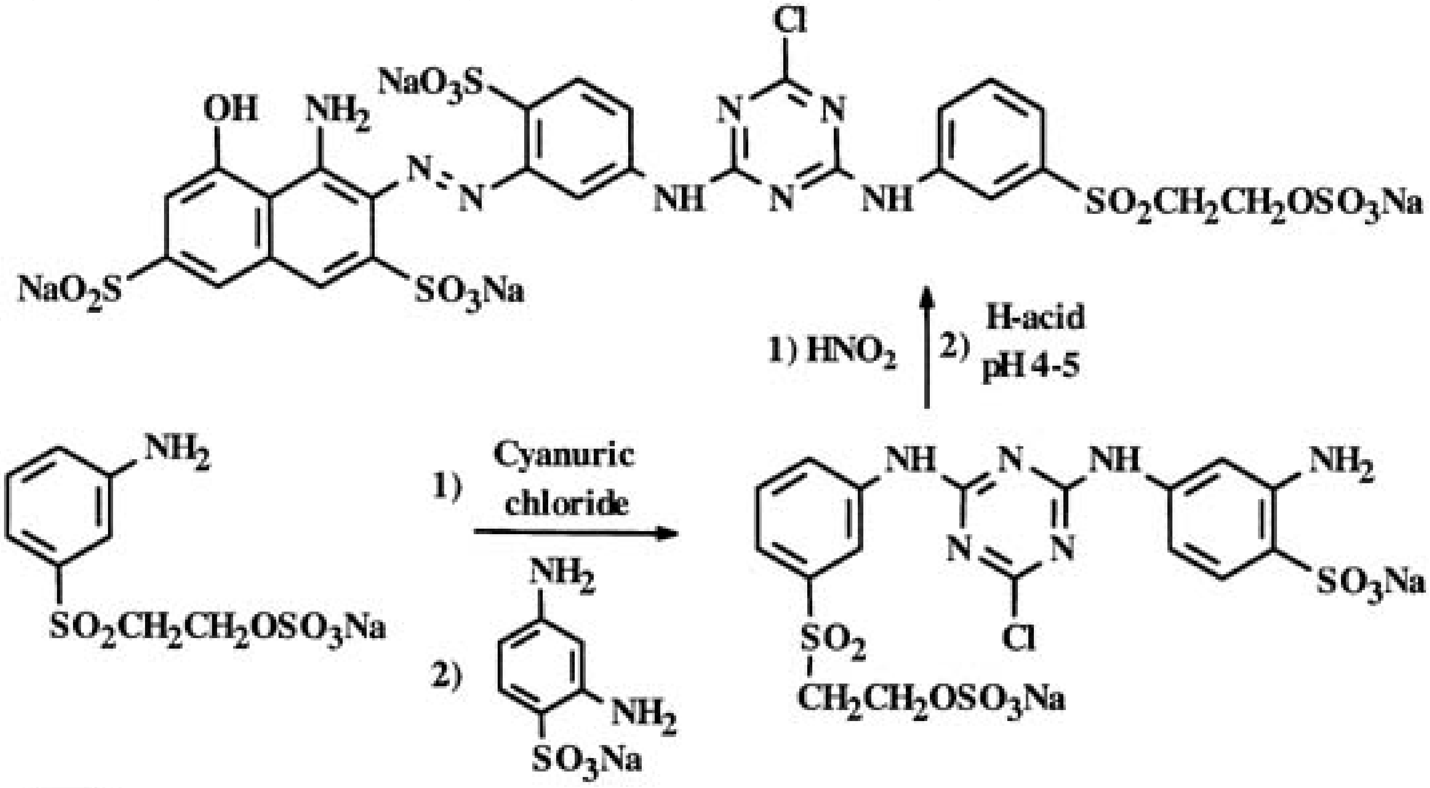
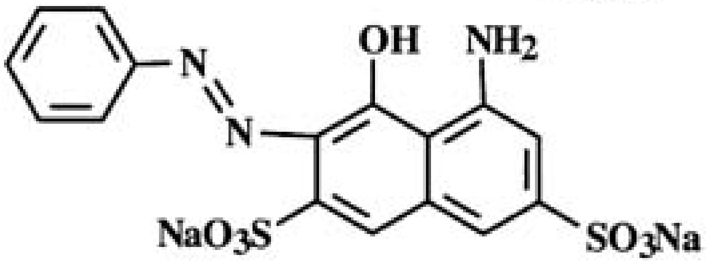


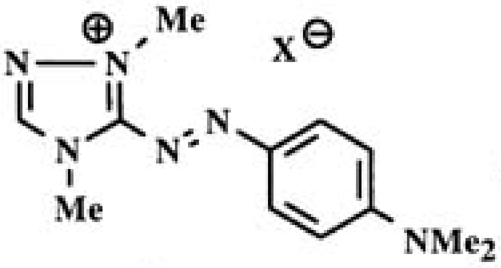
Fig. 23.90 Synthesis of a monoazo dye containing two different reactive groups.



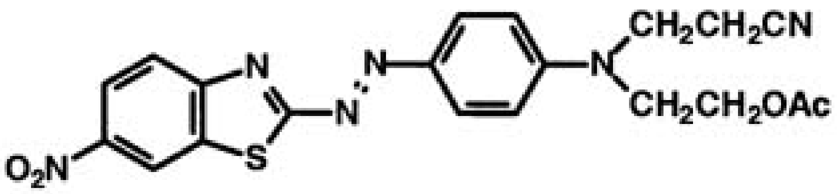


Other examples of monoazo dyes that are synthesized via an A → E process are shown in Fig. 23.91, further illustrating the wide range of structural types that can be manufactured this way.

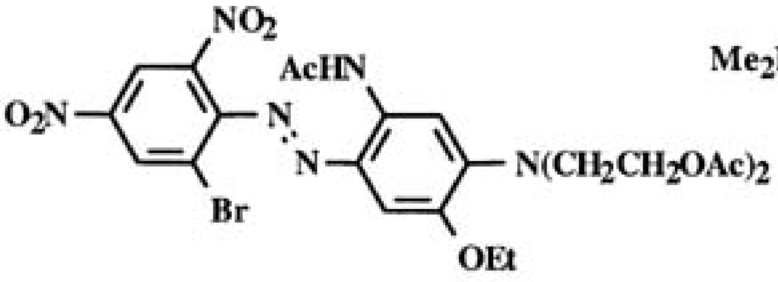
Fig. 23.91 Monoazo cationic (1), disperse (24), 1:1 chromium complexed (5), and mordant (6) dyes prepared via an A → E process.



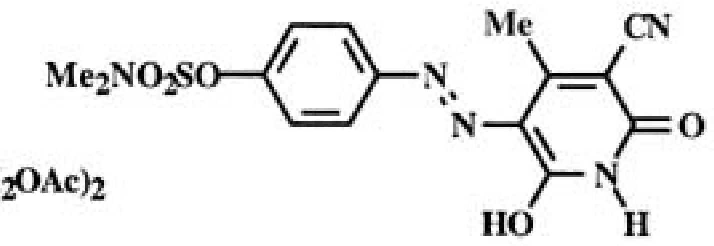
1



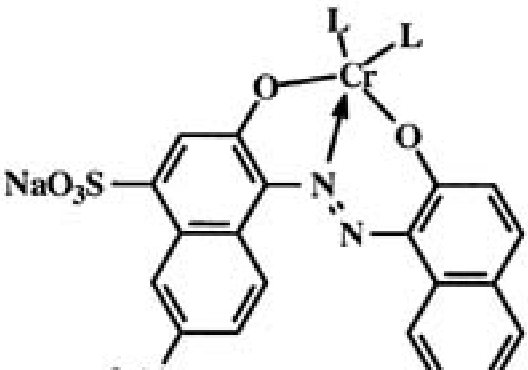
2



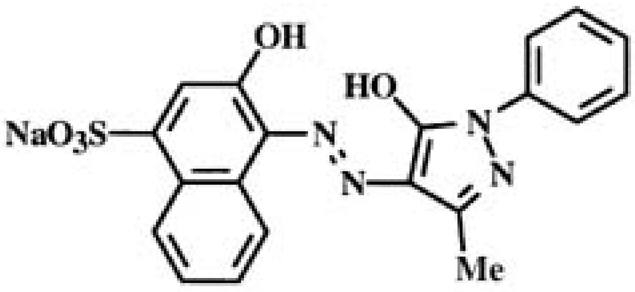
3



4



5



6

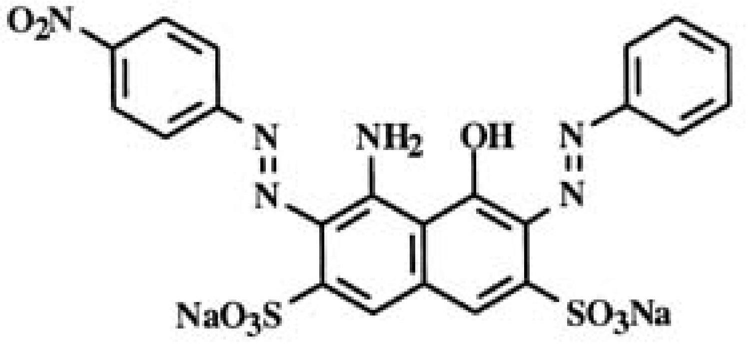
Disazo Dyes

There are four often used methods for synthesizing dyes containing two azo linkages, each of which requires two diazo coupling reactions. A non-traditional "disazo" dye involves 1:2 metal complex formation.

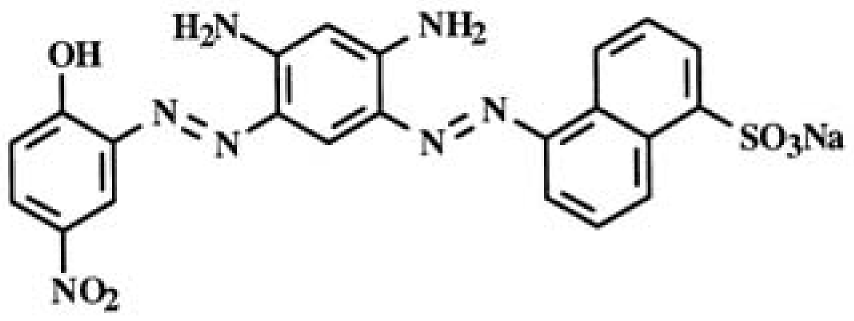
Type A1 → Z ← A2 synthesis

Dyes of this type include those shown in Fig. 23.92 (C.I. Acid Black 1 (7), C.I. Mordant Brown 1 (8), and C.I. Acid Black 17 (9), C.I. Direct Orange 18 (10)), in which couplers such as H-acid, resorcinol, and *meta*-phenylenediamine are coupled twice. Although A1 and A2 are different in the present examples, they need not be different. As pointed out in the section "Azo Dyes", coupling *ortho* to the amino group of H-acid is usually conducted first, under weakly acidic conditions, followed by coupling with diazotized aniline under alkaline conditions. This is also true for the structurally similar dye 9, which is prepared from S-acid. In the case of dye 10, however, coupling with aniline under slightly acidic conditions is the second step. For dye 8, both couplings are conducted under acidic conditions, with 2-amino-4-nitrophenol introduced first.

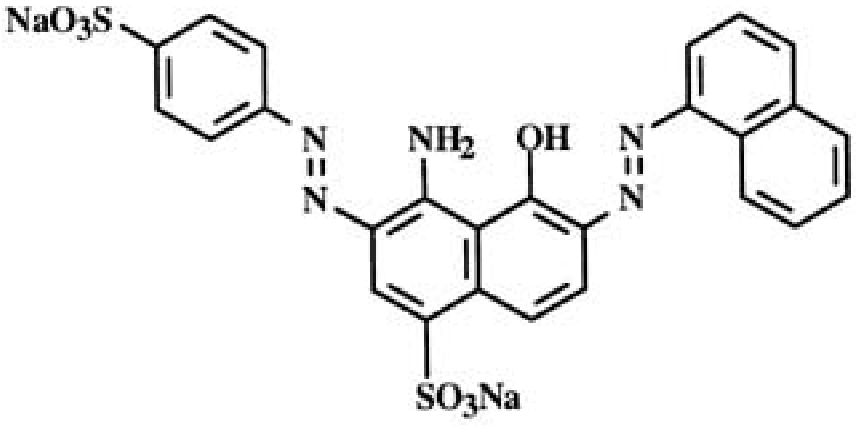
Fig. 23.92 Disazo dyes prepared by coupling twice to H-acid (7), *meta*-phenylenediamine (8), S-acid (9), and resorcinol (10).



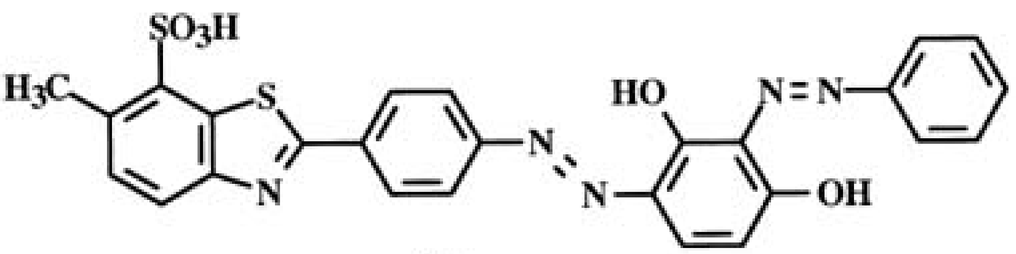
7



8



9



10

Type E1 ← D → E1 synthesis

Dyes of this type require the conversion of an aryldiamine to a tetrazonium compound (one that has two diazonium groups), in a process known as tetrazotization. (see Fig. 23.93), which involves environmentally friendly alternatives to benzidine. Following tetrazotization, one tetrazonium molecule reacts with two coupler molecules to produce the target dye, examples of which are provided in Fig. 23.94 (11: C.I. Direct Red 28 (Congo Red), 12: C.I. Direct Yellow 12, and 13: C.I. Acid Yellow 42). Disazo dyes prepared this way include dye 14 (C.I. Direct Blue 15), which is converted to the important bis-copper complexC.I. Direct Blue 218 (Fig. 23.95).

Fig. 23.93 Tetrazotization of (a) a di-n-propoxybenzidine and (b) a diaminostilbene disulfonic acid.

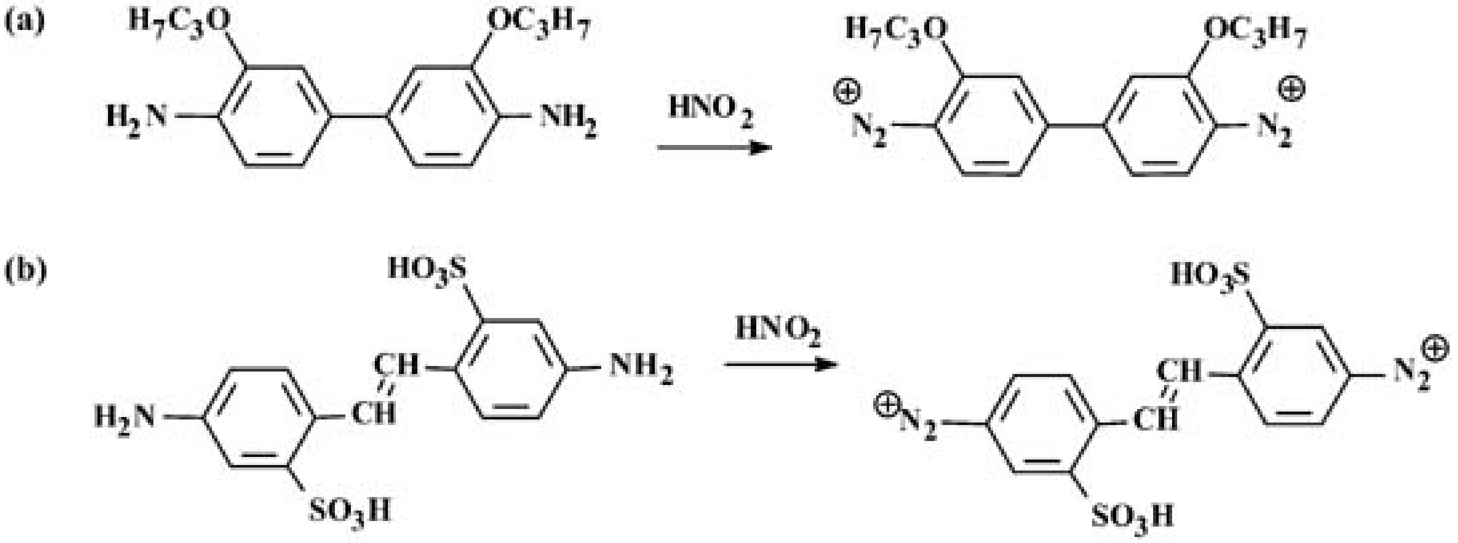
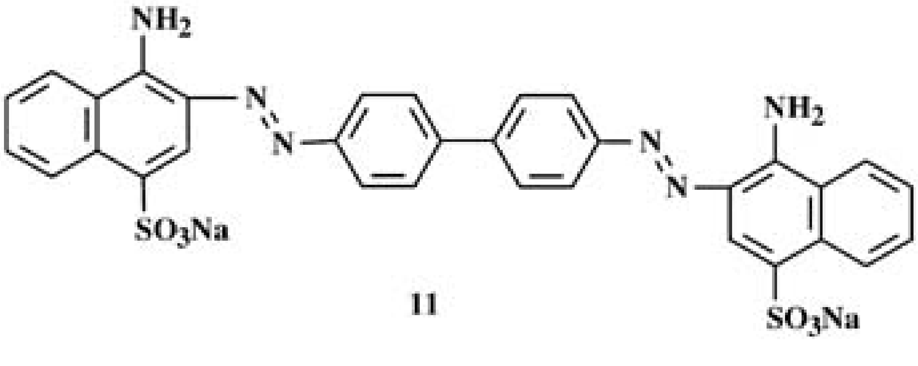
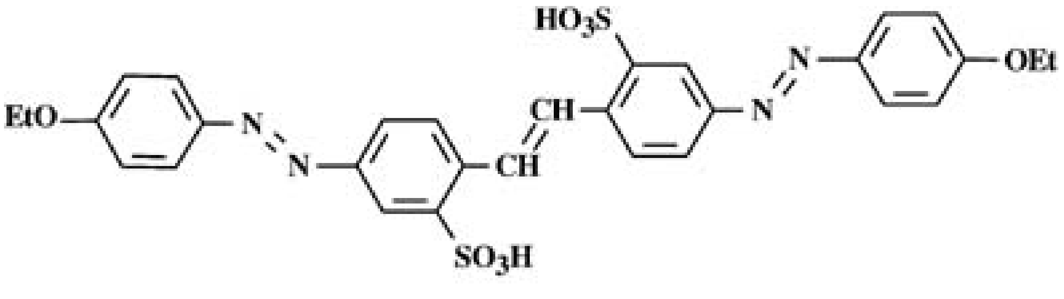


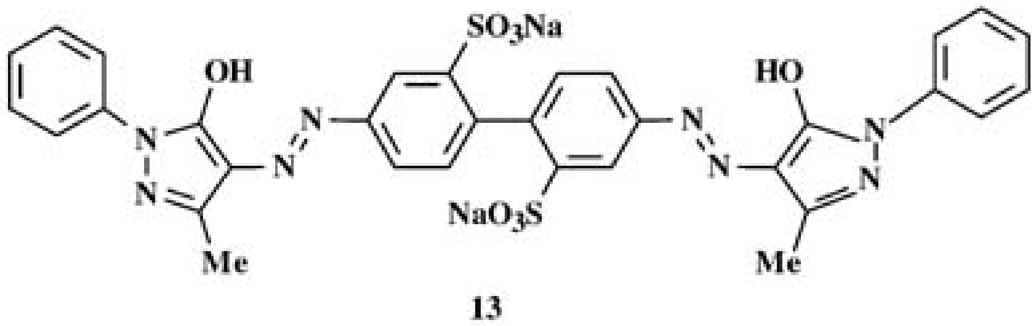
Fig. 23.94 Examples of disazo dyes (1113) prepared via a type E1 ← D → E1 synthesis.



11

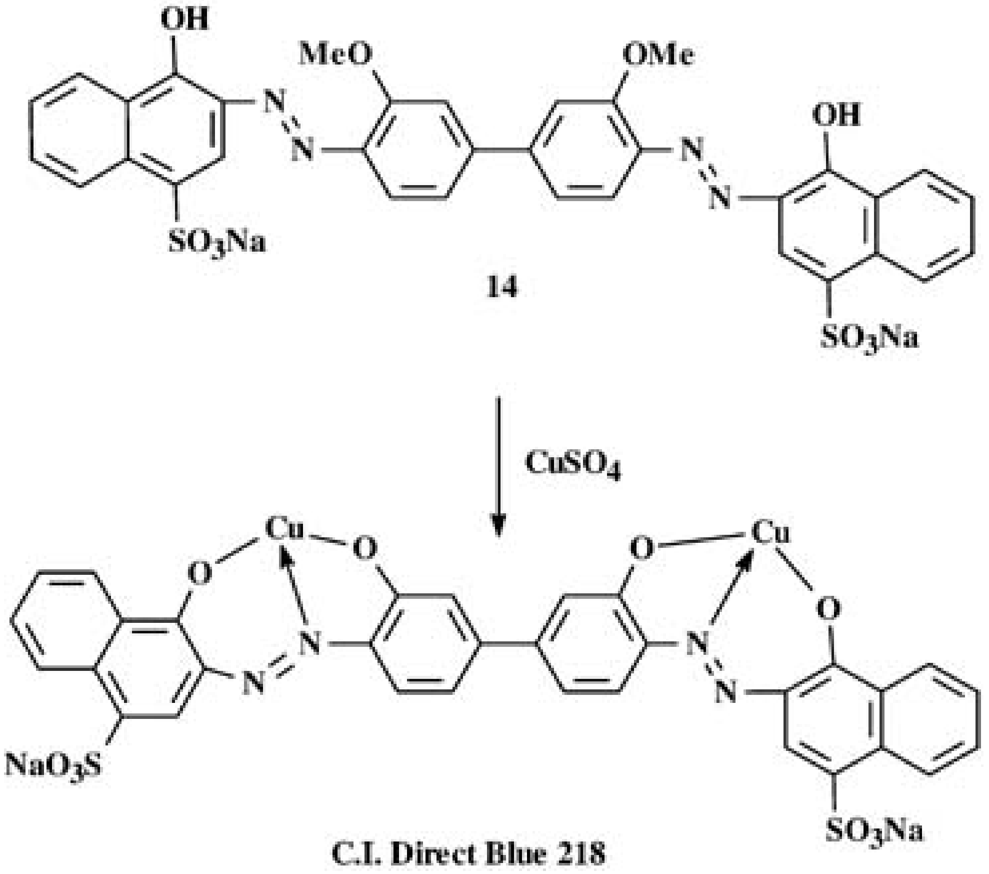


12



13

Fig. 23.95 Formation of Direct Blue 218.



Type A → M → E synthesis

This is one of the largest groups of disazo dyes, as they include acid, disperse, direct, and reactive dye structures. A representative synthesis is shown in Fig. 23.96. The second diazotization and coupling steps can be conducted inside certain textile fibers. For instance, disperse black dyes are produced in the presence of polyester by conducting the chemistry shown, in Fig. 23.97 after dyeing polyester with the monoazo dye.

Fig. 23.96 Disazo dye synthesis via an A → M → E process.

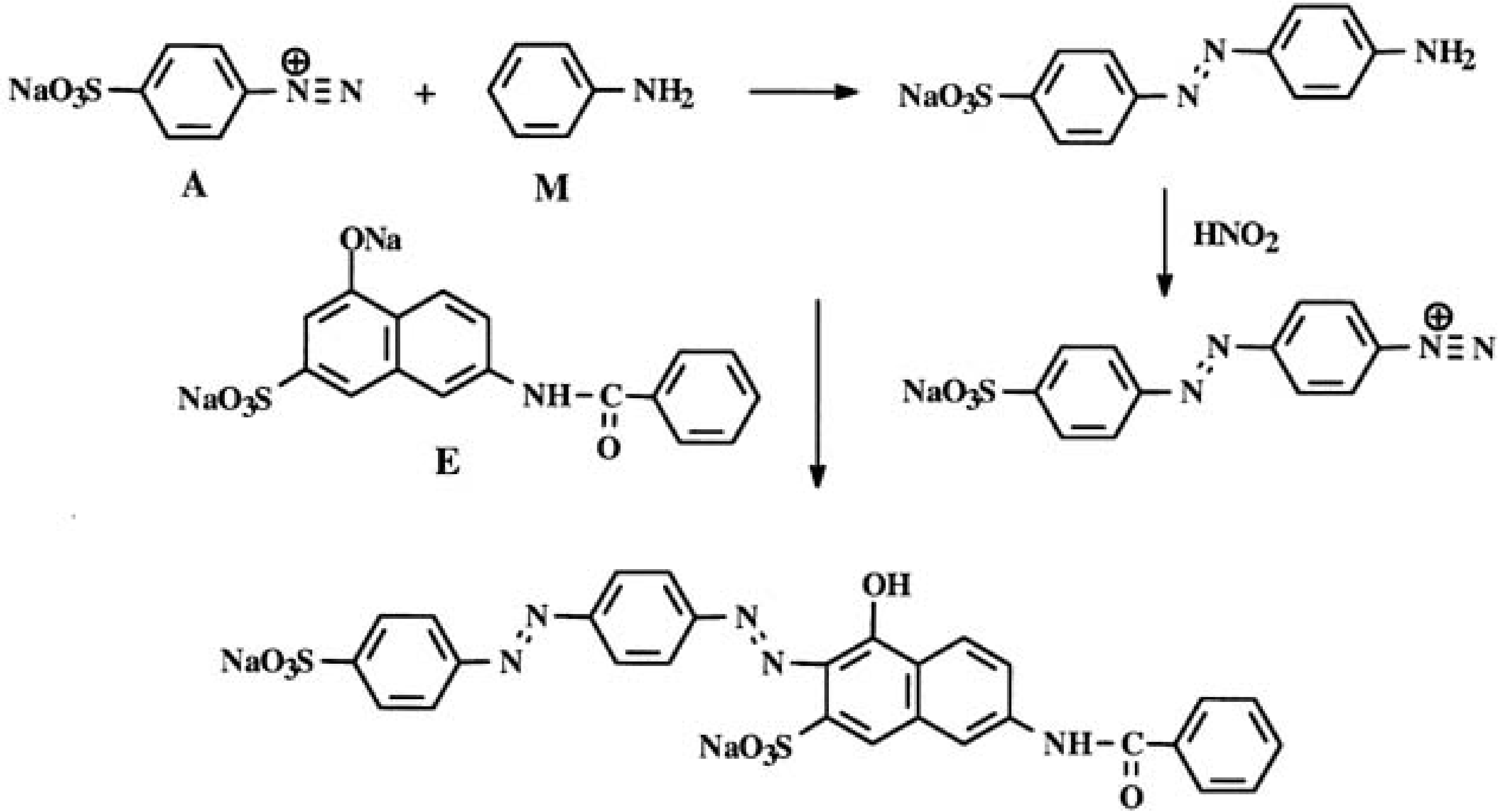
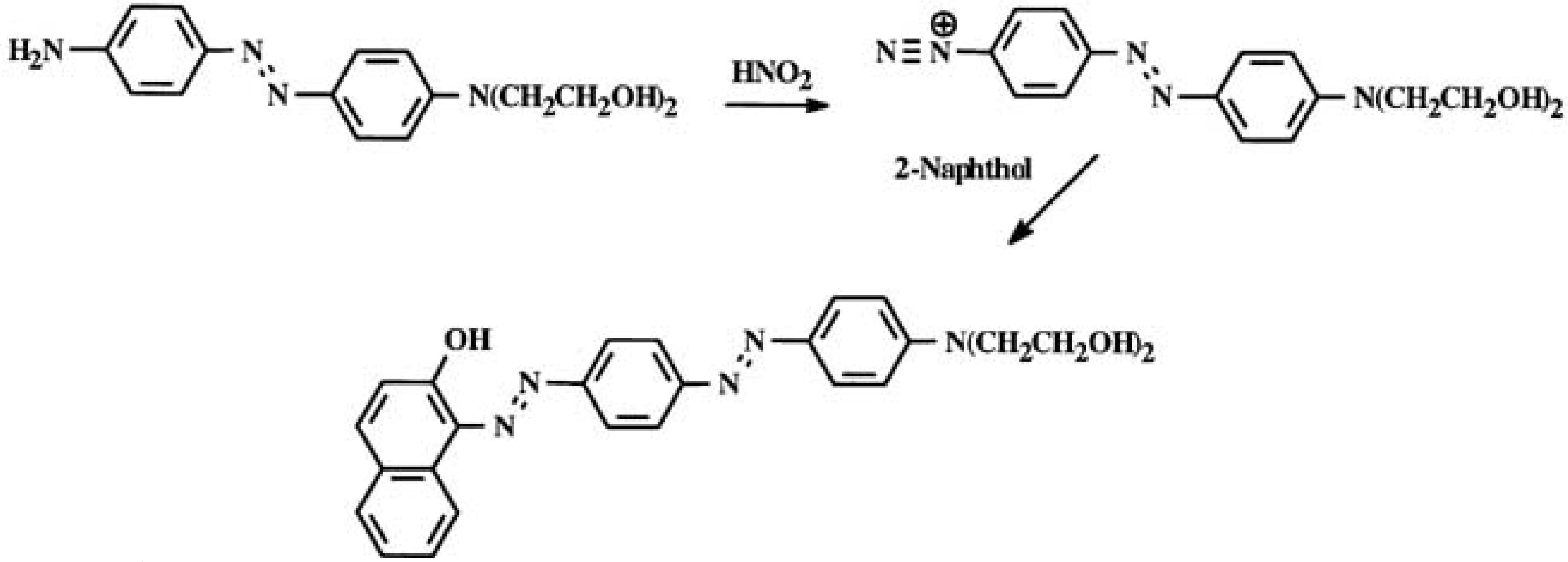
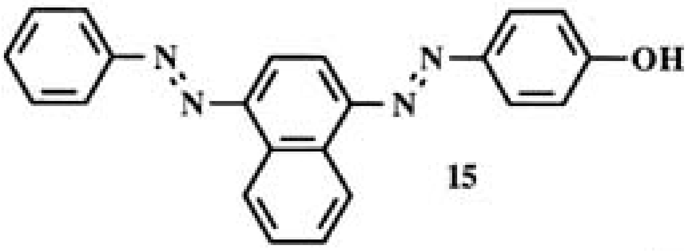


Fig. 23.97 Disazo disperse black dye synthesis conducted inside polyester fibers.

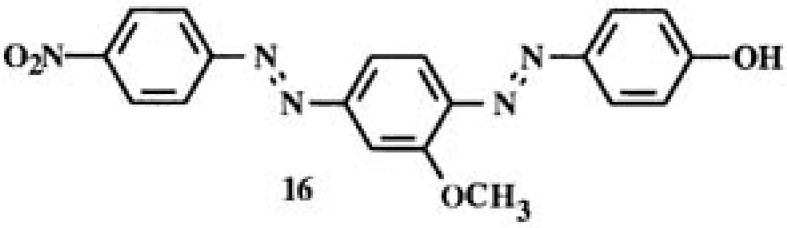


Examples of dyes made via an A → M → E synthesis are shown in Fig. 23.98. While most azo disperse dyes are based on monoazo structures, disazo structures such as 15 (C.I. Disperse Orange 13) and 16 (C.I. Disperse Orange 29) are manufactured. An important direct dye of this type is 17 (C.I. Direct Red 81), a reactive dye is 18 (C.I. Reactive Blue 40), and acid dyes include 19 (C.I. Acid Red 151) and 20 (C.I. Acid Blue 116).

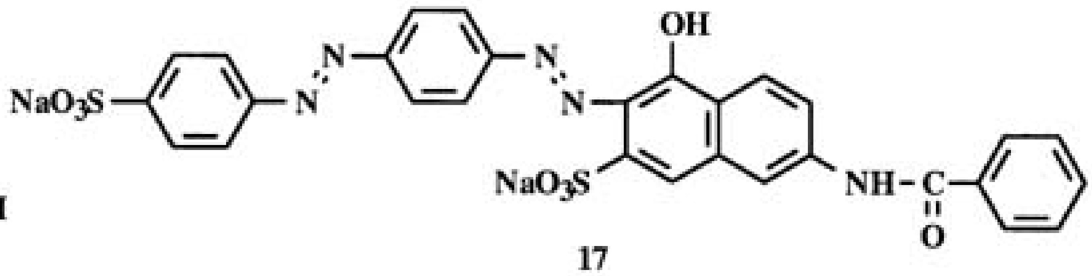
Fig. 23.98 Examples of disazo dyes prepared by a type A → M → E synthesis.



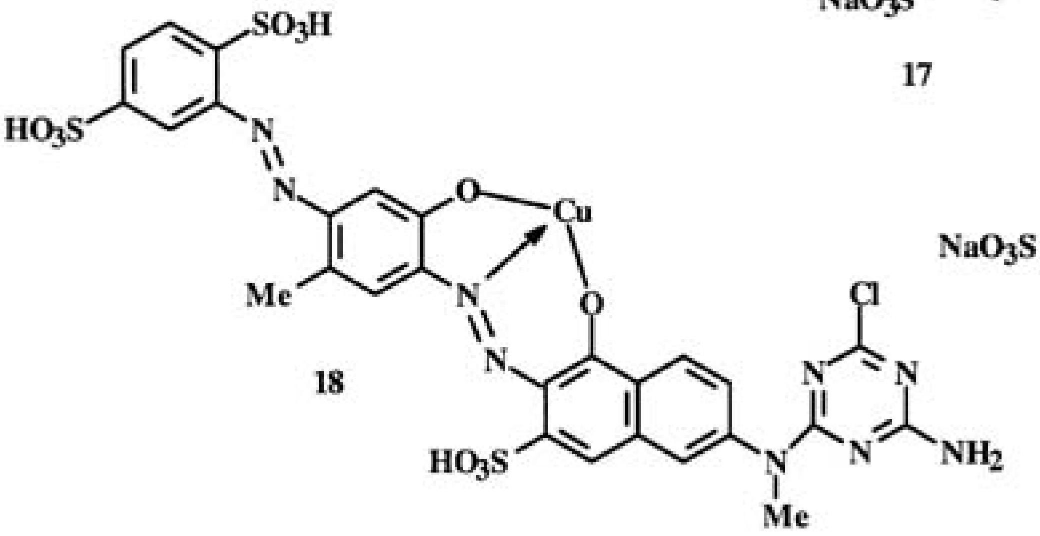
15



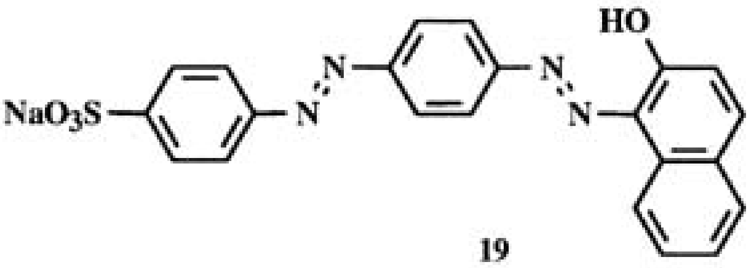
16



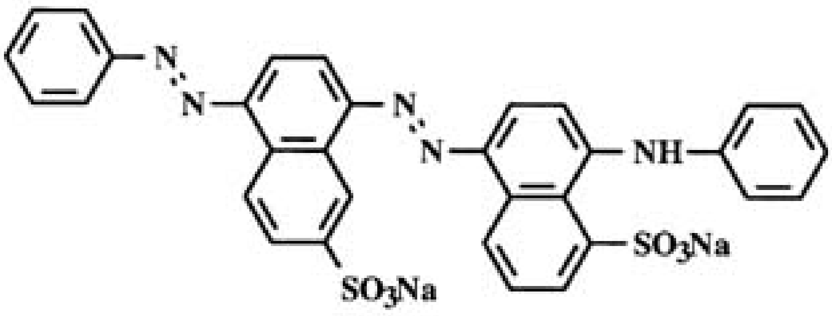
17



18



19



20

Type A1 → Z·X·Z ← A1 synthesis

Disazo dyes of this type are produced from coupling twice to dye intermediates such as those shown in Fig. 23.99, and are largely direct dyes for cotton. A representative synthesis is shown in Fig. 23.100, for C.I. Direct Red 83. In this case the target dye is prepared by metallization after the coupling step.

Fig. 23.99 Structures of (a) J-acid imide and (b) J-acid urea.

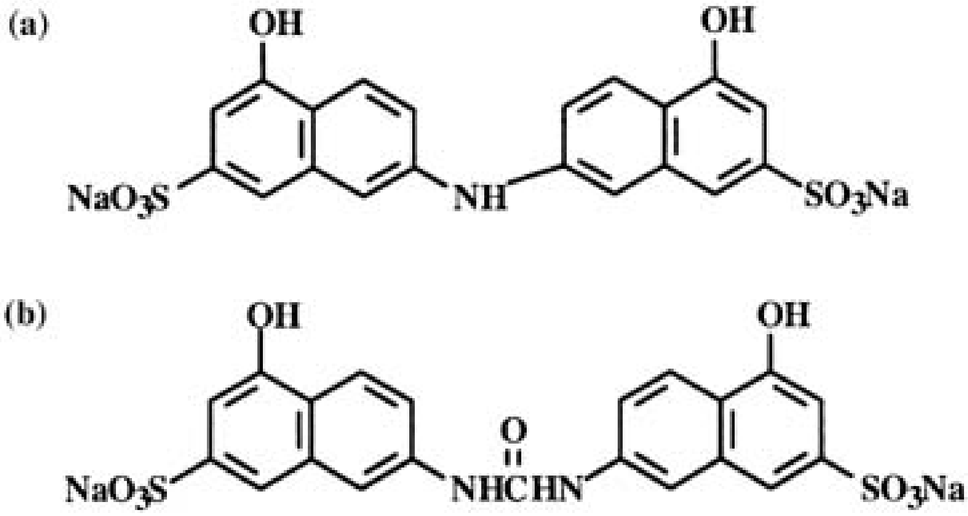
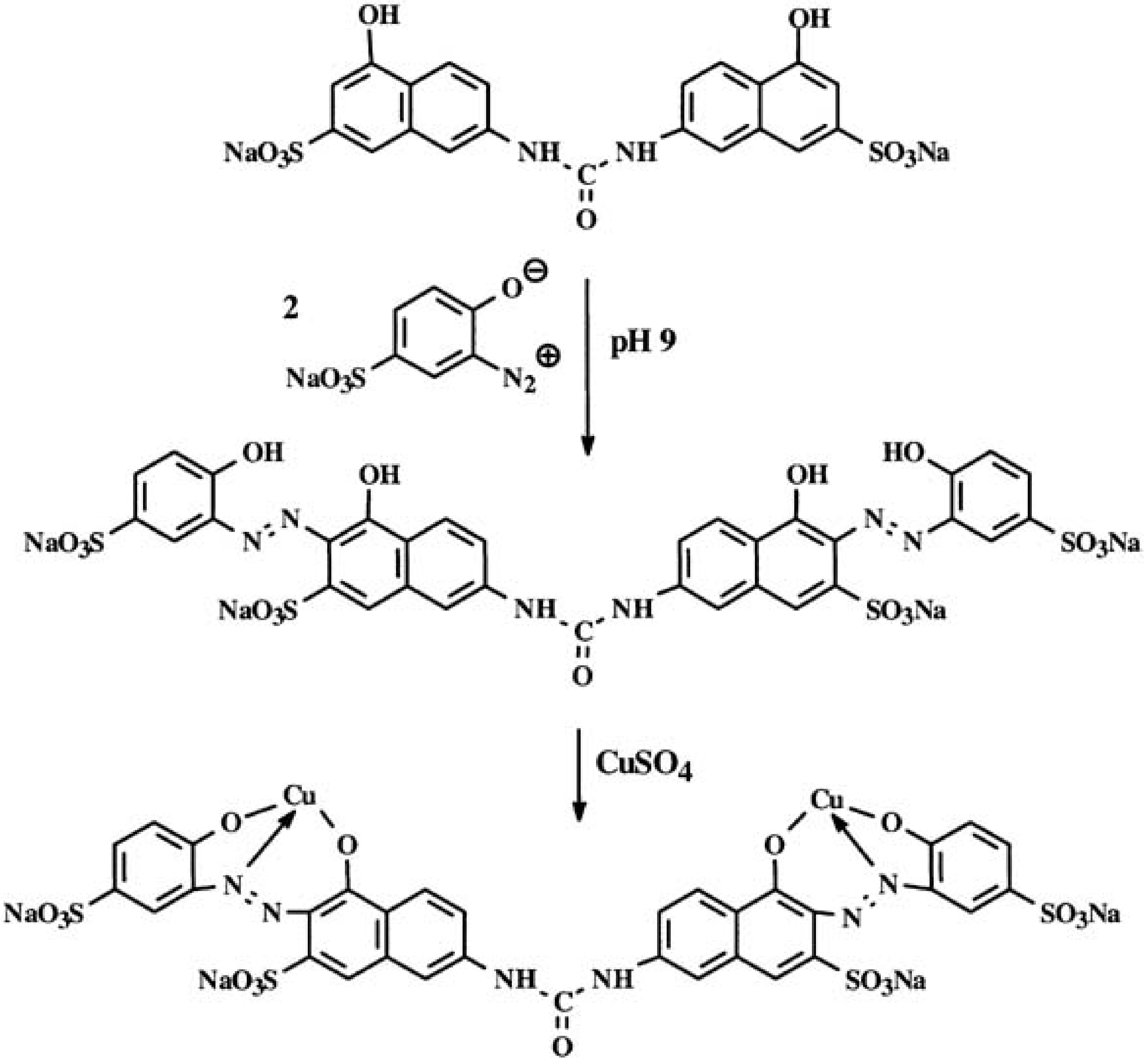
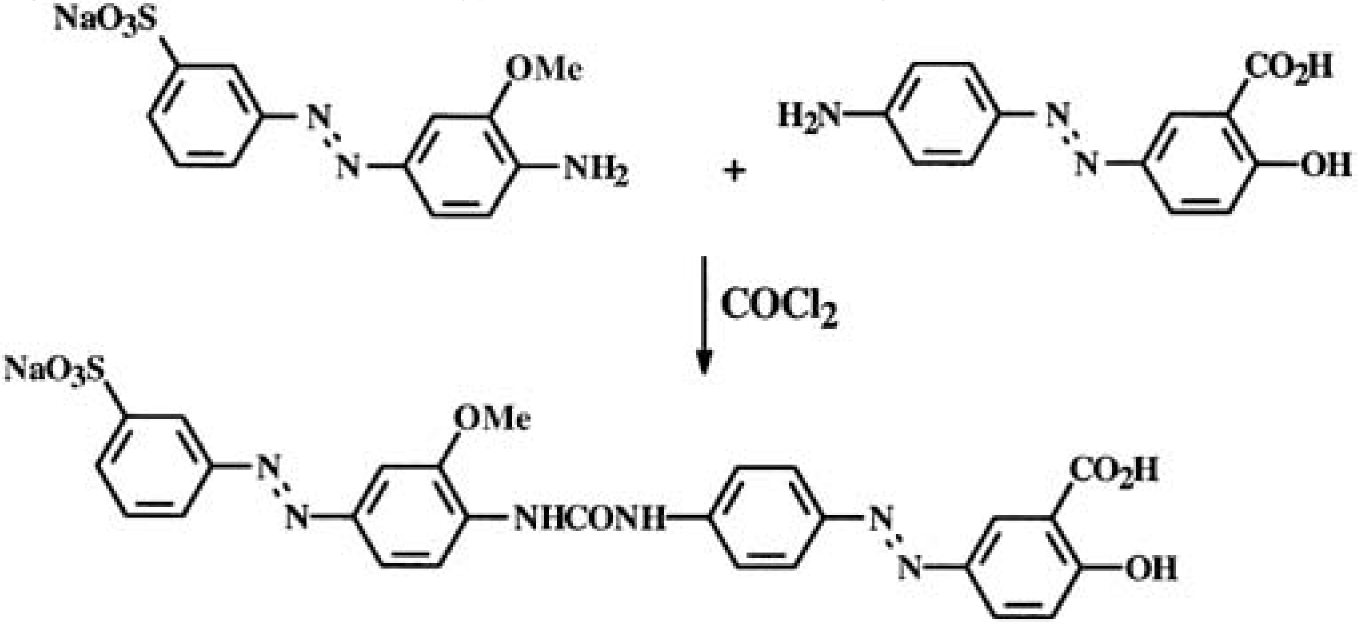


Fig. 23.100 Synthesis of Direct Red 83 via an A1 → Z·X·Z ← A1 synthesis and metallization.



Disazo dyes such as C.I. Direct Yellow 44 are prepared according to the sequence shown in Fig. 23.101. In this example, a pair of monoazo dyes is reacted with phosgene.

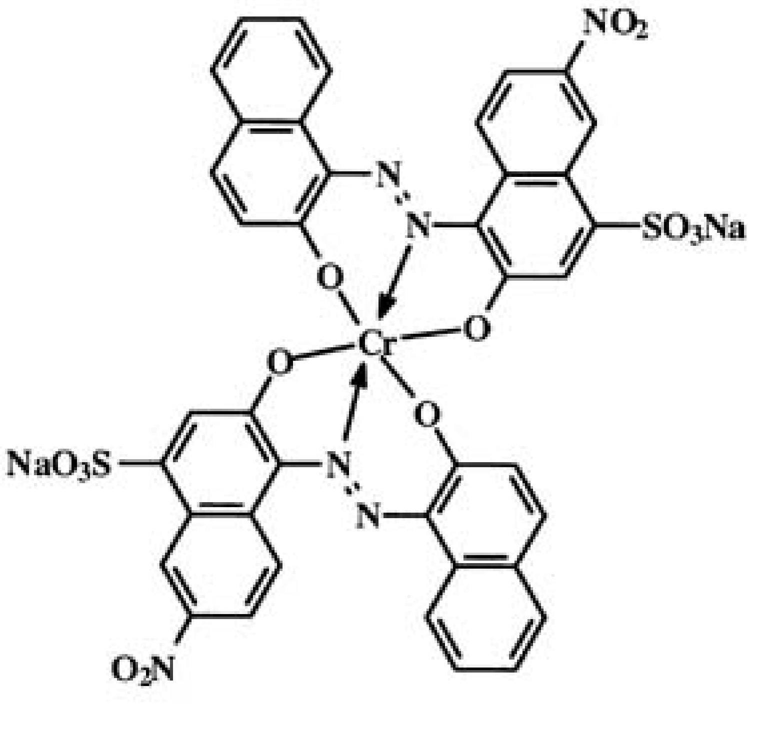
Fig. 23.101 An alternative route to type A1 → Z·X·Z ← A1 disazo dyes.



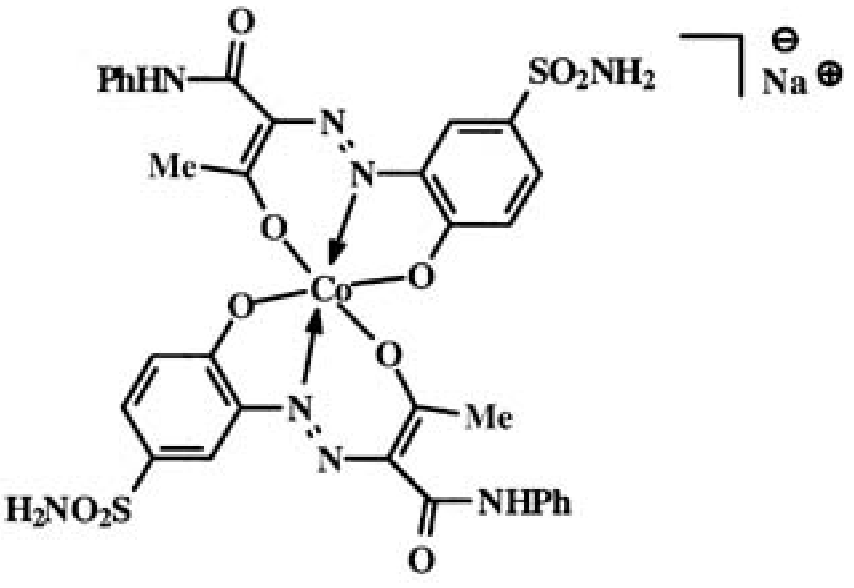
1:2 Metal complexes

Though somewhat different from the previous examples and methods, dyes containing two azo groups can also be synthesized by forming 1:2 metal complexes of suitably substituted monoazo dyes. The resultant dyes are mostly acid dyes for protein and polyamide substrates and the metals employed are Cr, Co, and Fe. Examples shown in Fig. 23.102 are for C.I. Acid Black 172 (21) and C.I. Acid Yellow 151 (22). In these examples, the corresponding monoazo dye is treated with one-half the molar amount of Cr2(SO4)3 or CoCl3, respectively.

Fig. 23.102 Representative 1:2 bisazo metal complexed dyes.



21

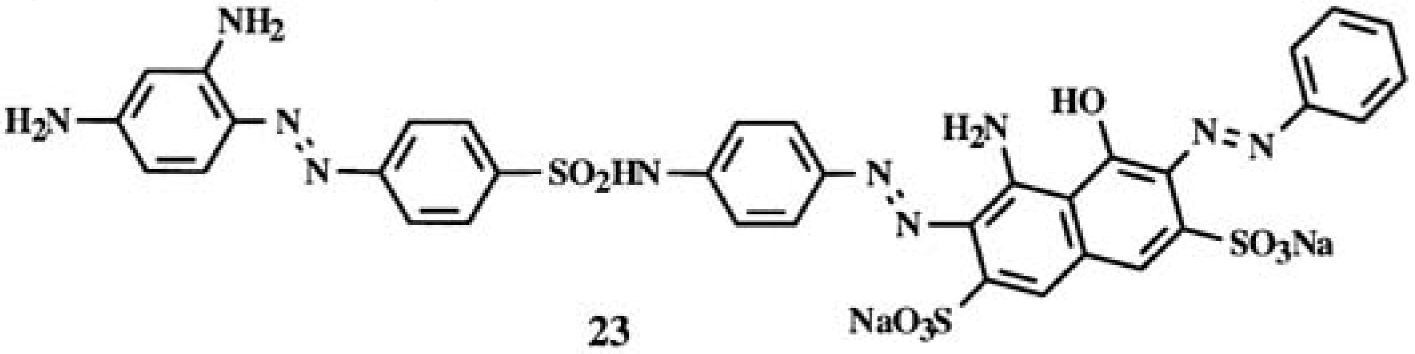


22

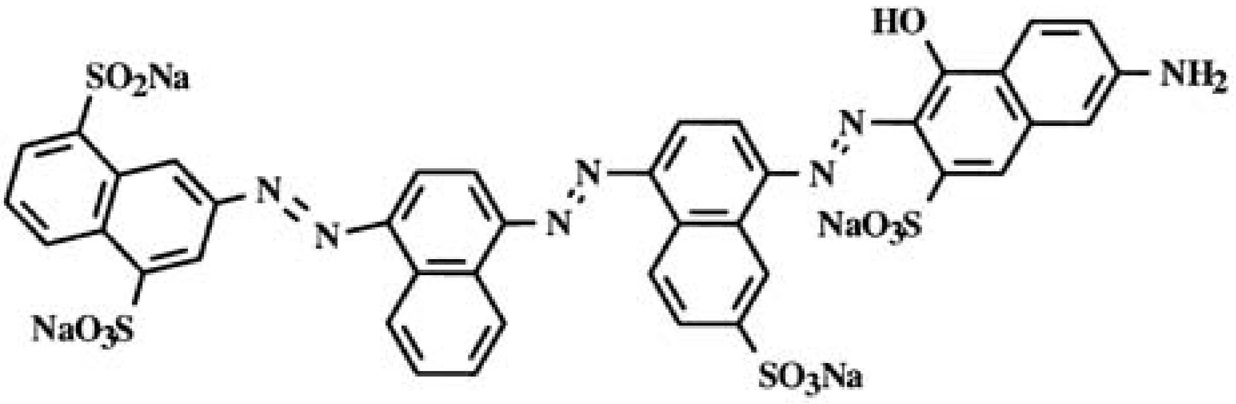
Polyazo Dyes

In this section, we cover the synthesis of dyes containing three or more azo linkages. In this regard, methods for producing trisazo dyes (those having three azo linkages) include E ← D → Z ← A and A → M1 → M2 → E syntheses. Examples are shown in Fig. 23.103 for C.I. Acid Black 234 (23) and C.I. Direct Blue 71 (24). In the synthesis of dye 23, an unsymmetrical dye can be made from diamine 25 because the end of the tetrazonium compound (cf. 26) that is *para* to the SO2 moiety is more reactive than the one that is *para* to the NH moiety (Fig. 23.104).

Fig. 23.103 Examples of trisazo dyes.

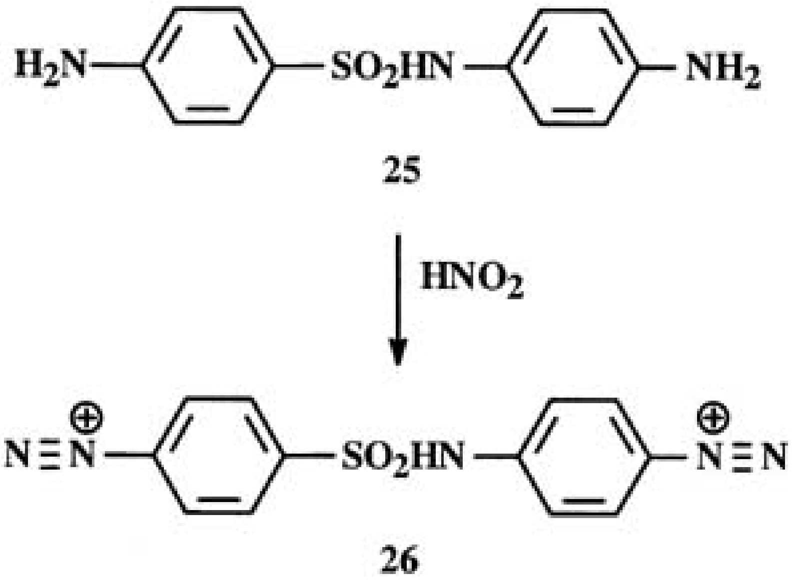


23



24

Fig. 23.104 Structures of compounds 25 and 26.

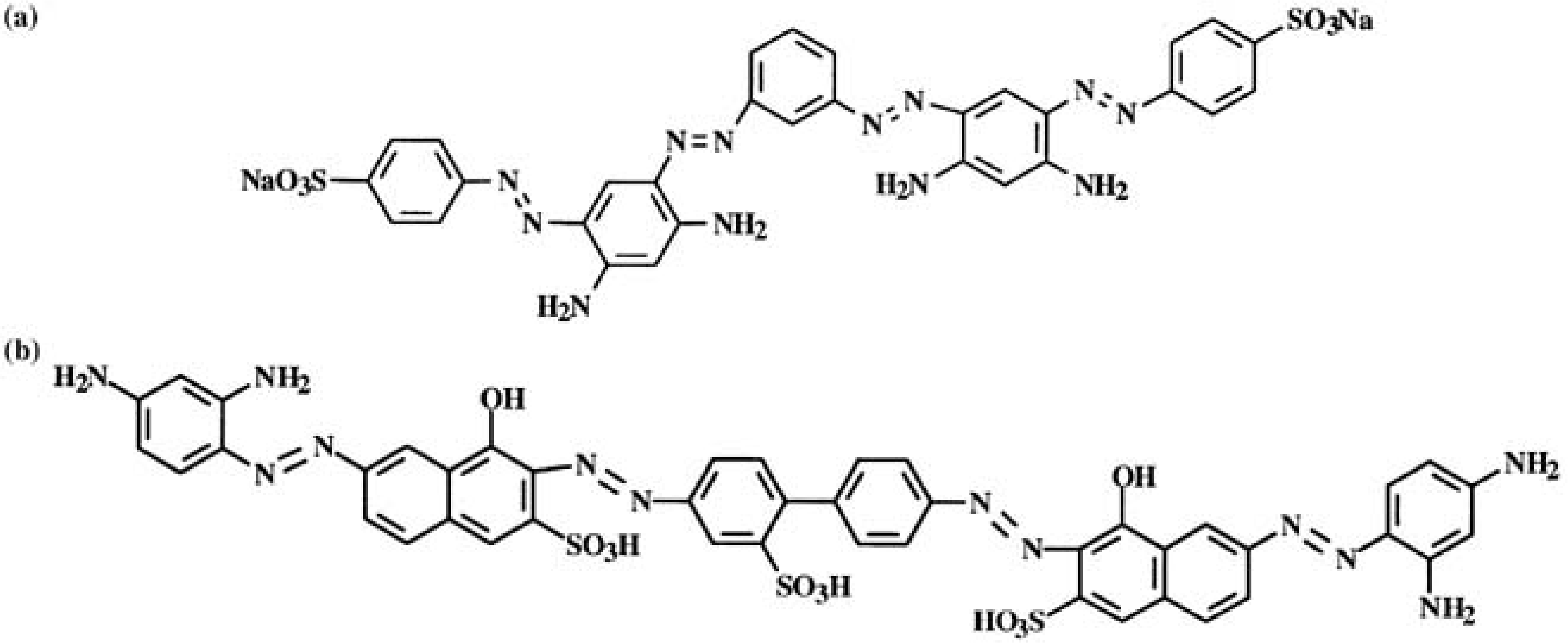


25

26

Dyes containing four azo linkages are direct dyes for cotton and can be prepared in several ways, including via A → M → Z ← D → E, A1 → Z1 ← D → Z2 ← A2, E1 ← M1 ← D → M2 → E2, E1 ← D1 → Z ← D2 → E2, and E1 ← D → M1 → M2 → E2 sequences. Examples of the second and third methods are shown in Fig. 23.105. Note that both are symmetrical molecules, the first of which (C.I. Direct Brown 44) employs *meta*-phenylenediamine as a type "Z" coupler and a type "D" diazo component. In the second example (C.I. Direct Black 22), gamma acid is twice used as the "M" moiety, and the dye is synthesized by (1) coupling tetrazotized benzidine sulfonic acid to two molecules of gamma acid, (2) diazotizing the amino groups on the gamma acid moieties, and (3) coupling to two molecules of *meta*-phenylenediamine.

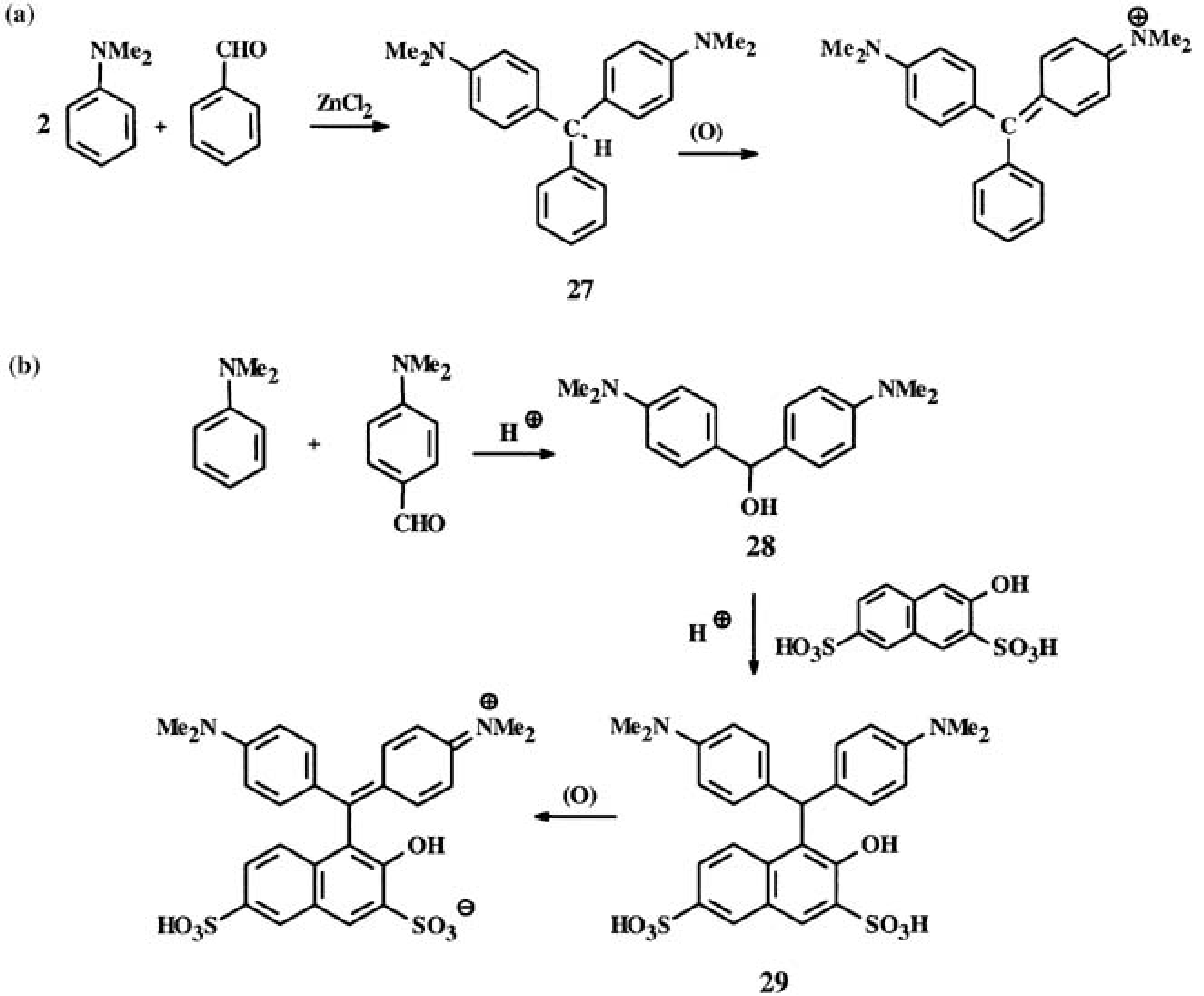
Fig. 23.105 Representative polyazo dyes prepared via the A1 → Z1 ← D → Z2 ← A2 (top) and E1 ← M1 ← D → M2 → E2 (bottom) methods.



Triphenylmethane Dyes

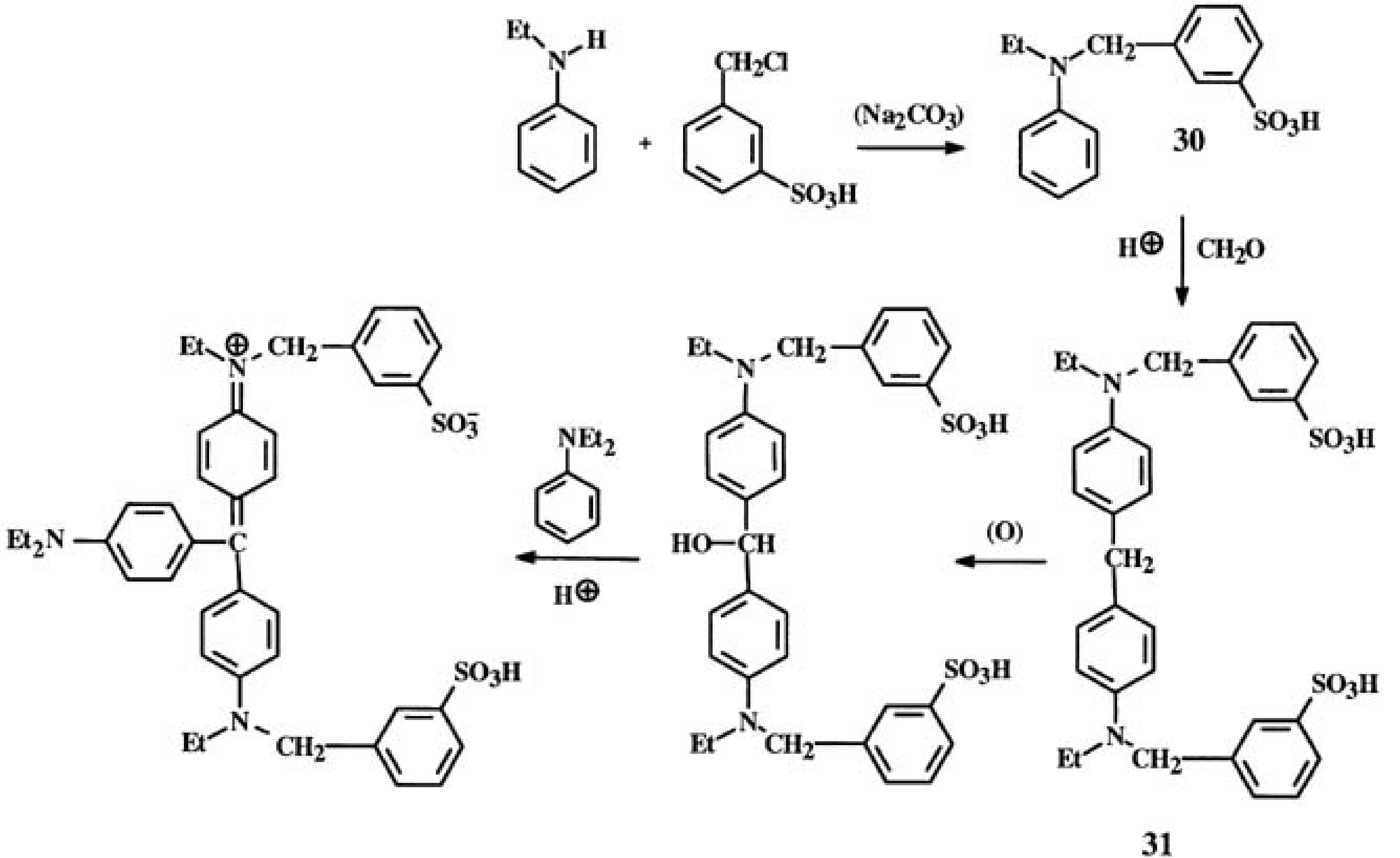
Triphenylmethane dyes are usually prepared in two steps: (1) condensation of an *N*, *N*-dialkylaniline with a benzaldehyde compound and (2) oxidation of the resultant leuco base (27). The synthesis of C.I. Basic Green 4 (Malachite Green) is given as an example in Fig. 23.106(a). Alternatively, C.I. Acid Green 50 is prepared in three steps (Fig. 23.106(b)): (1) condensation of *N*, *N*-dimethylaniline and *para*-(*N*, *N*-dimethylamino)benzaldehyde to produce Michler's hydrol (28), condensation with R-acid to give an intermediate leuco base (29), and (3) oxidation to give the target dye. Historically, PbO2 has been used as the oxidizing agent. However, concerns about its toxicity have led to the use of a more environmentally friendly agent such as tetrachloro-*para*-benzoquinone (chloranil).

Fig. 23.106 (a) Two-step synthesis of Malachite Green (Basic Green 4); (b) three-step synthesis of Acid Green 50.



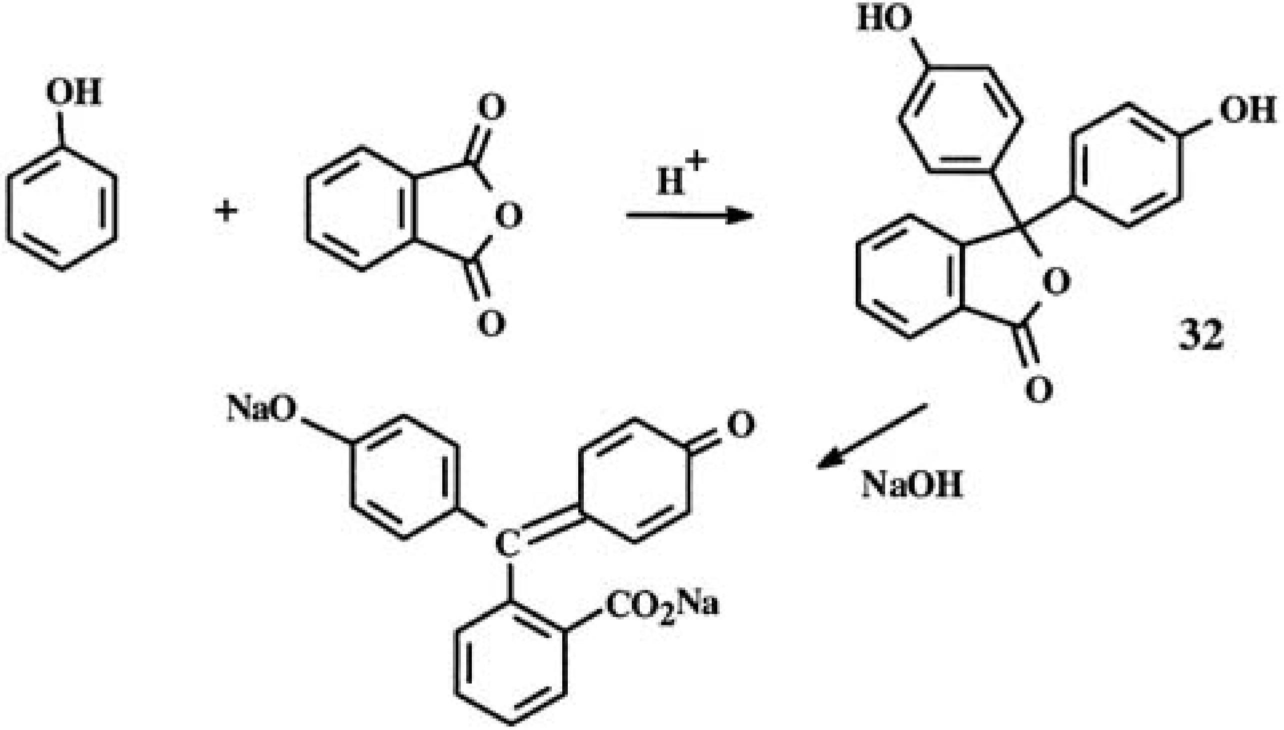
In another synthetic variation, C.I. Acid Violet 17 is prepared in the four steps shown in Fig. 23.107. The different steps in this process are the synthesis of the *N*-arylmethyl intermediate 30 and the diphenylmethane intermediate 31. Oxidation to the intermediate hydrol and condensation with *N*, *N*diethylaniline produce the target dye.

Fig. 23.107 Synthesis of Acid Violet 17 via diarylmethane intermediate 31.



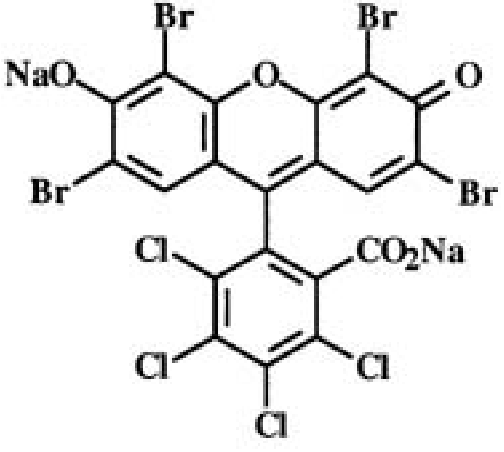
Structurally related dyes are synthesized by condensing phenols with phthalic anhydride to give a colorless intermediate lactone (32) that reacts with alkali to give the colored form. An example of this dye type is phenolphthalein, the synthesis of which is shown in Fig. 23.108.

Fig. 23.108 Synthesis of phenolphthalein via colorless lactone 32.



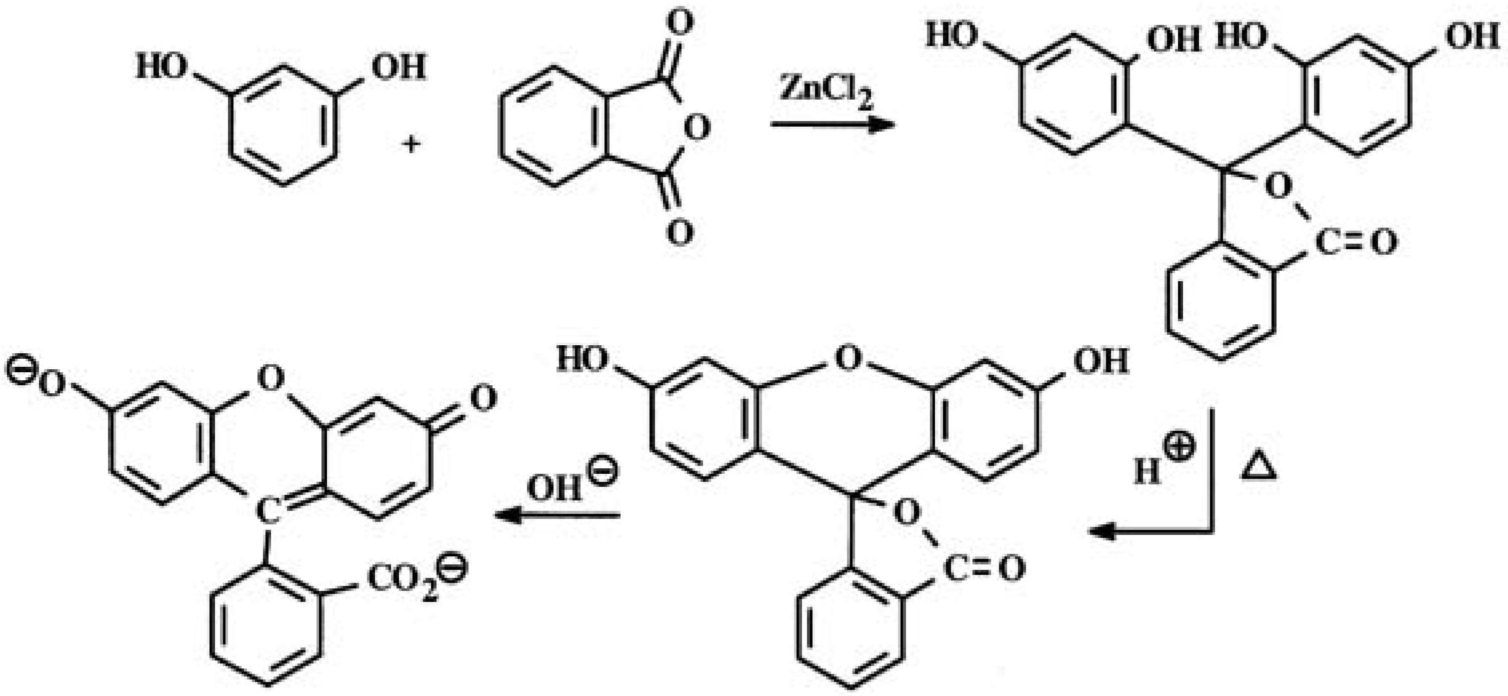
Xanthene Dyes

Like phenolphthalein, xanthene dyes are prepared in a condensation reaction involving phthalic anhydride. However, resorcinol is employed instead of phenol. The simplest representative of this family is C.I. Acid Yellow 73 (fluorescein), which is made via the sequence of step shown in Fig. 23.109. Similarly, C.I. Acid Red 92 is made by the condensation of tetrachlorophthalic anhydride and resorcinol followed by bromination.



Acid Red 92

Fig. 23.109 Synthesis of Acid Yellow 73, a xanthene dye.



Anthraquinone and Related Dyes

The commercial preparation of anthraquinone dyes begins with the synthesis of anthraquinone itself. In this regard, the three-step synthesis involves: (1) the oxidation of naphthalene to phthalic anhydride, (2) FriedelCrafts acylation of benzene to give a keto acid, and (3) cyclodehydration using H2SO4 (see Fig. 23.110). The preparation of 1,4-disubstituted anthraquinones utilizes the intermediates prepared in Fig. 23.111, where R = OH corresponds to quinizarin.

Fig. 23.110 Synthesis of the anthraquinone ring system.

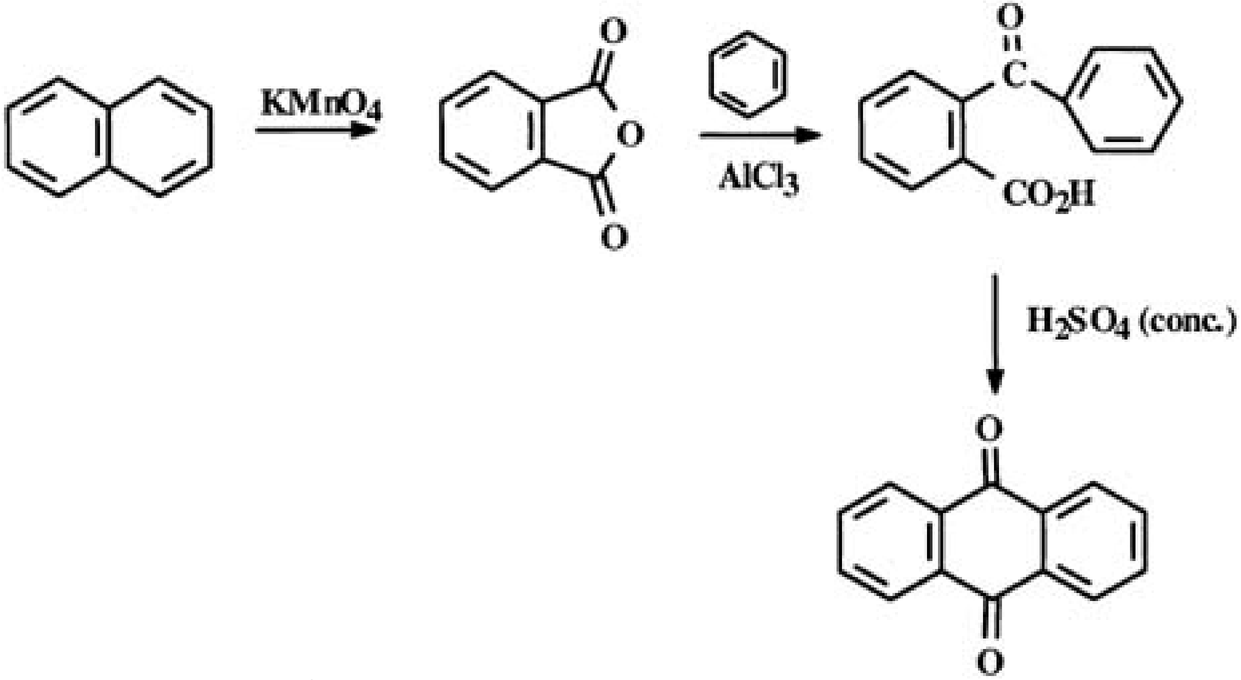
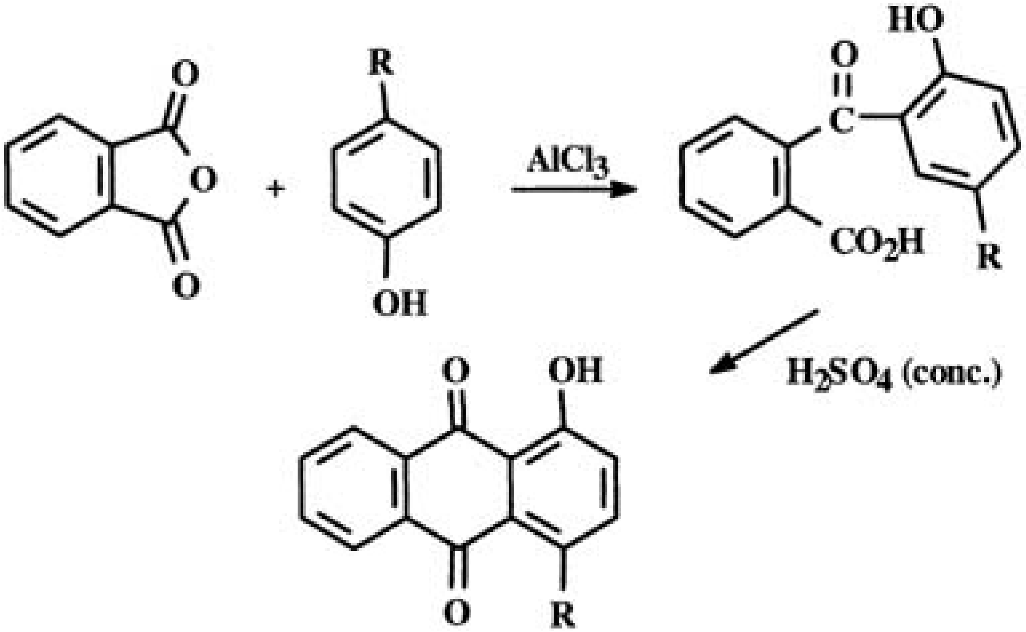


Fig. 23.111 Synthesis of key substituted anthraquinone intermediates, where R = Cl, OH.



The reduction of quinizarin using sodium hydrosulfite produces leuco quinizarin, which, in turn, undergoes condensation with alkyl- or arylamines and reoxidation to produce blue and green disperse and solvent dyes. While chemical oxidation can be used, air oxidation is normally sufficient. See steps (a) and (b) in Fig. 23.112 for the general reaction scheme. The use of boric acid in the reduction step follows the course outlined in Fig. 23.113, where the synthesis of C.I. Solvent Green 3 is given as an example.56

Fig. 23.112 General reaction scheme for the synthesis of 1,4-diaminoanthraquinone dyes.

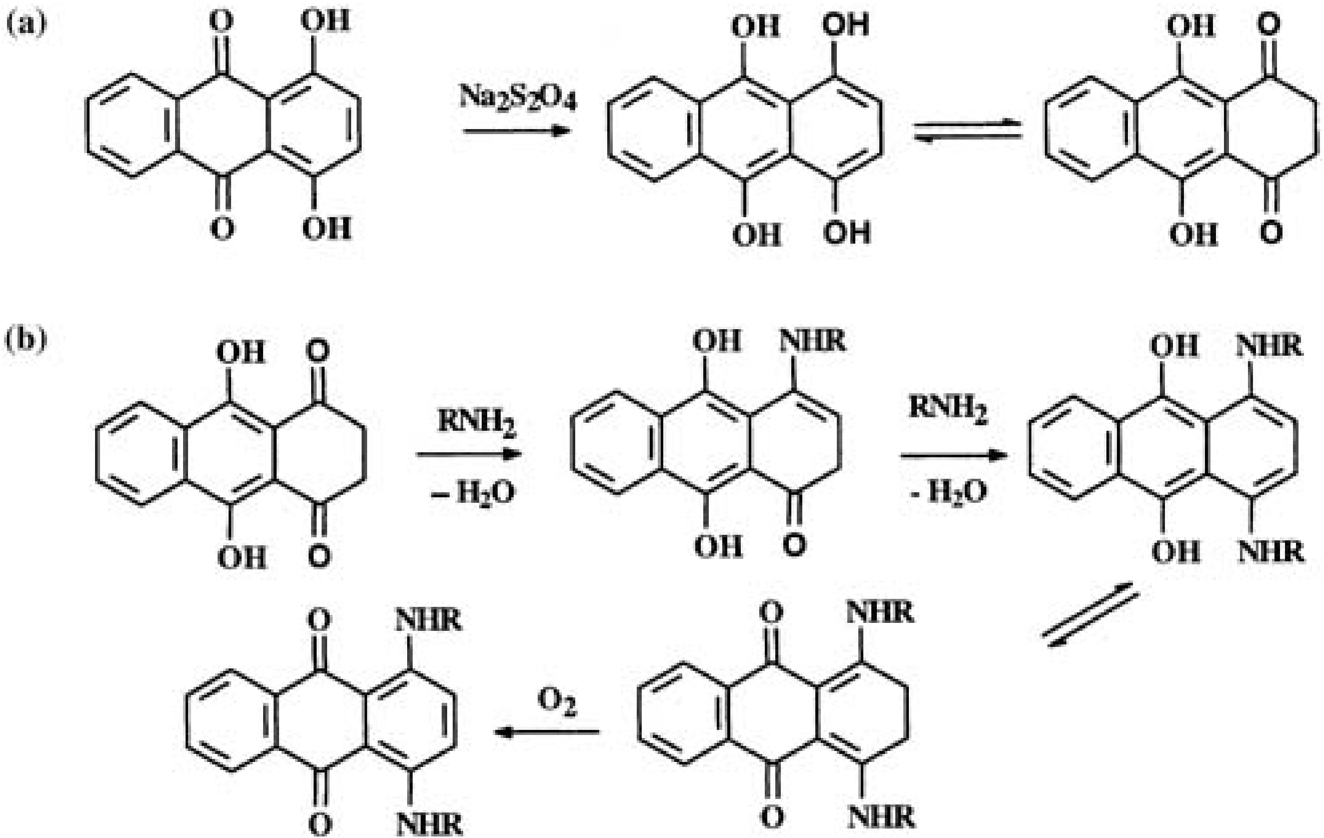
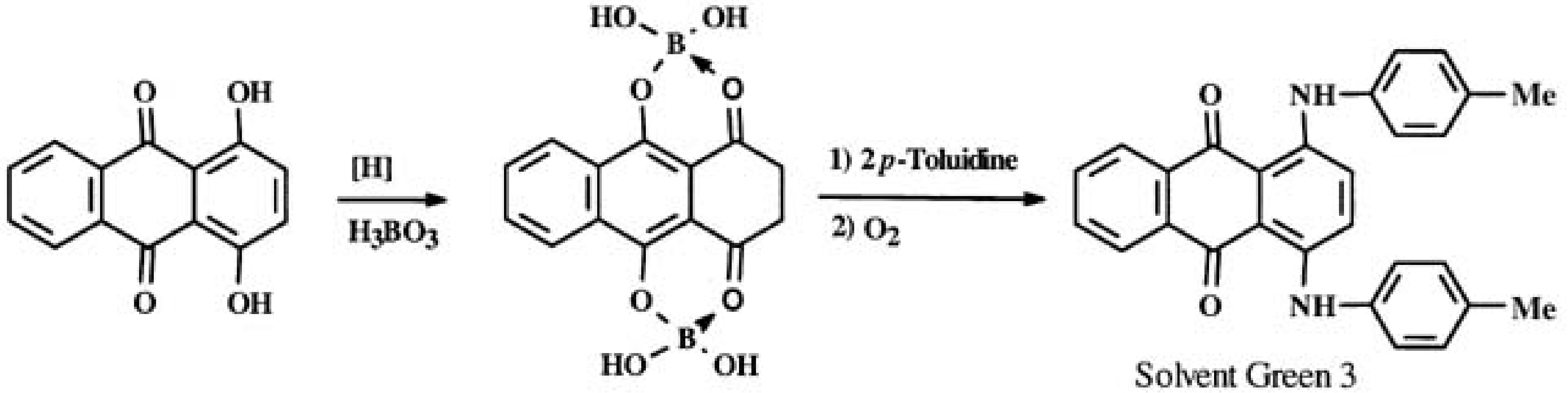


Fig. 23.113 Boric acid catalyzed synthesis of Solvent Green 3.

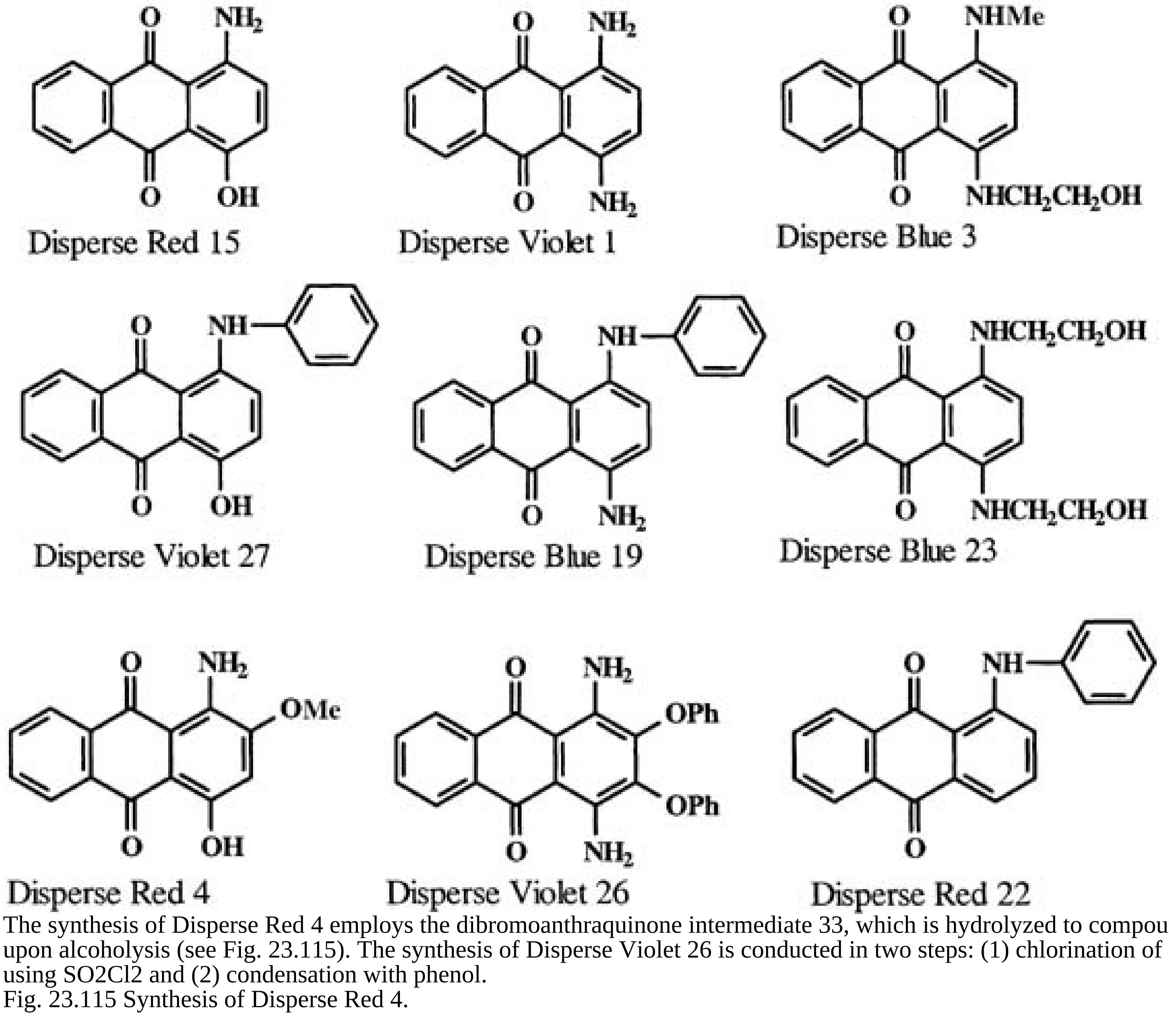


Solvent Green 3

Anthraquinone Disperse Dyes

Examples of dyes prepared using the above methods are shown in Fig. 23.114. The C.I. disperse dyes Red 15, Violet 1, Blue 3, Violet 27, Blue 19, and Blue 23, are prepared from leucoquinizarin. When unsymmetrical dyes such as Disperse Blue 3 are made, the use of a mixture of two amines in the condensation step gives the corresponding symmetrical dyes as by-products. In this case, Disperse Blue 23 would be one of the by-products.

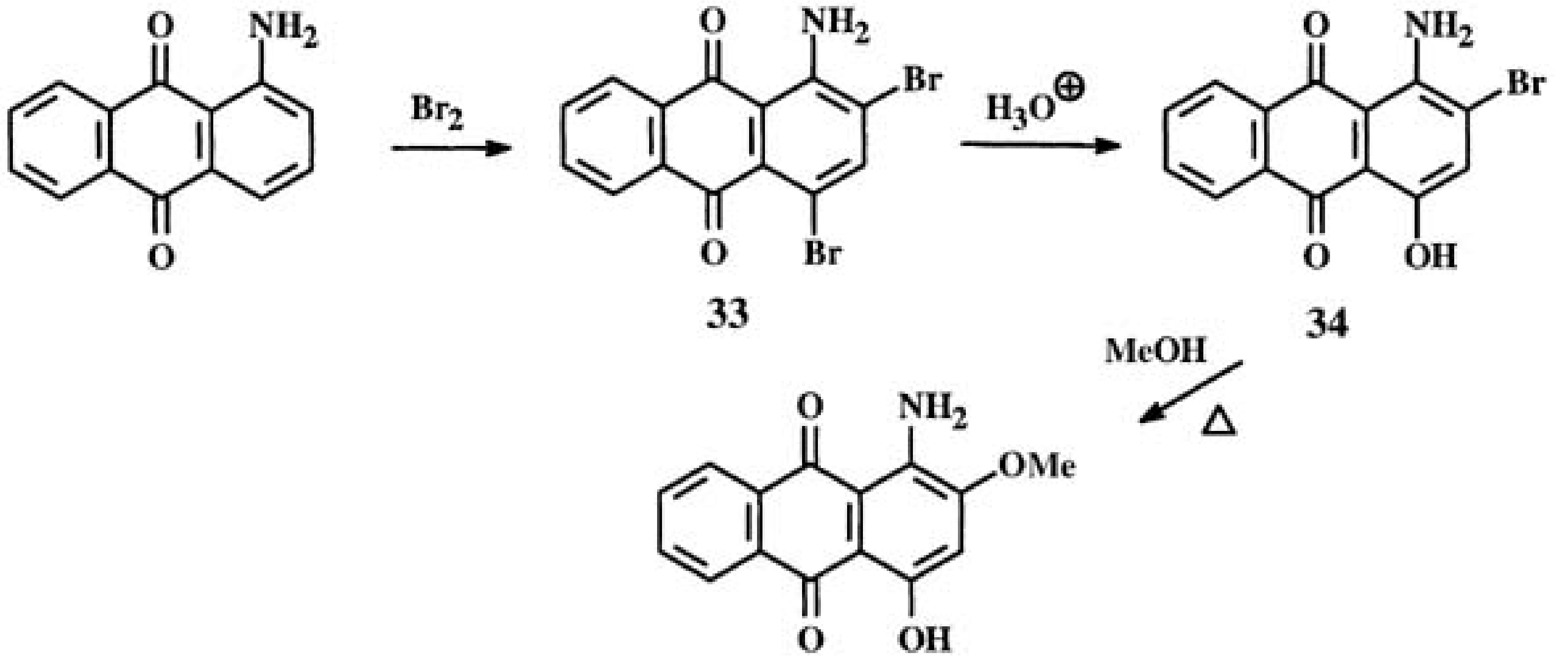
Fig. 23.114 Representative 1,4-disubstituted anthraquinone dyes.



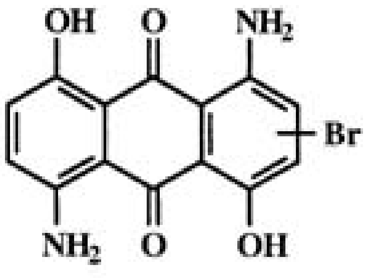
Disperse Red 15
Disperse Violet 1
Disperse Blue 3
Disperse Violet 27
Disperse Blue 19
Disperse Blue 23
Disperse Red 4
Disperse Violet 26
Disperse Red 22

The synthesis of Disperse Red 4 employs the dibromoanthraquinone intermediate 33, which is hydrolyzed to compound 34 and converted to the target dye upon alcoholysis (see Fig. 23.115). The synthesis of Disperse Violet 26 is conducted in two steps: (1) chlorination of Disperse Violet 1 in the 2,3-positions using SO2Cl2 and (2) condensation with phenol.

Fig. 23.115 Synthesis of Disperse Red 4.



Disperse dyes containing substituents in both anthraquinone rings are often prepared from dinitroanthrarufin (DNA) and dinitrochrysazin (DNC), the structures of which are shown in Fig. 23.116. Examples of these dyes are C.I. Disperse Blue 56 and Blue 77. The former dye is made by reduction of DNA followed by bromination, while the latter is made by condensing aniline with DNC. The DNC condensation shows that nitro groups in the α-position can be displaced like a halogen.


Disperse Blue 56

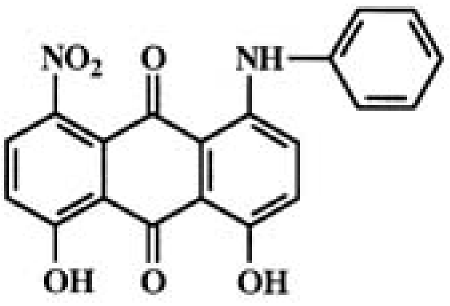
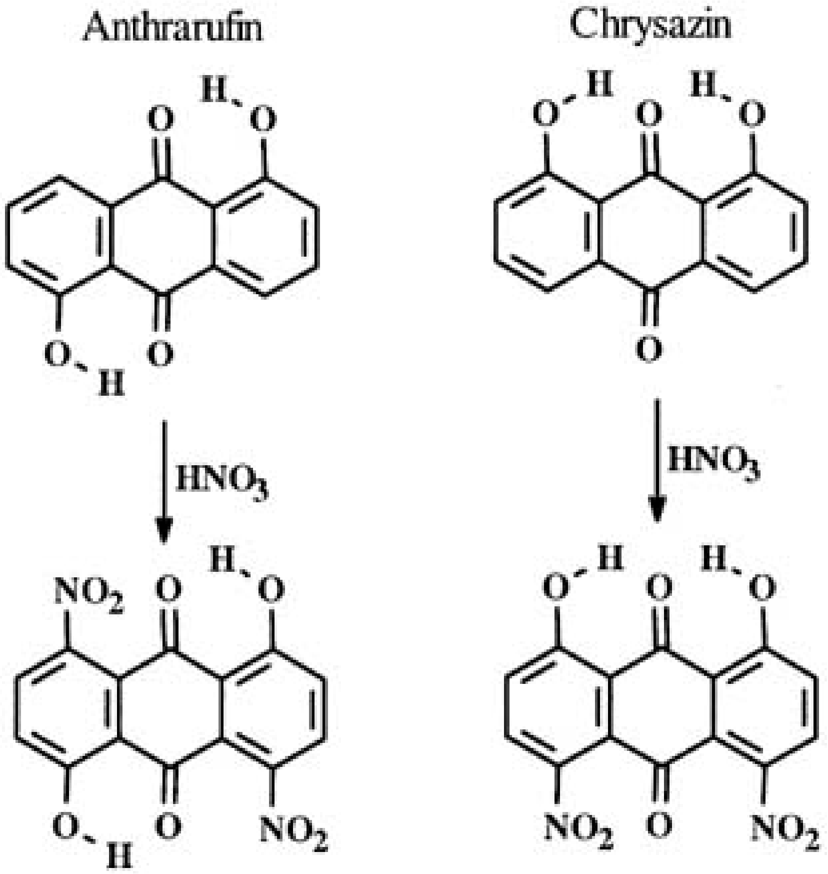

Disperse Blue 77

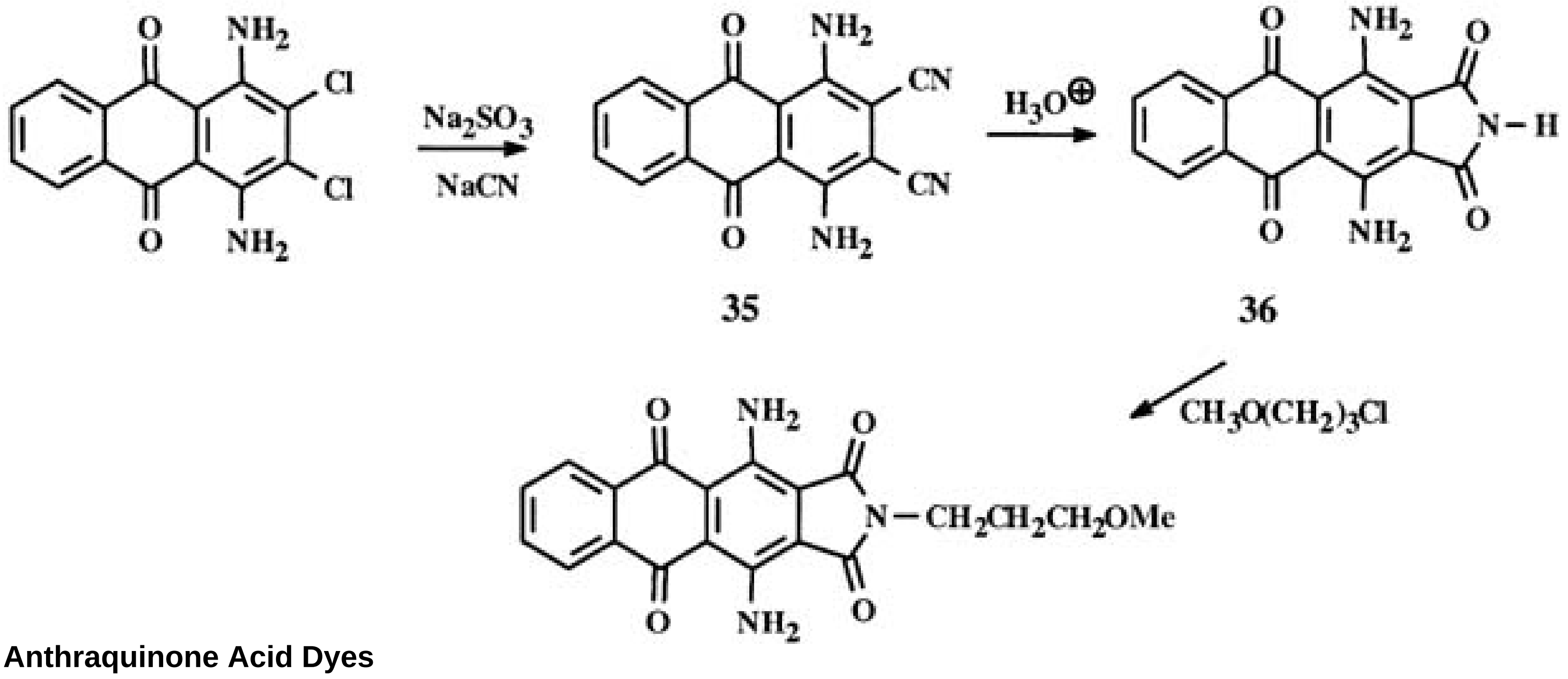
Fig. 23.116 Formation of DNA and DNC.


Anthrarufin

Chrysazin

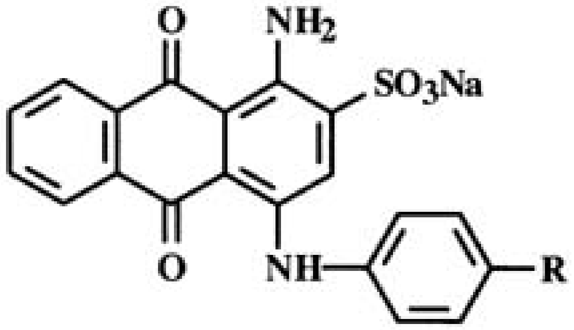
The dichlorinated precursor for Disperse Violet 26 can be used to make turquoise blue dyes such as C.I. Disperse Blue 60, as shown in Fig. 23.117. In this sequence, the chloro groups are replaced by cyano groups, using NaCN, and the resultant intermediate (35) is hydrolyzed to give the corresponding imide (36), which in turn is alkylated to give the target dye.

Fig. 23.117 A three-step synthesis of Disperse Blue 60.



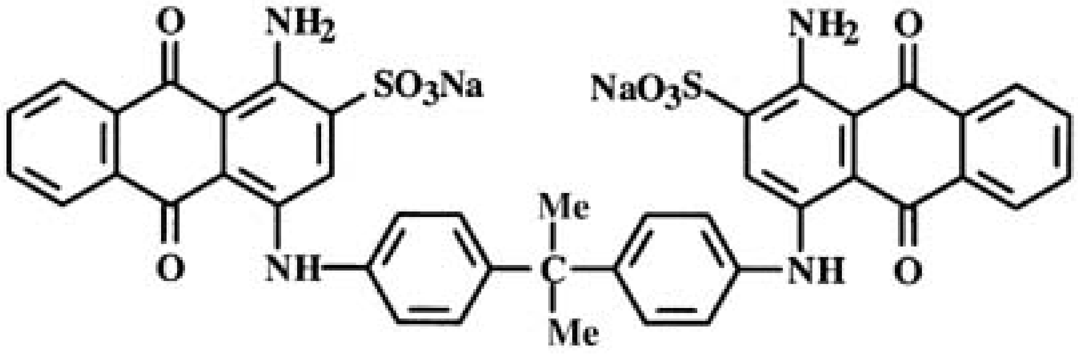
Anthraquinone Acid Dyes

A key intermediate in the synthesis of anthraquinone acid dyes is bromamine acid. This compound is made via the sequence shown in Fig. 23.118. Acid dyes made from this intermediate include C.I. Acid Blue 25, C.I. Acid Blue 40, and C.I. Acid Blue 127.



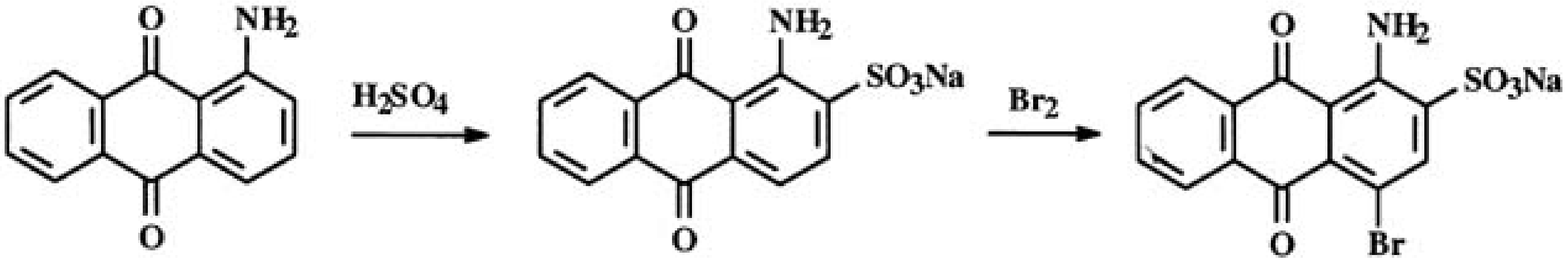
R = H (Acid Blue 25)

R = NHAc (Acid Blue 40)



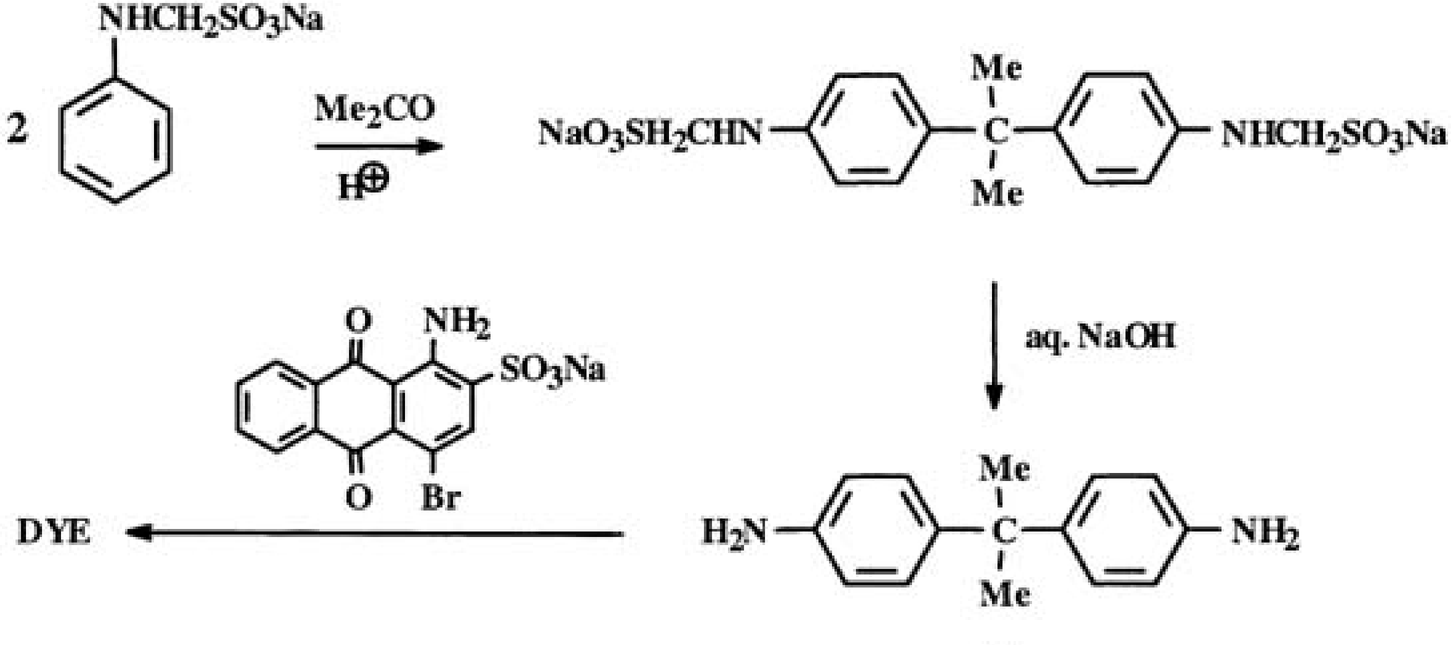
Acid Blue 127

Fig. 23.118 A two-step synthesis of bromamine acid.



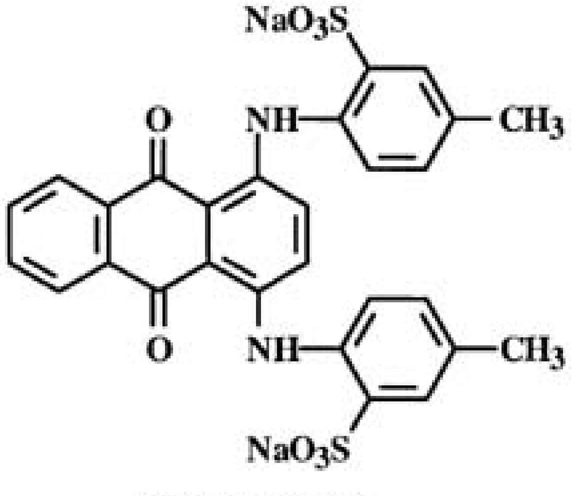
The synthesis of C.I. Acid Blue 127 takes place according to the route shown in Fig. 23.119. A key step in the synthesis is the formation of diamine 37, which is produced in two steps from *N*-sulfomethylaniline:(1) condensation with acetone and (2) hydrolysis to remove the protecting group. At this point, one molecule of diamine 37 is condensed with two molecules of bromamine acid to form the dye.

Fig. 23.119 Synthesis of Acid Blue 127.



37

Another important dye is C.I. Acid Green 25. This dye is made by the sulfonation of C.I. Solvent Green 3 (Fig. 23.113). Since the benzene rings are more reactive than the anthraquinone system, sulfonation occurs there preferentially.

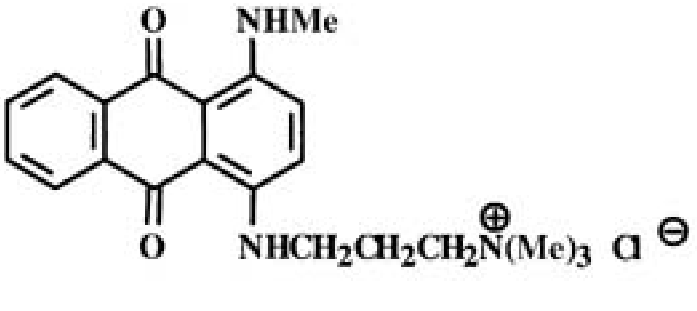


Acid Green 25

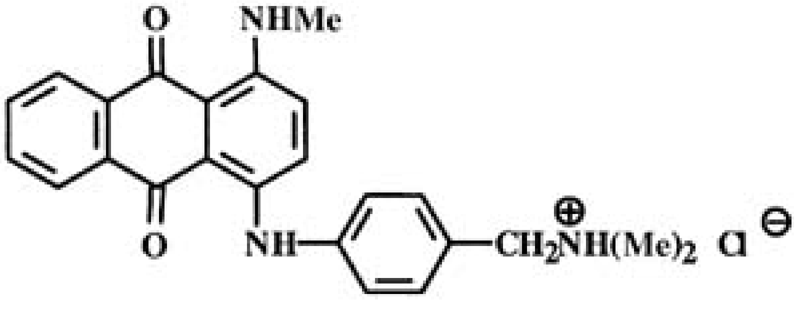
Anthraquinone Basic Dyes

Dyes of this type include C.I. Basic Blue 22 and Basic Blue 47. The synthesis of Basic Blue 22 is shown in Fig. 23.120, as an example of the type chemistry required. The sequence begins with the preparation of *N*, *N*-dimethylpropylenediamine, which in turn is combined with methylamine and condensed with leucoquinizarin. Oxidation gives the key intermediate 38, which is alkylated using methyl chloride to produce the dye.

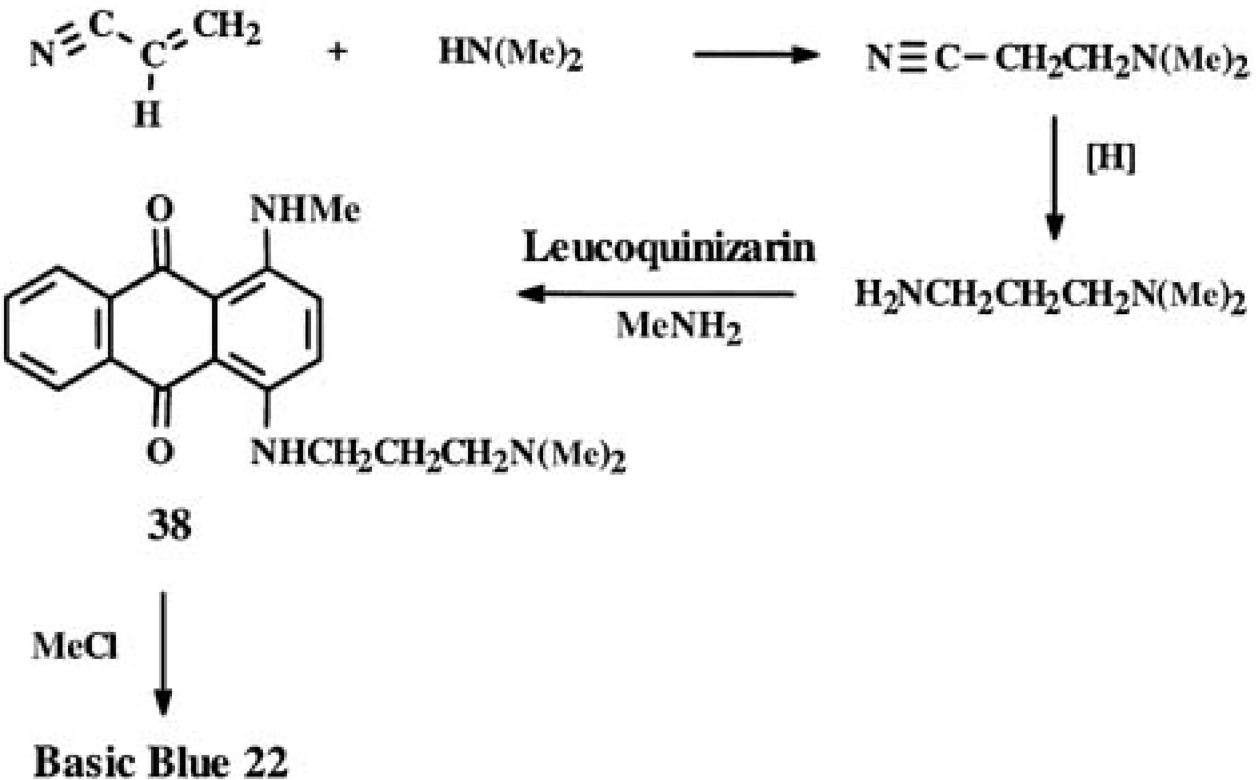
Fig. 23.120 Synthesis of Basic Blue 22.



Basic Blue 22



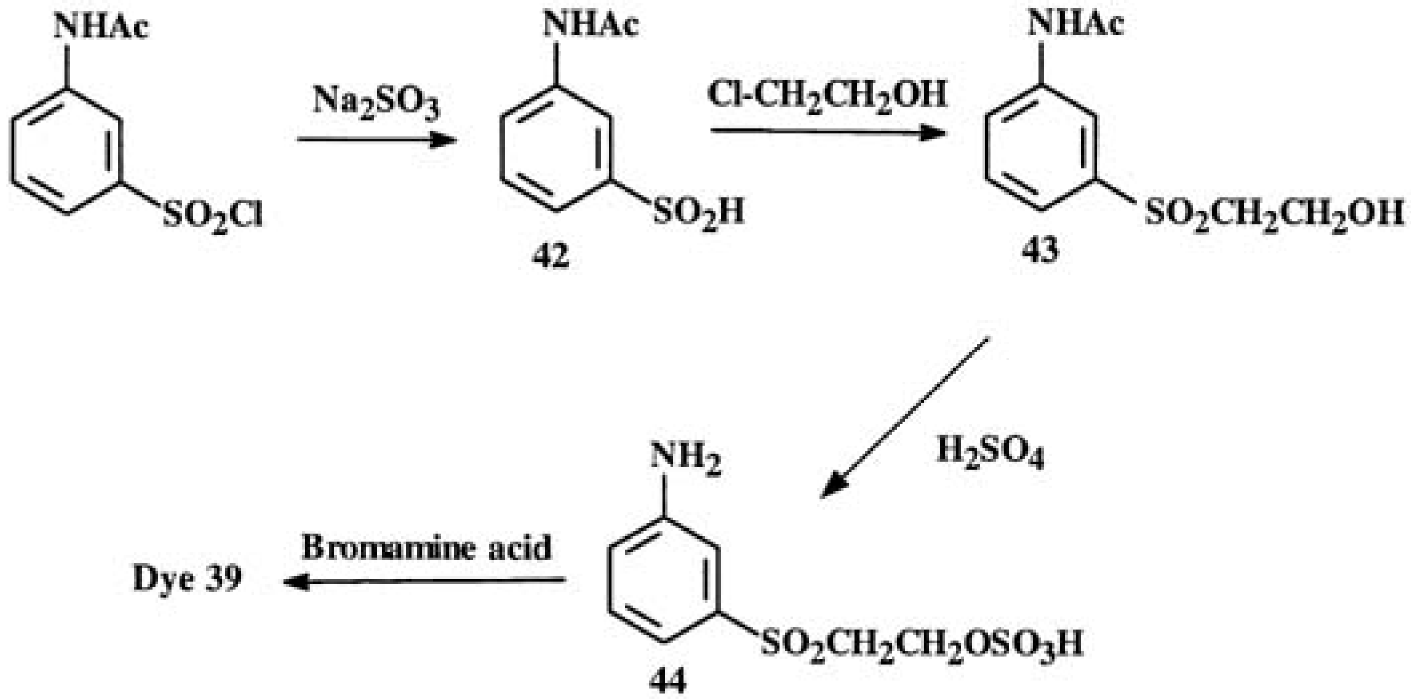
Basic Blue 47



Anthraquinone Reactive Dyes

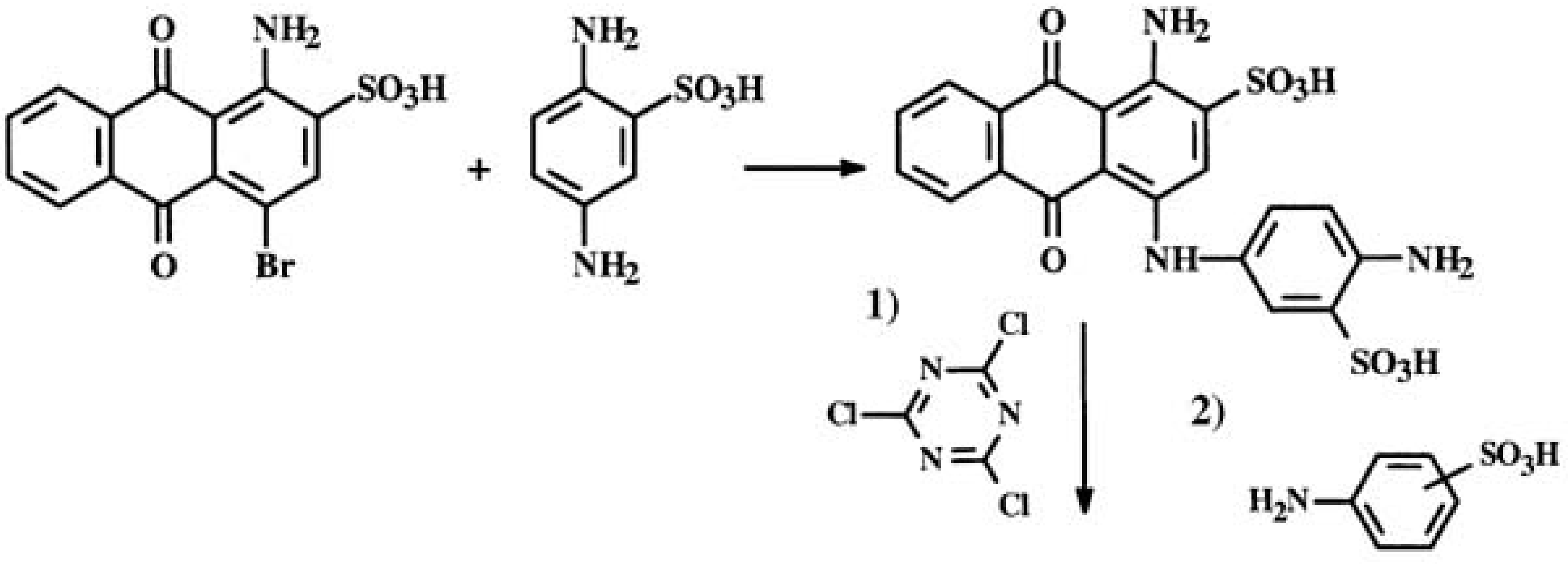
Three examples of dyes of this type are C.I. Reactive Blue 19 (39), Reactive Blue 2 (40), and Reactive Blue 4 (41). All three dyes can be synthesized by condensing the appropriate arylamine with bromamine acid. In the case of the high volume dye Reactive Blue 19, arylamine 44 is the key intermediate, and its synthesis is shown in Fig. 23.121. Chlorosulfonation and then reduction of the intermediate sulfonyl chloride produce the sulfinic acid 42. Alternatively, the reduction step can be conducted with Na2S2O4. Alkylation of the sulfinic acid with 2-chloroethanol or ethylene oxide (a more toxic agent) produces compound 43. Treatment of this compound with hot H2SO4 gives simultaneous hydrolysis of the acetamido (NHAc) group and sulfonation of the hydroxyethyl (CH2CH2OH) group to give key intermediate 44.

Fig. 23.121 Synthesis of key dye intermediate 44 and Reactive Blue 19 (39).

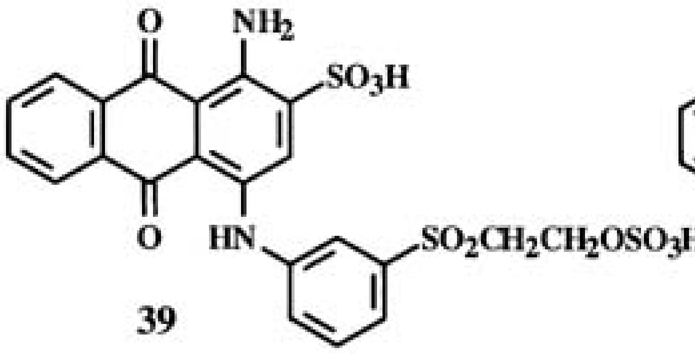


Similarly, dyes 40 and 41 are prepared by condensing 2,5-diaminobenzenesulfonic acid with bromamine acid, which reacts first at the less hindered amino group, followed by a reaction with cyanuric chloride to introduce the reactive group. These steps produce dye 41 and dye 40 is formed by reacting 41 with a mixture of sulfonated anilines (see Fig. 23.122).

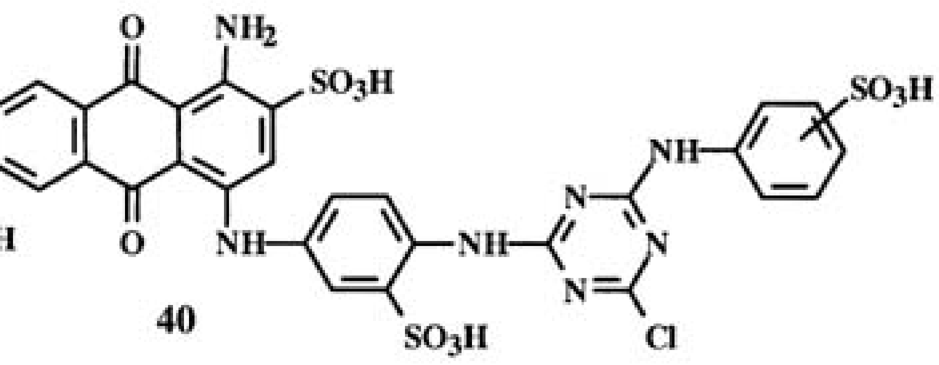
Fig. 23.122 Synthesis of chlorotriazine-based anthraquinone reactive dyes.



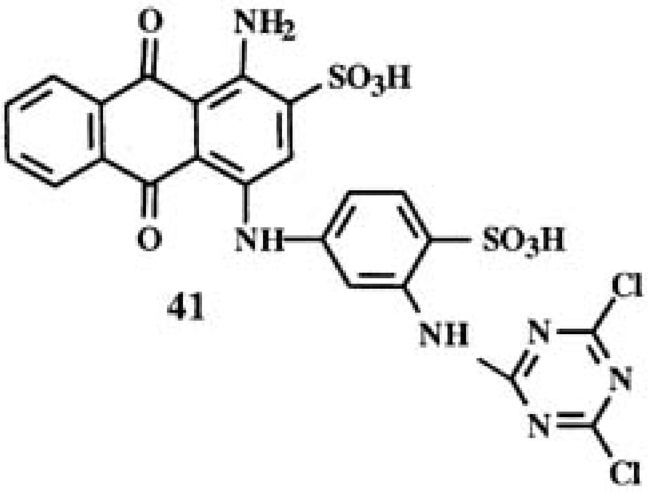
Reactive Blue 2



39



40



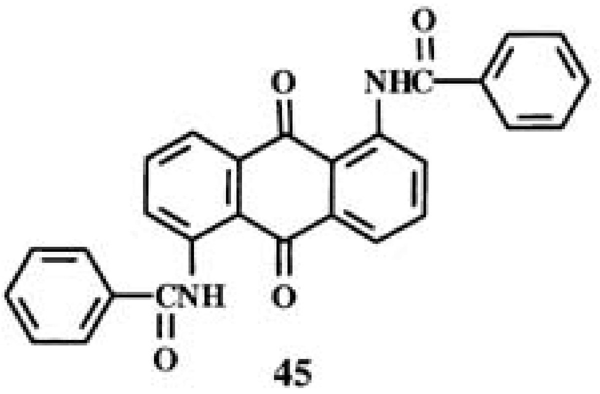
41

Vat Dyes

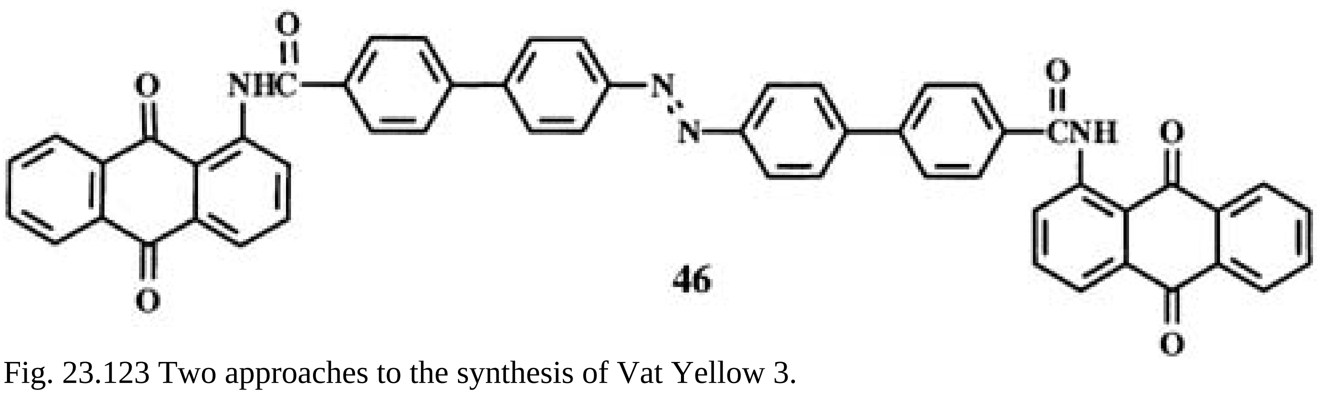
The synthesis of vat dyes covers the full gamut of simple to complex chemistry. We have chosen examples to illustrate the broad spectrum of possible structures and synthetic methods. Emphasis is placed on anthraquinone vat dyes, since they dominate the number of commercial dyes.

Anthraquinone

The simplest anthraquinone vat dyes are benzoylated amines such as C.I. Vat Yellow 3 (45) and Vat Yellow 33 (46). The syntheses are shown in Figs 23.123 and 23.124.



45



46

Fig. 23.123 Two approaches to the synthesis of Vat Yellow 3.

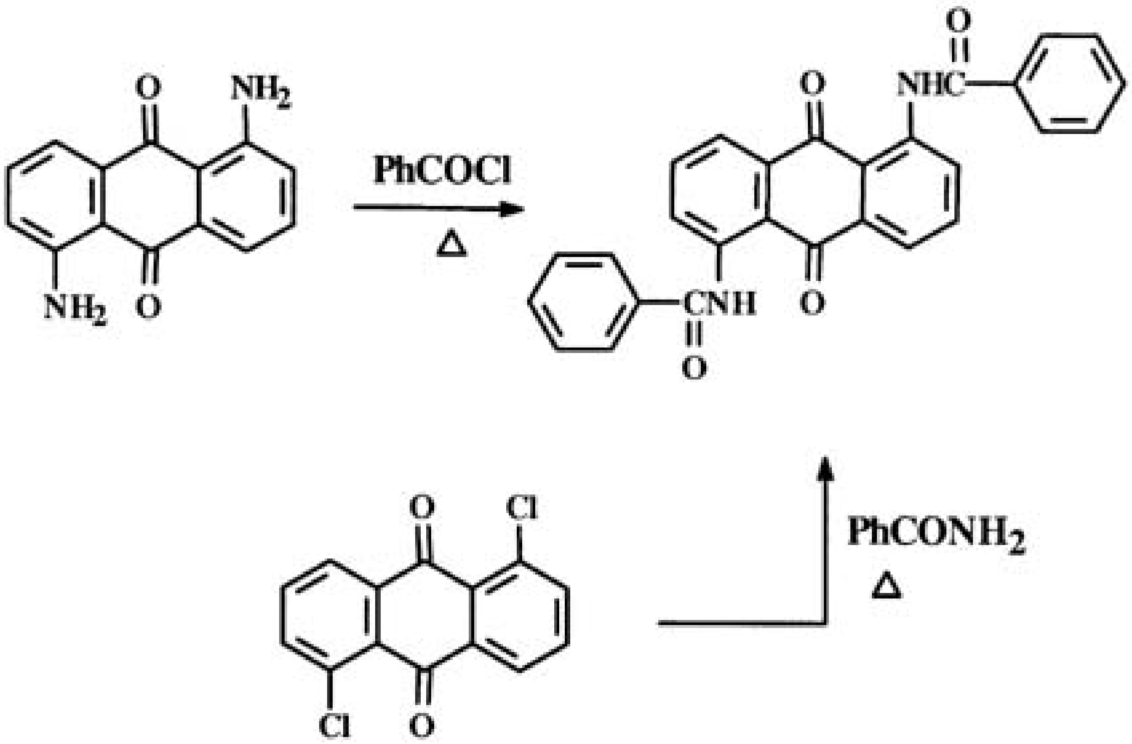
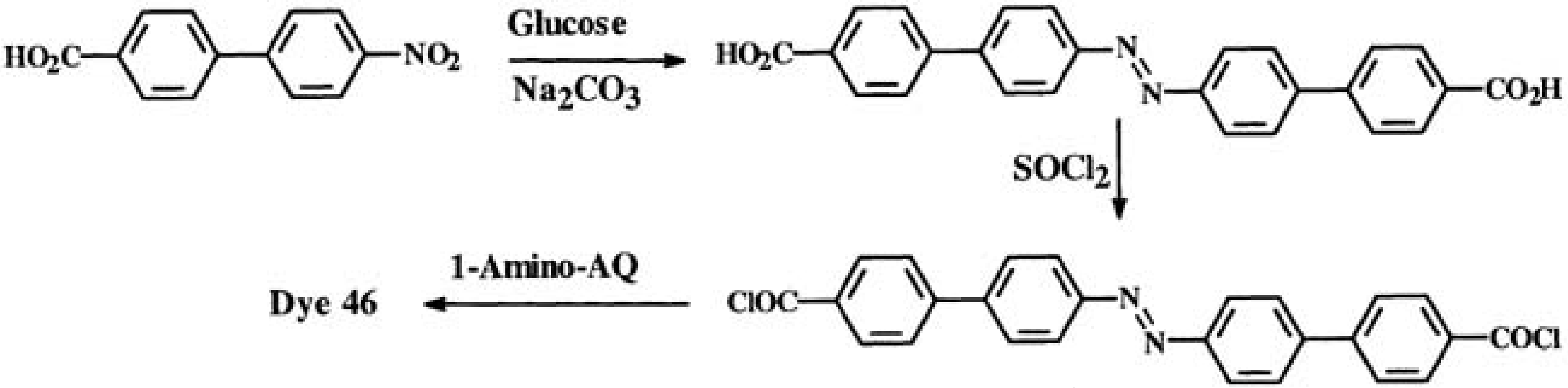
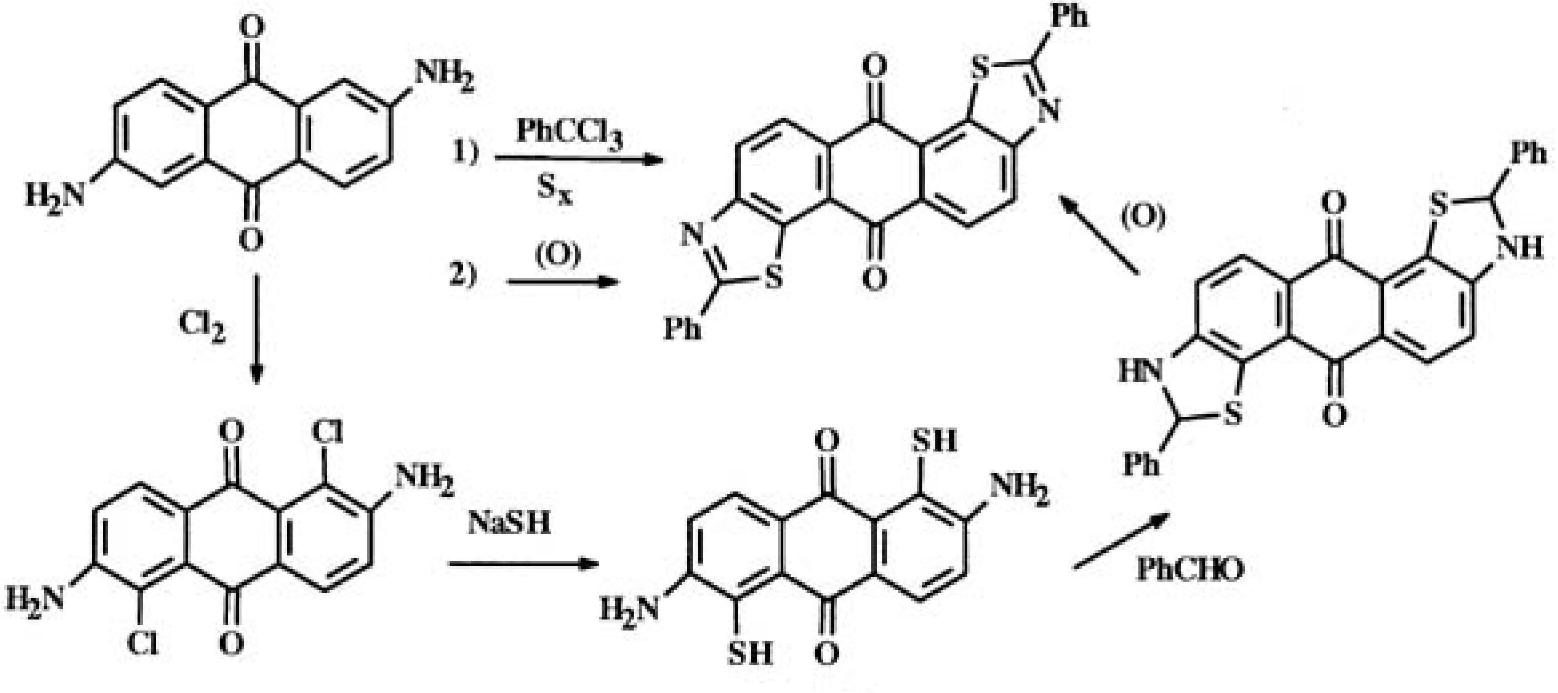


Fig. 23.124 Synthesis of Vat Yellow 33, where AQ = anthraquinone.



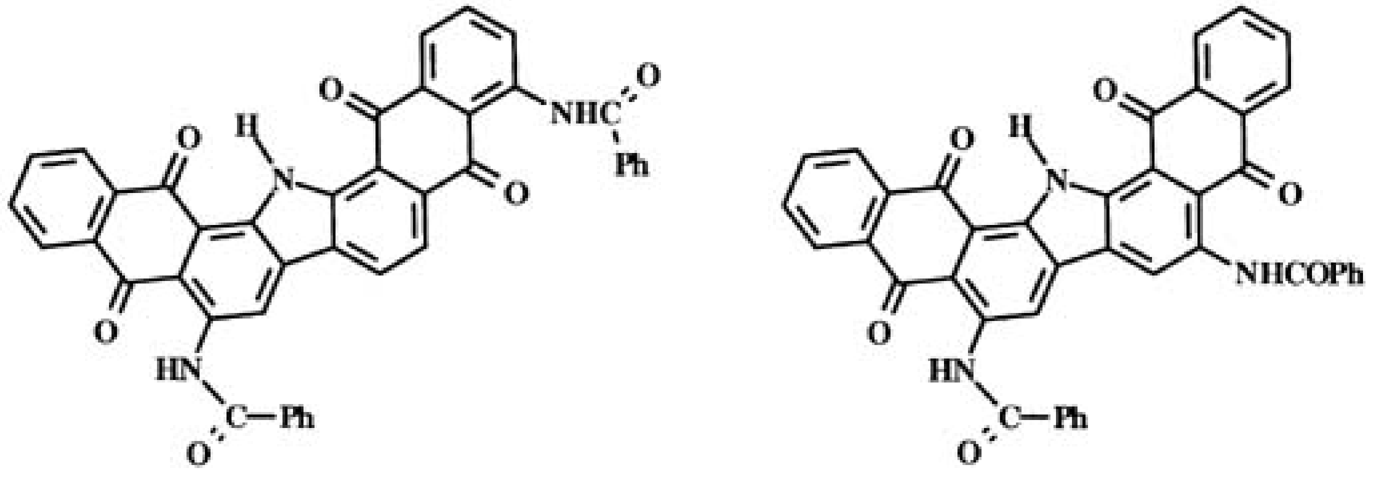
Anthraquinone vat dyes containing a thiazole ring include C.I. Vat Yellow 2, the synthesis of which is shown in Fig. 23.125. In this case, at least two approaches are possible. In the first, 2,6-diaminoanthraquinone is condensed with benzotrichloride in the presence of sulfur and the initial product is oxidized without isolation to give the target dye. Alternatively, the starting diamine can be chlorinated and converted to the corresponding dithiol (47). At this point condensation with benzaldehyde followed by oxidation (e.g., air or dichromate) gives the dye.

Fig. 23.125 Alternatives to the synthesis of Vat Yellow 2.



47

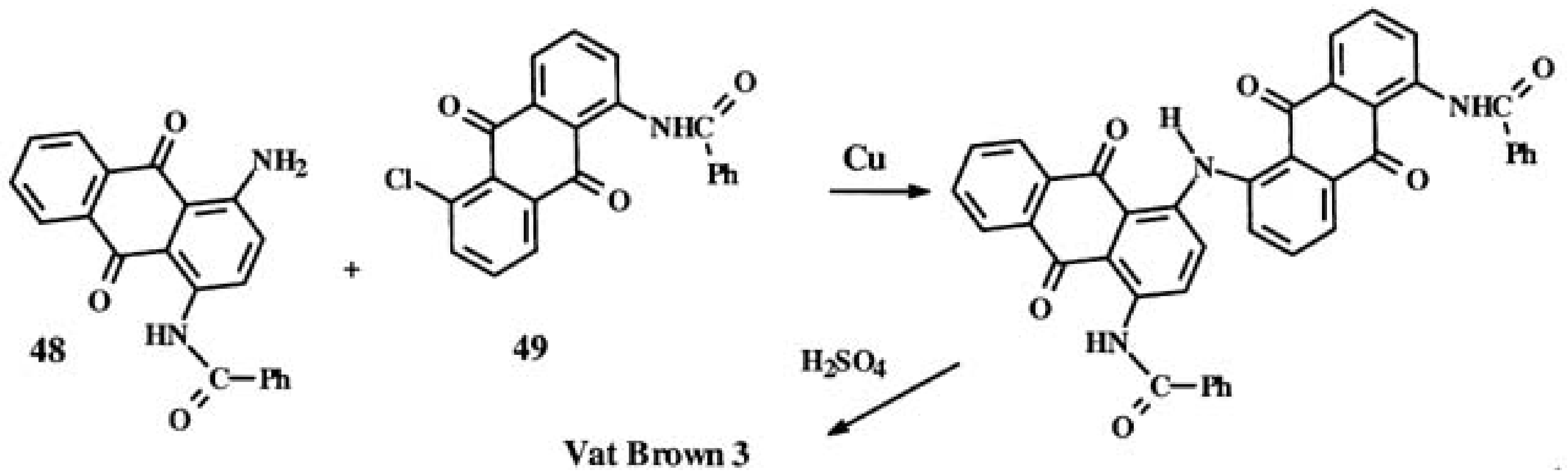
Important vat dyes containing a carbazole moiety include C.I. Vat Brown 3 and Vat Black 27. These dyes are made according to the method shown in Fig. 23.126 for Vat Brown 3. The synthesis employs an Ullmann type condensation reaction between compounds 48 and 49 followed by acid-induced cyclization using H2SO4.



Vat Brown 3

Vat Black 27

Fig. 23.126 Synthesis of carbazole-based anthraquinone vat dyes.



Vat dyes that do not contain all of the elements of the anthraquinone moiety include benzanthrone-based vat dyes such as C.I. Vat Orange 1 and Vat Green 1, which are made according to the routes shown in Figs 23.127 and 23.128. The first synthesis is a three-step process: (1) dibenzoylation of naphthalene, (2) Lewis acid-induced cyclization to the benzanthrone system, and (3) dibromination. The second synthesis is a four-step process: (1) oxidative-coupling of benzanthrone in the presence of alkali to give compound 50, (2) H_2SO_4-induced ring closure to give compound 51, (3) reduction to compound 52, and (4) methylation to give the target dye.

Fig. 23.127 Three-step synthesis of Vat Orange 1.

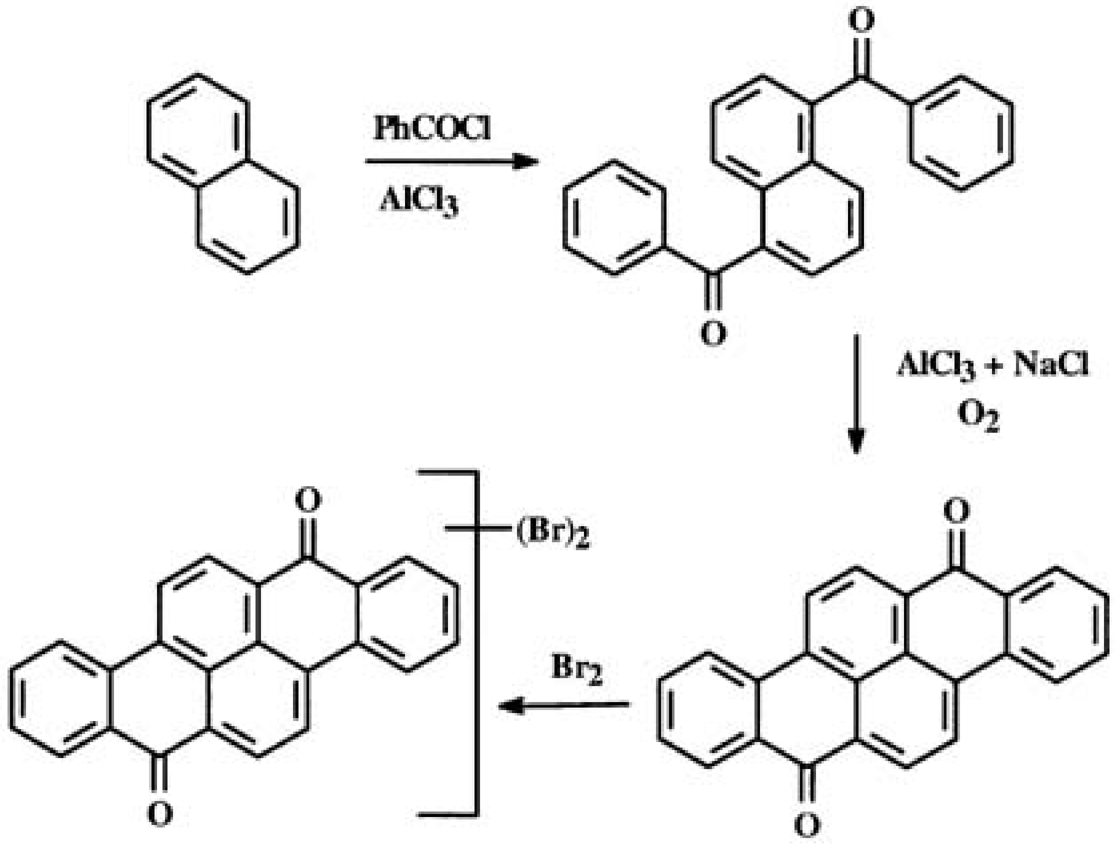
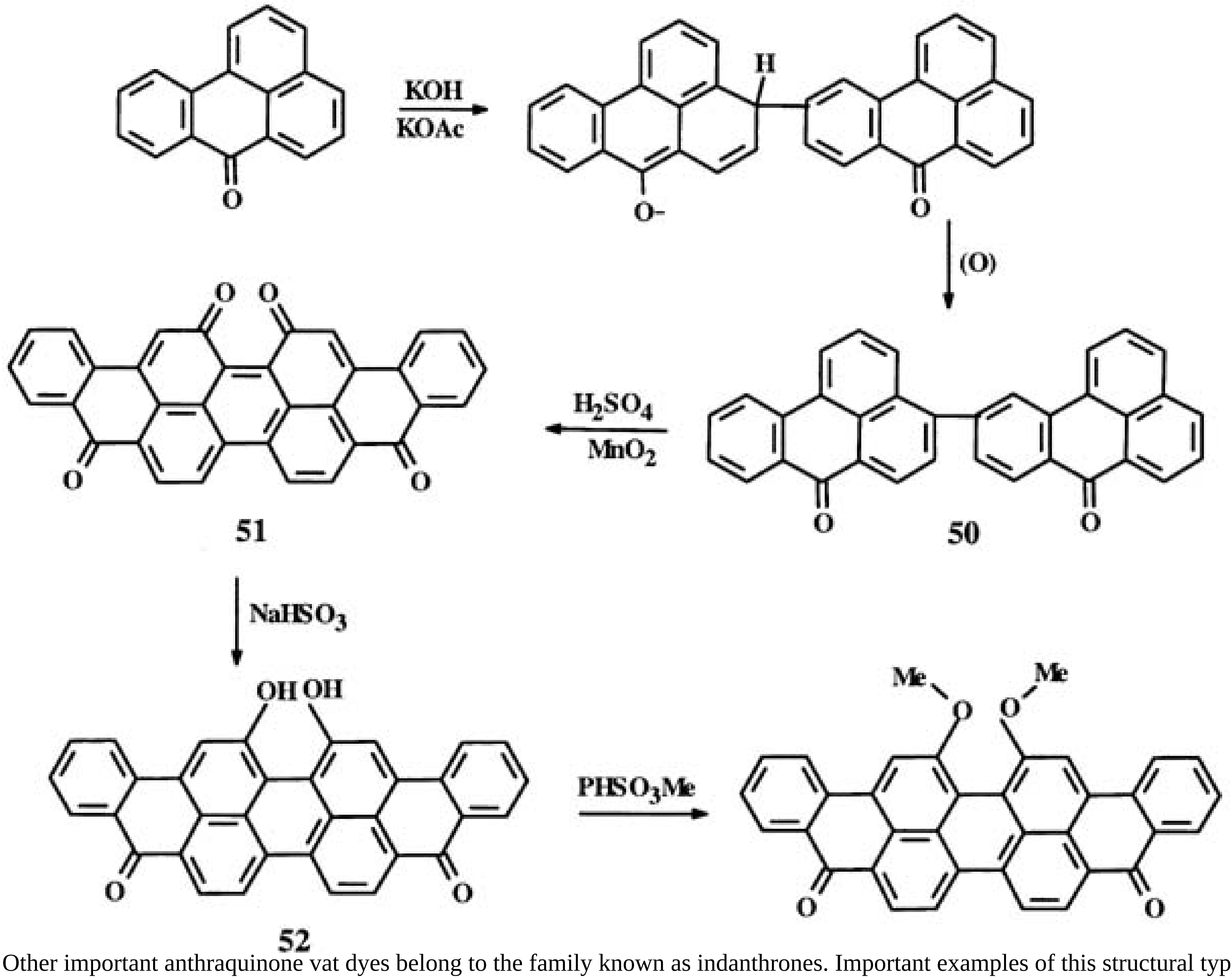
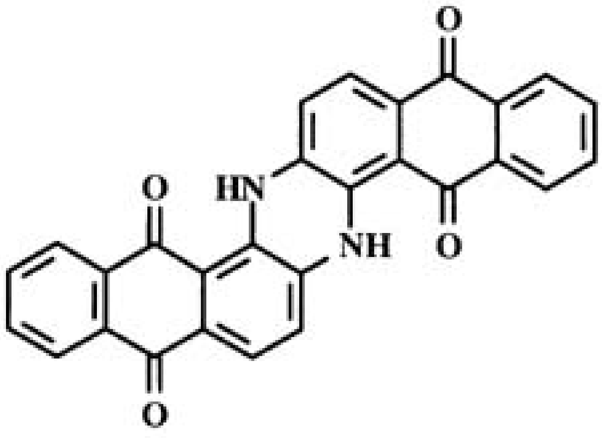
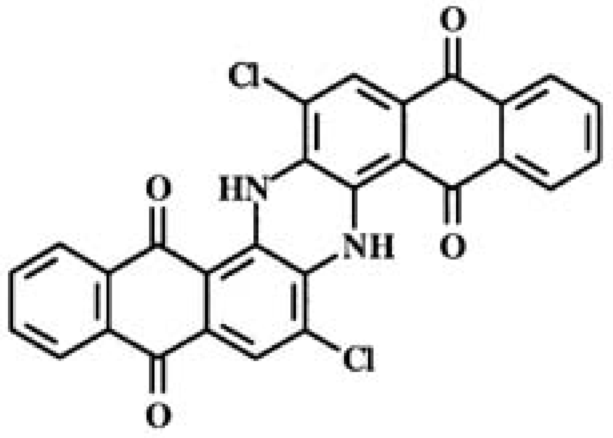


Fig. 23.128 Synthesis of Vat Green 1 from benzanthrone.



Other important anthraquinone vat dyes belong to the family known as indanthrones. Important examples of this structural type are C.I. Vat Blue 4 and Vat Blue 6. Vat Blue 4 is made by heating 1-amino or 2-aminoanthraquinone at 220230°C in a KOH/H_2O mixture. The Vat Blue 6 synthesis is a much longer process that requires the synthesis of 2-chloro-3-aminoanthraquinone.57 The resultant amine is brominated and converted to the target dye via an Ullmann reaction.

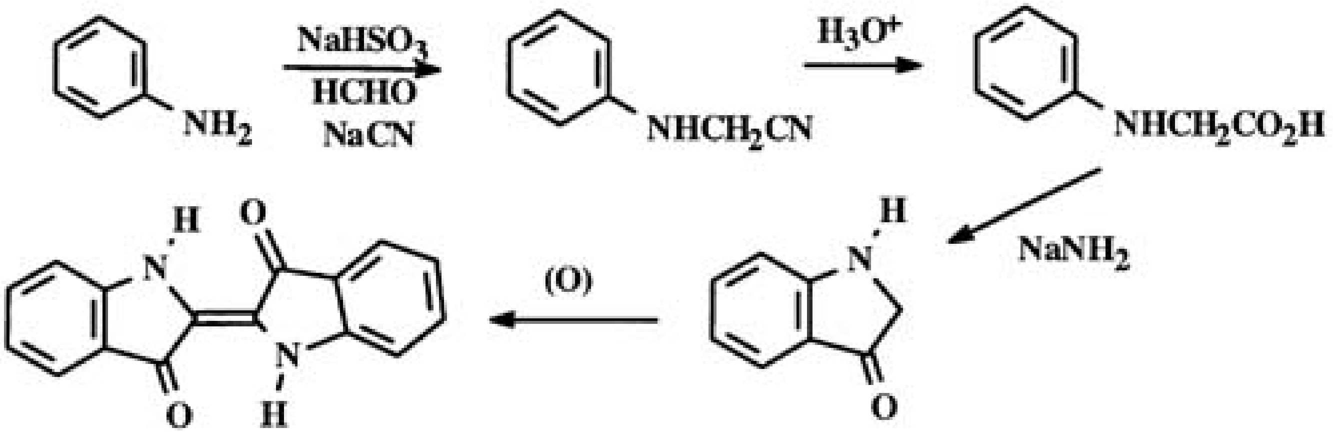

Vat Blue 4


Vat Blue 6

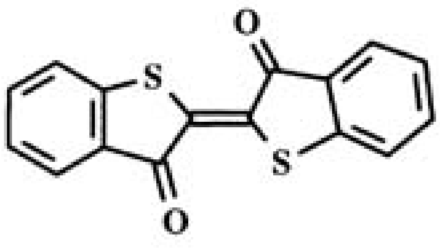
Indigoid and Thioindigoid

By far the most important member of these vat dye families is C.I. Vat Blue 1 (indigo). Its synthesis can be achieved via the four-step method shown in Fig. 23.129. The method shown is known as the HeumannPfleger synthesis,58 where the key intermediate, *N*-carboxymethylaniline, is fused with $NaNH_2$. The cyclic product of the fusion step undergoes air oxidation to give indigo.

Fig. 23.129 Commercial preparation of synthetic indigo.



Thioindigoids are similarly prepared, in that the synthesis of carboxymethyl intermediates is conducted (Fig. 23.130). The resultant cyclic ketones are much less air sensitive, making oxidation with a chemical agent important. However, this also means that unsymmetrical thioindigoid systems can be synthesized. While many have been made, few are in commerce today. Examples are C.I. Vat Red 1 and Vat Red 41.


Vat Red 41

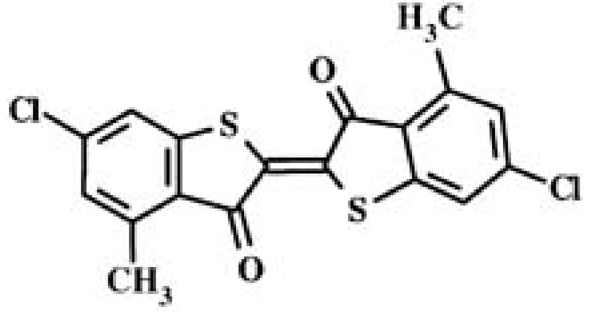
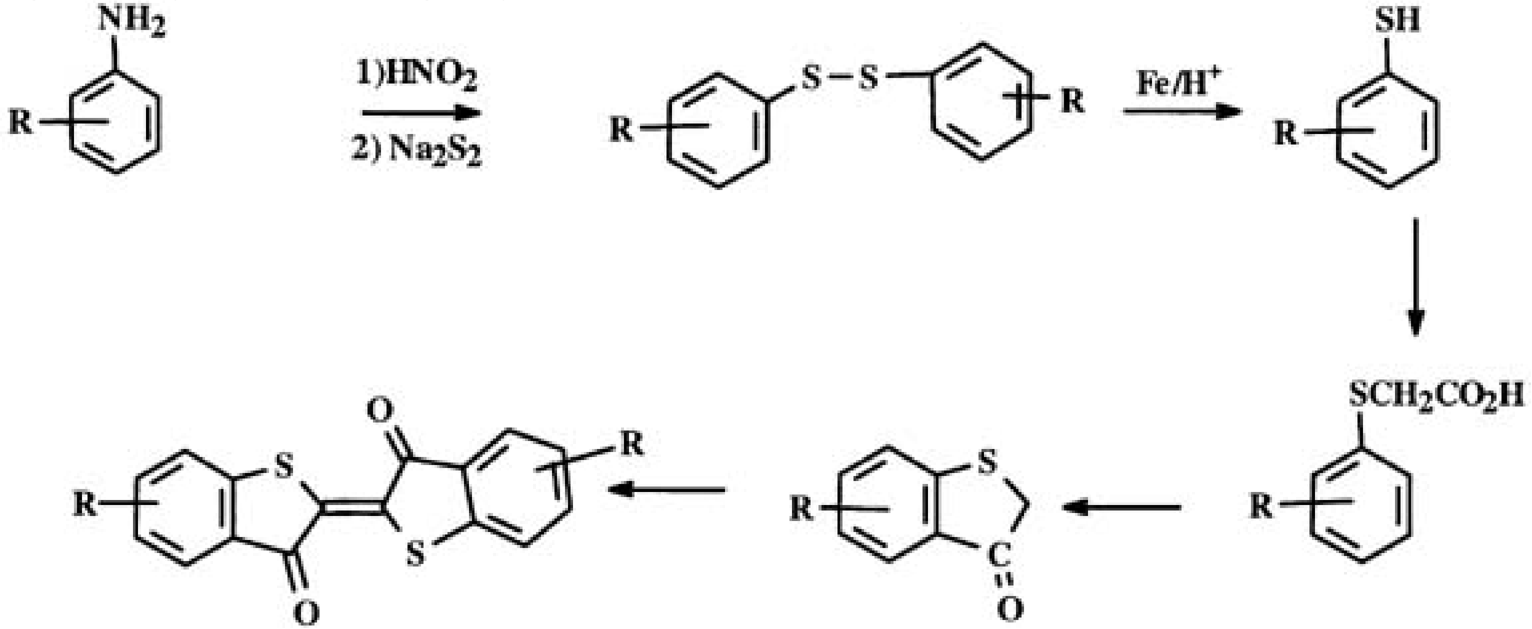

Vat Red 1

Fig. 23.130 Synthesis of thioindigoid dyes.



Sulfur Dyes

In the section "Sulfur Dyes" under "Classification of Dyes" we mentioned that sulfur dye chemistry, while quite old, is still much less well-defined than for the other classes of dyes. It is clear, however, that many sulfur dyes are produced by the sulfur bake process and that compounds containing the benzothiazole group (e.g., 53) are formed in route to the final dyes. For instance, the synthesis of C.I. Sulfur Yellow 4 follows a course of the type outlined in Fig. 23.131. In this regard, heating a mixture of *para*-toluidine and sulfur produces a 2(*para*-aminophenyl)benzothiazole. The sulfur bake process has also been used to make C.I. Sulfur Orange 1, where benzothiazone intermediate 54 is produced along the way59 (see Fig. 23.132).

Fig. 23.131 Synthesis of Sulfur Yellow 4 by the sulfur bake process.

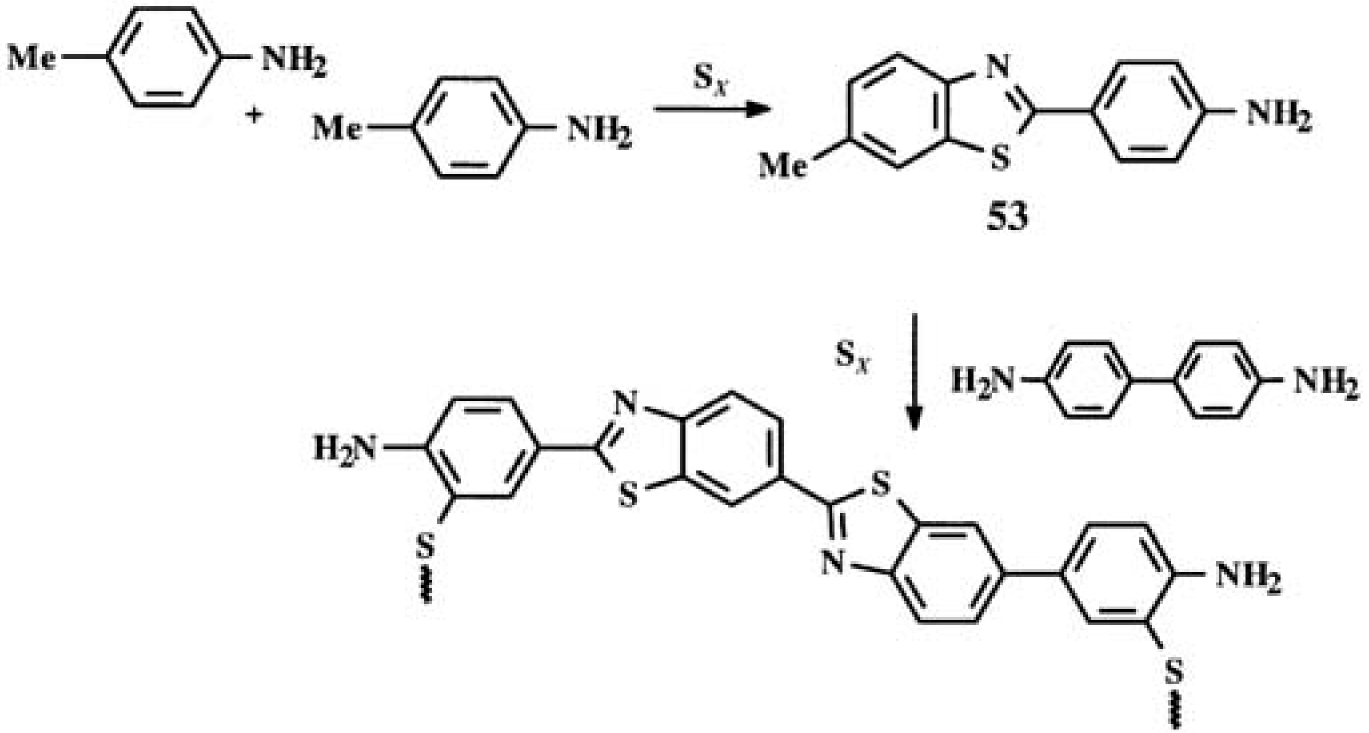
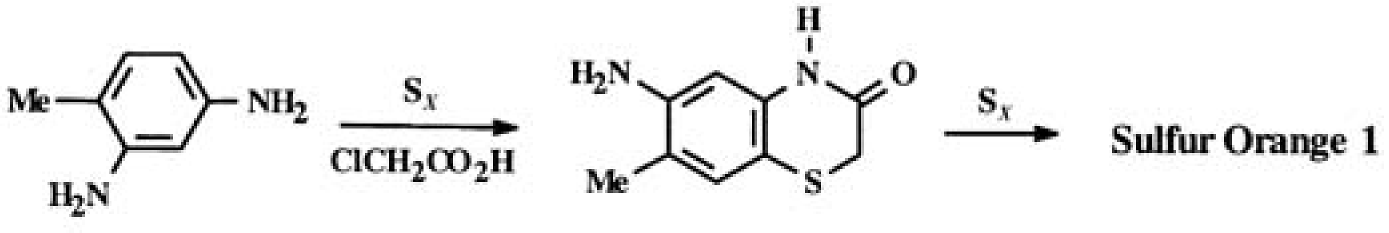


Fig. 23.132 Synthesis of Sulfur Orange 1 by the sulfur bake process.



54

Sulfur blue dyes are often made using an organic solvent such as n-butanol, in what is known as the solvent reflux process. Examples are C.I. Sulfur Blue 9 and Sulfur Blue 13. In this case, intermediate structures are indophenols (e.g., 55) (see Fig. 23.133). Similarly, sulfur dyes containing benzothiazine groups can be made from tetrahalogenated benzoquinones (see Fig. 23.134).

Fig. 23.133 Synthesis of blue dyes by the solvent reflux process.

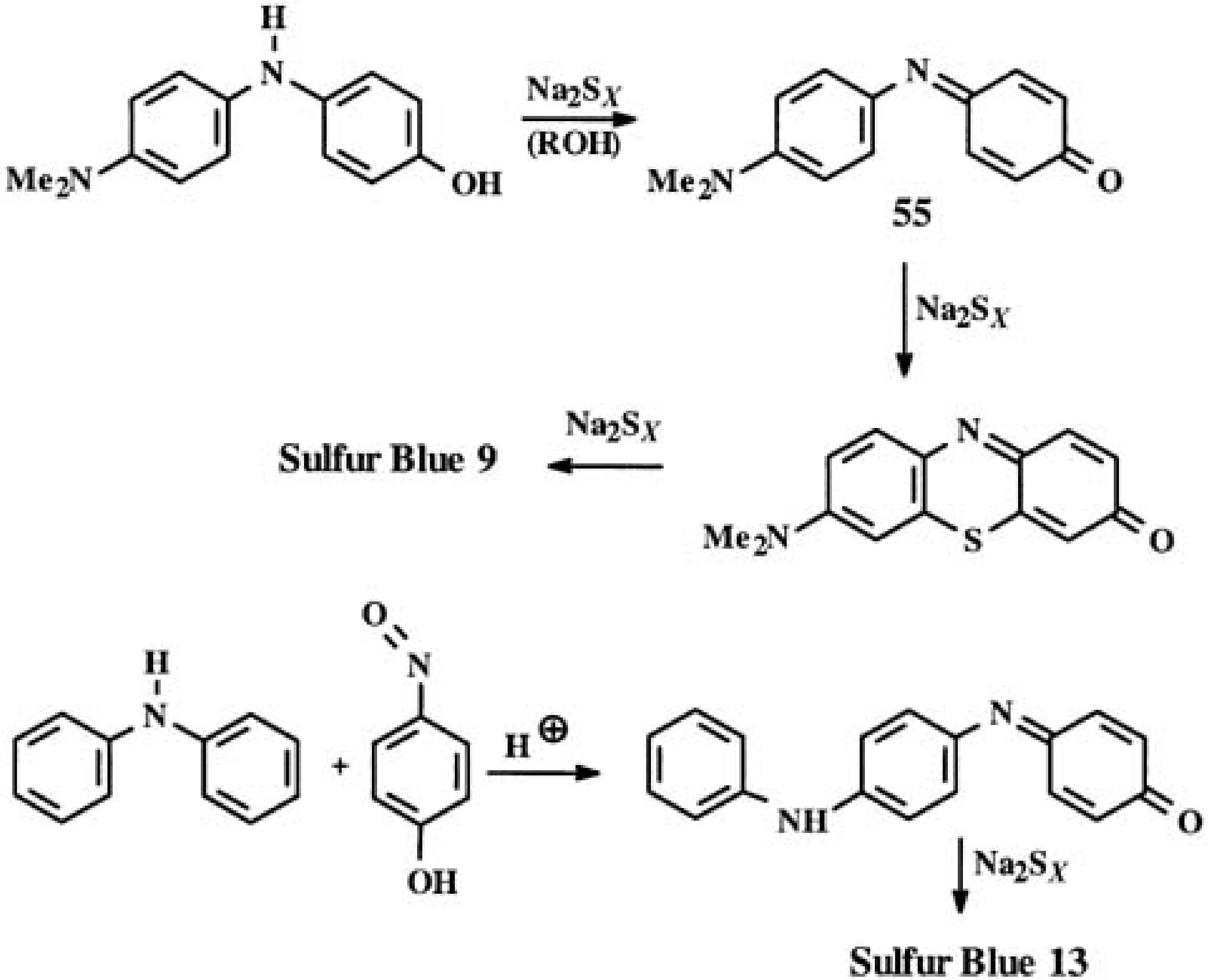
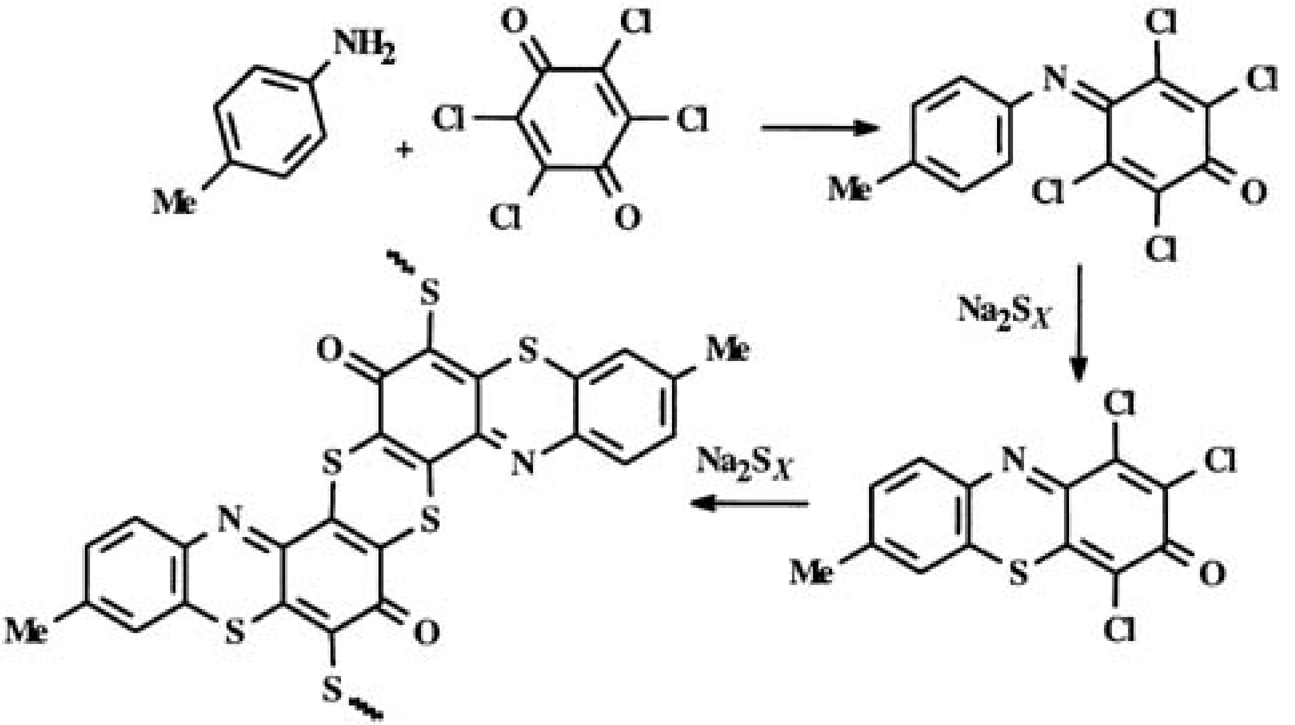
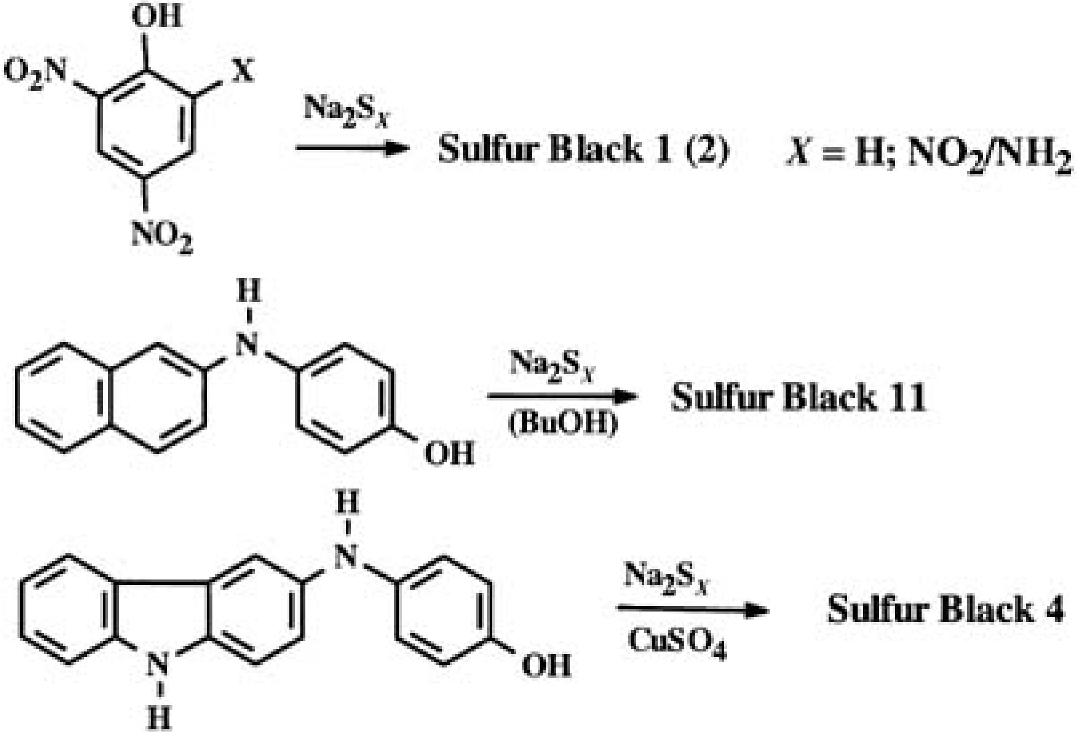


Fig. 23.134 Sulfur dye synthesis from tetrachlorobenzoquinone.



Sulfur black dyes are synthesized according to the methods shown in Fig. 23.135. In these examples sodium polysulfide is the sulfurizing agent employed.

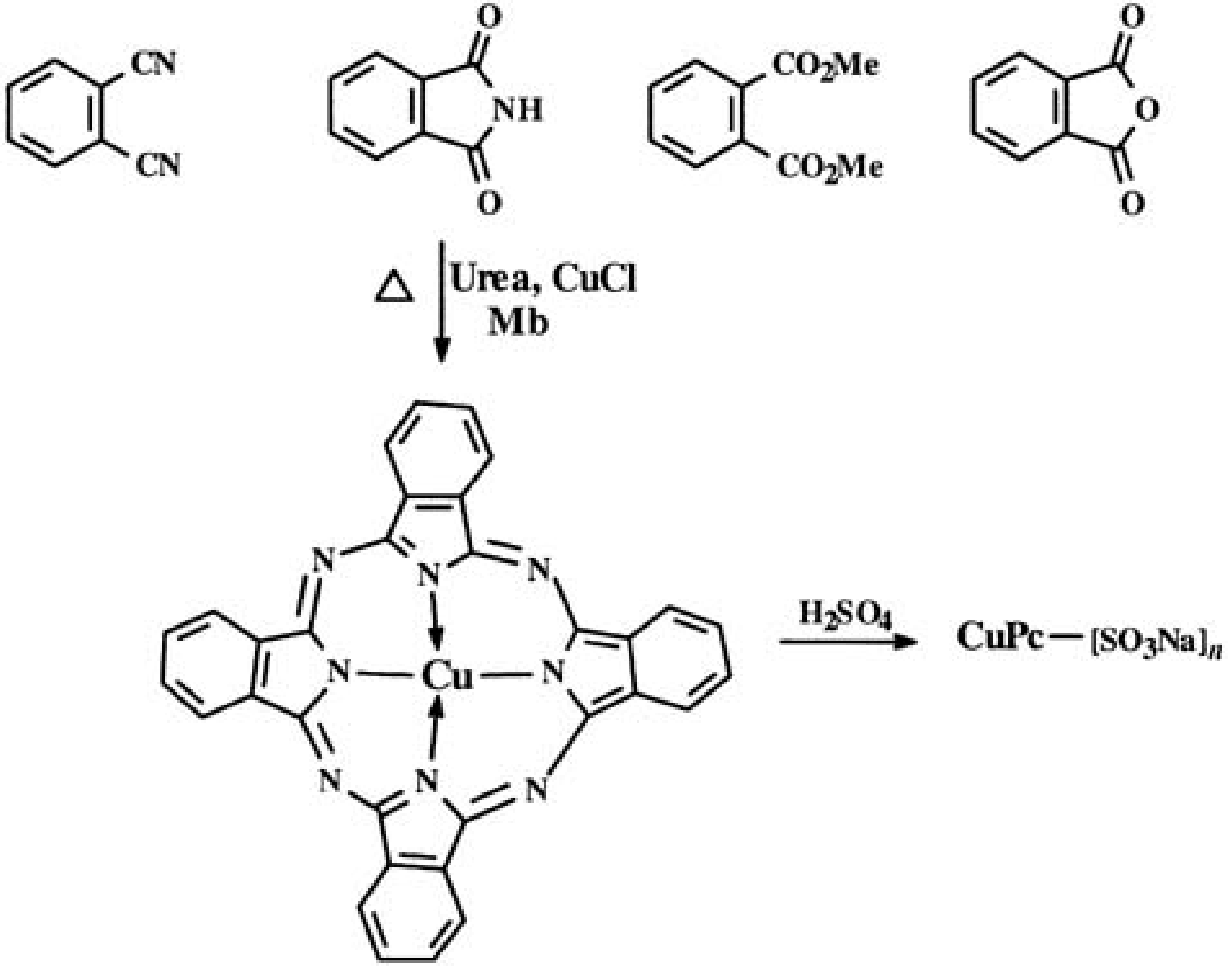
Fig. 23.135 Synthesis of sulfur black dyes.



Phthalocyanine Dyes

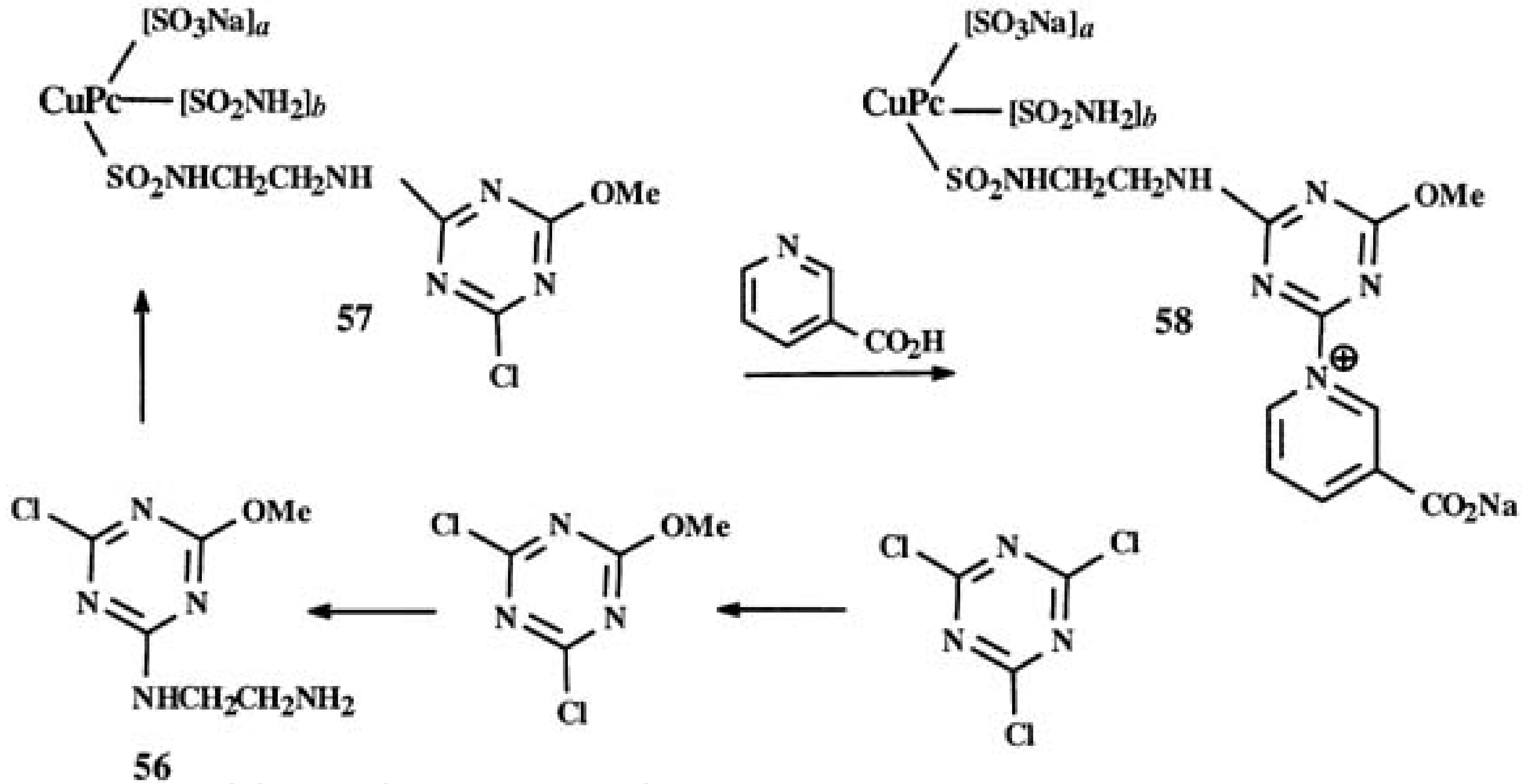
The synthesis of the copper phthalocyanine (CuPc) system is achieved as shown in Fig. 23.136. Here it can be seen that any of the four precursors can be used. Disulfonation gives C.I. Direct Blue 86 and tetrasulfonation gives C.I. Acid Blue 249.

Fig. 23.136 Synthesis of phthalocyanine and sulfonated derivatives.



The chlorosulfonation of the CuPc system opens the door to the synthesis of reactive dyes, as shown in Fig. 23.137. In this case, aminochlorotriazine 56 reacts with a CuPcSO2Cl intermediate to give a monochlorotriazine reactive dye (57), which in turn can be used to make the cationic reactive dye 58.

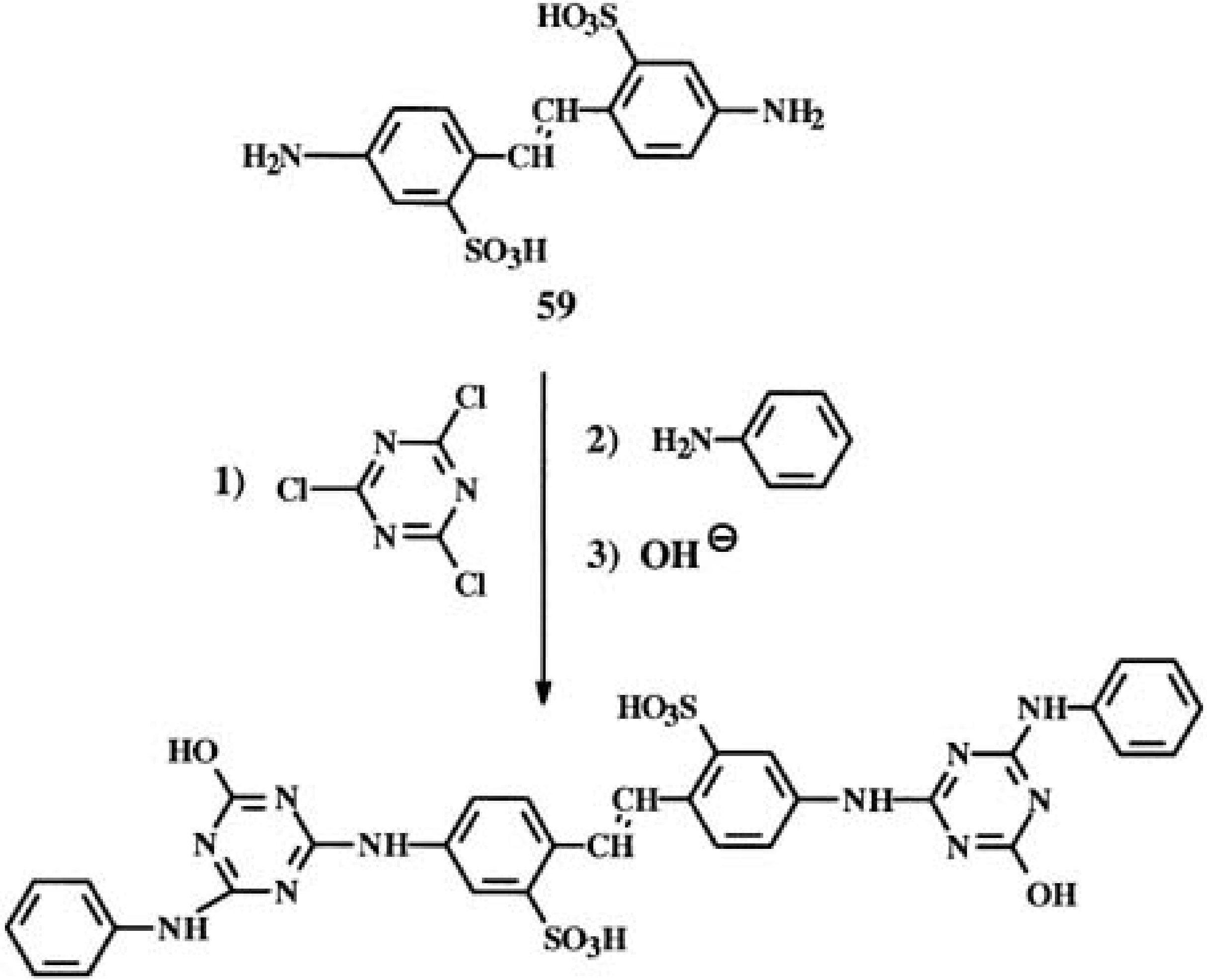
Fig. 23.137 Synthesis of CuPc-based reactive dyes.



Fluorescent Brighteners (Colorless "Dyes")

Many fluorescent brighteners are derivatives of 4,4′diamino-stilbene-2,2′-disulfonic acid (59), an example of which is C.I. Fluorescent Brightener 32 (Fig. 23.138). In this case, successive reactions involving diamine 59 with two molecules of cyanuric chloride and two molecules of aniline followed by hydrolysis of the final chloro groups give the target compound.

Fig. 23.138 Synthesis of Fluorescent Brightener 32.



Structurally related fluorescent brighteners containing a benzotriazole moiety are made according to the route shown in Fig. 23.139. In this case, diamine 59 is tetrazotized, coupled to 2 molecules of 1,6-Cleve's acid, and the intermediate disazo stilbene structure (60) is oxidized to C.I. Fluorescent Brightener 40. Nowadays, monosulfonated benzotriazole brighteners are more important.60 The synthesis of one example is shown in Fig. 23.140 for C.I. Fluorescent Brightener 46.

Fig. 23.139 Synthesis of Fluroescent Brightener 40.

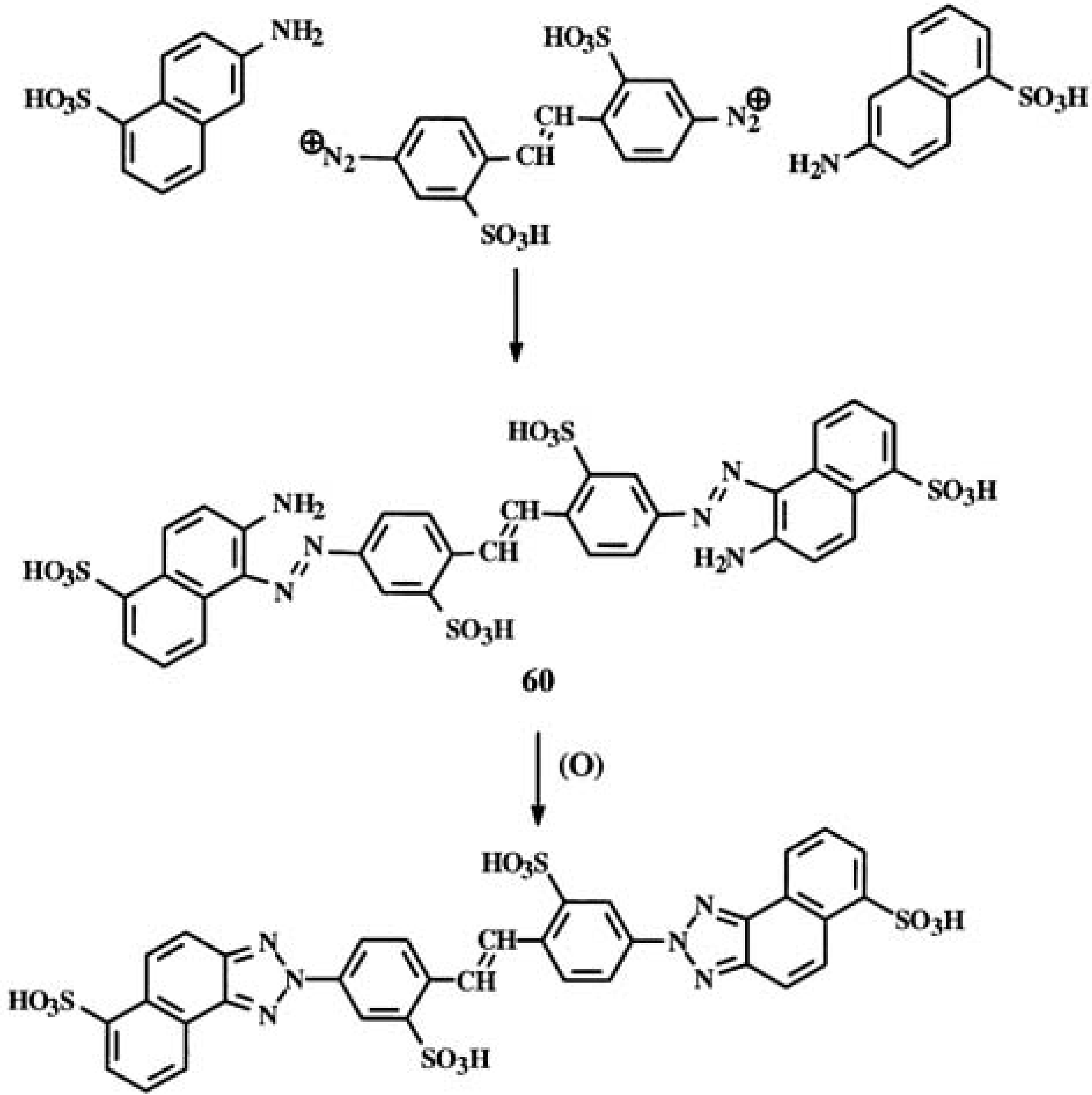
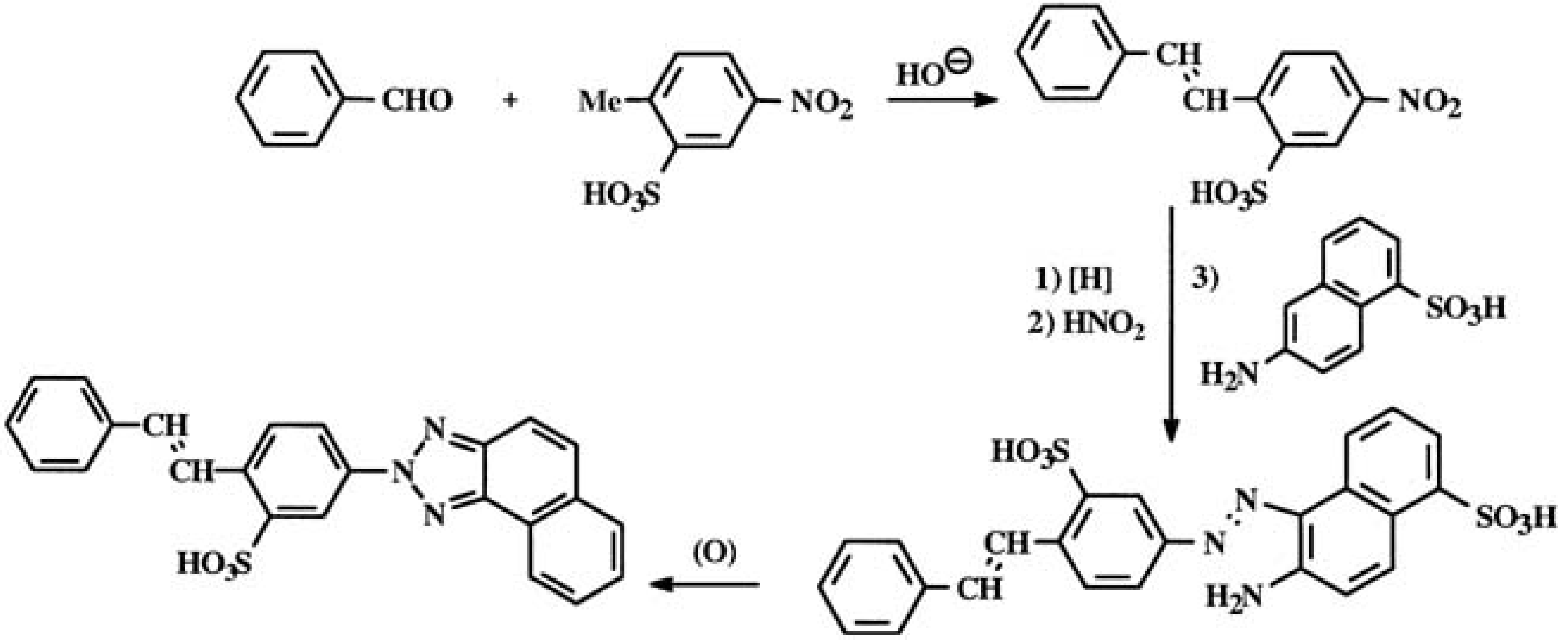


Fig. 23.140 Synthesis of Fluorescent Brightener 46.



Examples of hydrophobic fluorescent brighteners include C.I. Fluorescent Brighteners 199, 130, 236, and 162. The synthesis of these compounds is shown in Figs 23.14123.144. In the first of these examples, a bis-stilbene structure is made in two steps from bis-chloromethyl-xylene, using the traditional reaction of a phosphorus ylide with an aldehyde as the key step in the sequence.

Fig. 23.141 Synthesis of Fluorescent Brightener 199.

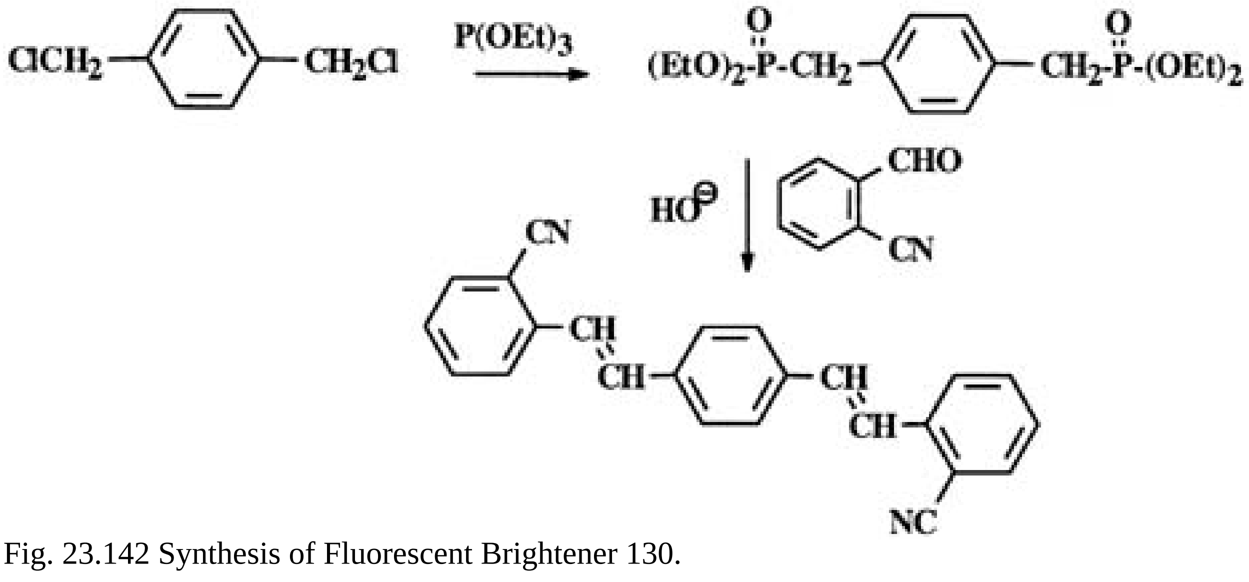


Fig. 23.142 Synthesis of Fluorescent Brightener 130.

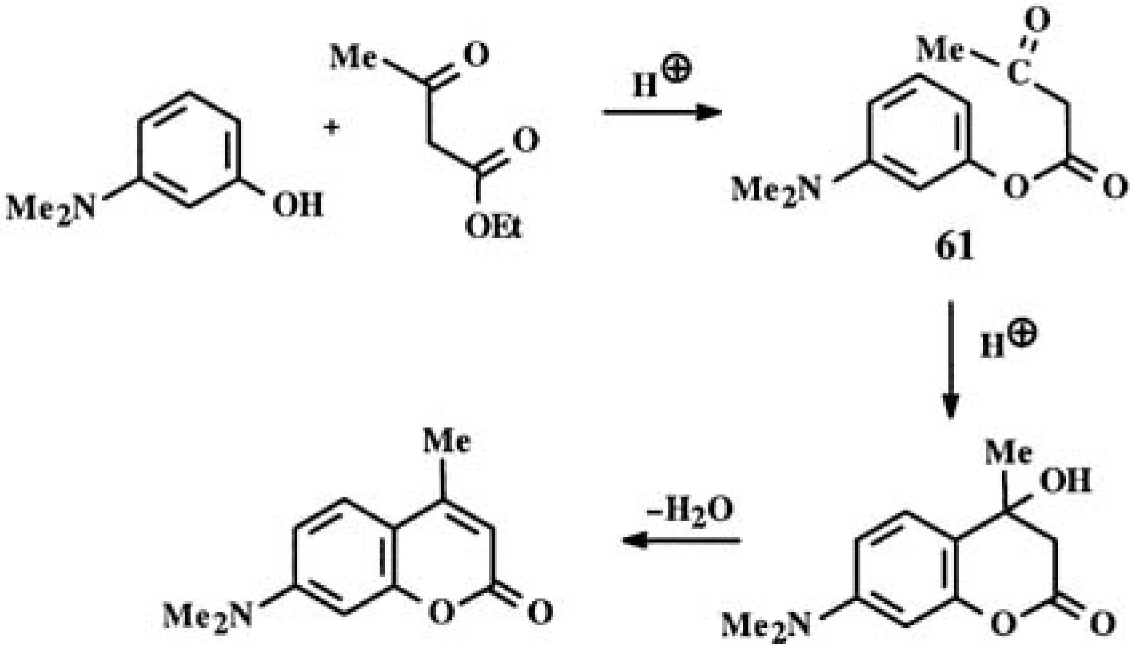


Fig. 23.143 Synthesis of Fluorescent Brightener 236.

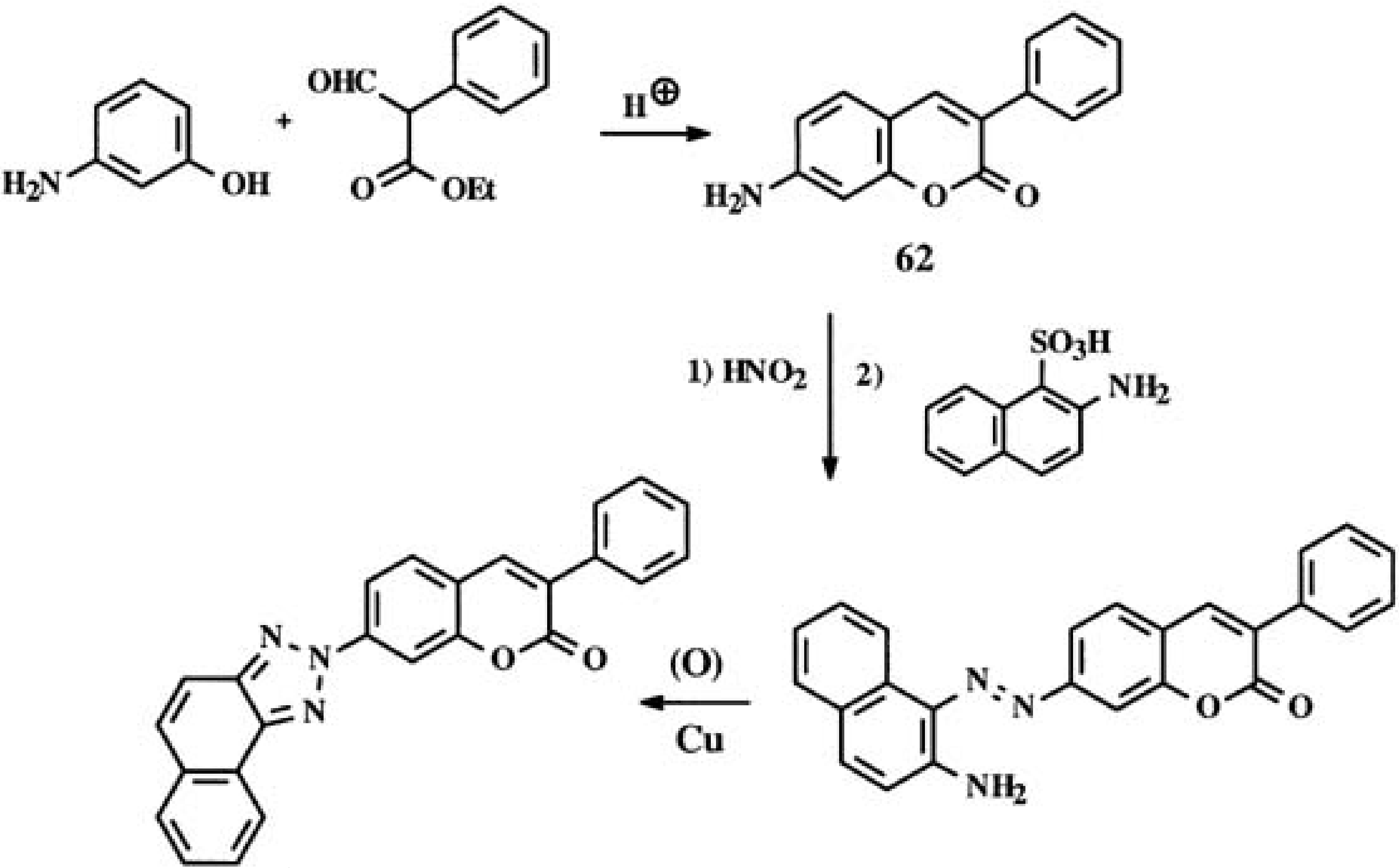
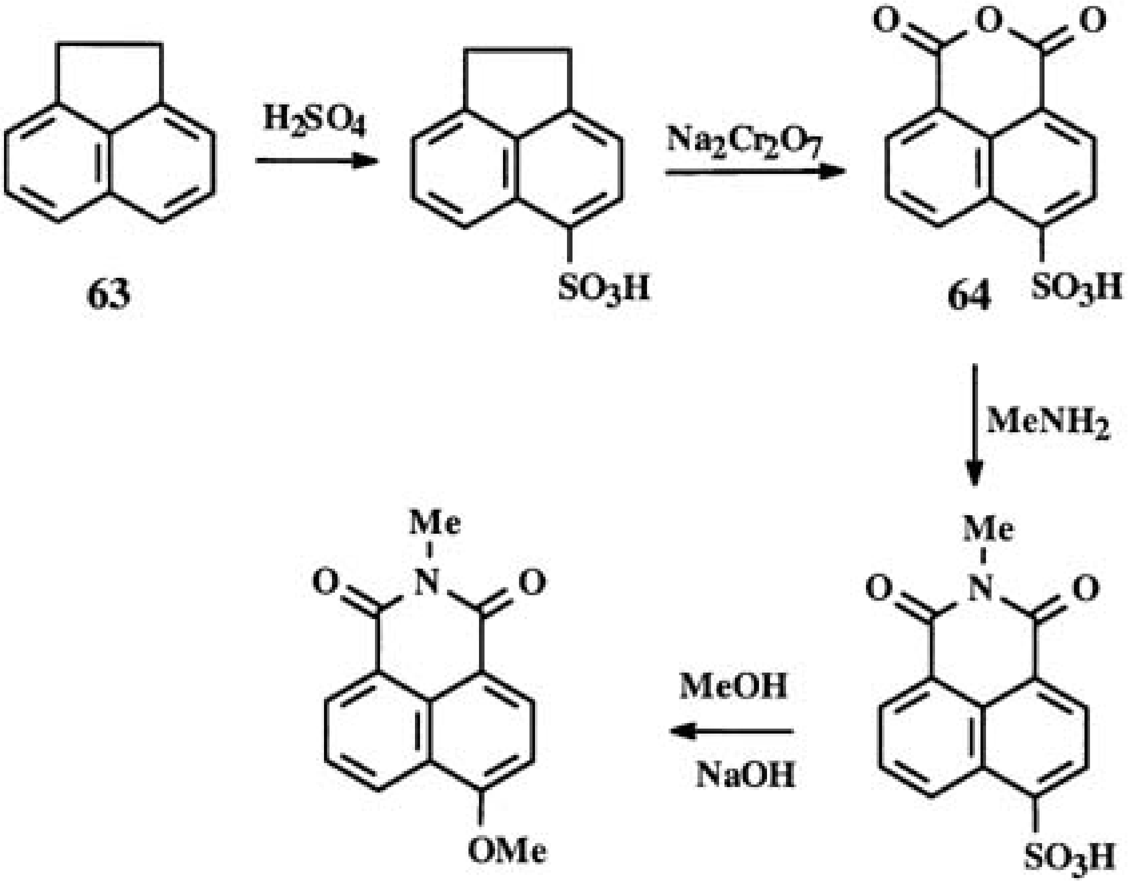


Fig. 23.144 Synthesis of Fluorescent Brightener 162.



In the second example, the synthesis of a coumarin-type fluorescent brightener is illustrated. Here, *meta*-hydroxy-*N*, *N*-diethylaniline is condensed with ethyl acetoacetate followed by cyclization of the intermediate keto ester 61. The latter compound undergoes acid-catalyzed cyclization and dehydration to give C.I. Fluorescent Brightener 130 (see Fig. 23.142).

A fluorescent brightener containing coumarin and triazole groups is made according to the method shown in Fig. 23.143. The synthesis begins with the preparation of amino-coumarin 62, which in turn is coupled to Tobias acid with concomitant loss of the SO3H group and then oxidized to give C.I. Fluorescent Brightener 236.

The final example is for a naphthalimide structure that is made from acenaphthene (63) in the four-step sequence shown in Fig. 23.144: (1) sulfonation, (2) chromate oxidation to give the naphthalic anhydride (64), (3) condensation with *N*-methylamine, and (4) replacement of the sulfonic acid group in a reaction with methoxide. This process gives C.I. Fluorescent Brightener 162.

23.10 Production and Sales

During the 1990s, the large international companies began to form alliances with producers around the world. Hoechst AG, which had done little research on disperse dyes since the 1970s, signed an agreement in 1990 with Mitsubishi of Japan and gained access to a strong line of disperse dyes. BASF AG and Mitsui signed agreements in vat dyes. ATIC resulted from a joint venture between ICI and Atul of India. Finally, a major break came in January 1995, when Bayer AG and Hoechst AG, the parent companies in Germany, announced the formation of DyStar, a worldwide consolidation of their textile dye businesses, which included the U.S. Hoechst Celanese, and Bayer. Within a short time, BASF acquired the textile dyes business of ICI/Zeneca. Swiss companies Ciba and Clariant (derived by consolidating Sandoz and portions of Hoechst in 1995) announced a merger of the textile dyes business but cancelled the venture in 1998. C&K emerged as the sole U.S. based major company, but the company struggled during the late 1990s and was sold to Yorkshire Group plc of the United Kingdom. Yorkshire Pat-Chem and C&K became Yorkshire Americas.

Globalization and establishment of NAFTA meant fewer textile dyes were needed and manufactured in the United States during the late 1990s. The market shrank from 232 million lb ($955 million) in 1994 to 214 million lb ($689 million) in 1998 with further cuts expected. Imported dyes expanded but prices fell. Some 1.1 million lb of disperse dyes were brought in with a value of $5 million in 1992. In 1999, 5.7 million lb with a value of $10 million were imported. For each class of dyes, you can find expansion of imports for fewer and fewer dollars. The latest year when consumption was publicly revealed is given in Table 23.6.

TABLE 23.6 Projected Sales and Market Share of Textile Dyes in the United States

Dye Classification	%
Fiber reactive	29
Disperse	22
Acid and premet.	12
Sulfur	11
Pigments	10
Direct	6
Basic	6
Vat	3
Naphthol	1

Note: Projected 1995 sales = $920 million.
Source: *Chemical & Engineering News*, January 15, 1996.

23.11 References

Edelstein, S. M., Historical Notes on the Wet-Processing Industry, Dexter Chemical Co., Bronx, NY, no date.
Zahn, J., Bayer Farben Revue, 12, 32; 11, 75 (1967).
Edelstein, S. M., and H. C. Borghetty, The Plictho of Gioanventura Rosetti, The M. I. T. Press, Cambridge, MA, 1969.
Robinson, R., "The Life and Work of Sir William Henry Perkin," Proceedings of the Perkin Centennial 18561956, AATCC, 41 (1957).
Lehner, S., "America's Debt to Perkin," ibid., 269.
"Buyer's Guide," AATCC Review, 1 (7), 17 (2001).
Raghavan, K. S. S., Textile Dyer Printer, XV (20), 21 (1982).
Mock, G. N., AATCC Review, 1 (12), 18 (2001).
Compiled from: The Colour Index, Vol. 5, 4th ed. (1992).
Crossley, M. L., Am. Dyestuff Rep., 27 (3), 124 (1938).
Crossley, M. L., Am. Dyestuff Rep., 28 (3), 487 (1939).
Welham, A. C., J. Soc. Dyers & Colourists, 102, 126 (1986).
Beffa, F., and Back, G., Rev. Prog. Colorat., 14, 33 (1984).
Allen, R. L. M., Colour Chemistry, Chapter 7, Thomas Nelson & Sons, London, 1971.
Hueckel, M., Textile Chemist & Colorist, 1 (11), 510 (1969).
Raue, R., Rev. Prog. Colorat., 14, 187 (1984).
Aspland, J. R., Textile Chemist & Colorist, 25 (6), 21 (1993).
Allen, R. L. M., Colour Chemistry, Chapter 5, Thomas Nelson & Sons, London, 1971.
Shore, J., Colorants and AuxiliariesOrganic Chemistry, Properties and Applications, Vol. 1 (Colorants), p. 20, Society of Dyers and Colourists, Manchester, England, 1990.
Aspland, J. R., Textile Chemist & Colorist, 21 (11), 21 (1991).
Shore, J., Rev. Prog. Colorat., 21, 23 (1991).
Haley, T. J., Clin. Toxicol., 8, 13 (1975).
Prival, M. J., Bell, S. J., Mitchell, V. D., Peirl, M. D., and Vaughan, V. L., Mutat. Res., 136, 33 (1984).
Ecological and Toxicological Association of Dyes and Organic Pigments Manufacturers, ETAD Information Notice No. 6, Textile Chemist Colorist, 28 (6), 11 (1996).
Freeman, H. S., Esancy, M. K., Esancy, J. F., and Claxton, L. D., Chemtech, 21 (7), 438 (1991).
Dawson, J. F., Rev. Prog. Colorat., 14, 90 (1984).
Annen, O., Egli, R., Hasler, R., Henzi, B., Jakob, H., and Matzinger, P., Rev. Prog. Colorat., 17, 72 (1987).
Weaver, M. A., and Shuttleworth, L., Dyes Pigments, 3, 81 (1982).
Egli, R., "The Chemistry of Disperse Blue Dyes, Past and Present," Chapter 1, in Colour Chemistry, A. T. Peters and H. S. Freeman (Eds.), Elsevier Applied Science, London, 1991.
Reinert, G., Rev. Prog. Colorat., 27, 32 (1997).
Renfrow, A. H. M., and Taylor, J. A., Rev. Prog. Colorat., 20, 1 (1990).
Aspland, J. R., Textile Chemist Colorist, 24 (5), 31 (1992).
Carr, K., "Reactive Dyes, Especially Bireactive Molecules," in Modern Colorants: Syntheses and Structures, Vol. 3, Chapter 4, A. T. Peters and H. S. Freeman (Eds.) Blackie Academic and Professional, London, 1995.
Wood, W. E., Rev. of Prog. in Colorat. 7, 80 (1976).
Guest, R. A., and Wood, W. E., Rev. of Prog. in Colorat. 19, 63 (1989).
Aspland, J. R., Textile Chemist & Colorist, 24 (4), 27 (1992).
Baumgarte, U., Rev. of Prog. in Colorat. 17, 29 (1987).
Jones, F., Rev. of Prog. in Colorat., 19, 20 (1989).
Egerton, G. S., Journal of the Society of Dyers & Colourists, 65, 764 (1949).
Freeman, H. S., and Peter, A. T. (Eds.), Colorants for Non-Textile Applications, Elsevier Science, Amsterdam, 2000.
Gregory, P., "Colorants for High Technology," in Colour Chemistry, A. T. Peters and H. S. Freeman (Eds.) Elsevier Applied Science, London, 1991.
Freeman, H. S., and Sokolowska, J., Rev. of Prog. in Colorat. 29, 8 (1999).
Bauer, W., and Ritter, J., "Tailoring Dyes for Ink-Jet Applications," in Z. Yoshida and Y. Shirota (Eds.), Chemistry of Functional Dyes, Chapter 8.1, p. 649, Mita Press, Tokyo, 1993.
Carr, K., "Dyes for Ink Jet Printing", Chapter 1, in H. S. Freeman and A. T. Peter (Eds.), Colorants for Non-Textile Applications, Elsevier Science, Amsterdam, 2000.
Gregory, P., Dyes & Pigments, 13, 251 (1990).
Bradbury, R., "Thermal Transfer Printing," Chapter 2, in H. S. Freeman and A. T. Peter (Eds.), Colorants for Non-Textile Applications, Elsevier Science, Amsterdam, 2000.
Bamfield, P., Chromic Phenomena: Technological Applications of Colour Chemistry, Chapter 1, p. 50, The Royal Society of Chemistry, Cambridge, UK, 2001.
Inokuchi, H., "Organic Semiconductors: A Still Fashionable Subject," in Z. Yoshida and Y. Shirota (Eds.), Chemistry of Functional Dyes, Chapter 7.1, p. 521, Mita Press, Tokyo, 1993.
Bamfield, P., Chromic Phenomena: Technological Applications of Colour Chemistry, Chapter 4, pp. 245256, The Royal Society of Chemistry, Cambridge, UK, 2001.
Pavlopoulos, T., "Laser Dyes," Chapter 7, in H. S. Freeman and A. T. Peter (Eds.), Colorants for Non-Textile Applications, Elsevier Science, Amsterdam, 2000.
Moura, J. C. V. P., "Biomedical Applications of Dyes," Chapter 5, in H. S. Freeman and A. T. Peter (Eds.), Colorants for Non-Textile Applications, Elsevier Science, Amsterdam, 2000.
Corbett, J. G., Rev. Prog. Colorat. 15, 52 (1985).
Corbett, J. G., Journal of the Society of Dyers & Colourists, 83, 273 (1967).
Corbett, J., "Hair Dyes," Chapter 10, in H. S. Freeman and A. T. Peter (Eds.), Colorants for Non-Textile Applications, Elsevier Science, Amsterdam, 2000.
Waller, D., Zbigniew, Hinz, J., and Filosa, M., "Dyes used in Photography," Chapter 3, in H. S. Freeman and A. T. Peter (Eds.), Colorants for Non-Textile Applications, Elsevier Science, Amsterdam, 2000.
Moury, D. T., Chemical Reviews, 42, 213 (1948).
Zollinger, H., Color Chemistry, 2nd ed., Chapter 7, VCH, Weinheim, 1991.
Fierz-David, H. E., and Blangey, L., Fundamental Processes in Dye Chemistry, 5th ed., p. 250, interscience, NY, 1979.
Gordon, P. F., and Gregory, P., Organic Chemistry in Colour, Chapter 2, Springer-Verlag, Berlin, 1983.
Zollinger, H., Color Chemistry, 2nd ed., Chapter 8, VCH, Weinheim, 1991.
Püntener, A., and Schlesinger, U., "Natural Dyes," Chapter 9, in H. S. Freeman and A. T. Peter (Eds.), Colorants for Non-textile Applications, Elsevier Science, Amsterdam, 2000.
Hallas, G., "Chemistry of Anthraquinoid-Polycyclic, and Miscellaneous Colorants," Chapter 6, in J. Shore (Ed.), Colorants and AuxiliariesOrganic Chemistry, Properties and Applications, Vol. 1 (Colorants), p. 20, Society of Dyers and Colourists, Manchester, England, 1990.
Siegrist, A. E., Heffi, H., Meyer, H. R., and Schmidt, E., Rev. of Prog. in Colorat. 17, 39 (1987).

24.1 Evolution of Modern Fermentation Processes

Humans were well aware of fermentations, even though they had little knowledge of what caused them, long before they were able to record such an awareness. Sometime in prerecorded history, people discovered that meat allowed to stand a few days was more pleasing to the taste than meat eaten soon after the kill. They also were aware that intoxicating drinks could be made from grains and fruits. The aging of meat and the manufacture of alcoholic beverages were human's first uses of fermentation.

Without even knowing that microorganisms existed, ancient people learned to put them to work. The ancient art of cheese-making involves fermentation of milk. For thousands of years, the soy sauces of China and Japan have been made from fermented beans. For centuries, the Balkan people have enjoyed fermented milk, or yogurt, and members of Central Asian tribes have found equal pleasure in sour camel's milk, or kumiss. Bread, which has been known almost as long as agriculture itself, involves a yeast fermentation process. Loaves of bread have been found in Egyptian pyramids built 6000 years ago.

The discovery of fruit fermentation was made so long ago that the ancient Greeks believed wine had been invented by the god Dionysus. The manufacture of beer is only slightly less ancient that that of wine. A Mesopotamian clay tablet written in Sumerian-Akkadian about 500 years before Christ tells us that brewing was a well-established profession 1500 years earlier. An Assyrian tablet of 2000 B.C. lists beer among the commodities that Noah took aboard his ark. Egyptian documents dating back to the Fourth Dynasty, about 2500 B.C., describe the malting of barley and the fermentation of beer. Kui, a Chinese rice beer, has been traced to 2300 B.C.

During the Middle Ages, experimenters learned how to improve the taste of wine, bread, beer, and cheese. Yet, after thousands of years of experience people still did not realize that in fermentations they were dealing with microorganisms. It was not until 1857 that Pasteur proved alcoholic fermentation was brought about by yeasts, and that yeasts were living cells. The discovery marked a turning point in medical history and the birth of microbiology.

The inheritors of Pasteur's knowledge sought to use microbes as production workers in industry. The production of bakers' yeast in deep, aerated tanks was developed toward the end of the nineteenth century. During World War I, Chaim Weizmann used a bacterial cousin of the gas gangrene microbe to convert maize mash into acetone, which was essential in the manufacture of the explosive cordite. In 1923, Pfizer opened the world's first successfull plant for citric acid fermentation. The process involved a fermentation utilizing the mold *Aspergillus niger* whereby ordinary sugar was transformed into citric acid.

Other industrial chemicals produced by fermentation were found subsequently, and the processes were reduced to commercial practice. These processes included butanol, acetic acid, oxalic acid, gluconic acid, fumaric acid, and many more.

In 1928 Alexander Fleming, working with *Staphylococcus aureus*, a bacterium that causes boils, observed a strange fact. A mold of the *Penicillium* family grew as a contaminant on a Petri dish inoculated with *Staphylococcus aureus*; a clear zone was observed where the *Staphylococcus* organism in the vicinity of the contaminating mold has been killed. Fleming nurtured the mold and then extracted a chemical from it that killed the bacteria. He named the extracted material penicillin.

Fleming's discovery received little notice until two Oxford University experimenters, under the stress of World War II, resolved to find an antibacterial agent of wider activity than the sulfa drugs. These two British workers, Dr. Howard Florey and Dr. Ernst Chain, set out to produce a therapeutic agent capable of saving the lives of war casualties. Their first candidate was the *Penicillium notatum* mold preserved from Fleming's studies. Penicillin turned out to be exactly what they were looking for. The American pharmaceutical industry helped them solve their difficulties in mass-producing the antibiotic. Three American companies led the wayMerck, Pfizer, and Squibb.

Initially, *Penicillium* was surface-cultured in flasks. A chance discovery in a Peoria market provided the major breakthrough. There, a government worker found a moldy cantaloupe on which was growing a new strain of penicillium, *Penicillium chrysogenum*, which would thrive when cultured in deep, aerated tanks and that gave 200 times more penicillin than did Fleming's mold.

Other antibiotics were quick to appear. Streptomycin was next, an antibiotic that was particularly effective against the causative organism of tuberculosis. The search was now on. Antibiotic prospectors combed the earth for organisms that produced different and more useful antibiotics. The list of these antibiotics is long today and includes such important antibiotics as chloramphenicol, the tetracyclines, bacitracin, erythromycin, novobiocin, nystatin, kanamycin, and many others.

Future Developments

Progress in fermentation is continuing at an ever-increasing pace. Each year, new products are added to the list of compounds derived from fermentation. Several vitamins now are produced routinely, employing fermentation steps in their synthesis. Outstanding examples are B2 (riboflavin), B12 (cyanocobalamin), and C (ascorbic acid). Some of the more interesting fermentation processes are the specific dehydrogenations and hydroxylations of the steroid nucleus. These chemical transformations are economical shortcuts used in the manufacture of the anti-arthritic cortisone and its derivatives. Fermentative syntheses of the amino acids L-lysine and L-glutamic acid also are being carried out commercially. Important agricultural uses are found for the new fermentation product gibberellin, a plant-growth regulator; and crystalline inclusions for a species of *Bacillus* are being used as specific insecticides in another agricultural application. A macrolide, avermectin, produced by a microorganism found near a golf course in Japan turned out to be a most potent and important antiparasitic agent. Microbial hydrolysis of cellulose and subsequent fermentation of the resultant sugar to alcohol is a promising source of liquid fuel. Research is in progress on chemical transformations utilizing fermentation techniques, new fermentative biosyntheses, continuous algal culture, and submerged mammalian-tissue culture. The advent of recombinant DNA technology since the late 1970s has added to the list of fermentation-derived products such complex molecules as insulin, human growth hormone, interferons, hepatitis vaccine, tissue plasminogen activator, erythropoietin, and colony stimulating factors. The list has been extended in the last decade to include products like interleukins, soluble receptors, and human or humanized monoclonal antibodies.

24.2 Fermentation Fundamentals

Introduction

The engineer's contribution to the development of the penicillin fermentation was a very important one. The outgrowth of the undertaking was the pure-culture technique, carried out in aerated and agitated deep-tank fermentors. This technique, similar to its antecedent used for yeast propagation, introduced to the biochemical process industry refined fermentation equipment capable of being maintained under aseptic conditions even when vigorously aerated. The technique has now been applied widely with minor modifications to the production of other antibiotics, amino acids, steroids, enzymes, and therapeutic proteins.

Of the operations auxiliary to those in the fermentor, engineers have made a major contribution in establishing the theoretical bases for the design of equipment to provide large volumes of sterile medium and air. Close cooperation between biologists and engineers is needed to devise logical methods for screening large number of strains, and translating the results of shake-flask and pilot-plant experiments to production vessels.

Fermentation Unit Operations

Fermentation processes have in common many of the familiar chemical engineering unit operations. For example, aerobic fermentations involve the "mixing" of three heterogeneous phasesmicroorganisms, medium, and air. Other unit operations include "mass transfer" of oxygen from the air to the organisms and "heat transfer" from the fermentation medium.

Analysis of fermentations by the unitoperation technique has added greatly to the understanding of their behavior. This understanding, however, is far from complete. The scale-up of fermentations, for instance, is still rather empirical although the sensitive oxygen probes and sensitive gas analysis techniques now available have enabled a more rational approach to scaling-up aerated, non-newtonian fermentations.

Fermentation Unit Processes

Analysis of the many industrial fermentation processes shows that they are common reactions from a chemical as well as a physical viewpoint. Fermentation processes can be classified by the reaction mechanims involved in converting the raw materials into products; these include reductions, simple and complex oxidations, substrate conversions, transformations, hydrolyses, polymerizations, complex biosyntheses, and the formation of cells.

Unit-process classification provides a ready catalogue of the chemical activities and abilities of microorganisms for the biochemist. More important, it offers a logical approach to an examination of fermentation reaction mechanisms.

Microorganisms

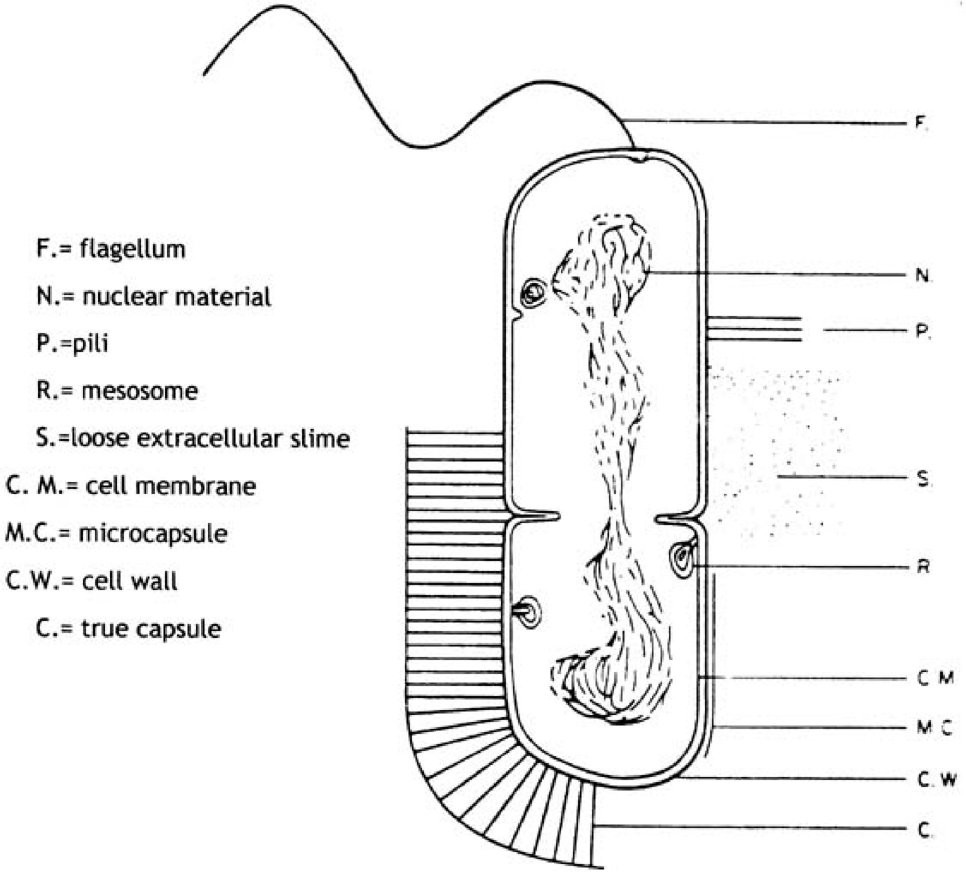
Microorganisms are chemically similar to higher plant and animal cells; they perform many of the same biochemical reactions. Generally microorganisms exist as single cells, and they have much simpler nutrient requirements than the higher life forms. Their requirements for growth usually are limited to air, carbon dioxide, and inorganic salts or inorganic salts supplemented by simple sugars. They occur in four main groups:

1. bacteria
2. viruses*
3. fungi, including yeast and actinomyces
4. protozoa and algae

*Not a microbial cell by strict definition.

Typically, *bacteria* are single cells, either cocci, rods, or spirals, which are 0.520 µm in size and capable of independent growth. They multiply by simple division. A typical bacterial cell is illustrated in Fig. 24.1.

Fig. 24.1 Detailed structure of a bacterial cell.



Viruses are the smallest microbes. They are obligate intracellular parasites of animals, plants, insects, fungi, algae, or bacteria. Bacterial viruses are called bacteriophage. Growth and multiplication take place intracellularly when the virus invades the host cell, taking over the genetic mechanism to direct the host cell to synthesize new viruses. The viruses consist primarily of genetic material, either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) and a protein coating. Because of the purity of the protein coat, it is possible to crystallize viruses and obtain X-ray diffraction patterns.

Fungi are widely spread in nature, in environments of lower relative humidity than those that favor bacteria. The metabolism of the fungi is essentially aerobic. They generally form long filamentous nucleated cells 420 µm wide that are highly branched. Many species have complicated life cycles during which both sexual and asexual spores are formed. These spores are usually quite resistant to environmental changes.

Generally, fungi are free-living saprophytes, but a few are parasites to animal cells, and many are serious pathogens of plants. They have very wide degradative and synthetic capabilities and have proved a fruitful source of industrially important organic acids, antibiotics, and enzymes.

Among the fungi are yeasts, which are elliptical cells 35 µm by 815 µm. They differ from most fungi in that growth is by budding. Growth of yeast on solid medium resembles a bacterial colony. Yeasts are industrially important in making alcohol and bakers' yeast.

Actinomycetes are a group of organisms intermediate in properties between bacteria and true fungi. Industrially, the group is extremely important as a source of powerful antibiotics for the control of microbial infections of humans, animals, and plants. Their growth in fermentation tanks is very similar to that of fungal fermentations; for this reason, they have been grouped with the fungi in this discussion.

The *protozoa* are widely distributed in fresh and salt water, and found in lesser frequency in soil and animals. They may be unicellular or multicellular and exhibit a wide variety of morphological forms. Usually, they are divided in two main groups: the *algae*, which are all capable of photosynthesis and contain chlorophyll; and the protozoa, which are not photosynthetic and resemble primitive animal cells.

The chemical composition of microorganisms can be quite varied depending upon such factors as the composition of the growth medium, the age of the culture, and the cell growth rate. Table 24.1 lists some of the compositions of different microbial forms.

TABLE 24.1 Chemical Analyses, Dry Weights, and the Populations of Different Microorganisms Obtained in Culture[a]

	Composition (% dry weight)					
Organism	Protein	Nucleic Acid	Lipid	Population in Culture (Numbers/ml)	Dry Weight of This Culture (g/100 ml)	Comments
Viruses	5090	550	<1	108109	0.0005[b]	Viruses with a lipoprotein sheath may contain 25% lipid
Bacteria	4070	1334	1015	2 × 1082 × 1011	0.022.9	*Mycobacterium* may contain 30% lipid
Filamentous fungi	1025	13	27		35	Some *Aspergillus* and *Penicillium* sp. contain 50% lipid
Yeast	4050	410	16	14 × 108	15	Some *Rhodotorula* and *Candida* sp. contain 50% lipid
Small unicellular algae	1060 (50)	15 (3)	480 (10)	48 × 107	0.40.9	Figure in () is a commonly found value but the composition varies with the growth conditions

[a] A. Aiba, A. E. Humphrey, and N. Millis, *Biochemical Engineering*, 2nd ed., p. 26, University of Tokyo Press, Tokyo, 1973.

[b] For a virus of 200 µm diameter.

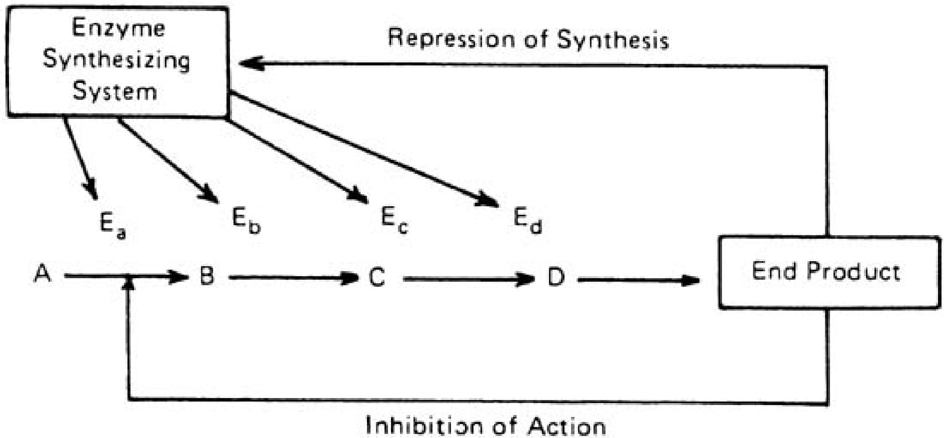
Mammalian Cells

Mammalian cells have been used for a long time in the manufacture of vaccines for rubella, adenovirus (types 4 and 7), rabies, poliomyelitis, and FMDV (foot-and-mouth disease virus). They have taken on greater importance in industrial fermentation since the advent of recombinant DNA technology. The commercial production of macromolecular proteins, particularly those requiring proper glycosylation and correct folding (e.g., tissue plasminogen activator and erythroprotein) for their biological activities, has accelerated the technical development and understanding of mammalian cell culture. Mammalian cells are quite different from their microbial counterparts, whose general characteristics were described earlier. Mammalian cells, typically 10100 µm in size, are much larger than microbes. They grow much more slowly, doubling in 1260 hr versus 0.52 hr for the microbes, and they usually require a complex and rich medium, often containing some sort of serum component, for growth. This makes the mainte-nance of asepsis extremely difficult, but most critical, in culturing mammalian cells. Mammalian cells differ from the microbes in another very important manner, in that they do not have a rigid cell wall to protect them against shear and osmotic pressure. This represents an additional challenge to engineers when scaling up a mammalian cell for commercial production. Some mammalian cells require a suitable surface for attachment and growth. Methods and bioreactors that have been developed in the last decade for the growth of the anchorage-dependent cells represent an important technical achievement by the cell culture practioners. The principles, processes, and products of industrial mammalian cell culture are presented in the next chapter.

Microbial Activity

All organisms contain the genetic information to produce a wide variety of enzymes and hence produce a great number of chemicals. However, only some enzymes are produced at all times, whereas others are greatly influenced by the substrate. Certain compounds interact with the substrate to repress the translation of genetic information for synthesis. (This process is called repression.) Also, the substrate sometimes reacts with a compound that is a genetic-mechanism repressor and removes its action. (This process is referred to as derepression.) This control is depicted in Fig. 24.2.

Fig. 24.2 System for the control of enzyme formation in biosynthesis.



These processes allow cells to regulate their enzyme content in direct response to the environment. They prevent the formation of excess end product and superfluous enzymes.

Mutants can be genetically engineered that lack these controls. Such mutants have been changed so that the genetic mechanism is no longer sensitive to a particular controlling metabolite. For industrial processes, such strains with faulty regulation, altered permeability, or metabolic deficiencies may be used to accumulate products.

Strain Development

Although in a normal bacterial population about one mutant arises in 106 cells, the rate of mutation is far higher than this because many mutations are lethal and so are not detected, and others involve base changes that do not change the amino acid read from the m-RNA codon. This low rate of spontaneous mutation is unsatisfactory when one is attempting strain improvement. However, mutagens are available that markedly increase the rate of mutation. Recently, tremendous strides have been made in genetic engineering, which allow the creation of mutants through such techniques as:

1. transduction
2. transformation

3. protoplast fusion

We now know that exchange of genetic information between cells is much more common than originally thought. The trick is not only to inject genetic material with the desired characteristics into a cell, but to ensure that it is "expressed" and that an environment is maintained that permits these "new species" to dominate.

Although it is relatively easy to produce mutants, strain improvement requires painstaking effort, ingenuity in devising screening tests, and luck. The difficult problem is to select desirable mutants, and this is accentuated when the parent strain is already a high producer. In this situation, any improvement is likely to be minor and quantitative rather than major and qualitative; major changes are most useful in the very first stages of a development program and in basic research.

Improvement in product formation can be achieved in a number of ways. Mutants can be selected that have the following properties:

1. Defective enzymes in a biosynthetic pathway so that useful intermediates accumulate.
2. Lack of feedback control by the end product itself or by a key intermediate on the pathway to the end product.
3. Defective cell membranes so that intracellular metabolites are excreted rapidly.
4. Lack of some undesirable by-product, such as a pigment.
5. Reduced toxicity to a precursor molecule such as phenyl acetate for penicillin production.

Stock Cultures

In all the industrially important bacteria and fungi, except some yeasts, the main vegetative phase of growth is haploid, so that variation in genetic material arises only from mutations, provided transduction does not occur in bacteria, and provided sexual spores are not formed in fungi. In maintaining stock cultures, genetic change must be minimized; this is best achieved by preventing nuclear divisions because most mutations occur as errors in DNA replication.

The method of choice is to store cells or spores (if these are produced) in sealed ampoules at very low temperatures (200°C) in liquid nitrogen. This method has the great advantage that the culture can be stored almost indefinitely, thawed, and used immediately as an inoculum without loss of viability or diminution in metabolic rate. Cultures kept at -20°C to -60°C are satisfactory but less active than those kept in liquid nitrogen. Although storage at 04°C allows some growth, this is better than storage at room temperature. Lyophilization (freeze-drying) is widely used and is very convenient because freeze-dried cultures retain viability without genetic change for years when stored at room temperature. It may be noted that all of these methods are, in effect, techniques to immobilize intracellular water and yet retain viability.

Microbial Kinetics

The kinetics of microbial systems may be expressed at four different system levels:

1. molecular or enzyme level
2. macromolecular or cellular component level
3. cellular level
4. population level

Each level of expression has a unique characteristic that leads to a rather specific kinetic treatment. For example, biological reactions at the molecular level invariably involve enzyme-catalyzed reactions. These reactions, when they occur in solution, behave in a manner similar to that of homogeneous catalyzed chemical reactions. However, enzymes can be attached to inert solid supports or contained within a solid cell structure. In this case, the kinetics are similar to those for heterogeneous catalyzed chemical reactions.

Enzyme Kinetics

In their simplest form, enzyme-catalyzed reactions, occurring in a well-mixed solution, are characterized by the well-known Michaelis-Menten kinetic expression. This relationship depicts the substrate, *S*, combining reversibly with the enzyme, *E*, to form an enzymesubstrate complex, *ES*, that can irreversibly decompose to the product and the enzyme, that is:

$$E + S \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} ES \underset{k_{+2}}{\rightarrow} E + P \qquad (24\text{-}1)$$

This leads to a kinetic expression for the velocity of the reaction, ν, of the following form:

$$v = \frac{v_{\max} S}{K_{\mathrm{M}} + S} = \frac{k_{+2} E_0 S}{\dfrac{k_{-1} + K_{+2}}{k_{+1}} + S} \qquad (24\text{-}2)$$

where νmax = *k*+2*E*0 is the maximum observable reaction rate at high substrate concentration and, hence, is only limited by the initial enzyme concentration, *E*0. *K*M is the dissociation constant when

$$k_{+2} \ll k_{-1}, \quad \text{then} \quad K_{\mathrm{M}} = K_{\mathrm{s}} \qquad (24\text{-}3)$$

and

$$ES \xrightarrow{k_2} E + P \quad \text{is limiting} \qquad (24\text{-}4)$$

The saturation constant, *K*s = *k*-1/*k*+1, is an indication of the affinity of the enzyme active site for the substrate.

Basically two kinds of catalytic poisoning or inhibition are considered: inhibition by competition for the active site by a nonreactive substrate, and inhibition by a substance that modifies the enzyme activity but does not compete for the active site. The competitive inhibitor adversely affects the binding of substrate and enzyme, and thus has an effect in increasing *K*M. The noncompetitive inhibitor, on the other hand, only affects *k*+2 and νmax.

Cell Growth Kinetics

Enzyme kinetic concepts have been utilized to express the kinetic behavior of cell growth on a single limiting substrate. Monod postulated that the growth of cells by binary fission on a single limiting substrate probably had a single limiting reaction step and therefore behaved in a manner analogous to the Michaelis-Menten enzyme kinetics, that is:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu_{\max} X \frac{S}{K_{\mathrm{s}} + S} \qquad (24\text{-}5)$$

where *X* is the cell concentration, *S* is the growth limiting substrate concentration, *t* is time, and μmax is the maximum growth rate. Since the growth rate of cells, increasing by binary fission, is defined by:

$$\mu = \frac{1}{X} \frac{\mathrm{d}X}{\mathrm{d}t} \qquad (24\text{-}6)$$

The growth kinetics can be expressed as:

$$\mu = \mu_{\max} \left(\frac{S}{K_{\mathrm{s}} + S} \right) \qquad (24\text{-}7)$$

For a complete theory of cell growth and substrate utilization, it is necessary to know the relationship between the growth of cells and the utilization of substrate.

This relationship is expressed as a yield, *Y*, defined as:

$$Y - \frac{\mathrm{d}X}{\mathrm{d}S} \qquad (24\text{-}8)$$

The simplest assumption is that *Y* is constant. That this is essentially true only at high growth rates will be shown later. From the kinetic equation and the yield expression one obtains the following relationship for substrate utilization:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{1}{Y_{\mathrm{G}}} \frac{\mathrm{d}X}{\mathrm{d}t} + mX \qquad (24\text{-}9)$$

or:

$$\frac{1}{X} \frac{\mathrm{d}S}{\mathrm{d}t} = \frac{1}{Y_{\mathrm{G}}} \mu + m \qquad (24\text{-}10)$$

where *m* is the maintenance requirement for substrate per unit of cell biomass per unit of time, and *Y*G is a true yield constant representing the substrate utilized only for growth.

Recently it has been suggested that the rate of cell increase should be expressed as a net growth rate that involves both growth, μ, and death, δ, that is:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \mu X - \delta X \qquad (24\text{-}11)$$

where:

$$\mu = \mu_{\max} \left(\frac{S}{K_{\mathrm{s}} + S} \right) \qquad (24\text{-}12)$$

and:

$$\delta = \delta_{\max} \left(1 - \frac{S}{K'_{\mathrm{s}} + S} \right) \qquad (24\text{-}13)$$

Here, cell death is depicted as having a first-order kinetic behavior and as maximal, that is, δmax, when the growth limiting substrate is zero, and minimal when the substrate is in excess.

Over the years, numerous models for depicting cell growth have evolved. Several will be discussed here. One such model is that for growth under multiple substrate limitation. It can be expressed as:

$$\mu = \mu_{\max} \left(\frac{S_1}{K_{\mathrm{s}1} + S_1} \right) \left(\frac{S_2}{K_{\mathrm{s}2} + S_2} \right) \cdots . \qquad (24\text{-}14)$$

Multiple substrate limitation frequently occurs in batch growth systems.

Another model for cell growth is that for situations of extremely low substrate concentrations, that is, *S* < *K*s. In this case, the growth kinetic equation reduced to:

$$\mu = KS \qquad (24\text{-}15)$$

where *K* is constant and approximately equal to μmax/*K*s. This kinetic behavior is frequently observed in large single reactor waste treatment systems.

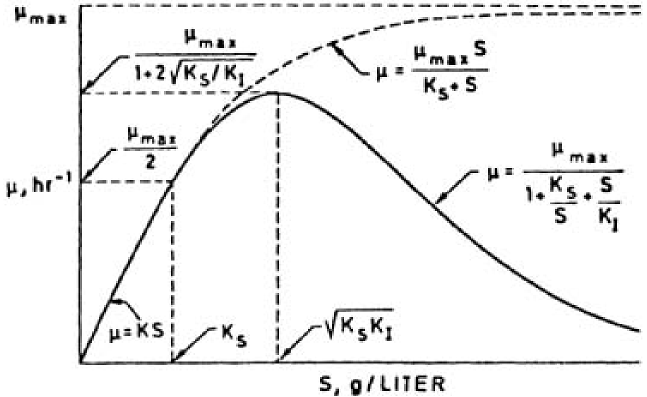
Another situation that commonly occurs in waste treatment is growth under conditions of "shock loading," that is, conditions in which the substrate rises to a level in which it becomes inhibitory to growth. This situation is frequently modeled by waste treatment designers by the following kind of kinetic expression:

$$\mu = \frac{\mu_{\max}}{1 + \dfrac{K_{\mathrm{s}}}{S} + \dfrac{S}{K_1}} \qquad (24\text{-}16)$$

The three basic growth relationships are depicted in Fig. 24.3. In passing, it is interesting to note that the observed μmax under conditions of substrate inhibition is a fraction of the true μmax if there were no inhibition, that is: at dμ/d*S* = 0

$$\frac{\mu}{\mu_{\max}} = \frac{1}{1 + 2\sqrt{K_{\mathrm{s}}/K_1}} \qquad (24\text{-}17)$$

Fig. 24.3 Kinetic expressions for growth.



Another common kinetic expression for growth is one that takes into account a substrate diffusional limitation. It has been observed in various waste treatment systems that the *K*s for "floc" or sludge growth is greater than that for single cell systems. Further, the growth rate of flocs or sludges appears to be a function of their size. This situation is depicted in Fig. 24.4. The growth kinetics in this case can be expressed as:

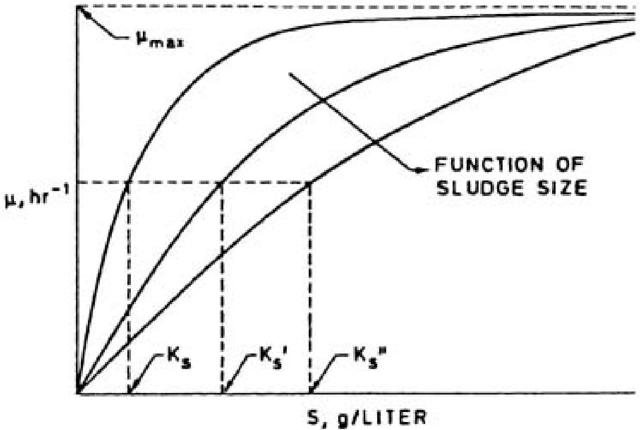
$$\mu = \frac{\mu_{\max} S}{K_s^{\text{app}} + S} \qquad (24\text{-}18)$$

and:

$$K_s^{\text{app}} = K_s + K_D \qquad (24\text{-}19)$$

where $K_D = f$ (sludge or cell size, membrane permeability, diffusion coefficient, and so on). When $K_D = 0$, there is no diffusional limitation to substrate uptake.

Fig. 24.4 Effect of sludge size on growth.



Temperature Effects

Absolute reaction rate theory has been found applicable to both cell growth and death, that is:

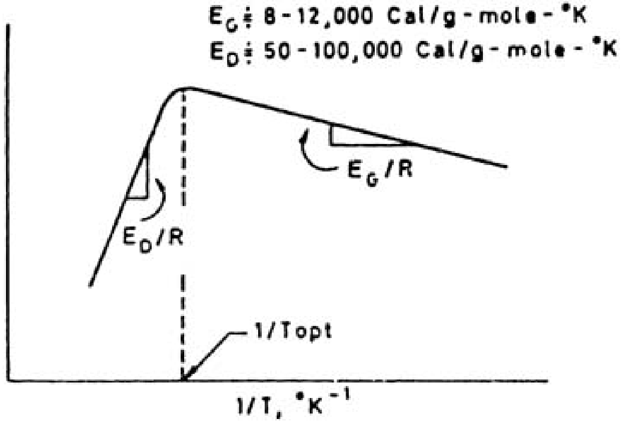
$$\frac{\mathrm{d}nk}{\mathrm{d}T} = \frac{E_a}{RT^2} \qquad (24\text{-}20)$$

or:

$$\mathrm{d}nk = -(E_a/R)\,\mathrm{d}(1/T) \qquad (24\text{-}21)$$

where k is the rate constant for growth, that is, μmax or death, that is, δmax, T is the absolute temperature, E_a is the activation energy for the process, and R is the gas law constant. Generally speaking, E_a for growth, E_G, is the order of 812,000 cal/g-mol, °K and E_a for death, E_D, is the order of 50100,000 cal/g-mol, °K. This behavior leads to an optimal temperature for growth for a given cell species. This is depicted in Fig. 24.5.

Fig. 24.5 Effect of temperature on growth.



pH Effects

The model for pH effect on growth has been based on the behavior of enzymes. Because enzymes are composed of amino acids, they exhibit "zwitterion" behavior, that is, they have an acid, base, and neutral form. Cells have been depicted to have an active or neutral form and an inactive base or acid form, that is:

$$X^{+-} \underset{K_1}{\rightleftharpoons} X \underset{K_2}{\rightleftharpoons} X^{-} \qquad (24\text{-}22)$$

where K_1 and K_2 are the equilibrium constants of the above reaction. If one defines the growing or active cell fractions as y, then:

$$y \;—\; \text{active cell fraction} = \frac{X}{X_0} \qquad (24\text{-}23)$$

where:

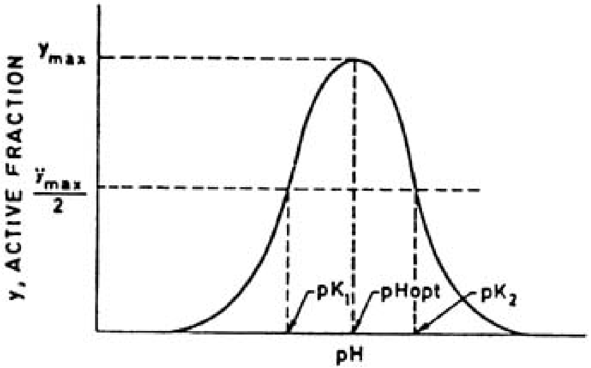
$$X_0 = X^{+} + X + X^{-} \qquad (24\text{-}24)$$

By analogy to enzyme behavior:

$$y = \frac{1}{1 + \frac{[\mathrm{H}^+]}{K_1} + \frac{K_2}{[\mathrm{H}^+]}} \qquad (24\text{-}25)$$

where [H+] is hydrogen ion concentration. Figure 24.6 depicts the effect of pH on the active cell fraction.

Fig. 24.6 Effect of pH on growth.



For most cell systems, pHopt = 6.5 ± 1 and pK_2 pK_1 = 2 ± 1.

Fermentor Configurations

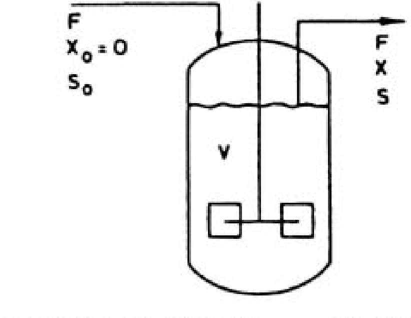
Numerous fermentor configurations may be found in microbial processes. These include:

1. batch
2. fed-batch
3. repeated draw-off
4. continuous
    1. single stage
    2. multiple stage
5. continuous with recycle
6. continuous with step feeding

For the most part, the antibiotic industry uses batch-type processes. The reason for this stems from the fact that most efficient antibioticproducing organisms are highly mutated and are readily replaced by fast-growing, less efficient antibiotic producers in a continuous culture. In order to avoid substrate repression or inhibition, some batch processes are continuously fed concentrated substrate on demand during the course of the batch cycle. This is referred to as a "fed-batch" fermentation. The production of bakers' yeast is an example of a fed-batch process. In some highly mycelial antibiotic fermentation, 2040 percent draw-off followed by fresh media makeup is practiced. In the trade, this is referred to as a "repeated draw-off" process. Strict continuous processes are practiced only in processes for the production of biomass for feed or food and the treatment of wastes. Continuous biomass-producing systems are usually single-stage reactors without recycle. Waste treatment systems usually use multistage systems with recycle. They frequently use step feeding of several stages to improve system stability.

A typical single-stage continuous fermentor is depicted in Fig. 24.7.

Fig. 24.7 Typical single-stage continuous fermentor.



F = flow rate, liters/hr
V = volume, liters
D ≡ F/V = dilution rate, hr⁻¹
X = cell conc., g/liter
S = substrate conc., g/liter

A biomass material balance around the reactor yields the following relationship. At steady state:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = 0 \text{ (steady state)} = \frac{F}{V}(0 - X) + \mu X \qquad (24\text{-}26)$$

from which:

$$\mu = \frac{F}{V} - D \qquad (24\text{-}27)$$

This means that the dilution rate or the nominal residence time of the fermentor sets the growth rate of the biomass in the fermentor. A change in the dilution rate causes a change in the growth rate.

A similar balance around the fermentor at steady state for the growth limiting substrate, assuming the overall yield, *Y*, is constant, gives:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = 0 \text{ (steady state)} = \frac{F}{V}(S_0 - S) - \mu\frac{X}{Y} \qquad (24\text{-}28)$$

from which:

$$(S_0 - S) = \frac{X}{Y} \qquad (24\text{-}29)$$

Substituting D for μ in Monod growth kinetics yields:

$$S = \frac{K_s D}{\mu_{\max} - D} \qquad (24\text{-}30)$$

and substituting *S0 X/Y* for *S* gives:

$$X = Y\left(S_0 - \frac{K_s D}{\mu_{\max} - D}\right) \qquad (24\text{-}31)$$

Strictly speaking, Y is not a constant, particularly at conditions of low growth rate because the maintenance requirement, *m*, of the biomass utilizes substrate without producing biomass. The steady-state substrate balance equation for a single-stage fermenter can be restated as:

$$D(S_0 - S) = \frac{\mu X}{Y_G} + mX \qquad (24\text{-}32)$$

Table 24.2 illustrates the effect of maintenance and cell growth or reactor dilution rate on the overall cell yield.

TABLE 24.2 Effect of Maintenance on Overall Cell Yield in Continuous Culture

		Cell Yield, Y, g/g	
D, h-1	m = 0.01	m = 0.02	m = 0.05
0.01	0.333	0.250	0.143
0.02	0.400	0.333	0.222
0.05	0.455	0.416	0.333
0.10	0.476	0.455	0.400
0.20	0.488	0.476	0.444
0.50	0.495	0.490	0.476

Because most single-stage continuous fermentors are used to produce biomass, they are usually operated to optimize the biomass productivity. The unit volume biomass productivity of such a reactor is defined as *DX*. This unit volume productivity can be expressed as

$$DX = DY\left(S_0 - \frac{K_s D}{\mu_{\max} - D}\right) \qquad (24\text{-}33)$$

By taking the first derivative of the productivity expression with respect to the dilution rate and setting it equal to zero, the dilution rate of maximum productivity, *D*m, can be found as:

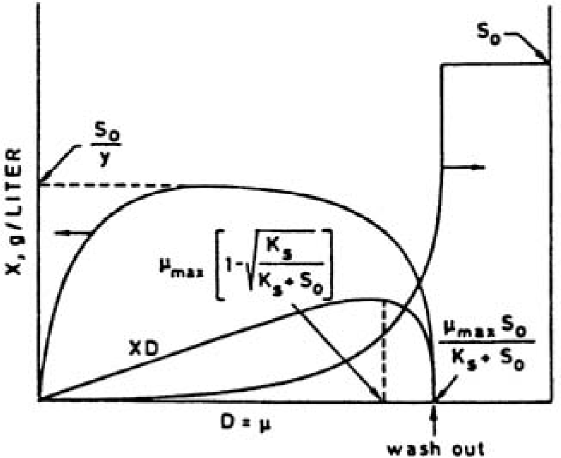
$$D_m = \mu_{\max}[1 - \sqrt{K_s/(K_s + S_0)}] \qquad (24\text{-}34)$$

and the maximum productivity, *D*m *X*, as:

$$\begin{aligned} D_m X \quad &= \quad \mu_{\max}[1 - \sqrt{K_s/(K_s + S_0)}] \\ &\quad \cdot Y[S_0 - K_s D/(\mu_{\max} - D)] \qquad (24\text{-}35) \end{aligned}$$

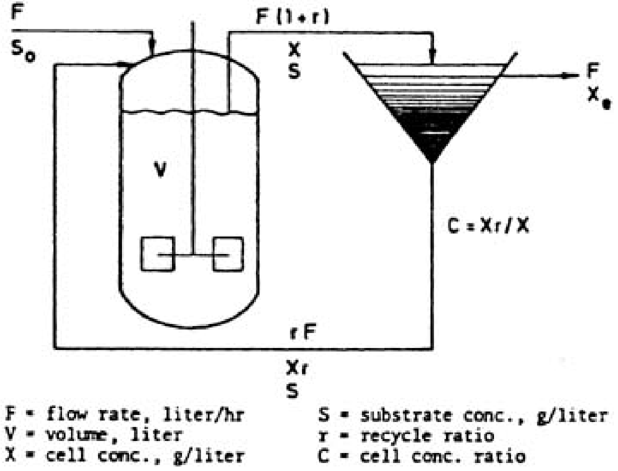
The behavior of these relationships is depicted in Fig. 24.8. Note that for this system "wash out" of the biomass occurs when *D* approaches the maximum growth rate, that is, $D \to \mu_{max}$.

Fig. 24.8 Relationships for a continuous single-stage fermentor.



Biomass recycle is frequently used in fermentors as a way of increasing the unit productivity. In order for the system to operate successfully, a biomass concentrator must be connected to the unit (see Fig. 24.9).

Fig. 24.9 Single-stage continuous fermentor with recycle.



Usually a centrifuge or settling tank is used as the concentrator. The increased productivity achieved with a recycle fermentor depends upon the recycle ration, *r*, and the cell concentration factor, *C* = *X*r/*X*, achieved in the concentrator. Equations expressing the recycle system behavior are derived from material balances around the reactor, that is, for the cell biomass balance at steady state:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = 0 = \frac{F}{V}(0) + \frac{rF}{V}X_r - \frac{F}{V}(1 + r)X + \mu X \qquad (24\text{-}36)$$

from which:

$$\mu = D\left(1 + r - r\frac{X_r}{X}\right) = D(1 + r - rC) \qquad (24\text{-}37)$$

Because (1 + *r rC*) < 1, it is possible to operate the system at dilution rates greater than the maximum growth rate. Utilizing the Monod equation, the following expressions for the growth limiting substrate, *S*, system effluent biomass concentration, *Xe*, and bio-mass concentration in the exit stream from the bioreactor, *X*, are obtained:

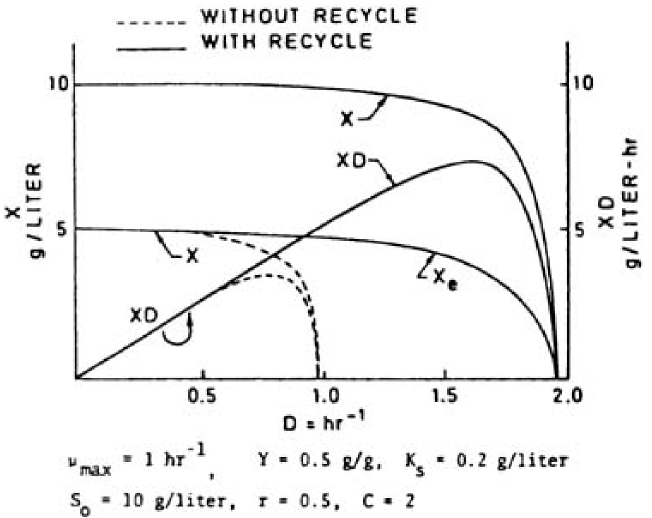
$$S = \frac{K_s D(1 + r - rC)}{\mu_{\max} - D(1 + r - rC)} \qquad (24\text{-}38)$$

$$X_e = Y(S_0 - S) \qquad (24\text{-}39)$$

$$X = \frac{Y(S_0 - S)}{(1 + r - rC)} \qquad (24\text{-}40)$$

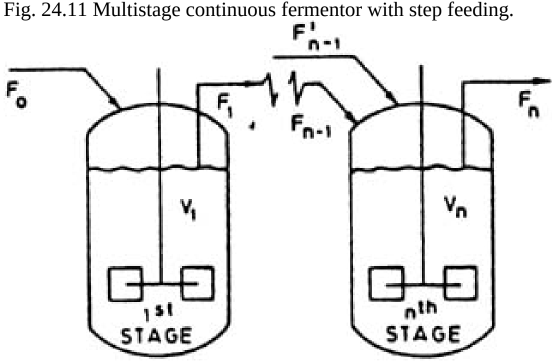
These relationships are illustrated in Fig. 24.10 for a particular set of conditions.

Fig. 24.10 Relationships for a single-stage continuous fermentor with recycle.



Another continuous fermentor system encountered in the microbial world is the multistage system with step feeding. This configuration is depicted in Fig. 24.11.

Fig. 24.11 Multistage continuous fermentor with step feeding.



In this system, fresh feed is step-fed, that is, fed to all reactor stages. This configuration is used in waste treatment systems to provide greater stability and to minimize the effects of substrate "shock loading" or inhibition through distribution of the concentrated feed along the system. A two-stage step feeding system is also used as a research tool to look at the inhibitory effects of high substrate loading, which cannot be achieved in the single stage without washout. The material balance relationships for step feeding systems are straightforward. What is usually troublesome is the selection of the proper kinetic expression for high substrate conditions.

Novel Bioreactors

The 1980s witnessed some innovative bioreactor designs and applications, largely stimulated by the need to commercialize therapeutic proteins whose production in large quantity was made possible by recombinant DNA and hybridoma technologies. Particularly worth mentioning are hollow fiber perfussion reactors used for the production of secreted products such as monoclonal antibodies, microcarrier culture for growth of anchorage dependent cells, and in situ rotating filters (shelves of screens) used for retention of cells and microcarriers and exchange of medium and sometime gases.

Oxygen Transfer in Fermentation Systems

In aerobic fermentations, oxygen is a basic substrate that must be supplied for growth. As in enzymatic reactions, the relationship between oxygen concentration and growth is of a Michaelis-Menton type. The specific rate at which cells respire (Q_{O_2}) increases rapidly with an increase in the dissolved oxygen concentrations (*C*) up to *C*crit, which typically ranges from 0.5 to 2.0 ppm for well-dispersed bacteria, yeast, and fungi growing at 2030°C. Beyond *C*crit, the specific oxygen uptake increases only slightly with increasing oxygen concentrations.

If Q_{O_2} is plotted against the specific growth rate (μ) of the microbe, a linear correlation is obtained. An intersection of this straight line with the ordinate is designed as $(Q_{O_2})_m$, which represents the oxygen required for cellular maintenance.

In mathematical terms, this relation is given by

$$Q_{\mathrm{O_2}}X = (Q_{\mathrm{O_2}})_{\mathrm{m}}X + (1/Y_{X/\mathrm{O_2}})(\mathrm{d}X/\mathrm{d}t) \qquad (24\text{-}41)$$

where Y_{X/O_2} is the yield based on cell mass, and *t* is the time.

This type of correlation can be applied to almost any substrate involved in cellular energy metabolism. It is justified from an energetic point of view and is supported by experimental steady-state data. Care must be used in applying this equation to transient conditions. It can be used only for situations at or near the steady-state equilibrium conditions.

When a product other than cell biomass is involved, the situation can be considerably more complex. Oxygen can be utilized for maintenance and growth as well as product formation. The simplest possible expression to represent the rate of oxygen utilization in this case is

$$\begin{aligned} Q_{\mathrm{O_2}}X \quad &= \quad (Q_{\mathrm{O_2}})_{\mathrm{m}}X + (1/Y_{X/\mathrm{O_2}})(\mathrm{d}X/\mathrm{d}t) \\ &\quad + (1/Y_{P/\mathrm{O_2}})(\mathrm{d}P/\mathrm{d}t) \end{aligned} \qquad (24\text{-}42)$$

where *P* is the product concentration.

Of course, the oxygen transfer rate under steady-state conditions must be equal to the oxygen uptake rate, that is:

$$Q_{\mathrm{O_2}}X = k_{\mathrm{L}}\mathrm{a}(C^* - C)_{\mathrm{mean}} \qquad (24\text{-}43)$$

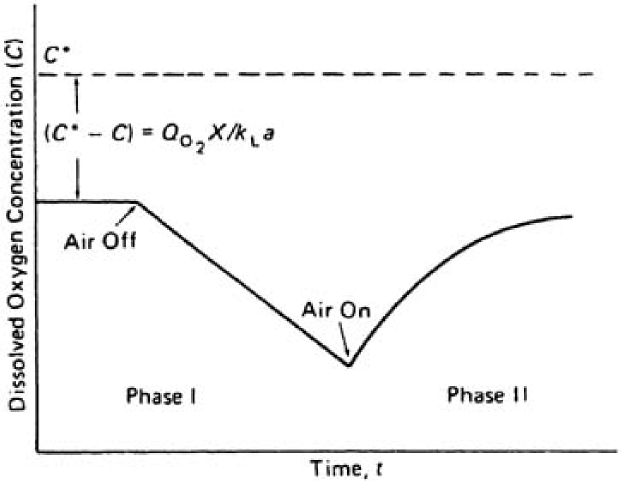
where *C** = concentration of oxygen in the liquid that would be in equilibrium with the gas-bubble concentration, and *k*La = the volumetric oxygen transfer rate.

Utilizing this relationship, the volumetric oxygen transfer coefficient can be estimated, that is:

$$k_{\mathrm{L}}\mathrm{a} = (Q_{\mathrm{O_2}}X)/(C^* - C)_{\mathrm{mean}} \qquad (24\text{-}44)$$

In practice, the dissolved oxygen concentration (*C*) is monitored by a membrane-covered probe. When the fermentation is equipped with a fast-responding probe (that is, the response time is less than 6 sec), then the fermentor can be used as a respirameter by making dynamic measurements of the oxygen concentration under aeration and nonaeration conditions. This is illustrated in Fig. 24.12.

Fig. 24.12 Determination of oxygen consumption and mass transfer rate by dynamic techniques.



The following equations represent these conditions. Air off, Phase I:

$$\mathrm{d}C/\mathrm{d}t = -Q_{\mathrm{O_2}}X \qquad (24\text{-}45\mathrm{a})$$

Air on, Phase II:

$$\mathrm{d}C/\mathrm{d}t = K_{\mathrm{L}}a(C^* - C)_{\mathrm{mean}} - Q_{\mathrm{O_2}}X \qquad (24\text{-}45\mathrm{b})$$

Once $Q_{O_2}X$ is obtained from d*C*/d*t* in the "air-off" phase, *k*L*a* can be calculated from the dissolved oxygen profile of the "air-on" phase.

Scale-Up

Many fermentation companies have the problem of scaling-up new fermentations as well as the translation of process-improvement data for well-established fermentations from laboratory operations to existing plant equipment. In general, fermentations are scaled up on the basis of achieving similar oxygen transfer capabilities in the plant equipment that proved to be optimal on the bench scale.

Fermentation biomass productivities usually range from 2 to 5 g/L/hr. This represents an oxygen demand in the range of 1.54 g O_2/L/hr. In a 500 m^3 fermentor this means achievement of a volumetric oxygen transfer coefficient in the range of 250400/hr. Such O_2 transfer capabilities can be achieved with aeration rates of the order of 0.5 VVM and mechanical agitation power inputs of 3.34.4 hp/m^3 (1.2l.6 hp/100 gal).

On scale-up, however, it is usually heat removal that causes design problems. With the aforesaid mechanical agitation power inputs, between 40 and 50 BTU/gal/hr of turbulent heat is generated. This, coupled with a peak metabolic heat release from the growing biomass of 120180 BTU/gal/hr, means that the fermentor must be capable of removing up to 160230 BTU/gal/hr. If the fermentation is a penicillin fermentation operated at 25°C, and if the best cooling water temperature that can be achieved is around 18°C, and if the overall heat transfer coefficient is in the range of 100 BTU/hr/°F/ft^2, there is no way that the heat will be removed in large fermentors without external heat exchange or extensive cooling coils in the fermentor. In highly viscous fermentations, internal cooling coils are usually not desirable because of interference with mixing patterns.

As a consequence, numerous schemes exist for heat removal in large fermentors including half coil baffles and draft tubes. No common design seems to have emerged. Manufacturers each have their own design scheme that they promote.

Power input to the liquid phase in a sparged system (*P*g) can be calculated from:

$$\frac{P_{\mathrm{g}}}{V} = \frac{Q_{\mathrm{g}}P_{\mathrm{g}}}{V}\,\frac{\eta\nu_o^2}{2} + \frac{RT}{M}\left(\ln\frac{\pi_o}{\pi}\right) \qquad (24\text{-}46)$$

The first term represents the energy dissipated at the sparger holes. The second term represents the energy involved to move the gas through the static liquid head. For well-designed spargers the first term can usually be neglected.

Unaerated mechanical power input (*P*mo) can be estimated from:

$$\frac{P_{\mathrm{mo}}}{V} = \frac{f\left(\rho N^3 D^5\right)}{V} \qquad (24\text{-}47)$$

The power factor in this correlation can be obtained from the paper of Rushton et al.

For estimation of mechanical power input to a sparged agitated system (*P*m), the correlations of Michel and Miller are widely used:

$$\frac{P_{\mathrm{m}}}{V} = \frac{0.706}{V}\left[\frac{P_{\mathrm{mo}}^2 N D^3}{Q_{\mathrm{g}}^{0.56}}\right]^{0.45} \qquad (24\text{-}48)$$

Although the separate effects of sparged and of mechanically agitated power input can be estimated, their effect is not simply additive. In general, a correlation of the following type is used to estimate their combined effect:

$$\frac{P_{\mathrm{e}}}{V} = \frac{P_{\mathrm{m}}}{V} + \frac{CP_{\mathrm{g}}}{V} \qquad (24\text{-}49)$$

The *C* correlation term usually varies from 0.05 to 0.4 in value. The data of Miller can be used to estimate the value of the constant.

There are various correlations between *kLa* and power inputs. Some design engineers prefer to scale-up on the following basis:

$$k_L a = k(P_m/V)^{\alpha}(v_s)^{\beta} \qquad (24\text{-}50)$$

where α = 0.95 and β = 0.67 for pilotplant equipment and α = 0.4 and β = 0.5 for large-scale plant equipment. When such a basis is used for scale-up, the designer attempts also to maintain constant tip speed on scale-up. Generally, these run from 600 to 1200 ft/min in most fermentors, with 1000 ft/min a typical value.

Once a plant is built, the conditions of agitation, aeration, mass (oxygen) transfer, and heat transfer are more or less set. Therefore, it has been suggested that the problem of translating process improvements is one not of scale-up but rather of scale-down. Those environmental conditions achievable in plant-scale equipment should be scaled down to the pilot-plant and screen-size equipment (shake-flask) to ensure that the studies are carried out under conditions that can be duplicated.

Scale-up of mammalian cells creates different dimensions of challenges for engineers. First of all, most mammalian cells with their lack of protecting cell wall are very sensitive to shears that are generated with high-speed agitation. Some of the cells are so sensitive to shear that they are damaged upon the collapse of gas bubbles. Sparging, a common aeration mechanism used in the fermentation industry, thus becomes unsuitable for "aerating" these cells. So, how should one go about mixing and gassing mammalian cell cultures? Low-shear marine propellers or pitched blades generally are used to provide gentle mixing. A hemispherical, instead of the conventional, dished bottom, is sometimes incorporated into the bioreactor design. Gas-permeable silicone tubing is used for "bubbleless" aeration or oxygenation. And the rotating filters (shieves), mentioned earlier, used for cell retention and medium exchange, can be conveniently used as a gassing cage. An external loop including an oxygenation chamber is just another means of introducing oxygen to the system without directly sparging on the cells.

Air Sterilization

Submerged aerobic fermentation processes require a continuous supply of large quantities of air. Sterilization of this air is mandatory in many fermentations. For pure culture operation, incomplete destruction or inadequate removal of the microorganisms carried in the air may preclude successful operation. Many ways have been suggested for sterilizing air. Only filtration through beds of fibrous and membrane-type materials have found widespread usage on an industrial scale. Examples would be filtration through beds of fibrous materials such as glass wool or plate-type filters made of polyvinyl alcohol several centimeters thick. In recent years sufficient research has been carried out to permit the design of these fibrous filters on a rational basis. Major requirements that every air sterilization system must satisfy are as follows:

1. The system should be simple in design.
2. It should not be inordinately costly to operate.
3. It should remove or destroy airborne contamination to the extent necessary for satisfactory fermentation performance.
4. It should be stable to repeated steam or chemical vapor sterilization.
5. It should condition the air.
6. Its ability to maintain a sterile air supply should not be jeopardized by power failure or compressor surges.

This last requirement is frequently over-looked. Its consideration is paramount in the design of filters compounded from fibrous materials. For a particular filter there is an intermediate air velocity at which filtration efficiency is a minimum. If the filter design is based upon a performance observed at an operating velocity other than that at which minimum efficiency occurs, surges or brief power failures could create periods of operation at lower than designed-for efficiencies. Hence the filter design should be based on satisfactory operation at this point of minimal efficiency.

Minimal efficiency at an intermediate air velocity occurs because different forces act to collect airborne particles at different velocities. At low velocities, gravitational, diffusional, and electrostatic forces act on the particle. Their effect is inversely proportional to air velocity. At high velocities, inertia forces come into play, and they are directly proportional to air velocity. The nature of inertial effects is such that below a certain air velocity, collection due to inertial forces is zero.

This velocity can be estimated by the following relation:

$$V_{\text{min eff}} = \frac{1.125\,\mu d_f}{C\rho_p d_p^2} \qquad (24\text{-}51)$$

where μ = air viscosity; *df* = fiber diameter; *C* = Cunningham correction factor; ρp = particle density; and *d*p = particle diameter.

For the collection of unit-density, 1 μm sized bacterial particles from air streams at room temperature and pressure. This velocity is equal to

$$V_{\text{min eff}} = 0.066 d_f \qquad (24\text{-}52)$$

where velocity *V* is in feet per second and the fiber diameter, *df*, is in micrometers.

Regardless of air velocity, some collection always occurs because airborne particles possess a finite size and will be intercepted by some fiber blocking an air stream along which a particle moves.

Collection must always be greater than that due to interception because it represents the minimum collection physically possible. In the absence of experimental data, a reasonable estimate of the minimal collection efficiency of a particular filter is that estimated solely from interception effects calculated at a velocity just below which inertia effects are zero.

The theoretical expression for the collection efficiency of a single isolated fiber, due solely to interception effects, can be estimated from

$$\begin{aligned} n_o \quad &= \quad 0.5[1/(2 - \ln N_{\text{Re}})][2(1+R)\ln(1+R) \\ &\qquad - (1+R) + 1/(1+R)] \qquad (24\text{-}53) \end{aligned}$$

where *no* = collection efficiency of a single isolated fiber; *R* = *d*p/*d*f; and *N*Re = Reynolds number.

From experimentation with aerosols it has been found that the effectiveness of a filter can be expressed by

$$\ln\frac{N_1}{N_2} = \frac{1.27 n_o(1+4.5\alpha)\alpha L}{(1-\alpha)d_f} \qquad (24\text{-}54)$$

where: *N*1 = total number of particles entering the filter; *N*2 = total number of particles penetrating the filter; *L* = filter thickness; and α = volume fraction of fibers in the filters.

This equation is a reasonable basis for air sterilizing filter design. The single fiber efficiency, *no,* to use in this equation when experimental data are lacking is that based on collection due solely to interception effects estimated at an air velocity where minimal filtration efficiency is expected.

Although depth filters are still in use, because of the interest in maintaining absolute asepsis particularly for long-term culture and for ease of operation, membrane-type absolute filter cartridges are becoming more popular. In this case, a prefilter for removal of particulates and liquid droplets is usually placed upstream of the absolute filter. Parallel installation of the filters prevents a total shutdown of the fermentation in the event of filter clogging.

Medium Sterilization

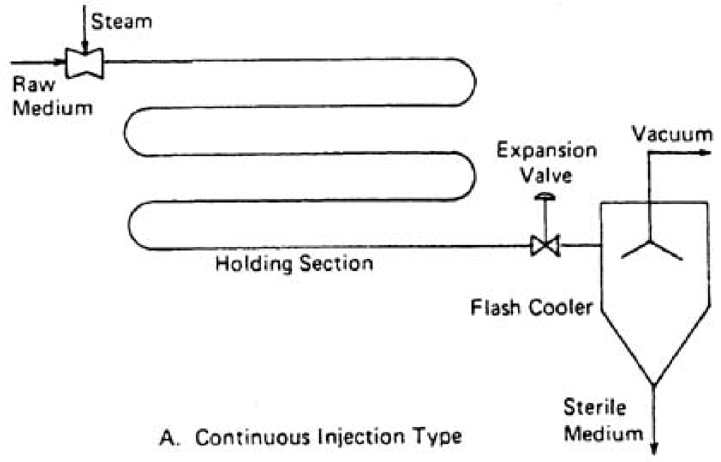
Microorganisms can be removed from fluids by mechanical methods, for example, by filtration, centrifugation, flotation, or electro- static attraction. They also may be destroyed by heat, chemical agents, or electromagnetic irradiation. Although cells may be disrupted or killed by mechanical abrasion on a small scale, this method is not satisfactory industrially. Similarly, X-rays, β rays, ultraviolet light, and sonic irradiations, although useful in the laboratory, are not applicable to the sterilization of large volumes of fluids.

Although antibacterial agents have an important place in the fermentation industry, particularly for the production of a pure water supply, they have little application for the sterilization of fermentation media. Despite the fact that heat sterilization of media is the most common method, little attention has been paid until recently to the engineering aspects of heat sterilization.

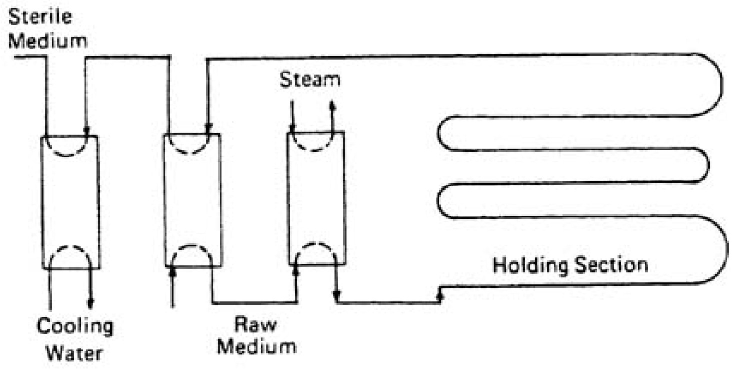
Interest in continuous methods of sterilizing media is increasing, but for the successful operation of a continuous sterilizer, foaming of the media must be carefully controlled and the viscosity of the media must be relatively low. Figure 24.13 illustrates two types of continuous media sterilizers that have been utilized in the fermentation industry. The advantages of continuous sterilization of media are as follows:

1. Increase of productivity because the short period of exposure to heat minimizes damage to media constituents.
2. Better control of quality.
3. Leveling of the demand for process steam.
4. Suitability for automatic control.

Fig. 24.13 Two types of continuous sterilizers.



A. Continuous Injection Type



B. Continuous Plate Exchanger Type

At present, most media in the fermentation industry are sterilized by batch methods. Overexposure of the medium to heat is inherent in batch sterilization processes. But continuous sterilization, when properly operated, can minimize damage to the medium.

Design and operation of equipment for sterilizing media are based on the concept of thermal death of microorganisms. Consequently, a understanding of the kinetics of the death of microorganisms is important to the rational design of sterilizers. The destruction of microorganisms by heat implies loss of viability, not destruction in the physical sense. The destruction of organisms by heat at a specific temperature follows a monomolecular rate of reaction:

$$\frac{dN}{dt} = -KN = -(Ae^{E/RT})N \qquad (24\text{-}55)$$

where *k* = reaction rate constant, time-1; *N* = number of viable organisms/unit volume; *t* = time; *T* = absolute temperature; *E* = energy of activation for death; *R* = gas law contact; and *A* = constant.

This equation can be integrated to yield the design equation:

$$\ln\frac{N_o}{N_f} = A\int_0^{t_s} e^{-E/RT}\,dt \qquad (24\text{-}56)$$

where *No* = number of contaminating organisms in the total fermentation medium to be sterilized; *N*f = level of contamination that must be achieved to produce the desired degree of apparent sterility; and *t*s = sterilization time.

In estimating the medium sterilization time, one must define the contamination, the desired degree of apparent sterility, and the timetemperature profile of the medium, that is, *T* = *f*(*t*). For typical bacterial spore contaminants, the constants used in most designs have the following values:

$$\begin{aligned} E \quad &= \quad 68{,}700\,\text{cal/g-mol} \\ R \quad &= \quad 1.987\,\text{cal/g-mol, °K} \\ A \quad &= \quad 4e^{+87.82},\ \text{min}^{-1} \end{aligned}$$

The medium used for culturing mammalian cells must be filter-sterilized. For mammalian cell culture where viral attacks are more of a concern than for microbial fermentations, filters of 0.45 and 0.1 μm, placed in series, are commonly used.

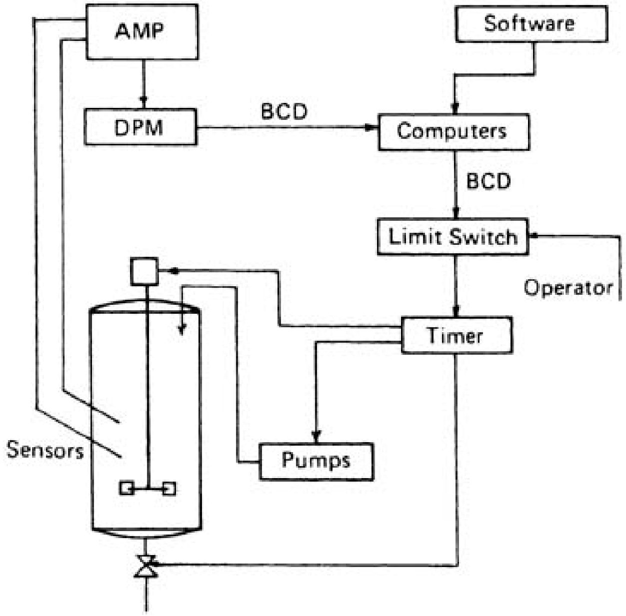
Instrumentation and Control

In scaling up any successful fermentation, the molecular biologist would suggest the necessity for controlling the environment and hence regulating the fermentation. However, two problems arise. First, knowledge of the regulatory mechanisms of metabolic pathways that produce the desired product may not exist. In fact, the metabolic pathways may not be fully known or understood. Second, even if the pathways and the regulatory mechanisms are known, the necessary instrumentation to detect regulatory metabolites may not exist. The design engineer, thus, is presented with a dilemma in developing a new fermentation or improving an existing process.

Should the engineer focus on researching the mechanisms of product regulation and control, developing needed analysis and sensing instrumentation to provide this control, or should he or she use trial-and-error development procedures for strain and medium selection to evolve an apparent optimal environment for plant-scale production? The answer is that the engineer must do both.

In the past, metabolic controls were simply bred into or out of fermentation organisms through mutation and strain selection. However, tremendous strides have been, and are being, made in sensor development. The design engineer will soon be able to rely more on environmental control than before in order to gain economical fermentation results. Figure 24.14 shows how a highly instrumented fermentor, which is designed to secure basic information on environmental control, can be coupled to a computer for data analysis and control.

Fig. 24.14 General layout of a highly instrumented computer-coupled fermentor.



To achieve meaningful fermentation control, the following three steps are necessary:

1. Carry out fermentation research on fully monitored environmental systems.
2. Correlate the environmental observations with existing knowledge of cellular control mechanisms.
3. Reproduce the desired environmental control conditions through continuous computer monitoring, analysis, and feedback control of the fermentation environment.

Until recently fermentation control was limited to that of temperature, pH, and aeration. With the development of numerous sensors and inexpensive personal and mini-computers, the engineer can think in terms of sophisticated control systems for fermentation processes.

Indeed, the engineer can even think of utilizing computers to perform on-line dynamic optimization.

New instruments are still needed. The most important sensor needed is a reliable biomass monitoring device that can be sterilized.

An IR fiber-optic cell density probe has enjoyed some success. It has been used to monitor cell growth directly (without dilution) in high-cell-density bacterial fermentations. Still, there is a need for a glucose, a nitrogen substrate, and a phosphate sensor that can withstand repeated system sterilization. However, the ability to do on-line sample filtration through the use of hollow fibers or rotating filters has made possible the continuous, on-line measurement of glucose, lactate, and so on.

Indirect measurement via computers appears to be a viable alternative to measurement involving sampling. Certain sensor information can be combined to give additional information such as oxygen-uptake rate, carbon-dioxide evolution rate, and respiratory quotient. These measurements can be thought of as "gateway" measurements because they make possible the calculation of additional information (Table 24.3).

TABLE 24.3 Gateway Measurements[a]

Measurement	Result
pH	Acid product formation rate
Air flow rate In and out O_2 concentration	O_2 uptake rate
Air flow rate Out CO_2 concentration	CO_2 evolution rate
CO_2 evolution rate O_2 uptake rate	Respiratory quotient
Power input Air flow rate	O_2 transfer rate

[a] S. Aiba, A. E. Humphrey, and N. Millis, *Biochemical Engineering*, 2nd ed., p. 332, University of Tokyo Press, Tokyo, 1973.

Further, the indirect measurement of a given component can be achieved by material-balancing that component around the fermentor. If a model for utilization of that component for biomass or product formation is known, then either the biomass or the product level can be estimated by computer summation or integration of the data using the model.

Besides these uses, the computer has application in fermentation processes for continuous nonprejudicial monitoring and (most important) continuous feedback control and dynamic optimization of the process.

Recovery of Fermentation Products

From the amount of space devoted to fermentor design and scale-up, one might gather that the recovery processes of fermentation products are rather straightforward and relatively simple. Nothing could be further from the truth. In one case of an antibiotic production plant, the investment for the recovery facilities is claimed to be about four times greater than that for the fermentor vessels and their auxiliary equipment. As much as 60 percent of the fixed costs of fermentation plants are attributable to the recovery portion in organic acid and amino acid fermentations.

Figure 24.15 shows a typical recovery process for antibiotics, and Fig. 24.16 presents a flowsheet for an enzyme plant. It is apparent from these diagrams that most recovery processes involve combinations of the following procedures:

1. Mechanical separations of cells from fermentation broth.
2. Disruption of cells.
3. Extraction.
4. Preliminary fractionation procedures.
5. High-resolution steps.
6. Concentration.
7. Drying.

Fig. 24.15 Basic flow sheet for the recovery of antibiotics. (*Ind. Eng. Chem.*, 49, 1494, 1957. Copyright American Chemical Society.)

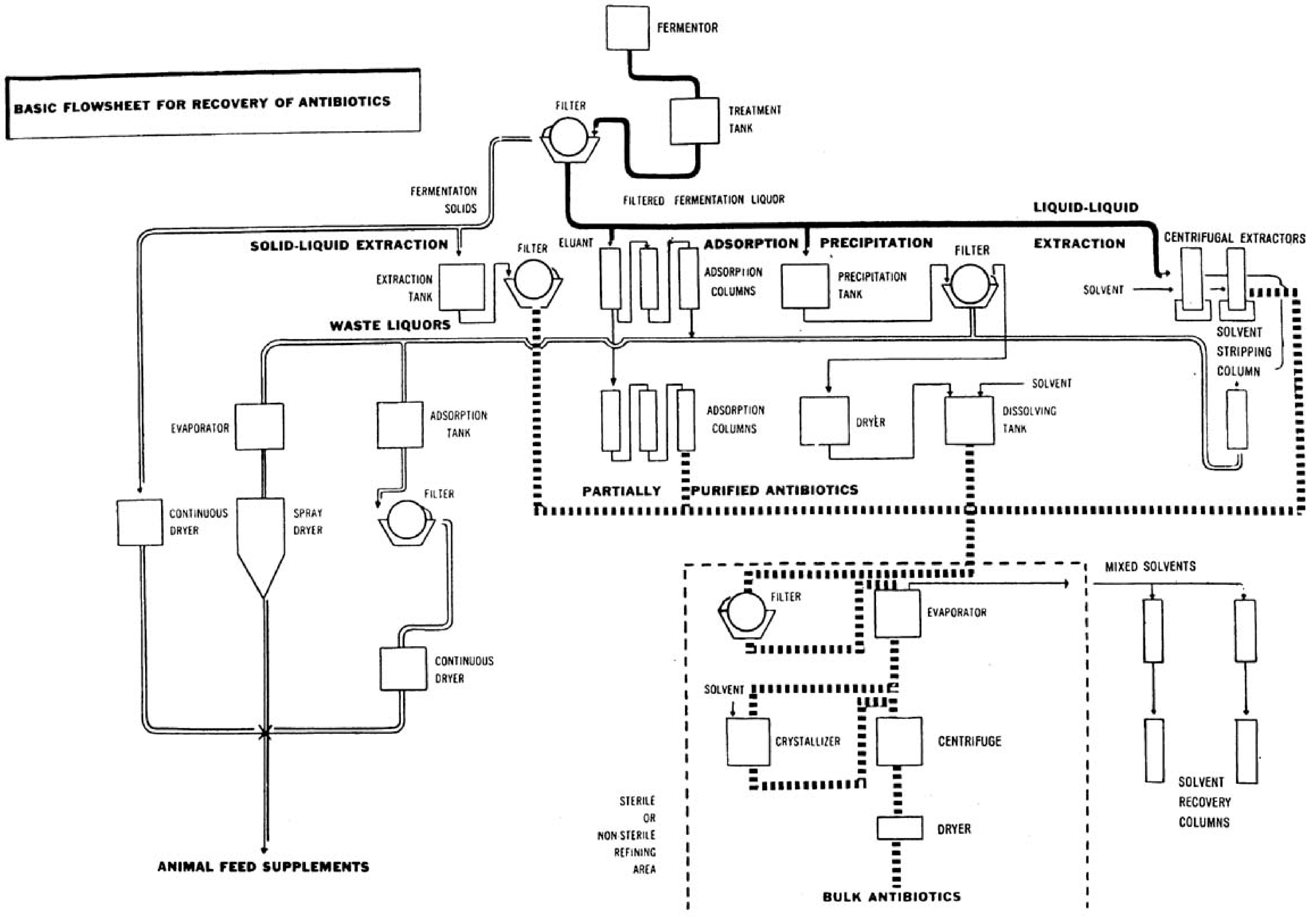
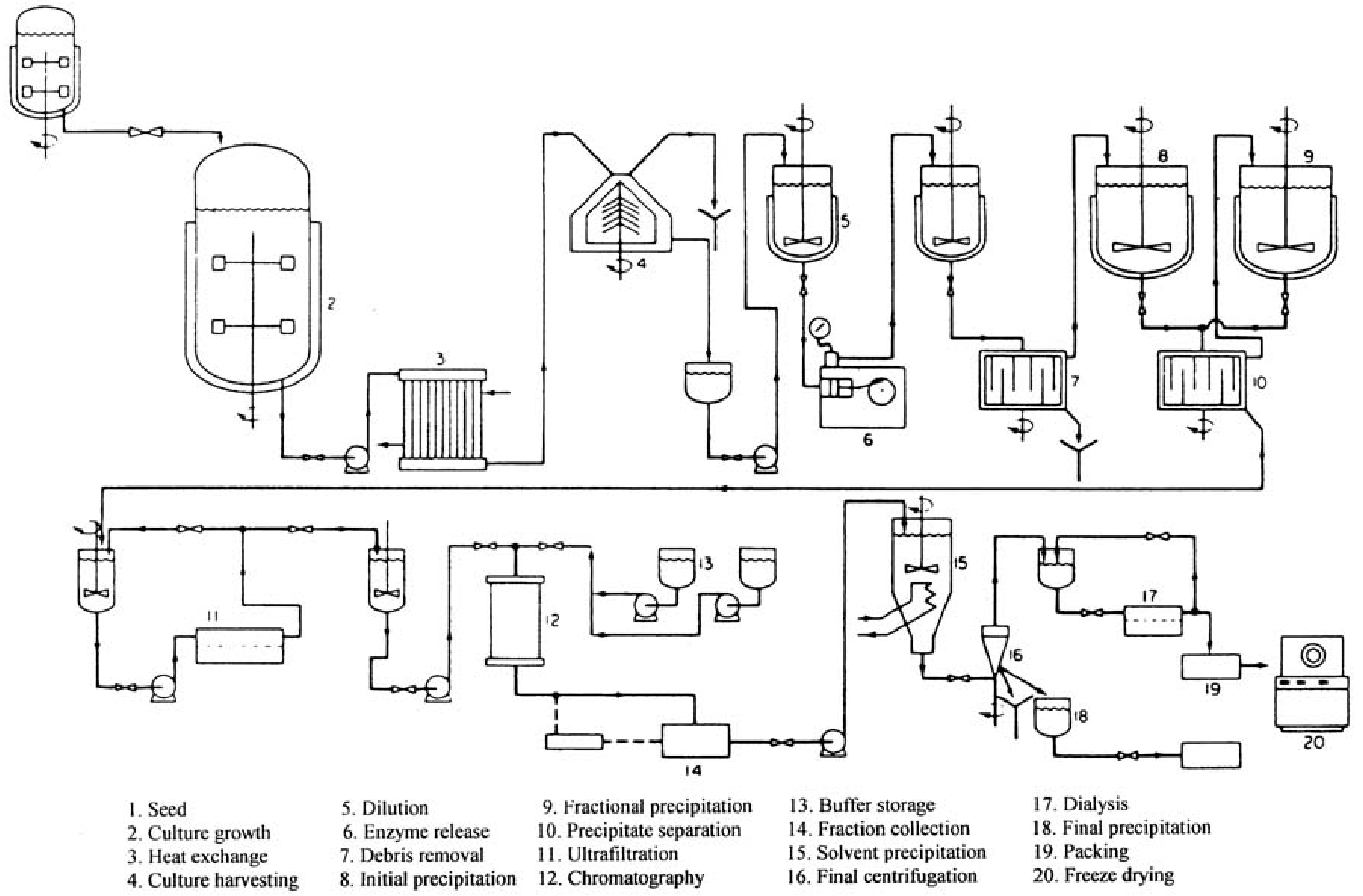


Fig. 24.16 Basic flow sheet for an enzyme production plant. (*Process Biochem.*, 5, 22, 1970.)



Chromatography technology has advanced rapidly in the last decade. Various separation principles, gel filtration, size exclusion, ion exchange, hydrophobic interaction, and affinity binding have been applied in the chromatographical mode. Many different pressure-sustaining support matrices have been developed for high-throughput applications. All these techniques have greatly expanded the tools now available for high-resolution separation.

Immobilized Enzymes and Cells

Most of the soluble hydrolytic enzymes in common use in industry are formed extra-cellularly; the wide range of intracellular microbial enzymes are virtually unexplored commercially. In order to exploit these enzymes, it is necessary to develop economic methods of purification. Reports of continuous methods for harvesting, breaking cells, and fractionating their protein represent major advances toward this end. Recently, enzymes as well as whole cells have been immobilized by adsorption, encapsulation, or inclusion in gels. They may also be covalently bound with a bifunctional linking agent to an insoluble polymer, covalently cross-linked with themselves, or bound directly in an enzymepolymer complex. The preparation can be packed into a column or complexed onto porous sheets, and the substrate can then react with the enzyme in a batch or a continuous process. Enzymes can also be contained within an ultrafiltration membrane. For instance, starch can be fed continuously into a membrane structure containing α-amylase and the products of hydrolysis collected outside the membrane. Whole cells or crude extracts also can be placed in capsules that allow substrate and products to diffuse to the enzymes. Some properties of the enzyme, such as the pH optimum, Michaelis constant, or stability, may change when it is immobilized, but the economies achieved by retaining the enzyme for reuse and obtaining products free of contaminating enzyme greatly outweigh any loss of stability or reduction in the rate of reaction. In many cases, the life of the immobilized enzyme is greater than that of the soluble enzyme. If substrates or products are sensitive to pH, it may be possible to select a polymer that changes the optimum pH of the immobilized enzyme to a range more favorable to the stability of the reactants.

24.3 Organic Acids

There are many organic acids that can be produced by microbial or biochemical means. The major ones are listed in Table 24.4. However, only acetic acid (as vinegar), citric acid, itaconic acid, gluconic acid, and 2-ketogluconic acid are produced industrially by fermentation now. Other organic acids, such as fumaric, gallic, lactic, malic, and tartaric acids, once produced by fermentation or enzyme processes, are now produced commercially, predominantly by the more economic means of chemical synthesis.

TABLE 24.4 Major Organic Acids that can be Produced by Fermentation

Organic Acids	Producing Microorganisms	Substrates	Yields %
Acetic	*Acetobacter acetic*	Ethanol	95
	Clostridium thermoaceticum	Glucose	90
Araboascorbic	*Penicillium notatum*	Glucose	45
Citric	*Aspergillus niger*	Sucrose	85
	Candida lipolytica	*n*-Paraffin	140
Fumaric	*Rhhizopus delemar*	Glucose	58
Gluconic	*Aspergillus niger*	Glucose	95
Isocitric	*Candida brumptii*	Glucose	28
Itaconic	*Aspergillus terreus*	Glucose	60
2-Ketogluconic	*Pseudomonas fluorescens*	Glucose	90
5-Ketogluconic	*Gluconobacter suboxydans*	Glucose	90
α-Ketoglutaric	*Candida hydrocarbofumarica*	*n*-Paraffin	84
Kojic	*Aspergillus oxyzae*	Glucose	50
Lactic	*Lactobacillus delbriickii*	Glucose	90
Malic	*Lactobacillus brevis*	Glucose	100
Propionic	*Propionibacter shermanii*	Glucose	60
Pyruvic	*Pseudomonas aeruginosa*	Glucose	50
Salicylic	*Pseudomonas aeruginosa*	Naphthalene	94
Succinic	*Bacterium succinicum*	Malic acid	57
Tartaric	*Gluconobacter suboxydans*	Glucose	27
Xylonic	*Enterobacter cloacae*	Xylose	90

Acetic Acid and Vinegar

Acetic acid as a chemical feedstock is manufactured by chemical synthesis. Acetic acid in the form of vinegar (at least 4% acetic acid by law) is produced largely via the oxidation of ethanol by bacteria of the *Acetobacter* genus.

Synthetic acetic acid is currently produced by several different routes:

1. Wood distillation (being phased out).
2. Oxidation of acetaldehyde (Wacker process).
3. Liquid-phase oxidation of *n*-butane.
4. Carbonylation of methanol (Monsanto process).

The last method is the most economical among the four, and is employed in most new acetic acid manufacturing facilities throughout the world. Monsanto was awarded the Kirkpatrick Chemical Engineering Merit Award in 1976 for the development of this process.

The U.S. annual production of acetic acid was about 3.4 billion lb in 1987, and the selling price around $0.30/lb. The world production totaled 10 billion lb. The major U.S. producers are Celanese, Monsanto, USI, Union Carbide, Eastern, and Borden.

The price for acetic acid has been in the range of 2347 cents/lb (19952001), technical grade, by tanks delivered. Worldwide demand was about 13 billion lb in 1997. Demand for acetic acid in the United States has been strong (6.2 billion lb in 1999 and going up to an estimated nearly 7 billion lb in 2003), but the Asian markets remain depressed due to the downturn in the economy. The major producers include Celanese, Millennium, Sterling, and Eastern in the United States, BP and Celanese in Europe, Kyodo Sakussan and Showa Denko in Japan, Samsung-BP in Korea, and China Petrochemcial and Chang Chun Petrochemcial in Taiwan.

Process routes via synthesis gas using the Monsanto methanol carbonylation technology, and natural gas and methanol as main raw materials mostly account for worldwide capacity. Some recent, drastic cost improvements have been achieved through the Celanese "low water" process and the BP iridium catalysis process. The new technologies allowed plants to double or triple in capacity without significant capital outlay. As the fixed cost decreases, the cash cost for acetic acid has been brought down to less than 10 cents/lb, particularly at the new mega plants. Other technologies remain competitive. Celanese's Pampa, Texas, plant uses the butane liquid-phase oxidation technology. Showa Denko, in 1997, launched a 100,000-metric tons/year plant based on direct oxidation of ethylene.

High-tonnage production of acetic acid for industrial purposes was earlier done through fermentation. In 1915, USI of Baltimore, Maryland constructed and operated two of the largest acetic acid fermentation plants ever built, with a combined tank capacity of about 20 million gal. The acetic acid was converted with lime to calcium acetate, and the latter pyrolyzed to acetone. The plant operated for a number of years until competition from acetone produced by direct fermentation made the process uneconomical. Today, fermentive acetic acid is limited to vinegar production only.

Vinegar is one of the oldest known fermentation products, predated only by wine and possibly by certain foods from milk. First derived from the spoilage of wine, vinegar has been used as a condiment, food preservative, medicinal agent, primitive antibiotic, and even today as a household cleansing agent. The present production of vinegar is used almost entirely in foods.

Vinegar may be defined as the product of a double fermentation: an alcoholic fermentation of a sugary mash by a suitable yeast (usually a selected strain of *Saccharomyces cerevisiae* or *ellipsoidens*) and a second fermentation to oxidize the alcohol to acetic acid by a suitable culture of *Acetobacter* organisms. The theoretical maximum of acetic acid yield on glucose is 67 percent (two moles of acetic acid produced from every mole of glucose consumed) by this route. A homo-fermentative culture, *Clostridium thermo-aceticum*, is known to be capable of fixing CO2 and yielding three moles of acetic acid from one mole of glucose under anaerobic conditions. The technology for this process has not been commercialized, however.

Vinegar can be made from a variety of fermentable substances, the essential requirements being that they are satisfactory and economical sources of ethanol. The commonly used substances include fruits and their juices, cereals, sugar syrups, and synthetic ethanol. There are four major types of vinegar: white distilled, cider, wine, and malt vinegar. Almost the entire U.S. production of white distilled vinegar is derived from synthetic ethanol. Cider vinegar is made by the alcoholic and subsequent acetous fermentation of the juice of apples or its concentrate; wine vinegar is made of grape wine; and malt vinegar is made of a mash containing malt, corn, and/or barley.

Several vinegar manufacturing processes are commercially used, including the following:

1. Circulating generators (trickling genertators).
2. Frings Acetators (submerged culture generators).
3. Yeomans Cavitators.
4. Tower fermentors (column fermentors).

The circulating, trickling generator is still widely used. It is a large tank constructed in a variety of dimensions, generally of wood, including redwood and fir but preferably cypress. The vertical timbers are held in place with steel hoops. A false bottom supports curled beachwood shaving above the lower one-fifth of the tank, which serves as a collection reservoir. Air is supplied by a simple fan-type blower, and is distributed to the generator by a number of equally spaced inlets just beneath the false bottom. An airflow rate of about 0.015 vvm (volume of air per volume of packing per minute) is adequate. A pump circulates the ethanolwateracetic acid mixture from the collection reservoir up through a cooler to a distributing sparger arm in the top of the tank. The liquid trickles down through the packing and returns to the bottom reservoir. Cooling water to the cooler is regulated to maintain the temperature of the generator around 29°C at the top and below 35°C at the bottom. A portion of the finished vinegar is periodically withdrawn from the reservoir, and replaced with the ethanol-containing charge. The ethanol concentration in the generator should not exceed 5 percent or fall below 0.2 percent. If ethanol is depleted in the generator, the *Acetobacter* will die, and the generator becomes inactive.

The Frings Acetator (produced by the Heinrich Frings Company of Bonn, Germany) consists of a stainless steel tank with internal cooling coils, a high-speed, bottom-entering agitator, and a centrifugal foam breaker. The unique feature of this Acetator is its highly efficient method of supplying air. This is accomplished by means of a high-velocity self-aspirating rotor that pulls air in from the room to the bottom of the tank. The equipment is operated batchwise. When the ethanol content falls to 0.2 percent by volume, about 3540 percent of the finished product is removed. Fresh feed is pumped in to restore the original level, and the cycle starts again. Cycle time for 12 percent vinegar is about 35 hr. The rate of production can be as much as ten times as great as that obtained in the trickling generator of equivalent size. The yield on ethanol is higher. Values of 94 and 85 percent have been reported, respectively, for the Acetator and the trickling generator. But much more extensive refining equipments are necessary for filtering vinegar produced by the submerged process because the mash contains the bacteria that produced it.

The Yeomans Cavitator (no longer being manufactured) is a submerged culture system somewhat similar to the Acetator but differing in the way in which air is supplied. This generator has a top-driven rotor-agitator. It withdraws liquid and air from a centrally located draft tube. The system can be easily installed in existing wooden vats and adapted to continuous production. Technical difficulties have forced its abandonment. Some units are still in use to produce vinegar in the continuous mode. A 98 percent efficiency of ethanol to acetic acid has been achieved in commercial operations.

The tower fermentor is a relatively new aeration system applied to vinegar production. The fermentor is constructed of polypropylene reinforced with fiber glass. Aeration is accomplished through a plastic perforated plate covering the cross section of the tower and holding up the liquid. The cost of the tower fermentor is said to be approximately half that of a Frings Acetator of equivalent productive capacity. It has been reported that the tower fermentor is satisfactory for producing all types of vinegar.

Vinegar clarification is accomplished by filtration, usually with the use of filter aids and/or fining agents such as diatomaceous earth or bentonite, respectively. After clarification, vinegar is bottled, sealed tightly, pasteurized at 6065°C for 30 min, and then cooled to 22°C. Vinegar can be concentrated by a freezing process. Vinegar of 200-grain strength is readily obtainable from 120-grain raw vinegar. There are a number of advantages of 200-grain vinegar. The acid strength of brine solutions that would have had to be discarded because of dilution by pickle juice can easily be increased by pickle processors. Transportation costs are substantially reduced, as well.

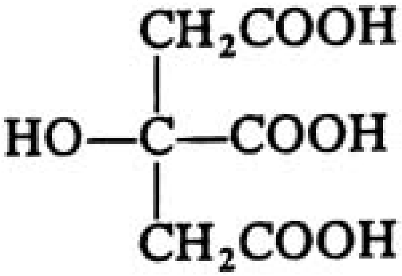
Nakano Vinegar is the largest manufacturer of food vinegar in Japan. The company is famous for its brand name Mitukan. Nakano, in collaboration with Professor T. Beppu, University of Tokyo, has cloned alcohol dehydrogenase and aldehyde dehydrogenase, and established phenotypic transformation of *Acetobacter aceti*. The acetic acid bacteria with the aldehyde dehydrogenase gene produced acetic acid more rapidly than those with the alcohol dehydrogenase, and were more resistant to high concentrations of acetic acid.

Since 1993, a two-stage submerged fermentation has been used to produce vinegar of more than 20 percent acetic acid. In the first fermenter, alcohol is added slowly to a total concentaration of about 18.5 percent. After the acetic acid concentration has reached 15 percent about 30 percent of the liquid from the first fermenter is transferred to a second fermenter. In the second fermenter the fermentation carried out at a reduced temperature of 1824°C continues until the alcohol is almost depleted.

The world production (excluding China and Russia) of 10 percent vinegar is estimated to be about 2 billion L a year or about 200,000 metric tons of acetic acid. The last time the U.S. vinegar industry published vinegar production figures was in 1987. The production volumes were 581 million L of distilled vinegar, 54 million L of cider vinegar, and 126 million L of other types of vinegar. In 1992, the European Union produced 325 million L of alcohol vinegar, 143 million L of wine vinegar, and 45 million L of other kinds of vinegar. And, in 1992, Japan produced a total of about 400 million L of vinegar.

Citric Acid

Citric acid, whose structure is shown below, is the most important organic acid produced by means of fermentation.



$$\begin{array}{c} CH_2COOH \\ | \\ HO-C-COOH \\ | \\ CH_2COOH \end{array}$$

In the food and beverage industries, citric acid is used in soft drink mixes, in carbonated and still beverages, candies, wines, desserts, jellies, and jams, as an antioxidant in frozen fruits and vegetables, and as an emulsifier in cheese. These uses represent 64 percent of the use of citric acid. As the most versatile food acidulant, citric acid accounts for about 70 percent of the total food acidulant market.

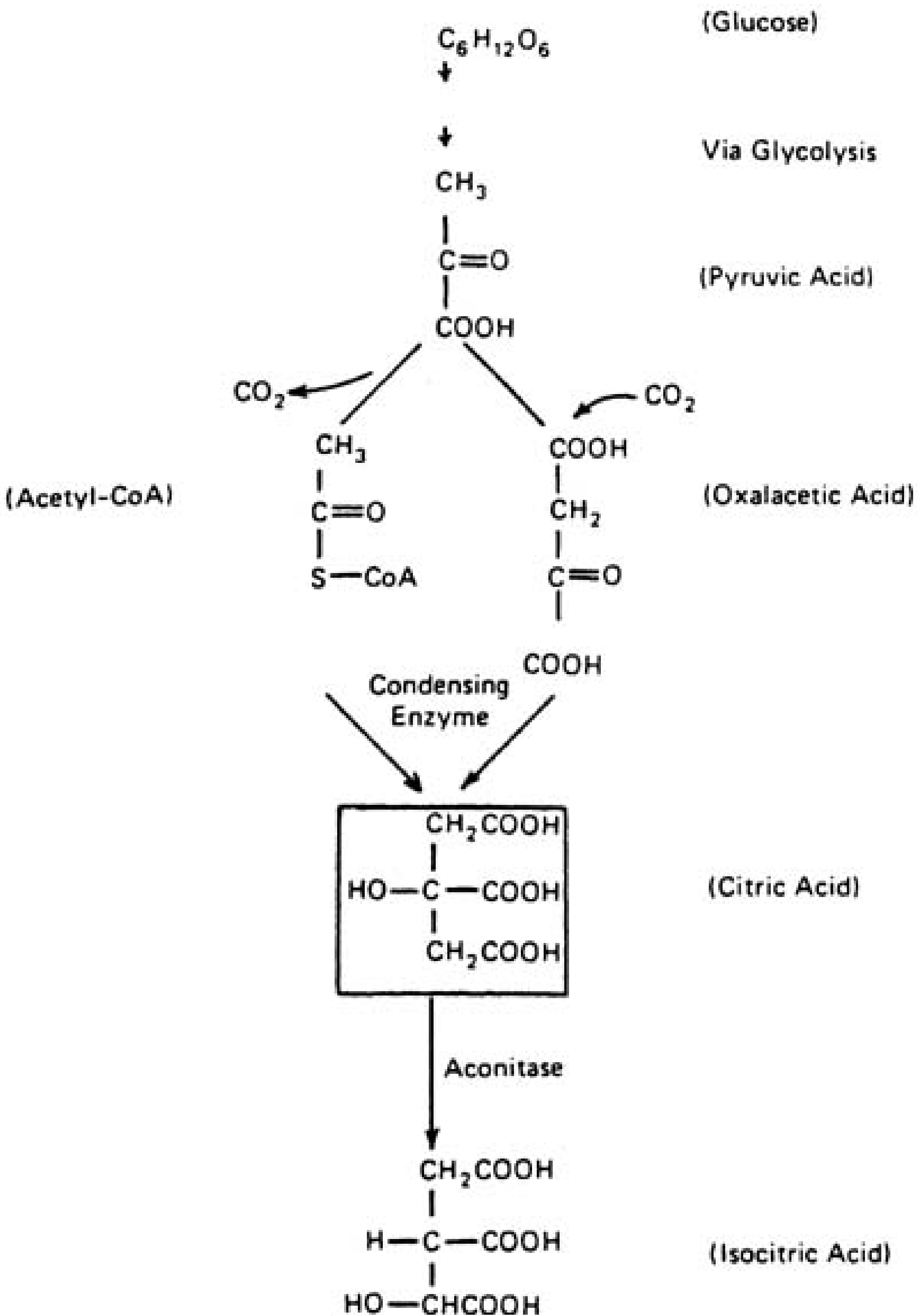
In terms of pharmaceutical and cosmetic applications, citric acid provides effervescence in oral dosage forms and in products for external use by combining the citric acid with a biocarbonate/carbonate source to form carbon dioxide. Citric acid and its salts are also used in blood anticoagulants to chelate calcium, block blood clotting, and buffer the blood. Citric acid is contained in various cosmetic products such as hair shampoos, rinses, lotions, creams, and toothpastes. These areas of application account for 10 percent of citric acid use.

More recently, citric acid is being used for metal cleaning, substituting for phosphate in detergents as plasticizers in its ester forms, for secondary oil recovery, and as a buffer/absorber in stack gas desulfurization. The use of sodium citrate in heavy-duty liquid laundry detergent formulations is responsible for the rapid increase in citric acid use.

Citric acid was first isolated from lemon juice and crystallized as a solid in 1784 by Scheele. In 1893, Wehmet first described citric acid as a product of mold fermentation. In 1919, fermentation processes based on sucrose were developed commercially. Many organisms have been shown to produce citric acid from carbohydrates as well as from *n*-paraffins; however, *Aspergillus niger* has always given the best results in industrial production of citric acid.

The generally accepted pathway from sugar to citric acid is assumed to follow the glycolysis pathway to pyruvate. Pyruvate then enters the TCA cycle. It is decarboxylated to acetate as acetyl-CoA, or adds on CO2 to form oxalacetate. Acetyl-CoA and oxalacetate then react to form citrate. It is generally considered that aconitase, the enzyme responsible for the conversion of citrate to isocitrate in the TCA cycle, is inhibited by depriving it of the necessary metal (iron) coenzymes. Therefore, the TCA cycle does not operate, and citric acid accumulates. Figure 24.17 summarizes the reactions leading to citric acid from glucose. It is worthwhile to note that one mole of glucose yields one mole of citric acid with no consumption of oxygen. The overall reaction is actually energy-yielding. It yields one mole of ATP and two moles of NADH2 per mole of citric acid produced. Both ATP and NADH2 are high energy-containing molecules. This aspect makes the process a good candidate for the process of cells immobilization.

Fig. 24.17 Pathway leading to citric acid from glucose.



Microbiological production of citric acid can be implemented by three techniques:

1. solid state fermentation
2. liquid surface fermentation
3. submerged culture fermentation

In solid state or Koji fermentation, *Aspergillus niger* grows on moist wheat bran (7080 percent water) and produces citric acid in 58 days. This process is practiced only in Japan and accounts for about one-fifth of Japanese citric acid production.

In liquid surface or shallow tray fermentation, beet molasses (containing 4852% sugar) or cane molasses of blackstrap (containing 5257% sugar) or high test (containing 7080% sugar) is introduced into a mixer. Dilute sulfuric acid is added to adjust the pH to about 6.0. Phosphorus, potassium, and nitrogen in the form of acids or salts are added as nutrients for proper mold growth and optimal citric production. The mix is then sterilized and finally diluted with water to a 1520 percent sugar concentration. The medium flows by gravity into shallow aluminum pans or trays arranged in tiers in sterile chambers. Most chambers have provisions for regulation and control of temperature, relative humidity, and air circulation. One plant has 80 trays per chamber. Each tray holds about 400 L of solution at a depth of 76 mm. When the medium has cooled to about 30°C, it is inoculated with spores of *Aspergillus niger*. The tray fermentation requires 812 days. The pH drops to about 2 at the end of the fermentation, and the acid content varies from 10 to 20 percent. Some oxalic and gluconic acids are also formed. The temperature is maintained at 2832°C during the fermentation. Sterile air is circulated through the chambers, and the relative humidity is controlled between 40 and 60 percent.

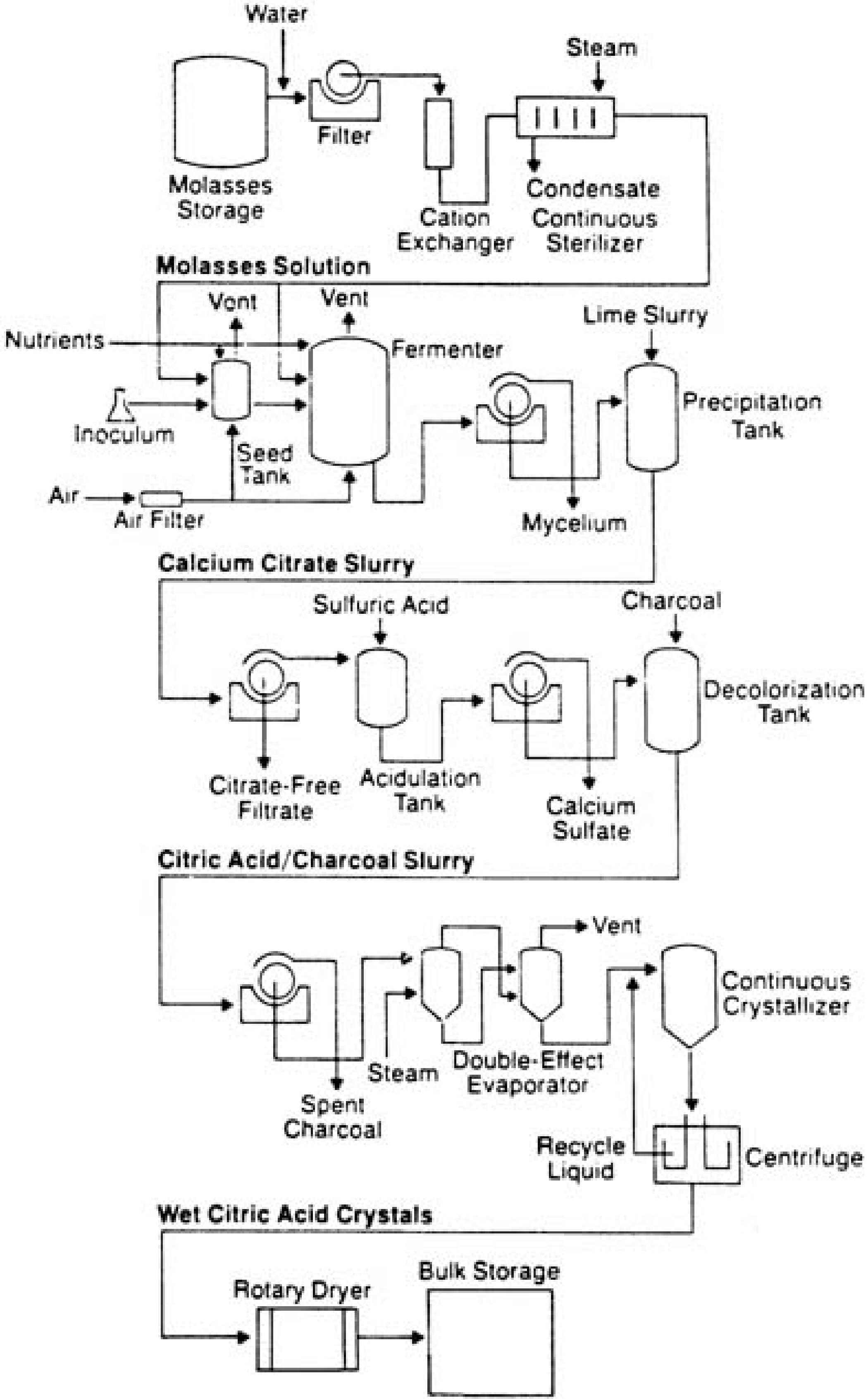
The submerged culture or deep fermentation process has been adopted by most newly constructed plants. The fermentation medium consists of sucrose (around 200 g/L) and mineral salts to provide a balanced supply of iron, zinc, copper, magnesium, manganese, and phosphate. The provision of a suitable culture medium is the most critical factor in obtaining a high yield of citric acid. The fermentation is carried out at 2527°C. Continuous aeration is provided by bubbling air at a rate of 0.51.5 volumes of air/volume of solution/minute. Mechanical stirring is not necessary. It is generally accepted that the formation of pellets between 1 and 2 mm in diameter in the fermentation mash is most desirable. Pelleting reduces both viscosity, increases oxygen transfer, and simplifies mycelium separation in the recovery scheme. The submerged fermentation has a time cycle of 69 days.

The yield of citric acid on sugar varies from process to process and from manufacturer to manufacturer. The theoretical maximum is 112 percent on sucrose. The liquid surface fermentation has a yield of 9095 percent, and the submerged culture fermentation 7585 percent. Improvements have been made by reducing the formation of by-products, mainly oxalic acid, and yields of the submerged culture process are reaching those of the surface culture process.

The fermentation broth from the solid state, surface culture, or submerged culture process is treated similarly for recovering and refining citric acid. Two recovery methods are being used: (1) precipitation and filtration and (2) extraction.

A process flowsheet including the fermentation section and the refining section using the first method is shown in Fig. 24.18. The mycelium is filtered out of the fermentation liquor first. The mycelium may be used as fertilizer after proper weathering and processing. The clarified liquor flows to precipitating tanks fitted with stirrers, where it is heated to a temperature of 8090°C. The oxalic acid present is separated by preferential precipitation through the addition of a small amount of hydrated lime. The resulting calcium oxalate is worked up separately in a manner similar to the following process described for citric acid recovery. Approximately one part of hydrated lime for every two parts of liquor is added slowly over a one-hour period while the temperature is raised to about 95°C. The precipitated calcium citrate is filtered on a vacuum filter, and the filtrate free of citrate is run to waste. The calcium citrate cake is run to acidulation tanks, where it is acidified with dilute sulfuric acid. Then it is filtered, and the citric acid mother liquor is decolorized by a charcoal treatment. Next the purified liquor is concentrated in a vacuum evaporator, and then it is run into a crystallizer where, on cooling, citric acid crystallizes, generally in the form of the monohydrate. The resulting acid is of USP grade.

Fig. 24.18 Citric acid flow sheet.



The extraction method treats the filtered fermentation liquor with a highly selective solvent, tri-*n*-butyl phosphate, and then recovers free citric acid by counterextraction with water. The aqueous solution, which is further concentrated and crystallized, yields 92 percent citric acid with 8 percent soluble impurities.

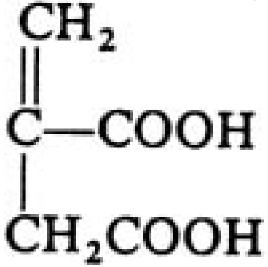
Takeda developed a citric fermentation process utilizing paraffins as substrate. Pfizer used a paraffin-based process for several years too, when the cost of paraffin was relatively low. But raw material availability and cost no longer favor this route.

At one time Pfizer and Miles were the only two major citric acid producers in the United States. Up to 1967 there was a third U.S. producer, the original Stephan Fermentation in Fieldsboro, N.J. The plant was sold to Bzura Chemical and was subsequently shut down because of operating difficulties. To successfully produce citric acid and compete with the established producers, a manufacturer must have extensive process know-how. The citric acid industry in the last two decades has seen some major changes in ownership and expansions in the Far East.

The list price of citric acid (anhydrous) has been in the range of 6090 cents per lb, USP, 100-lb bags. More buyers have been sourcing the product from China. Construction of large-scale facilities in the Far East has boosted worldwide capacity. The worldwide production has more than doubled in the past decade, and stands at about 1.4 billion lb a year. The major producers include the Swiss-based Jungbunzlauer, La Citique Belge/ Roche in Belgium, Bharat Starch Industries in India, and ADM, Cargill, and A.E. Staley in the United States. Cargill became a U.S. producer of citric acid in 1989. ADM acquired Pfizer' Southport, S.C. and Ringaskiddy, Ireland facilities in 1991. A.E. Staley, a subsidiary of Tate and Lyle, acquired the Bayer/Miles citric acid business in 1998. Jungbunzlauer constructed a new plant in Port Colborne, Ontario, Canada in 1999 to expand its presence in the North America. In 2000, Roche Zhongya, a joint venture between Roche and China's Wuxi Zhongya Chemical, brought on stream a 40,000 tons/year plant in Wuxi near Shanghai, China.

Itaconic Acid

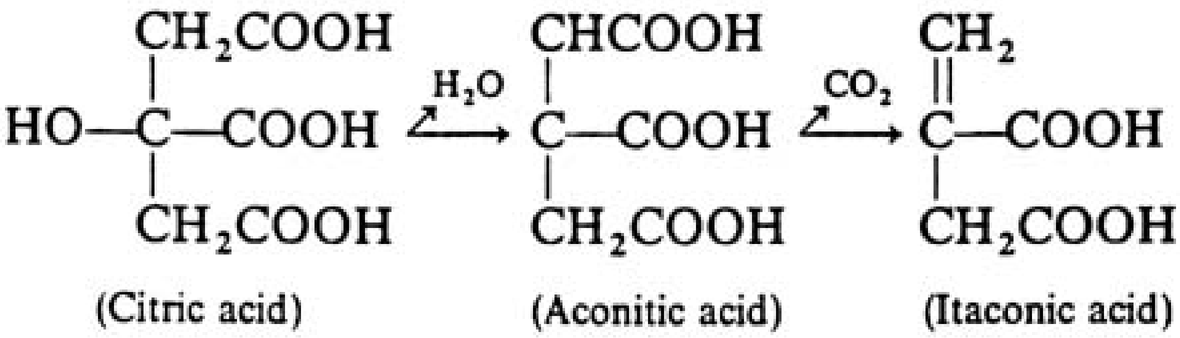
Itaconic acid (methylene succinic acid) is an unsaturated dibasic acid:



$$CH_2{=}C(COOH)CH_2COOH$$

It is a structurally substituted methacrylic acid. Consequently, its principle use is as a copolymer in acrylic or methacrylic resins. Acrylic fibers, by definition, contain at least 85 percent acrylonitrile. Because pure acrylic fibers are dye-resistant, it is necessary to include other components to make the fibers susceptible to dyes. An acrylic resin containing 5 percent itaconic acid offers superior properties in taking and holding printing inks and in bonding. In addition to its main application as a component of acrylic fibers, itaconic acid is used in detergents, food ingredients, and food shortenings.

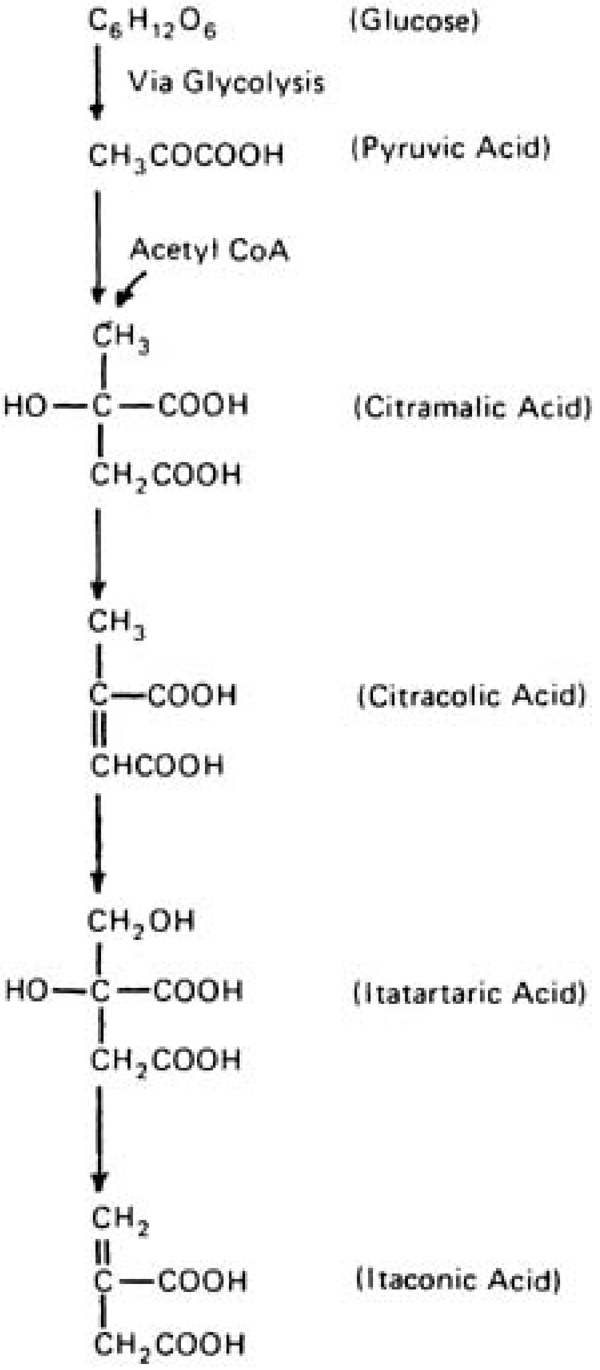
Previously, itaconic acid was isolated from pyrolytic products of citric acid or produced by converting aconitic acid present in sugar cane juice.



It is now produced on a commercial basis predominantly by direct fermentation of molasses.

The biosynthesis of itaconic acid once was believed to follow the decarboxylation of aconitic acid of the citric acid cycle. Now it is thought to follow the metabolic sequence shown in Fig. 24.19.

Fig. 24.19 Proposed metabolic sequence for biosynthesis of itaconic acid.



Although both *Aspergillus itaconicus* and *Aspergillus terreus* are known producers of itaconic acid, the latter is superior to the former and is believed to be used industrially. Either surface (shallow-pan) or submerged (deep-tank) fermentation can be used. The medium contains molasses, cornsteep liquor, ammonium sulfate, and mineral salts of calcium, zinc, magnesium, and copper. The fermentation, similar to that of citric acid, is very sensitive to concentrations of copper and iron. Copper iron favorably restricts growth and product destruction, but excessive concentration of iron results in reduction of product accumulation. Ten to twenty percent (by volume) inoculum is used. The fermentation is carried out at around 40°C and a pH of 2.04.0. Vigorous agitation is employed. Moderate, but continuous, aeration is required. Air failure of very brief duration is enough to damage the fermentation. The batch cycle is 36 days. The highest known product concentration is 180200 g/L from a medium containing 30 percent sugar. The yield of itaconic acid on sugar is typically 5070 percent. The itaconic acid recovery scheme involves the following:

1. Acidification of itaconic precipitates, if present.
2. Filtration to remove mycelium and other suspended solids.
3. Activated carbon treatment (this step and the next can be omitted for industrial-grade product).
4. Filtration to remove carbon.
5. Evaporation and crystallization.

If a high-purity acid is desired, further purification steps such as solvent extraction, ion exchange, and carbon decolorization are required.

The current (2002) price of itaconic acid is about $2/lb in 50-lb bags, and is discounted for purchases in 40,000-lb railroad cars. The world production is estimated to be about 20,000 tons/year. The major producers include Cargill (using Pfizer technology) in the United States, San Yuan in Taiwan, Rhodia in France, and Merck in Germany.

Gluconic Acid Aldehyde

Gluconic acid aldehyde is produced by the oxidation of the aldehyde group of glucose. Gluconic acid may be prepared from glucose by oxidation with a hypochlorite solution, by electrolysis of a solution of sugar containing a measured amount of bromine, or by fermentation of glucose by fungi or bacteria. The latter method is now preferred for economic reasons.

Gluconic acid is marketed in the form of 50 percent aqueous solution, calcium gluconate, sodium gluconate, and glucono-δ-lactone. Gluconic acid finds use in metal pickling, in foods as an acidulant, in tofu (soybean curd) manufacture as a protein coagulant, in detergent formulations as a calcium sequestrant, in the pharmaceutical area in mineral (calcium and iron) supplements, and in the construction area as a cement viscosity modifier. Calcium gluconate is widely used for oral and intravenous therapy. Sodium gluconate, a sequestering agent in neutral or alkaline solutions, finds use in the cleansing of glassware. Glucono-δ-lactone is used as a food flavor and an acidulant in baking powders and effervescent products.

The organism commonly used in gluconic acid fermentation is *Aspergillus niger* or *Gluconobacter suboxydans*. The largervolume production uses the fungal process. Most of the *Gluconobacter* production is marketed as glucono-δ-lactone.

During gluconic acid fermentation, glucose is first oxidized (or, more correctly, dehydrogenated) to glucono-δ-lactone. This is carried out by glucose oxidase. Hydrogen peroxide is also produced in this step, but is decomposed by catalase.

The fermentation can be by either surface or submerged culture, the latter more generally practiced in industry. Horizontally rotating fermentors have also been used.

Calcium gluconate fermentation, in which calcium carbonate is used for neutralization of the product, is limited to an initial glucose concentration of approximately 15 percent because of the low solubility of calcium gluconate in water (4% at 30°C). The addition of borate or boric acid allows the use of up to 35 percent glucose in the medium. However, borogluconate was found deleterious to blood vessels of animals, and the product was withdrawn from the market.

The recovery of calcium gluconate from fermentation broth involves the following:

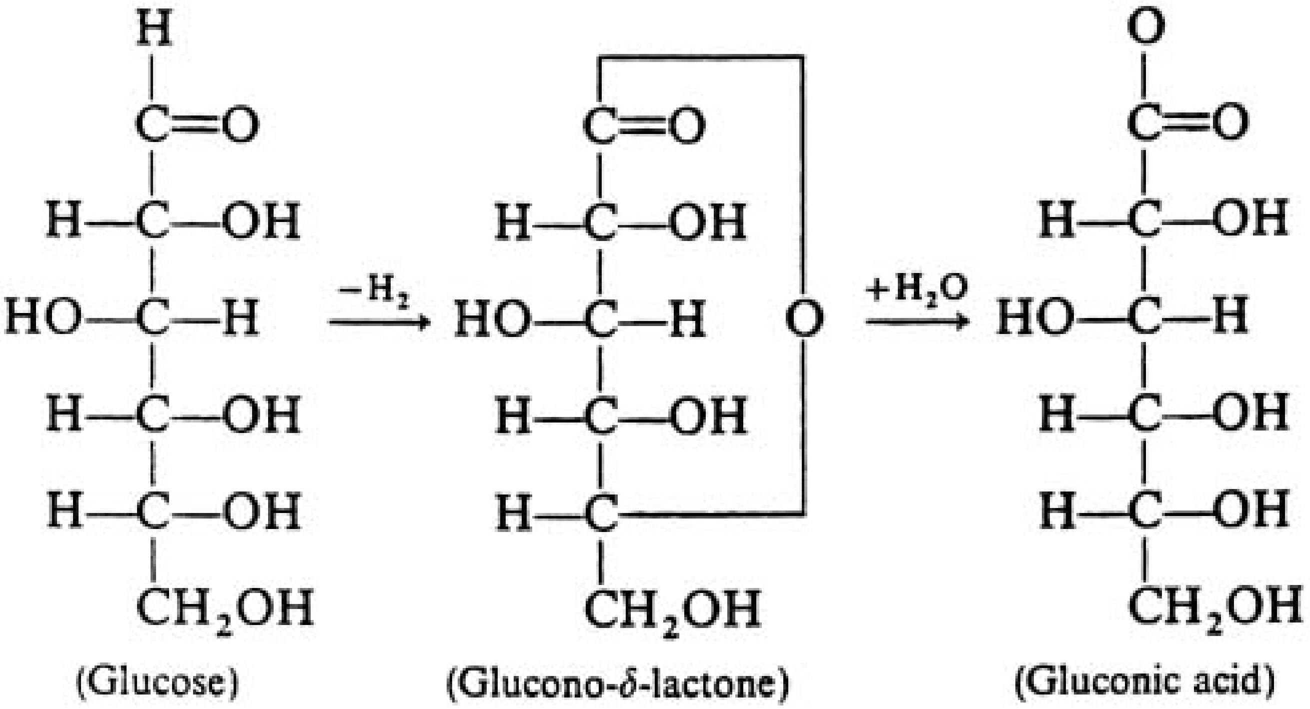
1. Filtration to remove mycelium and other suspended solids.
2. Carbon treatment for decolorization.
3. Filtration to remove carbon.
4. Evaporation to obtain a 1520 percent calcium gluconate solution.
5. Crystallization at a temperature just above 0°C.

In sodium gluconate fermentation, sodium hydroxide is used to control pH. Sodium gluconate is much more soluble than calcium gluconate. The addition of sodium hydroxide provides an easy and precise means of neutralizing the acid as it is produced. Much higher concentration of glucose (up to 35%) can be used in this fermentation. The medium also contains cornsteep liquor, urea, magnesium sulfate, and some phosphates. The pH is controlled above 6.5 by addition of sodium hydroxide. One to 1.5 volumes of air per volume of solution per minute (vvm) are supplied for efficient oxygenation. High back pressure (up to 30 psig) is desirable. The fermentation cycle is 23 days. Continuous fermentation is used in Japan to convert 35 percent glucose solution to sodium gluconate with a yield higher than 95 percent. The continuous process doubles the productivity of the usual batch system.

Sodium gluconate can be recovered from fermentation filtrate by concentrating to 4245 percent solids, adjusting to pH 7.5 with sodium hydroxide, and drum-drying.

In glucono-δ-lactone fermentation, *Gluconobacter suboxydans* converts a 10 percent glucose solution to glucono-δ-lactone and free gluconic acid in about 3 days. Approximately 40 percent of the gluconic acid is in the form of glucono-δ-lactone. Aqueous solutions of gluconic acid are in equilibrium with glucono-δ-lactone and glucono-γ-lactone. Crystals separating out of a supersaturated solution below 30°C will be predominantly free gluconic acid; from 30°C to 70° C the crystals will be principally glucono-δ-lactone; and above 70°C they will be mainly the γ-lactone.

Prices of gluconates vary. Sodium gluconate lists for about 70 cents/lb for truckload quantities, gluconic acid for about 45 cents/lb, and liquid gluconate 40 and 60 percent for about 30 and 45 cents/lb, respectively. Dry sodium gluconate is the main form of gluconic acid/gluconate consumed in the United States. Gluconic acid is used in applications where the sodium part of sodium gluconate is not desired, such as in acid cleaners. The primary applications of glucono-δ-lactone are as an acidulant in cheese and tofu manufacturing, and in making bakery goods, refrigerated dough, and salad dressings. The world market of gluconic acids and salts is about 60,000 tons/year. The major producers include Glucona America (a subsidiary of Aveba B.A.) and PMP Fermentation Products (acquired by Fujisawa) in the United States, Benckister, Jungbunzlauer, and Roquette Freres in Europe, and Fujisawa and Kyowa Hakko in Japan. In 1998, Glucona America made a major expansion of its facility in Janesville, WI, to double its fermentation capacity and triple its production of glucono-δ-lactone.



2-Ketogluconic Acid

2-Ketogluconic acid may be produced by a bacterial fermentation involving various strains of *Gluconobacter* or *Pseudomonas*. Selected strains of *Pseudomonas fluorescens* have been reported as giving the highest yield (up to 90%) when glucose and gluconate is used in the medium in highly aerated processes. Gluconic acid is an intermediate in the process.

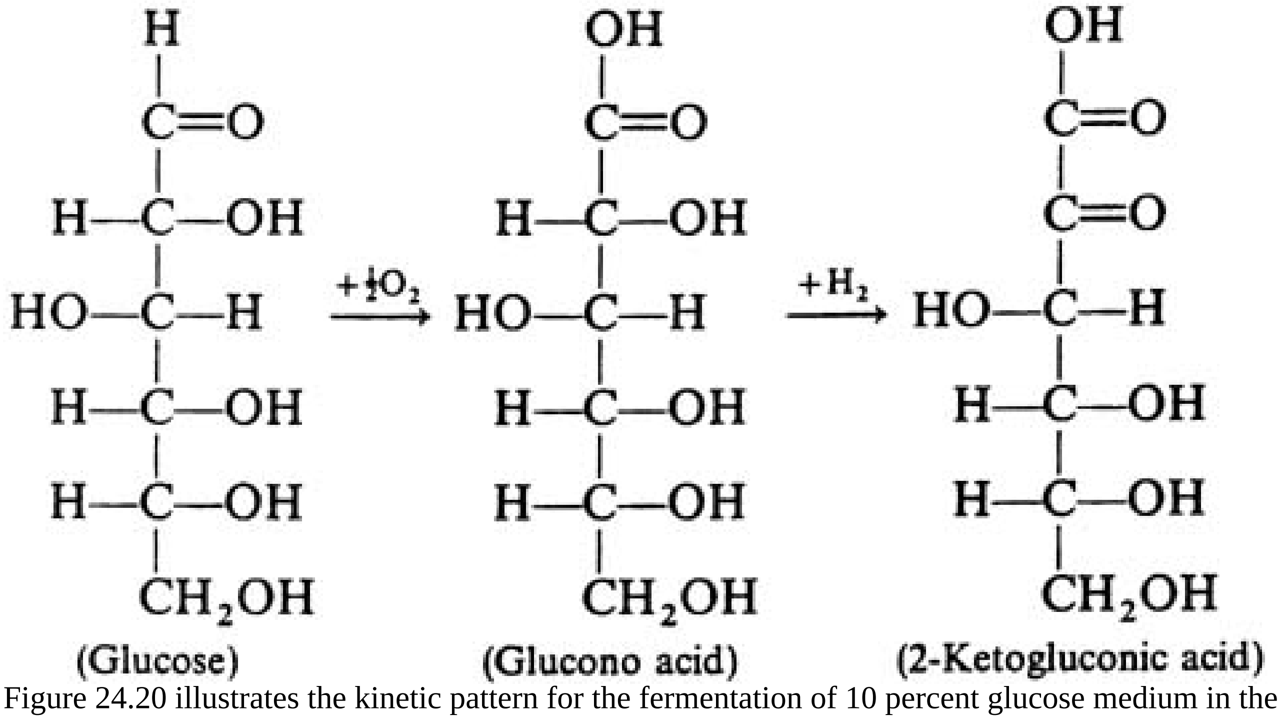
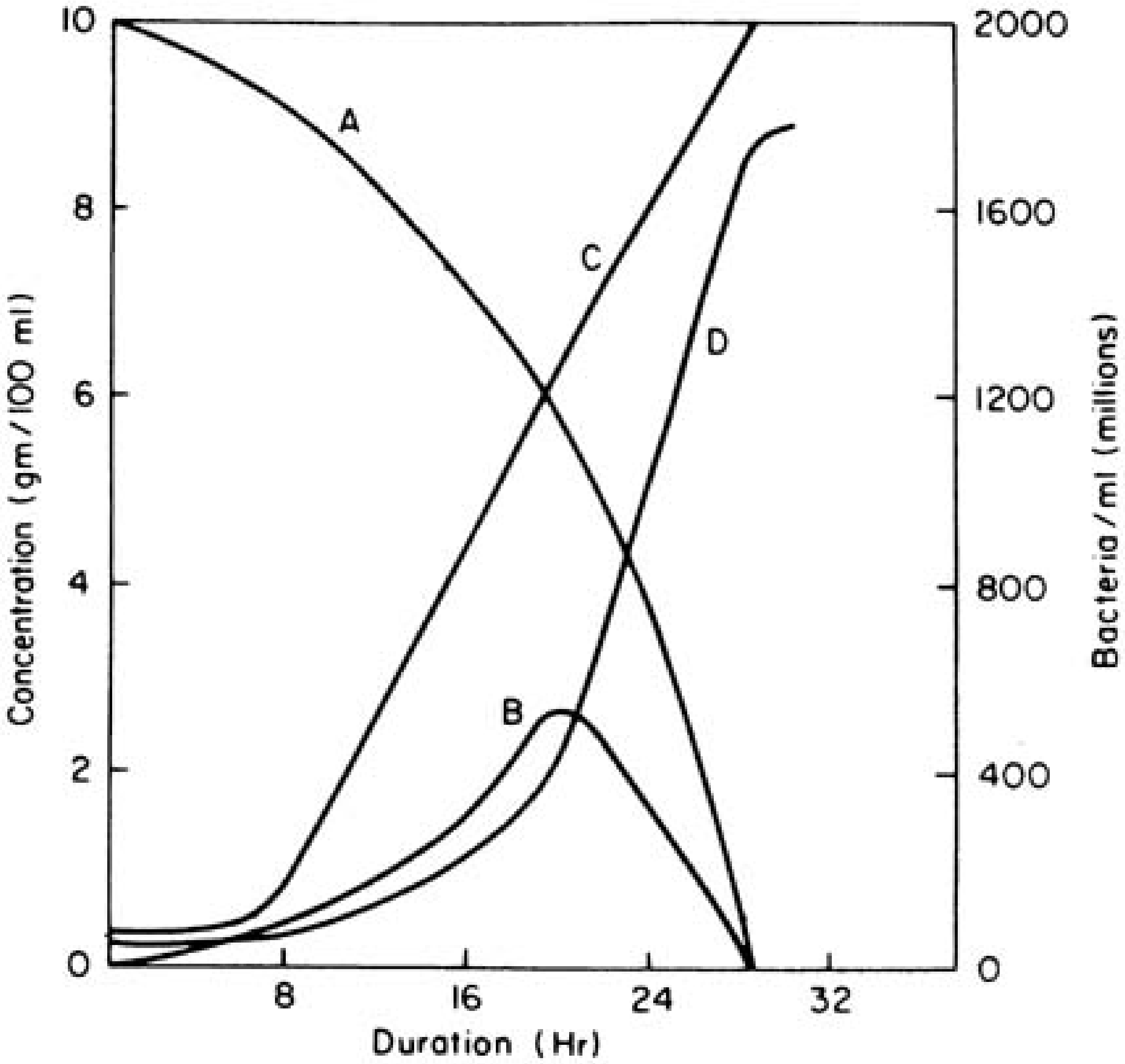


Figure 24.20 illustrates the kinetic pattern for the fermentation of 10 percent glucose medium in the presence of an excess of calcium carbonate. A 20 percent glucose medium is used commercially.

Fig. 24.20 Kinetics of 2-ketogluconic acid fermentation from glucose. (A, glucose; B, gluconic acid; C, bacteria; D, 2-ketogluconic acid.) (*Microbial Technology*, 2nd ed., Vol. 1, p. 384, Academic Press, New York, 1979.)



2-Ketogluconic acid is usually recovered and shipped in the free acid state after centrifugation or filtration to remove cells. Calcium is removed by precipitation with sulfuric acid. The filtered acid may be shipped as a syrup or as a crystalline material after evaporation under reduced pressure and below 50°C.

The principal use of 2-ketogluconic acid is as an intermediate in the preparation of isoascorbic acid, now known to the trade as erythorbic acid. Erythorbic acid, its esters, and its salts are used as water- or fat-soluble antioxidants to retain color, flavor, and nutritive values in canned fruits and vegetables and in meats and meat products.

Fujisawa Pharmaceutical (Japan) is a major producer of 2-ketogluconic acid. Its yearly production is about 2,000 tons of 2-ketogluconic acid, with most of that exported to the United States, Canada, and European countries.

24.4 Organic Solvents

The organic chemicals that fall into this category and can be produced by fermentation include ethanol, butanol, acetone, 2,3butanediol, and glycerol. 2,3-Butanediol and glycerol fermentations have been developed at laboratory and pilot-plant scales, but have not been commercialized. Ethanol, butanol, and acetone have been produced industrially by fermentation, but chemical synthesis is the manufacturing practice of choice for economic reasons. However, as price and availability of ethylene and propylene as feedstocks for the synthetic processes become subjects of concern, there is renewed interest in examining the fermentation processes as means of producing all or a portion of the future needs of ethanol, butanol, and acetone.

As the cost of oil continues to go up while the price of renewable resource such as agricultural crops remains relatively stable, there has been increasing interest to produce chemical feedstock from renewable resource by biological means. A most telling example is the Cargill-Dow Chemical joint venture that plans to set-up a $300 million plant at Blair, NE, capable of producing about 140,000 tons/year of polylactide polymers from corn sugar. It employs a fermentation process to produce from glucose two chiral isomers of lactic acid, which are then cracked to form three lactide isomers. The isomers are then polymerized to polylactide. The other interesting example is the joint venture between DuPont and Tate & Lyle to demonstrate the large-scale feasibility of DuPont's fermentation process to produce 1,3-propanediol using a genetically engineered organism that combines metabolic capabilities from bacteria and yeast. A 200,000-lb/year pilot plant has already been started at the Tate & Lyle subsidiary A.E. Staley's Decatur, IL, site. Commercialization is planned for 2003.

Ethanol

Ethanol is used in industrial solvents, thinners, detergents, toiletries, cosmetics, and pharmaceuticals, and, most important, as an intermediate for manufacturing other chemicals such as glycol ethers, ethyl chloride, amines, ethyl acetate, vinegar, and acetaldehyde. With the ever-increasing price and dwindling supply of crude oil, ethanol fermented from grains and other renewable organic resources is in close competition with synthetic ethanol produced from ethylene.

In 1984, 150 million gal of ethanol were produced synthetically in the United States, and around 500 million gal of ethanol were produced by fermentation. Since industrial markets have been mostly mature and ethanol has been replaced in many applications, total US capacity for synthetic ethanol is still about 170 million gal a year. The major producers are Union Carbide and Equister Chemicals. However, total fermentation ethanol capacity is now more than 2 billion gal a year. The capacity is expected to go up to 3.5 billion gal or even 4 billion gal a year by 2003. This is because of the industry's efforts to expand fermentation ethanol's use in fuels, particularly as a replacement for methyl *tert*-butyl ether as an oxygenate. Demand for industrial ethanol, including both synthetic and fermentation types, was about 270 million gal in 1999, 60 percent of which was used in solvents and 40 percent as chemical intermediatesof which the largest were ethyl acrylate and distilled vinegar. Fermentation ethanol is produced from sugar-containing materials such as grain products, fruits, molasses, whey, and sulfite waste liquor. ADM, A.E. Staley, and Cargill are the biggest among a large number of producers who typically have 2040 million gal/year capacity.

Yeasts, particularly strains of *Saccharomyces cerevisiae*, are almost exclusively used in industrial ethanol fermentation. *Saccharomyces cerevisiae* tolerates ethanol concentrations up to about 20 percent (by volume), and has relatively fast fermentation rates. It converts over 85 percent of the available carbohydrates to ethanol and carbon dioxide under anaerobic conditions. Air or oxygen suppresses the formation of ethanol (the Pasteur effect). under aerobic conditions, a major portion of the carbohydrates goes to cell growth.

Ethanol is formed via glycolysis (the EmbdenMeyerhofParnes pathway). The overall reaction starting from glucose can be written as follows:

$$C_6H_{12}O_6 \rightarrow 2C_2H_5OH + 2CO_2 + 31,200\,\mathrm{cal}$$

The ethanol yield from this equation is 51 percent by weight. Because carbohydrate is also used for cell growth and respiration, the overall yield of ethanol from total carbohydrate consumed is typically between 42 and 46 percent.

Ethanol fermentation can be conducted on nearly any carbohydrate-rich substrate. Molasses, which is the waste mother liquor remaining after the crystallization of sucrose in sugar mill operations, is widely used. Blackstrap molasses contains 3540 percent sucroses and 1520 percent invert sugars (glucose and fructose). High-test molasses contains 2227 percent sucrose and 5055 percent invert sugars. Most of the blackstrap molasses do not require the addition of other nutrients for ethanol fermentation. However, high-test molasses requires considerable quantities of ammonium sulfate and other salts such as phosphates. The nonsugar solids nutrient content of high-test molasses is about 7 percent, compared to 2835 percent in blackstrap molasses.

In the molasses process, blackstrap or high-test molasses is charged into a mixing tank, where it is diluted with warm water to give a sugar concentration of 1520 percent. Mineral acid is added to adjust the pH to between 4.0 and 5.0. The diluted and acidified molasses, called "mash," then is pasteurized, cooled, and charged into fermentor tanks, where about 5 percent yeast inoculum is added. The fermentation is carried out nonaseptically at 2332°C. Antibiotics may be added to control possible contaminations. Because the overall reaction is exothermic, cooling is required. The fermentation takes 2872 hr (averaging about 44 hr) to produce an ethanol concentration of 810 percent. Carbon dioxide normally is vented. If it is to be collected and recovered, the vent gas is scrubbed with water to remove entrained ethanol and then purified using activated carbon. The activated carbon is reactivated periodically using hot air or stream. The outgassing of the carbon dioxide from the fermentors provides sufficient agitation for small tanks. Mechanical agitation may be added for large fermentors. The fermentation may be conducted batchwise or continuously, with or without recycling yeast. Although continuous fermentation and/or cell recycle can significantly improve productivity and thus reduce required capital investment, it may have only a limited impact on lowering product costs. A major portion of the costs comes from raw materials.

After the fermentation is completed, the liquor, known as beer, is withdrawn from the fermentors, passed through heat exchangers, and pumped to the upper section of a beer still (or "whisky column"), where the alcohol and other volatiles such as aldehydes are distilled off as the overhead. The bottoms, known as slop or stillage, are processed into animal feeds, known as distillers' dried grains, which contain residual sugars, proteins, and vitamins.

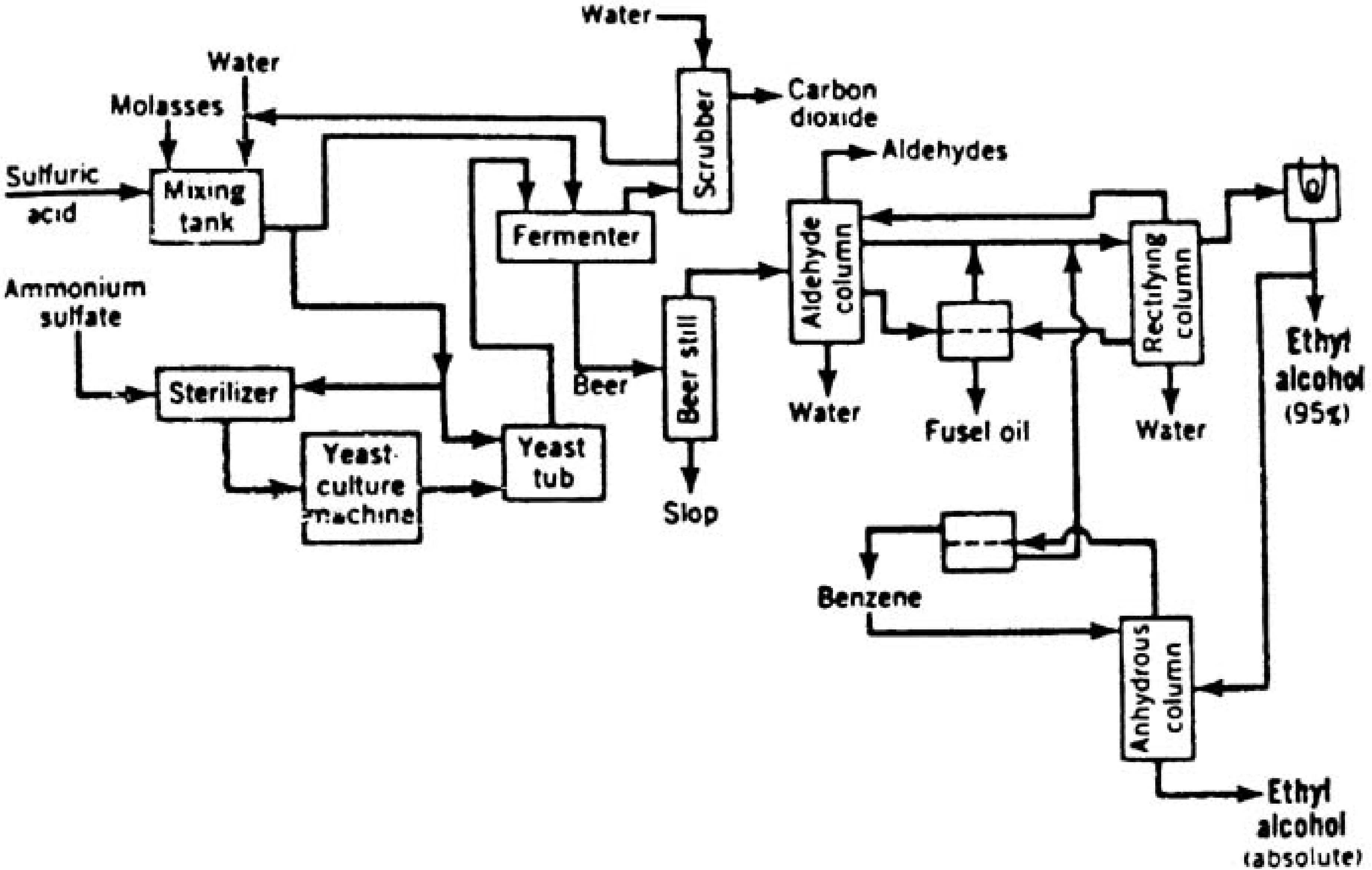
The overhead from the beer column is passed through a heat exchanger and condensed. This condensate, known as high wines, contains 5070 percent alcohol. It is charged into an aldehyde column (or "head column"), where aldehydes, esters, and other low-boiling impurities are separated as overhead. The stream from about the middle of the column is run into a refining column (or "rectifying column"). The tails from the aldehyde column and the weak fractions from the beer column, called low wines, are rerun with subsequent batches.

In the rectifying column, the heads containing a trace of aldehydes are returned to the aldehyde column. Near the top of the column, the azeotropic alcoholwater mixture of 95 percent alcohol is taken off, condensed, and run to storage. The higher-boiling alcohols, known as fusel oils, containing amyl, butyl, isobutyl, propyl, and hexyl alcohols, are withdrawn farther down the column. The fusel oils amount to 0.5 percent of total carbohydrates consumed. Water is discharged from the bottom of the column. In order to prevent the diversion of industrial alcohol to potable uses, it is denatured by the addition of some material that renders the alcohol so treated unfit for use as a beverage. The 95 percent alcohol from the rectifying column is stored in government-bonded warehouses. The alcohol is either denatured, dehydrated, or sold (tax-free or tax-paid).

Anhydrous or absolute alcohol is produced by several methods. A third component such as benzene may be added, and the mixture distilled. The ternary azeotrope thus formed carries over the water, leaving behind anhydrous alcohol. Another method uses countercurrent extraction with a third component such as glycerine or ethylene glycol. The added component depresses the vapor pressure of the water and allows anhydrous alcohol to be distilled from the top of the extraction column. Both these methods are run using continuous columns.

A process flow diagram of ethanol fermentation using molasses as substrate is shown in Fig. 24.21.

Fig. 24.21 Ethanol fermentation using molasses as a substrate. (*Industrial Chemicals*, Faith, Keyes and Clark, 4th ed., p. 357, 1975. Copyright John Wiley & Sons, Inc., New York, and used by permission of the copyright owner.)



Ethanol can also be produced by fermentation of starch, whey, and sulfite waste liquor. Grain fermentations require additional pretreatment because yeast cannot metabolize starch directly. The grain (usually corn) is ground and heated in an aqueous slurry to gelatinize or solubilize the starch. Some starch-liquefying enzymes may be added in this step at lower temperatures. The liquefied starch is then cooled to about 65°C, and treated with malt (germinated, dried barley, which is rich in starch-hydrolyzing enzymes), or fungal amylase (produced by *Aspergillus niger*) to convert starch to oligosaccharides. Then yeast is added along with amyloglucosidase (or glucoamylase), which breaks down oligosaccharides into glucose. The subsequent fermentation and refining procedures are the same as those using molasses as raw materials.

Research to improve ethanol fermentation has focused on the development of continuous fermentation and improved yeast strains. Other related developments include processes for the continuous hydrolysis of wood coupled with continuous fermentation, and for the use of other biomass feedstocks, such as whey and cellulosic waste material. Developments in genetic engineering have made possible the development of new organisms (yeasts and bacteria) that can survive in higher concentrations of ethanol, tolerate higher sugar concentrations, grow at higher temperatures, and utilize starch or cellulose feedstock directly.

Gasohol

Gasohol, a fuel mixture of 10 percent alcohol and 90 percent gasoline, is a going business in the United Statesand a growing one. It has become the most important use of fermentation ethanol.

Ethanol from grain fermentations has been made competitive as a liquid fuel in the United States because of technology improvements, especially in the area of efficient energy usage in production plants, and various government subsidies designed to reduce the nation's dependence on imported foreign oil. The incentives have included the following:

1. An extension through 2003 of the 5.4-cent federal tax exemption for every gallon of gasohol sold as fuel.
2. In more than 20 states, the elimination of state gasoline taxes for gasohol.
3. Tax credits, loans, and loan guarantees for biomass-based alcohol plant constructions.

These subsidies may become unnecessary in the future if the price of crude oil continues to inflate at a rate at least 5 percent higher than the prices of corn and other agricultural crops.

Other countries like Brazil, India, and China also have substantial fermentation ethanol production and gasohol use.

Fermentation ethanol also has the potential to become a major chemical feedstock. The U.S. production of ethylene was 29.2 billion lb in 1979. It would take the equivalent of 3 billion bushels of corn to produce the same quantity of ethylene via ethanol fermentation and subsequent dehydration. The use of other substrates, such as wheat, sugar cane, sweet sorghum, whey, waste paper, and fastgrowing trees, has to be explored.

To realize and expand the use of ethanol as chemical feedstock and liquid fuel, fermentation research and development on the following would be helpful:

1. Strains that tolerate higher ethanol concentrations.
2. Strains that ferment optimally at higher temperatures.
3. Strains capable of fixing CO2 to give higher yields.
4. Strains capable of utilizing a wider range of carbohydrates.
5. Utilization of cellulosic or fibrous components of corn and other crops.
6. Fermentations under vacuum conditions.
7. Continuous fermentations with or without recycling of yeast.
8. Simultaneous and continuous product removal by physical or chemical means such as extraction or ultrafiltration.
9. More energy-efficient ways of recovering and dehydrating ethanol.

Butanol/Acetone

The original observation of butanol production by *Clostridia* was made by Pasteur, and acetone formation was noted by Schardinger. Interest in commercializing butanolacetone fermentation occurred in 1909 primarily as a means of obtaining butadiene as raw material for synthetic rubber. In 1914, Weizmann established a working process to ferment starchy grains such as maize or corn to produce butanol, acetone, and ethanol using *Clostridium acetobutylicum*. With the outbreak of World War I, the production of acetone was of great interest for the manufacture of cordite. Large-scale operations were established in Canada, the United States, India, and elsewhere during the war period. Shortly after the war, Du Pont developed fast-drying nitrocellulose lacquers for the automobile industry, and butyl acetate was the solvent of choice for coating. Large quantities of butanol esters were needed as solvents, and butanol became the principal product of the butanolacetone fermentation. By the 1930s, some butanol and acetone were being produced by chemical synthesis, and the butanolacetone fermentation industry faced economic difficulties. This problem was solved by the discovery of the strains *Clostridium saccharo-acetobutylicum* and *Clostridium saccharo-butylacetonicumliquefaciens*, which would ferment molasses, a cheaper raw material than starchy grains, and the industry thrived until the end of World War II. The major U.S. producers at that time were Commercial Solvents and Publicker Industries. They operated plants at Terre Haute, Indiana; Baltimore, Maryland; and Philadelphia, Pennsylvania. The butanolacetone fermentations were conducted in large-scale equipment; fermentors of 50,000- to 500,000-gal capacity were commonly used. At the present time, the petrochemical processes dominate, and most fermentation processes have been closed down. Fermentation of molasses to acetone and butanol, however, is probably still carried out by National Chemical Products in South Africa and Usina Victor SA in Brazil. Similar operations existed in the past in several countries and probably could be started up under appropriate economic circumstances.

In the grain fermentation process, 810 percent corn mashes were fermented (corn contains 7072% starch on a dry basis). Fermentation yields were on the order of 2932 g mixed solvent per 100 g starch used, with a solvent ratio of approximately 603010 (butanolacetoneethanol, respectively). The organisms possessed good diastatic activity, so malting was not required. The cooked sterile cornmeal suspension was aseptically transferred to sterile fermentors, inoculated, and incubated for about 65 hr at 37°C, after which the solvents were recovered by distillation. The aqueous residue (slop or stillage) was concentrated in multiple-effect evaporators and drum-dried for use in animal and poultry feeds.

With the advent of molasses-fermenting strains, more rapid fermentations were attained (4048 hr), and the solvents produced contained as much as 6575 percent butanol, principally at the expense of ethanol. One hundred pounds of blackstrap molasses (containing 57 lb of sugar) was fermented into the following spectrum of products:

Product	Yield (lb)
Butanol	11.5 (solvent ratio = 68%) (yield on sugar = 20%)
Acetone	4.9 (solvent ratio = 29%) (yield on sugar = 9%)
Ethanol	0.5 (solvent ratio = 3%) (yield on sugar = 1%)
Total solvents	6.9 (yield on sugar = 30%)
Carbon dioxide	32.1 (yield on sugar = 56%)
Hydrogen	0.8 (yield on sugar = 1%)
Dried stillage	28.6

The anaerobic *Clostridium* yields energy by converting glucose to acetyl-CoA, formate, CO2, and H2. The reducing power is then used to produce butanol and acetone via acetoacetyl-CoA. The overall reaction for butanol and acetone production can be pictured as follows:

$$\begin{aligned}2C_6H_{12}O_6 \quad \rightarrow \quad & C_4H_9OH + CH_3COCH_3 \\ & + 5CO_2 + 4H_2\end{aligned}$$

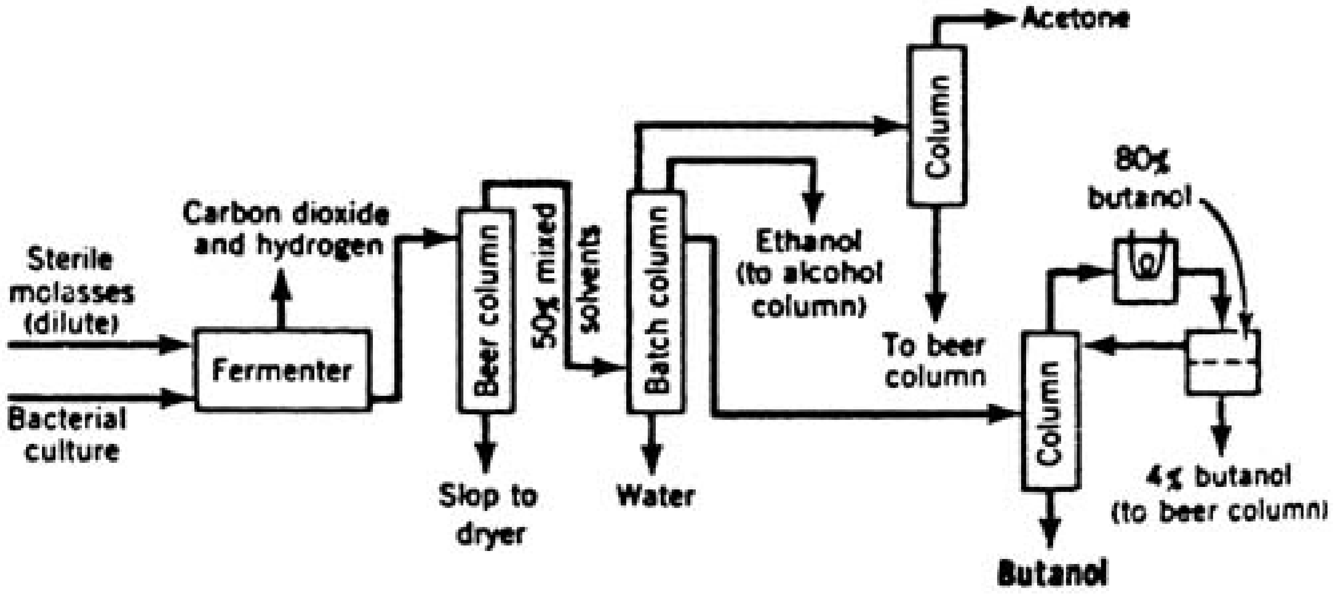
According to the equation the yields on sugar for butanol, acetone, carbon dioxide, and hydrogen are 21, 16, 60, and 2 percent, respectively.

Continuous fermentations were described by Russian workers. A plant in Dokshukin was operated in three batteries of seven to eight fermentors of 60,00070,000 gal. The continuous cycle was 4090 hr, and the flow rate through the battery was 5000 10,000 gal/hr. The feed contained 46 percent carbohydrates. A combination of raw material was used: molasses, flour, and hydrolysate containing pentoses. The continuous process gave a 20 percent productivity increase, and saved 142 lb of starch for every ton of solvents produced.

A considerable amount of care must be exercised in carrying out butanolacetone fermentation. It is biologically unstable and may fail completely when contaminated. Numerous instances of contamination by bacteriophage were encountered commercially, and on several occasions plants had to suspend operation until the entire plant could be decontaminated. Accordingly, absolute cleanliness, experienced personnel familiar with phage symptoms, and the maintenance of a vigorous rapid fermentation are necessities in butanolacetone fermentation.

The butanolacetoneethanol industrial fermentation of the 1950s made some remarkable advances in distillation efficiencies. Highlights included concentrating upon beer (fermentation mixture) stripping, early removal of water from solvents, thermo-compression on the beer still, continuous refining of butanol, and recovery of heat. A process flowsheet is shown in Fig. 24.22. When the fermentation is completed, the fermentor broth containing around 2 percent mixed solvents is pumped to a beer column where a 50 percent solvent mixture is taken off overhead, and distiller's slop is removed as bottoms. The slop may be dried and sold as animal feed. Another by-product is a mixture of carbon dioxide and hydrogen. The mixed-solvent vapors from the beer column are led to a batch fractionating column, from which three fractions (acetone, with BP 56.1°C, ethanol with BP 78.3°C, and butanol with BP 117.3°C) are removed overhead, leaving water as bottoms. The acetone and ethanol fractions are purified by conventional fractionation. The butanol fractions containing 70 percent butanol and 30 percent water are removed overhead. On condensation, two layers are formed. The top layer (80% butanol and 20% water) is returned to the butanol column, and the bottom layer (4% butanol and 96% water) is returned to the beer column. Approximately 35,000 lb stream is consumed for every ton of solvents produced.

Fig. 24.22 Process flow sheet of butanol-acetone-ethanol fermentation. (*Industrial Chemicals*, Faith, Keyes and Clark, 4th ed., p. 178. Copyright John Wiley & Sons, Inc., New York, 1975. Used by permission of the copyright owner.)



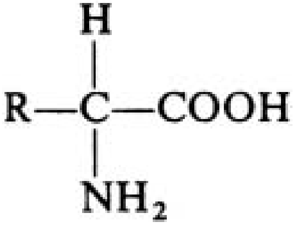
The U.S. production of butanol and acetone was 1.6 billion/lb and 4 billion/lb, respectively, in 1998. Butanol costs 25 cents/lb, and acetone 20 cents/lb. The fermentation industry has the potential to capture all or at least a portion of the market by concentrating on research and development along the following lines:

1. Develop strains to tolerate high concentrations of butanol and acetone.
2. Develop homofermentive strains to improve product yields.
3. Utilize cheaper raw materials such as waste carbohydrates.
4. Develop continuous processes to improve productivities.
5. Improve fermentation stability through both strain and equipment enhancements.
6. Develop more energy-efficient separation methods such as extraction or reverse osmosis.

24.5 Amino Acids

General

Amino acids, in general, can be represented by the following formula:



$$R-\underset{\underset{NH_2}{\vert}}{\overset{\overset{H}{\vert}}{C}}-COOH$$

Because the amino group is on the α-carbon, the amino acids with this general formula are known as α amino acids. The α-carbon atom becomes asymmetric when R is not an H atom. Naturally occurring amino acids have an L-configuration. Amino acids are the building blocks of proteins, and the elementary composition of most proteins are similar; the approximate percentages are:

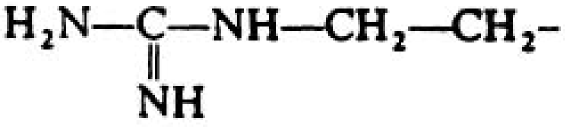
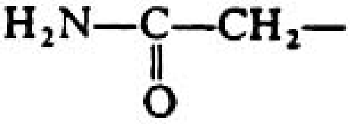
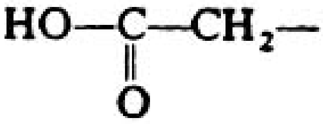
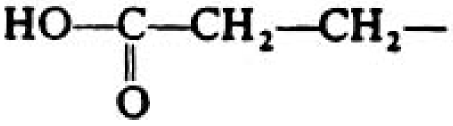
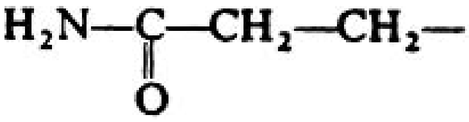
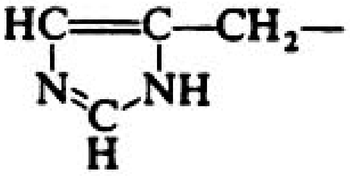
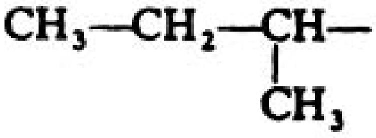
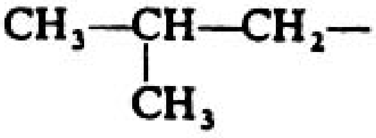
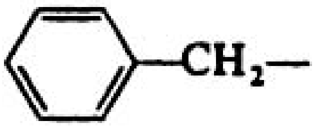
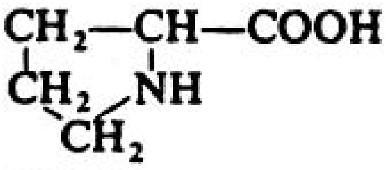
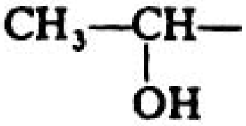
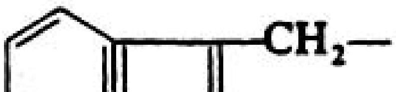
$C = 50–55$ $N = 15–18$

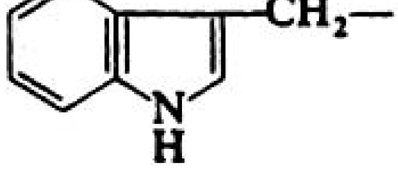
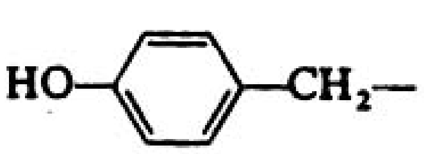
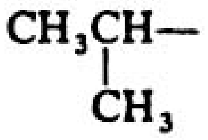
$H = 6–8$ $S = 0–4$

$O = 20–23$

Table 24.5 gives the structure of R, molecular weight, and elementary composition or each of the 20 amino acids commonly obtained on hydrolysis of proteins.

TABLE 24.5 Twenty Common Amino Acids

			Elemental Composition (% wt)				
Amino Acids	*R-*	*M.W.*	*C*	*H*	*O*	*N*	*S*
Alanine	CH_3-	89	40	8	36	16	0
Arginine	 $H_2N-\underset{\underset{NH}{\Vert}}{C}-NH-CH_2-CH_2-$	174	41	8	18	32	0
Aspargine	 $H_2N-\underset{\underset{O}{\Vert}}{C}-CH_2-$	132	36	6	36	21	0
Aspartic acid	 $HO-\underset{\underset{O}{\Vert}}{C}-CH_2-$	133	36	5	48	11	0
Cysteine	$HS-CH_2-$	121	30	6	26	12	26
Glutamic acid	 $HO-\underset{\underset{O}{\Vert}}{C}-CH_2-CH_2-$	147	41	6	44	10	0
Glutamine	 $H_2N-\underset{\underset{O}{\Vert}}{C}-CH_2-CH_2-$	146	41	7	33	19	0
Glycine	$H-$	75	32	7	43	19	0
Histidine		155	46	6	21	27	0
Isoleucine	 $CH_3-CH_2-\underset{\underset{CH_3}{\vert}}{CH}-$	131	55	10	24	11	0
Leucine	 $CH_3-\underset{\underset{CH_3}{\vert}}{CH}-CH_2-$	131	55	10	24	11	0
Lysine	$H_2N-CH_2-CH_2-CH_2-$	146	49	10	22	19	0
Methionine	$CH_3-S-CH_2-CH_2-$	149	40	7	22	9	2
Phenylalanine		165	66	7	19	8	0
Proline		115	52	8	27	12	0
Serine	$HOCH_2-$	105	34	7	46	13	0
Threonine	 $CH_3-\underset{\underset{OH}{\vert}}{CH}-$	119	40	8	40	12	0
Tryptophan		204	65	6	16	14	0

							
Tyrosine		181	60	6	26	8	0
Valine	 $CH_3CH—$ / CH_3	117	51	9	12	27	0

Amino acids can be obtained from purified proteins by chemical or enzymatic hydrolysis. They can also be isolated from industrial by-products, extracted from plant or animal tissues, or synthesized by organic, enzymatic, or microbiological means. Amino acids that are produced industrially by fermentation include arginine, citrulline, glutamic acid, glutamine, histidine, isoleucine, leucine, lysine, ornithine, phenylalanine, proline, threonine, tryptophan, tyrosine, and valine.

Alanine and aspartic acid are produced commercially utilizing enzymes. In the case of alanine, the decarboxylation of aspartic acid by the aspartate decarboxylase of *Pseudomonas dacunhae* is commercialized. The annual world production of alanine is about 200 tons. Aspartic acid is produced commercially by condensing fumarate and ammonia using aspartase from *Escherchia coli*. This process has been made more convenient with enzyme immobilization technique. Aspartic acid is used primarily as a raw material with phenylalanine to produce aspartame, a noncaloric sweetener, tradenamed Nutrasweet. Production and sales of aspartame have increased rapidly since its introduction in 1981. World production of aspartic acid in 2000 was 12,000 tons. Tyrosine, valine, leucine, isoleucine, serine, threonine, arginine, glutamine, proline, histidine, citrulline, L-dopa, homoserine, ornithine, cysteine, tryptophan, and phenylalanine also can be produced by the enzymatic or fermentation method.

The worldwide amino acids market amounts to about $5 billion. The animal feed additives, glutamic acid, methionine, and lysine, account for about 75 percent of this figure. The other amino acids are used as precursors in pharmaceuticals, food additives, and animal feed. The worldwide demand for glutamic acid is about 800,000 tons/year, 300,000 tons/year for methionine (synthesized chemically), and 500,000 tons/year for lysine. Other significant amino acids include aspartic acid, phenylalanine, threonine, and glycine (synthesized chemically), each with a worldwide market of about 10,00020,000 tons/year. Tryptophan and cysteine command a global market in the thousands of tons as well. The December 2000 European Union ban, due to fear of mad cow disease (bovine spongiform encephalopathy), on feeding animal proteins to any animal raised for human consumption has also helped to increase the demand for amino acids used in animal feed.

The two fermentative amino acids with the largest worldwide demand and production are glutamic acid and lysine.

Glutamic Acid

MSG was first produced using acid hydrolysis of wheat gluten or soybean protein in 1909 by Ajinomoto. In 1957, a glutamate-producing bacterium was isolated, and subsequent research and development brought about the economic production of glutamic acid by fermentation.

A large number of glutamic acid-producing microorganisms are known. A partial list is given below.

Bacteria

1. *Corynebacterium glutamicum*
2. (Synonym *Micrococcus glutamicus*)
3. *Brevibacterium flavum*
4. *Brevibacterium divaricatum*

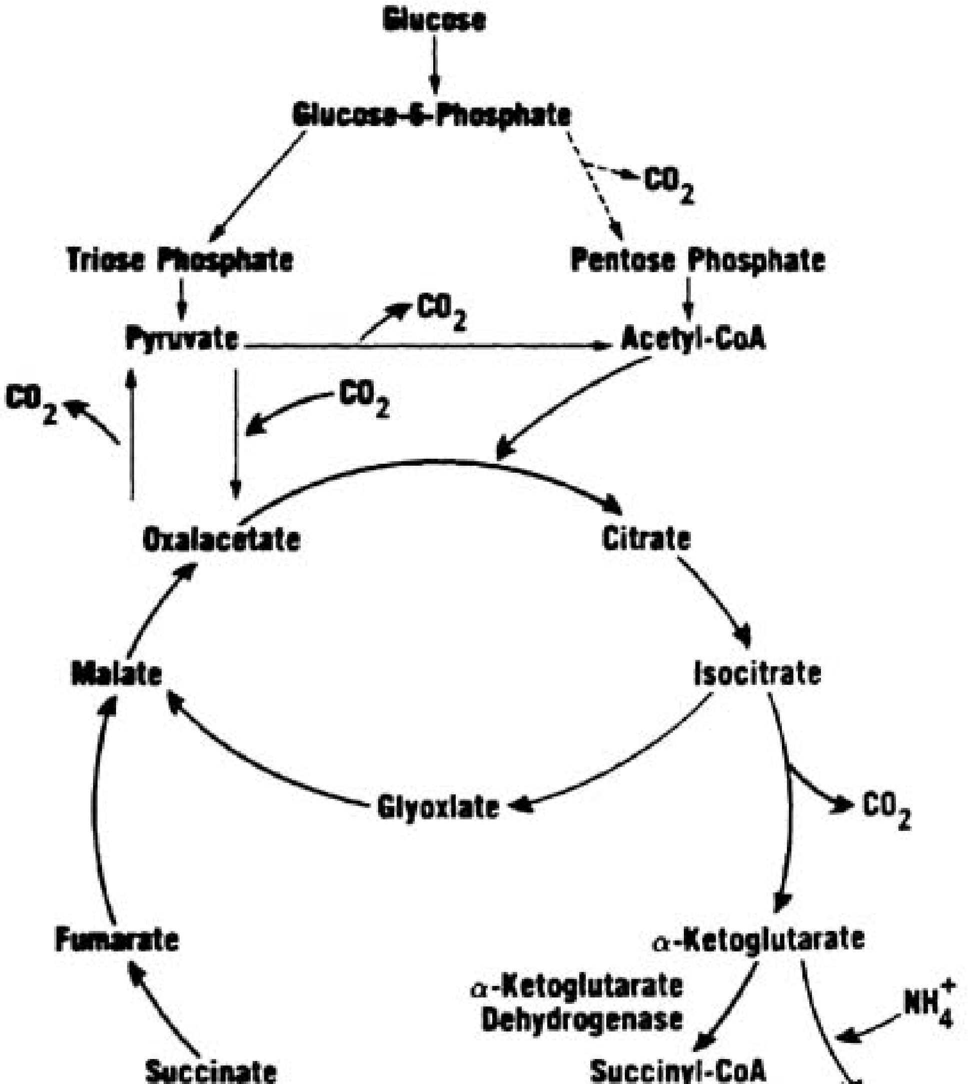
Fungi

1. *Aspergillus terreus*
2. *Ustilago maydis*

Among them, *Corynebacterium glutamicum* is used most commonly in industry.

The fermentation medium contains a carbon source (glucose, molasses, or acetic acid), a nitrogen source (urea, ammonium sulfate, cornsteep liquor, or casein hydrolyzate), small amounts of mineral salts that supply potassium, phosphorus, magnesium, iron, and manganese, and a few (less than 5) micrograms of biotin per liter. The biotin requirement is the major controlling factor in the fermentation. When too much biotin is supplied for optimal growth, the organism produces lactic acid. Under conditions of suboptimal growth, glutamic acid is excreted. The metabolic pathway involved in the biosynthesis of glutamic acid from glucose is shown in Fig. 24.23. The lack or very low content of α-ketoglutarate dehydrogenase is a special characteristic of glutamic acid-producing microorganisms.

Fig. 24.23 Pathway to glutamic acid formation from glucose.



The fermentation is conducted aerobically in tanks with *k*L*a* (volumetric oxygen transfer coefficient) values around 300 mmol O2/L/hr/atm. If aeration is not adequate, lactic acid is produced, and the yield of glutamic acid is poor. Too much air is not good either; it gives even more lactic acid, plus some α-ketoglutaric acid. The fermentation temperature is usually 2833°C, and the pH optimal for growth and glutamic acid production is 7.08.0. Continuous feeding of liquid ammonium hydroxide or gaseous ammonia controls pH and supplies ammonium ions to the fermentation. Modern digital computers are in some cases used for event sequencing, control, and optimization of the fermentation operation.

The fermentation cycle is 2448 hr. The final concentration of glutamic acid is about 150 g/L if glucose is used, and about 120 g/L if molasses is used. The overall yield of glutamic acid on sugar is about 65 percent on a weight basis. A portion of sugar is used for cellular growth with about 50 percent yield, whereas the sugar used for glutamic acid production has a yield of 86 percent according to the following equation:

$$C_{12}H_{22}O_{11} + 2NH_3 + 3O_2$$
$$\rightarrow 2C_5H_9O_4N + 2CO_2 + 5H_2O$$

The recovery of glutamic acid from fermen-tation broth is relatively easy. The broth is clarified by adding acid to pH 3.4, heating to 87°C, holding for a sufficient time to coagulate suspended solids, and filtering the coagulated broth to yield a clarified broth. The clarified broth is concentrated by evaporation, and then crystallized to yield glutamic acid. Other recovery schemes exist.

MSG is an important flavor enhancer for natural and processed foods. It is also good for protecting the flavor and color of preserved foods and suppressing off-flavors. MSG is selling for about $2/kg, and glutamic acid (99.5%) about $4/kg. Glutamic acid is not an essential amino acid. However, it has some pharmaceutical uses. It also improves the growth of pigs.

The major manufacturers of glutamic acid include Ajinomoto, Asahi Foods, Kyowa Hakko, Takeda (Japan), Orsan (France), Biacor (Italy), Cheil Sugar, Mi-Won (S. Korea), and Tung Hai Fermentation Industry, Ve Wong (Taiwan). The U.S. producers Merck, IMC, Great Western Sugar, and Stauffer ceased their MSG production in 1970, 1975, 1977, and 1984, respectively. To reduce production costs, most of the Japanese production is being moved to Southeast Asia or South America where raw material is available locally, in abundance and at lower costa trend for the manufacturing of all low-margin and high-volume fermentation products.

Kyowa Hakko has recently completed a genome map of its glutamic acid producing organism, and hopes to improve glutamic acid fermentation efficiency and to create a next-generation production system.

Lysine

Lysine, biologically active in its L-configuration, is an essential amino acid in human and animal nutrition. It is contained in good measure in fishmeal and in soybean meal. Plant products, particularly cereal grains such as corn, wheat, and rice, are usually low in lysine. Because of lysine deficiency, these grain proteins are of poorer quality than animal-derived proteins that contain higher levels of lysine and command much higher prices. For applications in human food, lysine in its salt forms can be added to cookies and bread, and in solution it can be used to soak rice. As most animal feed rations are based on maize and other grains, supplementing the feed-stuffs with lysine (plus methionine) significantly improves their nutritional value for breeding poultry and pigs.

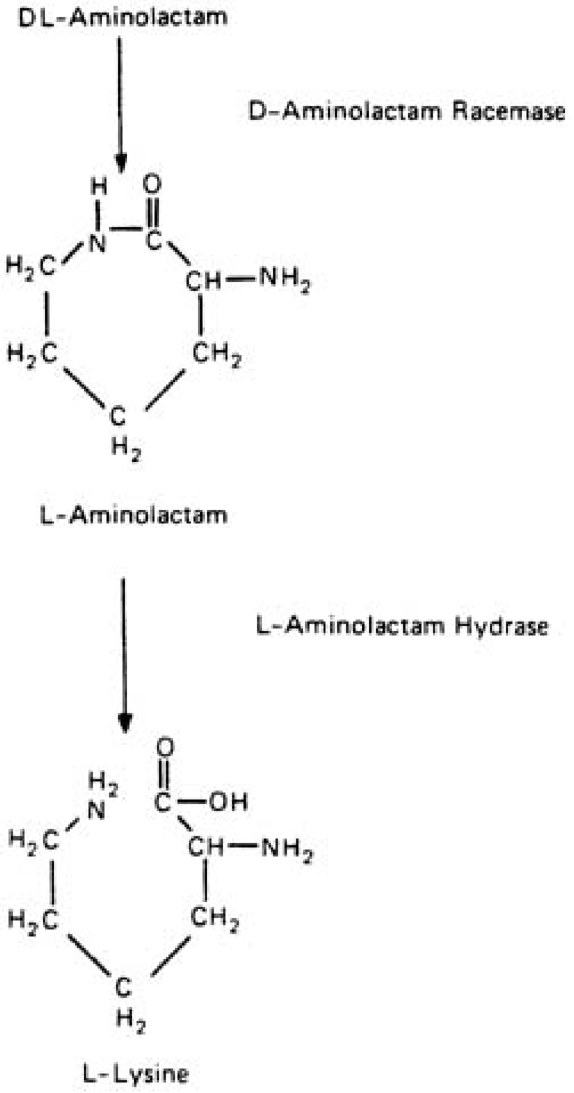
Lysine can be obtained by several different processes:

1. isolation from natural sources
2. chemical synthesis/enzymatic racemization
3. enzymatic conversion
4. fermentation

The chemical pathway to lysine, although perfectly practical and giving high yields, results in a DL-racemic mixture. The resolution of the mixture to the L-configuration is costly.

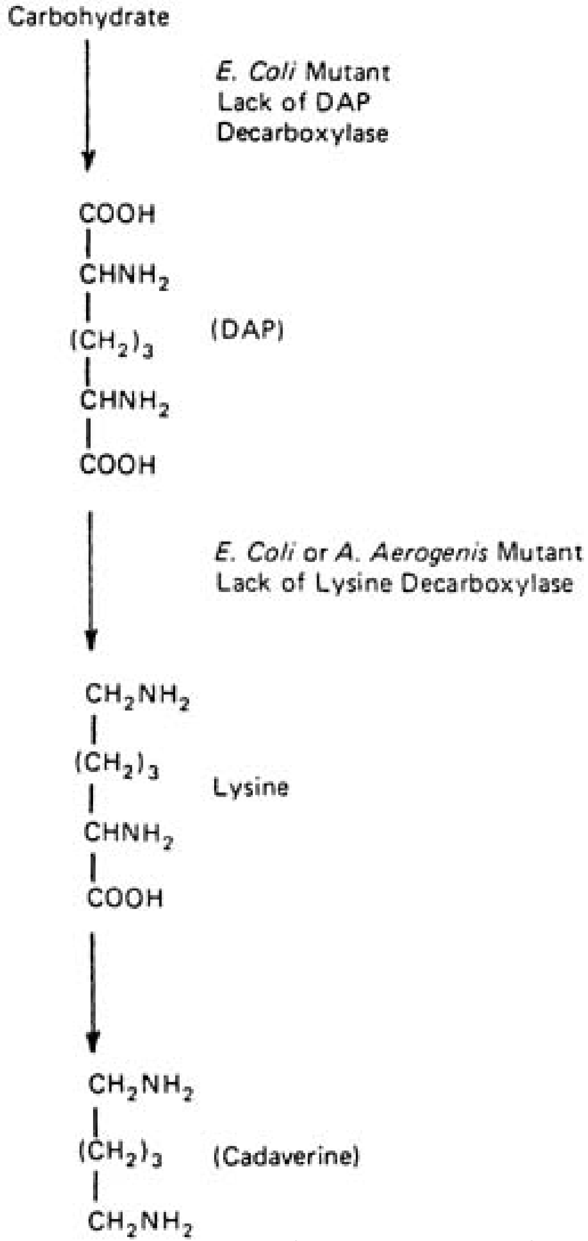
The enzymatic process uses DL-aminolactam as starting material. It consists of the racemization of D-aminolactam and the hydrolyzation of L-aminolactam to form lysine. The reactions are shown in Fig. 24.24. One hundred g/L of DL-aminolactam can be converted into lysine in 25 hr with 100 percent molar yield.

Fig. 24.24 Enzymatic steps to form lysine from DL-aminolactam.



The fermentation process to produce lysine was first developed by Pfizer. *Escherichia coli* synthesized its own lysine requirements by converting carbohydrate and ammonia to α,•-diaminopimelic acid (DAP). Its decarboxylation resulted in lysine. The industrial fermentation employed an *E. coli* mutant devoid of DAP-decarboxylase. It accumulated substantial DAP in the culture medium. After maximum DAP yields were attained, a second organism, another *E. coli* strain or *Aerobacter aerogenis*, was used as a source of DAP-decarboxylase to produce lysine. This second organism was devoid of lysine decarboxylase. Lysine, thus, accumulated in the broth. The reactions are shown in Fig. 24.25.

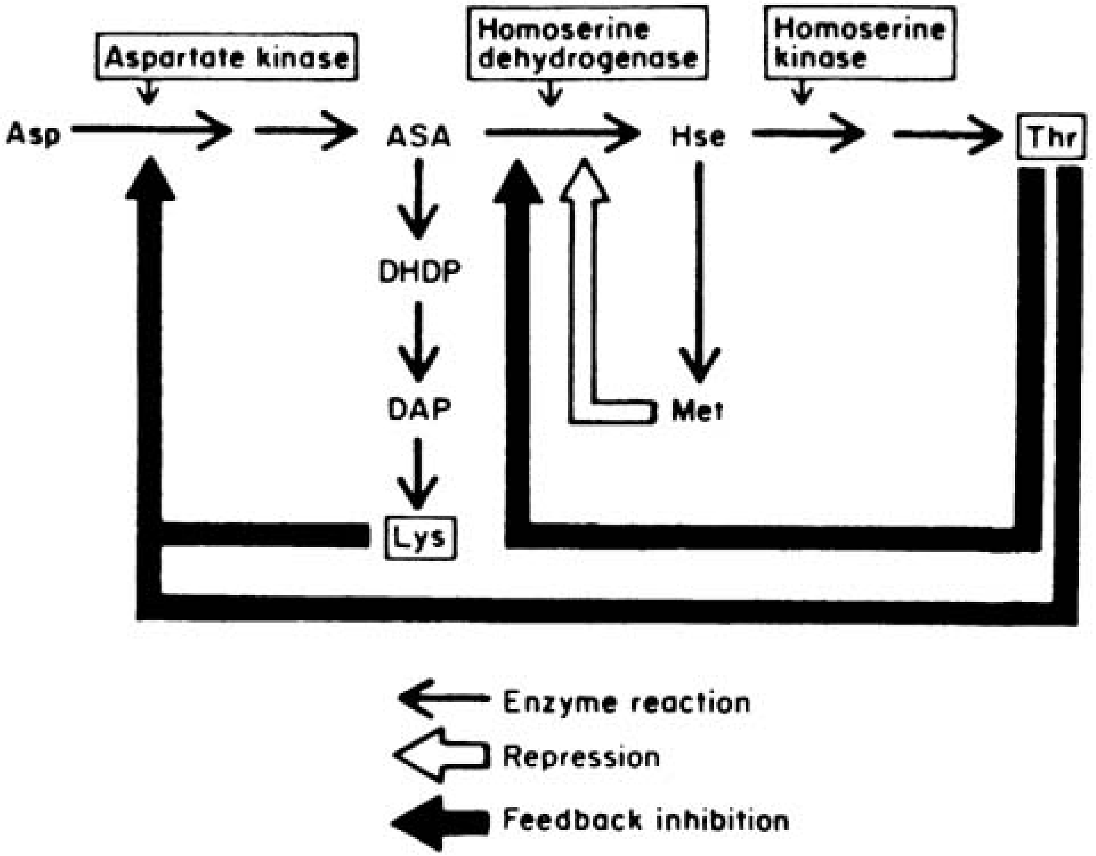
Fig. 24.25 Two-stage fermentation for lysine production.



A single-stage or directed fermentation process for producing lysine was later made possible by using mutants of *Micrococcus glutamicus (Corynebacterium glutamicum)* or *Brevibacterium flavum*.

Both auxotrophic and regulatory mutants have been obtained for overproduction of lysine. Figure 24.26 shows the biosynthetic pathway of lysine and its metabolic controls in *Corynebacterium glutamicum*. The key enzyme, aspartate kinase, is under the influence of the concerted feedback inhibition by lysine and threonine. Threonine, alone, also inhibits the activity of homoserine dehydrogenase. Excess homoserine increases the threonine pool within the cells, and aspartate kinase then is inhibited. An auxotrophic mutant that requires homoserine (or threonine plus methionine) and lacks homoserine dehydrogenase, when grown under suboptimal homoserine concentrations, alleviates the feedback inhibition on aspartate kinase and shunts ASA to form lysine.

Fig. 24.26 Regulation in lysine biosynthesis in *Brevibacterium flavum* and *Corynebacterium glutamicum*. (ASA, aspartate semialdehyde; DHDP, dihydrodipicolinate; Hse, hemoserine; DAP, diaminopimelate.) (Microbial Technology, 2nd ed., Vol. 1, p. 220, Academic Press, New York, 1979.)



In the case of *Brevibacterium flavum*, only aspartate kinase is sensitive to the feedback inhibition.

Regulatory mutants, whose aspartate kinase is desensitized to the concerted feedback inhibition, have been obtained by isolation of lysine analogueresistant mutants, whose growth was not inhibited by a lysine analogue, such as amino ethyl cysteine (AEC). The lysine analogue behaved as a false feedback inhibitor on aspartate kinase and did not allow growth of the parent strain. Only the mutants desensitized to the feedback inhibition grew.

The *Corynebacterium glutamicum* mutant is generally used industrially in direct fermentation of lysine. Molasses is the most common carbon source. Sufficient amounts (over 30 µg/L) of biotin must be included in the medium to prevent the excretion of glutamic acid. This biotin requirement is usually met by using molasses as the carbon source. The fermentation runs at temperatures of about 2833°C, and pH 68. High aeration is desirable. The final product concentration is around 60 g/L, and the fermentation cycle is 4872 hr. The yield of lysine on carbohydrate is about 40 percent. The formation of lysine from molasses can be represented as follows:

$$C_{12}H_{22}O_{11} + 5O_2 + 2NH_3$$

$$\rightarrow C_6H_{14}O_2N_2 + 6CO_2 + 7H_2O$$

Ion-exchange resins are used for isolation of lysine from fermentation broths. The eluted lysine then is crystallized from water. The most common commercial form of lysine used in animal feed is 98 percent lysine monohydrochloride. It is equivalent to 78.4 percent of the amino acid lysine, which can be metabolized by animals to body proteins. The supplementation level is about 0.5 percent lysine in feed. When the *Brevibacterium* mutant is used to produce lysine, the entire fermentation broth may be evaporated and dried, and the dried product used as animal feed supplement.

The standard commercial food-grade lysine (as 98.5% lysine monohydrochloride) is about $5/kg, and lysine as feed supplement is about $2/kg. Its price varies and ties in closely with the cost and availability of fishmeal and soybean cake. The annual world production of lysine is about 500,000 tons. The major producers include Ajinomoto and Kyowa Hakko of Japan, Sewon and Chiel Sugar of South Korea, and ADM, BioKyowa (a subsidiary of Kyowa Hakko), Heartland Lysine (a subsidiary of Ajinomoto), and Midwest Lysine in the United States. Midwest Lysine, a joint venture of Cargill and Degussa, brought on stream its 75,000 tons/year plant at Blair, NE, in 2000. Ajinomoto plans to increase its U.S. and European lysine capacity by 50 percent to a total of about 240,000 tons/year. ADM plans to increase lysine production to 160,000 tons/year at its Cedar Rapids, Iowa, and Decatur, IL, facilities.

Kyowa Hakko plans to increase its global capacity to 120,000 tons/year.

The world demand for lysine is expected to increase continually. The increasing ethanol fermentation capacity for gasohol production provides an opportunity for increased lysine production. Distillers' dried grains (DDG), a by-product of ethanol fermentation used as a protein source in animal feeds, is deficient in lysine and other essential amino acids. It needs to be supplemented with lysine for full-value use.

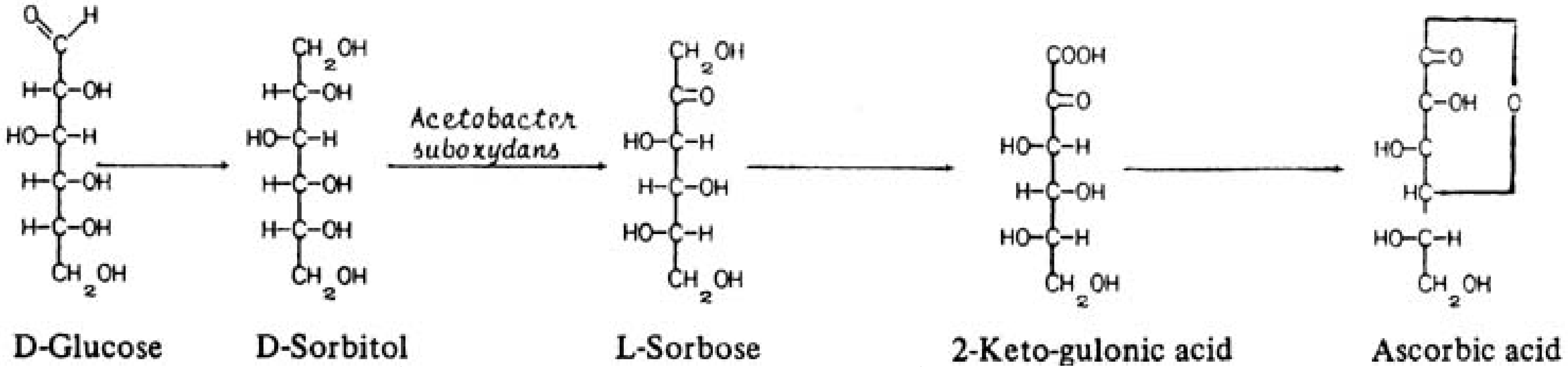
24.6 Vitamins

Many vitamins of medical importance can be synthesized by microorganisms. However, only "two and a half" vitamins are now produced by fermentation. Vitamin B2 (riboflavin) and vitamin B12 (cyanocobalamin) are products of fermentation. The remaining half vitamin produced by fermentation is vitamin C. L-Sorbose is the precursor of vitamin C (ascorbic acid), and L-sorbose is produced microbiologically from sorbitol.

Vitamin C (Ascorbic Acid)

Figure 24.27 gives the reactions involved in the combined microbiological and chemical conversions of glucose to vitamin C. Sorbitol is made by catalytic hydrogenation of glucose. L-Sorbose is produced from sorbitol by the action of several species of bacteria of the genus *Acetobacter*. The most commonly used microorganism is *Acetobacter suboxydans*. Because this organism is very sensitive to nickel ions, it is important that the medium and fermentor be free of nickel. The medium normally consists of 100200 g/L sorbitol, 2.5 g/L cornsteep liquor, and an antifoam such as soybean oil. The medium is sterilized and cooled to 3035° C, and about 2.5 percent inoculum is added. The tank is aerated and sometimes stirred. Yields of 8090 percent of the sugar used are commonly obtained in 2030 hr.

Fig. 24.27 Reactions leading to vitamin C from glucose.



The chemical steps in the conversion of sorbose to ascorbic acid involve the preparation of the diacetone derivative, which is then oxidized; the acetone groups are removed, and the resultant 2-keto-L-gulonic acid is isomerized to the enediol with ring closure.

About one half of the vitamin C is used in vitamin C supplements and multivitamin preparations, one quarter in food additives, 15 percent in beverages, and 10 percent in animal feed. The vitamin C market is about $600 million/year. Its unit price has been under pressure because of oversupply. It lists for about $15/kg, USP, 100 kg, and markets for about $10/kg or less. The two major U.S. manufacturers were Roche with 30 million lb/year capacity at its Belvidere, NJ, facility and Takeda with 11 million lb/year facility at Wilmington, N.C. And the major European manufacturers were Merck and BASF. Merck exited from the market and BASF took over Takeda's bulk vitamins operations in 2000. The four major vitamin C producers in China, Northeast Pharmaceutical, North China Pharmaceutical, Shijiazhong Pharmaceutical, and Jiangshan Pharmaceutical, have a combined capacity of about 100,000 tons/year.

Alternative vitamin C production technologies have been pursued by groups all around the world. One-step fermentation starting from glucose to produce 2-keto-L-gluconic acid was reported to be practiced in China, a microalgae process was developed in the United States and a two-stage fermentation via the intermediate 2,5-diketo-D-gluconic acid was developed in Japan. A U.S. group, on the other hand, succeeded in cloning the 2,5-diketo-D-gluconate reductase gene from a *Corynebacter* sp. and expressing it in an *Erwina* sp. that naturally produces 2,5diketo-D-gluconic acid via a glucose oxidation fermentation. The recombinant *Erwina* culture is thus capable of carrying out the complex oxidation and reduction steps to form 2-keto-L-gluconic acid in a single fermentation.

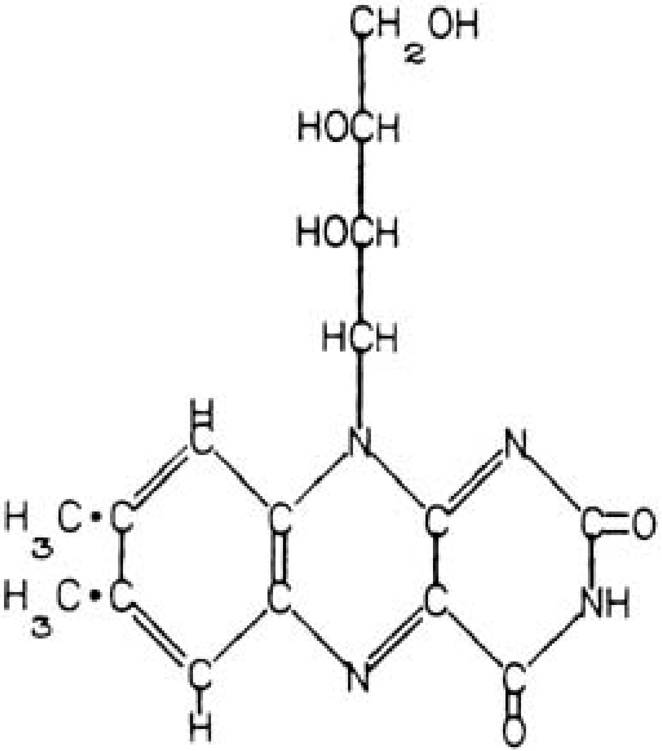
Genencor, a joint venture between Eastman Chemical and Cultor / Danisco, is said to be commercializing a single-step process using a genetically engineered bacteria to convert glucose to 2-ketogulonic acid directly. 2-ketogulonic acid will then serve as a precursor for ascorbic acid production. It's the result of a joint effort that also included Electrosynthesis, MicroGenomics, and Argonne National Laboratory, and was supported by a $15.6 million matching grant from the Commerce Department's Advanced Technology Program.

When ascorbic acid is used in cosmetics, it is phosphorylated to prevent oxidation. Kyowa Hakko has developed an enzymatic process to produce the ascorbic acid-2-phosphoric acid in a new plant with a 30 tons/year capacity.

Vitamin B2 (Riboflavin)

Microbiologically produced riboflavin (structure shown in Fig. 24.28) has long been available in yeast and related preparations in association with many other vitamins of the B-complex. Aside from yeast, the first organism employed primarily for riboflavin production was *Clostridium acetobutylicum*, the anaerobic bacterium used for the microbial production of acetone and butanol. Riboflavin was purely a by-product and was found in the dried stillage residues in amounts ranging from 40 to 70 µg/g of dried fermentation solids. Further research developed improvements, adaptable only to the fermentation of cereal, grains, and milk products by *Clostridium acetobutylicum* to yield residue containing as much as 7000 µg of vitamin B2 per gram of dried solids. This was effected principally by reducing the iron content of the medium to 13 ppm, and fermenting in stainless steel, aluminum, or other iron-free tanks.

Fig. 24.28 Riboflavin: vitamin B2.



Later investigations disclosed that riboflavin could be produced by species of a yeast, *Candida flareri* or *C. guilliermondi*, when grown under aerobic conditions in a medium containing a fermentable sugar, an assimilable source of nitrogen, biotin, and less than 100 µg of iron per liter of medium. Yields as high as 200 mg/L were obtained.

Other studies on a fungus, *Eremothecium ashbyii*, and a closely related organism known as *Ashbya gossyii* resulted in the production of much larger amounts of riboflavin. An aerobic process was used in which the iron content was not critical. Riboflavin was produced in large amounts by the fermentation industry using either the *Eremothecium* or *Ashbya* strains. Yields as high as 1015 g/L were possible.

Late in the 1960s, concurrently with the development of the riboflavin fermentation process, a synthetic means of producing riboflavin was discovered. This synthetic process dominated the production of riboflavin until 1972 when some major fermentation strain and process improvements were made with the *Ashbya gossypii* strain. Since then yields have been significantly improved. The fermentation method now accounts for essentially all the riboflavin produced.

The fermentation lasts 810 days. Cell growth occurs in the first 2 days, and enzymes catalyzing riboflavin synthesis are formed during the growth period. Glycine and edible oil stimulate the formation of riboflavin, but they are not its precursors. The additions of carbohydrate and oil permit the over-production of riboflavin. The C/C yield is about 50 percent on carbohydrate, and about 100 percent on oil.

Upon completion of the fermentation, the solids are dried to a crude product for animal feed supplement or processed to a USP-grade product. In either case, the pH value of the fermented medium is adjusted to pH 4.5. For the feed-grade product, the broth is concentrated to about 30 percent solids and dried on double-drum driers.

When a crystalline product is required, the fermented broth is heated for 1 hr at 121°C to solubilize the riboflavin. Insoluble matter is removed by centrifugation, and riboflavin recovered by conversion to the less soluble form. Both chemical and microbiological methods of conversion have been used. The precipitated riboflavin then is dissolved in water, polar solvents, or an alkaline solution, oxidized by aeration, and recovered by recrystallization from the aqueous or polar solvent solution or by acidification of the alkaline solution.

The major producers of riboflavin include Roche and BASF. There have been significant improvements made in the production process. Roche opened a new riboflavin production facility in Grenzach, Germany in 2000. This year, it notified the European Commission that it intends to market riboflavin made by a fermentation process with a genetically modified *Bacillus subtilis*, both as a nutrient and as a food color. On the other hand, BASF, working with the University of Salamanca, Spain, engineered an improved *Ashbya gossipii* that produces a larger amount of enzymes for riboflavin synthesis, and increased its production capacity by 20 percent at its Ludwigshafen facility without capital investment. Riboflavin is used as a dietary supplement in both human food and animal feed.

The yellow-orange riboflavin crystals are only sparingly soluble in water. To include riboflavin in watersoluble formulations, riboflavin-5′-phosphate sodium is used. The annual production of riboflavin is about 2 million kg. The feed-grade product sells for about $30/kg and the USP-grade product $50/kg.

Vitamin B12 (Cyanocobalamin)

Vitamin B12, cyanocobalamin, is an important biologically active compound. It serves as a hematopoietic factor in mammals and as a growth factor for many microbial and animal species. Its markets are divided into pharmaceutical (9698% pure) and animal feed (80% pure) applications. All vitamin B12 is now made commercially by fermentation.

The microbiological production of vitamin B12 arose from an interesting sequence of events. For many years, liver extract was used to check cases of pernicious anemia. Investigators at Merck discovered that crystalline extracts made from liver tissue contained the highly active compound responsible for the therapeutic action. Identity with the antianemia factor in liver then was established, and the compound was called vitamin B12. Later it was found that spent liquors from streptomycin and other antibiotic fermentations contained appreciable amounts of vitamin B12. Soon vitamin B12 derived from cultures of these fermentations supplanted beef liver as a practical source of the vitamin. Around 1950, materials rich in biomass such as *Actinomycetes* or bacteria broths of antibiotic fermentations or dried sewage residues of activated sludge processes were used for isolating vitamin B12, either in a crude form for animal feeding or in a pure state for medicinal uses. Later, high-producing bacterial strains were specially selected for commercial production. Today vitamin B12 is obtained from fermentations using selected strains of *Propionibacterium* or *Pseudomonas* cultures. A full chemical synthesis of vitamin B12 is known. However, it requires some 70 steps and is of little value for all practical purposes.

The *Pseudomonas denitrificans* strain is most often used for commercial production of vitamin B12. It only requires traditional components in the growth medium, such as sucrose, yeast extract, and several metallic salts. Dimethylbenzimidazole (1025 mg/L) and cobaltous nitrate (40200 mg/L) must be supplemented at the start of the culture in order to enhance vitamin production. Betaine stimulates the biosynthesis of vitamin B12, even though it need not be metabolized by the organism. Choline also has favorable effects in activating some biosynthesis steps or altering the membrane permeability. Glutamic acid, on the other hand, stimulates cellular growth. Because of its relative cheapness and high betaine and glutamic acid contents, beet molasses (60120 g/L) is preferentially used in industrial fermentations of vitamin B12. The fermentation is conducted with aeration and agitation. Temperature is optimal around 28°C, and pH optimal around 7.0. The yield reported in the literature was 59 mg/L in 1971, using a *Pseudomonas* strain. A yield of 200 mg/L was reported for vitamin B12 fermentations using *Propionibacteria* in 1974. It is believed that yields of vitamin B12 have been significantly improved since then.

The isolation of vitamin B12 from fermentation broth, where it is normally present in parts per million, is a brilliant achievement on the part of the chemist and chemical engineer. About 80 percent of the vitamin produced is outside the cells, and 20 percent inside the cells. The whole broth is heated at 80120°C for 1030 min at pH 6.58.5. The heated broth is treated with cyanide or thiocynate to obtain cyanocobalamin. The separation can then be accomplished by adsorption on a cation-exchange resin, such as Amberlite IRC 50. Extraction can also be done by using phenol or cresol alone or in a mixture with benzene, butanol, carbon tetrachloride, or chloroform; or it can be done by precipitation or crystallization upon evaporation with appropriate diluents such as cresol or tannic acid. Using the extraction method, 98 percent pure cyanocobalamin can be obtained with a 75 percent yield.

The total world market for cyanocobalamin is estimated to be about 1000 kg/year. The feed-grade cyanocobalamin is selling for about $10/g.

< previous page

next page >

24.7 Enzymes

All fermentation processes are the result of the enzyme activity of microorganisms. In fact, life itself, whether plant or animal, involves a complex myriad of enzyme reactions.

An enzyme is a protein that is synthesized in a living cell. It catalyzes or speeds up a thermodynamically possible reaction so that the rate of the reaction is compatible with the numerous biochemical processes essential for the growth and maintenance of a cell. An enzyme, like chemical catalysts, in no way modifies the equilibrium constant or the free energy change of a reaction. The synthesis of an enzyme thus is under tight metabolic regulations and controls sometimes that can be genetically or environmentally manipulated to cause the over-production of an enzyme by the cell.

Being a protein, an enzyme loses its catalytic properties when subjected to agents such as heat, strong acids or bases, organic solvents, or other materials that denature the protein. Each enzyme catalyzes a specific reaction or a group of reactions with certain common characteristics. The high specificity of the catalytic function of an enzyme is due to its protein nature; that is, the highly complex structure of a protein can provide both the environment for a particular reaction mechanism and the template function to recognize a limited set of substrates.

Enzymes are used quite extensively now as industrial catalysts. They offer the following advantages in comparison with chemical catalysts:

1. They are specific in action, and thus minimize the occurrence of undesirable side reactions.
2. They are relatively cheap when used in the crude form.
3. They are effective for chemical conversions within the physiological range of pH and at low temperatures and pressures.
4. They are relatively nontoxic and thus are acceptable for applications in food processes and medicinal treatments.
5. They are effective within a wide range of substrate concentrations.

The enzymes of most economical importance, such as amylases, glucose isomerase, and proteases, will be discussed individually. Altogether they account for almost 90 percent of the total sales of enzymes. Other industrially important enzymes with relative small sales volumes are listed in Table 24.6, along with their microbial sources and commercial applications. In addition, there are many microbial enzymes that are used for analytical, clinical, and research purposes; they include hexokinases, pyruvate kinase, uricase, glucose-6-phosphate dehydrogenase, amino acid oxidase, aminopeptidase, and others. Restriction enzymes, endonucleases, have been used widely in recombinant DNA research. Glycosyltransferases have received much attention recently as glycobiology research picks up steam.

TABLE 24.6 Some Industrially Important Enzymes Other than Amylases, Proteases, and Glucose Isomerase

Enzyme	Microbial Sources	Applications
Amino acylase	Bacteria	L-Amino acid production.
Asparginase	*Escherichia coli*	Treatment of acute lymphatic anaemia. Remission of lymphatic leukemia. Anticancer therapy.
Catalase	*Aspergillus niger*	Removal of trace of hydrogen peroxide. Used with glucose oxidase.
Cellulase and hemicellulase	*Trichoderma reesei* *Aspergillus niger*	As a digestive aid. Production of sugar syrup.
Dextranase	*Penicillium funiculosum* *Trichoderma sp.*	Prevention of dental plaque. Removal of dextran impurities in sugar refining.
Glucose oxidase (Notatin)	*Aspergillus niger* *Penicillium notatum*	As oxygen scavenger in food industry. Combined with catalase for removal of glucose or oxygen. Combined with peroxidase for quantitative determination of glucose.
Invertase	*Saccharomyces cerevisiase*	Sucrose hydrolysis to form glucose and fructose which are sweeter and have lower crystallinity. Used in jam making and chocolate manufacturing.
Lactase	*Kluyveromyces fragilis* *Kluyveromyces lactis* *Lactobacillus sp.* *Saccharomyces lactis*	Lactose hydrolysis to form glucose and galactose which are sweeter and more soluble.
Lipase	*Aspergillus niger* *Candida lipolytica* *Mucor javanicus* *Rhizopus arrhizus*	As a digestive aid. Used in waste treatment. Used in tanning. Used to improve flavor.
Pectinase	*Aspergillus niger* *Aspergillus sp.*	Clarification of fruit juice. Increased juice extraction. Pectin decomposition and viscosity reduction.
Penicillin acylase (Penicillin amidase)	*Bacillus megaterium* *Escherichia coli*	Formation of 6-APA from penicillin for production of semisynthetic penicillins.
Pullulanase	*Aerobacter aerogenes* *Klebsiella aerogenes*	Increase of fermentability of starch worts or syrups Structural determination of polysaccharides.
Rennets	*Bacillus polymyxa* *Bacillus subtilis* *Endothia parasitica* *Mucor miehei* *Mucor pusillus*	As coagulants in cheese curd making.

The world sales of enzymes is about $2 billion/year. About 50 percent of the sales comes from proteolytic enzymes for use by the detergent, dairy, and leather industries. The carbohydrases, mainly the amylases, isomerases, pectinases, cellulases, and hemicellulases, used in baking, brewing, distilling, starch, and textile industries, represent nearly 40 percent of the total enzyme market. Lipase, phytase, and other highly specialized enzymes make up the remaining 10 percent of the total enzyme sales. The major enzyme producers are Novo Nordisk and Genencor. Genencor acquired the enzyme business of Gist-Brocades and Solvay in 1995 and 1996, respectively.

Amylases

The commercial importance of amylolytic enzymes is rapidly increasing. These enzymes catalyze the hydrolytic reactions of amylose (the unbranched starch) and amylopectin (the branched starch). Amylases, according to their difference in modes of action, can be divided into:

1. α-Amylase, which hydrolyzes α-1,4-linkages randomly to yield a mixture of oligosacharides, maltose, and glucose.
2. β-Amylase, which cleaves away successive maltose units from the nonreducing end of starch to yield maltose quantitatively.
3. Glucoamylase (or amyloglucosidase), which chops off glucose successively to yield glucose.
4. The debranching enzymes such as pullulanase, which attacks the α-1,6-linkages at the braching point of amylopectin.

Amylases are used extensively in the following ways:

1. For production of grain syrup, glucose syrup, liquid glucose, and crystalline glucose.
2. For production of high-fructose corn syrup in connection with glucose isomerase.
3. To solubilize and saccharify starch for alcohol fermentation in brewing, distilling, and fuel industries.
4. To modify the viscosity of starch used in coating printing papers.
5. To remove starch sizes applied to cotton thread before weaving in the textile industry.
6. For production of maltose-containing syrups in brewing and baking industries.
7. To reduce the viscosity of sugar syrups used in various food and sugar products.
8. As a component in digestive aids.

9. To remove starch stains on clothes/ dishes.

α-Amylase is produced commercially by using both fungal and bacterial species. The fungal amylase has a relatively low heat stability, and its major application is in the baking industry to supplement the variable activity of the amylase present in wheat flour. The bacterial amylase is much more heatstable, and it is used in brewing, starch degradation, alcohol, and textile industries. The organisms commonly used for the commercial production of α-amylase include:

Fungi

1. *Aspergillus oryzae*

Bacteria

1. *Bacillus subtilis*
2. *Bacillus licheniformis*
3. *Bacillus amyloliquefaciens*

Aspergillus oryzae (the green fungus) can be grown in either semisolid or submerged culture. In semisolid culture it produces several enzymes, primarily α-amylase, glucoamylase, lactase, and protease. In submerged culture the production of α-amylase is increased, and the formation of other enzymes becomes minimal. The use of this fungal amylase in the baking industry speeds up the yeast (*Saccharomyces carlsbergenis*) fermentation, produces stiffer, more stable doughs, and improves the texture, porosity, digestibility, and shelf life of bread. The fungal α-amylase delivers its optimal activity at 57 pH and at 5055°C.

Different amylase-producing organisms may require different fermentation conditions for optimal enzyme production. When *Bacillus subtilis* is used, the fermentation medium may contain starch, cornsteep liquor, yeast, phosphate, and some mineral salts. The amylase treatment on starch is often short to prevent the significant accumulation of glucose, which is inhibitory to the *Bacillus* amylase fermentation. The fermentation is run at neutral pH and at around 35°C. Care must be taken to prevent contamination. The time cycle is about 48 hr. Whole mash may be used directly for starch liquefaction and saccharification, or the mash may be processed to produce liquid or crystal enzyme preparation with high purity. The processing, involving filtration or centrifugation of the bacterial fermentation broth, presents real problems to the recovery plant. Pretreatment with a coagulating or flocculating agents is often needed. The amylase produced by this *Bacillus* strain is relatively unstable, but the addition of calcium chloride improves the stability.

Following the solubilization of starch by α-amylase (often of a bacterial origin), further degradation is achieved using fungal glucoamylase. *Aspergillus niger* (the black fungus) is commonly used for the production of glucoamylase. The fungal fermentation starts with a medium containing 2530 percent starch and around 10 percent cornsteep liquor. Incremental or continuous feeding of concentrated nutrients may be used to circumvent the problems caused by a concentrated initial medium. The fermentation pH is about 4.0 and temperature around 28°C. The fermentation has a high oxygen demand. High oxygen tension, however, inhibits enzyme production. Zero D.O. (dissolved oxygen) is not atypical in this fermentation. After the completion of the fermentation in 45 days, the fermentor mash is cooled and filtered to remove cells and insoluble matters. Trans-glucosidase may be removed using clay, destroyed preferentially using proteases at certain pH's and temperatures, or inactivated by magnesium oxide. Contamination of glucoamylase by the transglucosidase activity may result in loss of 510 percent of glucose to form isomaltose and panose by a reversion process. These reversion products also impede the crystallization of glucose.

Fungal glucoamylase in combination with bacterial α-amylase make a complete enzymatic mixture for hydrolysis of starch to glucose. Prior to liquefaction, starch is gelatinized by heat treatment at temperatures above 100°C. The liquefaction to form maltodextrins is aided by the action of bacterial α-amylase. α-Amylase from *Bacillus amyloliquefaciens* functions at 5.57 pH and 90°C and α-amylase from *Bacillus licheniformis* functions at 5.59 pH and temperatures as high as 110°C. In the subsequent saccharification process, an appropriate amount of glucoamylase from *Aspergillus niger* is added to the thinned starch (3050% dry substance) with stirring at 5560°C and 45 pH for 4872 hr. This achieves a final D.E. (dextrose equivalence) of about 97, with about 94 percent of the dry weight being glucose. The equilibrium concentrations of the saccharides formed by resynthesis limit the maximum degree of hydrolysis obtainable.

Because the activity of glucoamylase toward the branching points (the α-1,6-linkages) is low, it may be advantageous to use a debranching enzyme such as pullulanase early in the hydrolysis process.

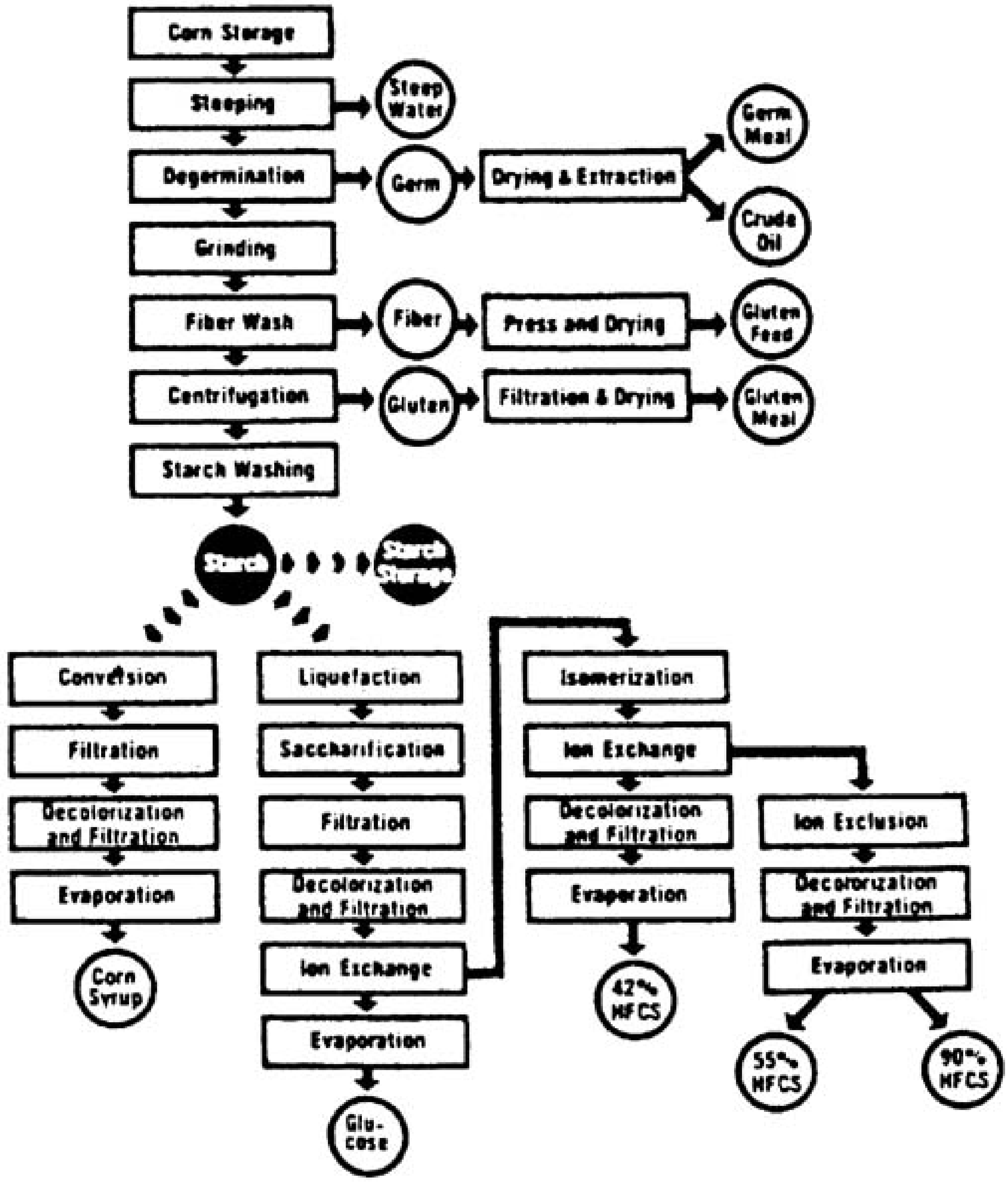
The enzymatic hydrolysis of starch to glucose is commercially preferred to the acid hydrolysis route using hydrochloric acid. The enzymatic process produces fewer side products, does not involves a corrosive acid, and allows the use of less pure starch products whose protein contaminants would, upon acid hydrolysis, give amino acids and browning reactions.

In the last decade, both protein engineering and recombinant DNA technology have been used to improve amylase's thermo-stability by increasing negative charges at its calcium binding site, and to improve the enzyme's resistance against oxidative compounds present in bleach-containing detergents by substituting the oxidation-sensitive methionine residue with leucine. Genecor and Novo Nordisk markets their improved enzymes under the trade names of Purafect OxAm and Duramyl, respectively.

Glucose Isomerase and HFCS (High Fructose Corn Syrup)

Starch degradation using α-amylase and glucoamylase produces corn syrup with glucose concentration up to 94 percent on a dry weight basis. The glucose then can be isomerized to a mixture of glucose and fructose by glucose isomerase, which is present in many microorganisms. The product, commonly known as "high fructose corn syrup" (HFCS), typically contains 42 percent fructose, 50 percent glucose, 6 percent maltose, and 2 percent maltotriose. HFCS is also available in the form of 55 percent syrup, 90 percent syrup, or 99 percent crystal. Figure 24.29 gives a block diagram that shows how a corn wet-milling processor produces its starch, corn syrup, glucose, and fructose syrups.

Fig. 24.29 Flowsheet of a corn wet-milling process.



Fructose, the monosaccharide commonly called fruit sugar, is about 50 percent sweeter than sucrose, the disaccharide familiarly known as table sugar. Sucrose can be hydrolyzed to invert sugar, which is a mixture of fructose and glucose. Liquid invert sugar has been the major sweetener used in soft drinks and as a food ingredient.

The 42 percent HFCS sells for about 20 cents/lb, which is 1520 percent cheaper than liquid invert sugar on a dry weight basis. The soft-drink industry, shifting away from liquid invert sugar, is the major user of 42 percent HFCS. The baking industry ranks as the second largest user.

The 55 percent HFCS sells for about 25 cents/lb, which is 510 percent cheaper than liquid invert sugar on a dry weight basis. Fifty-five percent HFCS, made by combining the 42 percent HFCS with 90 percent HFCS, has about the same degree of sweetness as sucrose. It is used as a sweetener and flavor-enhancer in fruit-flavored soft drinks. Fructose enhances flavors, whereas sucrose masks them.

The 90 percent HFCS, 4050 percent sweeter than sucrose, sells for about 45 cents/lb, which is 4050 percent more expensive than liquid invert sugar on a dry weight basis. Because it's a molecule far smaller than sucrose, a smaller amount of it is needed to sweeten a product to desired levels, and sweetener calories in the product can be reduced by about one-third. The major use of 90 percent HFCS is in dietetic foods and drinks.

The 99 percent fructose in crystal form is obtained by drying the 90 percent HFCS. It is about 70 percent sweeter than sucrose. Being an essentially pure sweetener, it allows the full taste of product flavors to develop. It is finding uses in diabetic and low calorie foods and drinks.

Many organisms are glucose isomerase producers. Most of them produce xylose isomerase with low specificity, and glucose can be an alternative substrate for the enzyme. Table 24.7 lists microorganisms believed to be used industrially for glucose isomerase production, along with their reported yields.

TABLE 24.7 Glucose Isomerase (GI) Producing Microorganisms

Microorganism	Patent Assignee	Yield (GIU[a]/L)
Arthrobacter sp.		
NRRL-B-3726	R. J. Reynolds	4720
NRRL-B-3728		4440
Streptomyces welmorensis		
ATCC-21175	Standard Brands	4640/g
Mutant-1		7540/g
Mutant-2		6680/g
Mutant-3		6000/g
Streptomyces olivaceus		
NRRL-3583	Miles	2560
Mutant NRRL-3916		2960
Streptomyces olivochromogenes		
Mutant CPC-3	CPC	480011,440
CPC-4		57009680
CPC-8		39604440
Actinoplanes missouriensis	Anheuser-Busch	250035,200
Bacillus coagulans	Novo	not known

[a] 1 GIU = The amount of enzyme which converts 1 µmol glucose to fructose/minute.

The desirable application conditions of glucose isomerase are 7.08.0 in pH and around 65°C in temperature.

Glucose isomerase fermentation typically has a cycle time of about 2 days. The fermentation conditions vary from producer to producer. Many glucose isomeraseproducing organisms require xylose and cobalt for full enzyme induction. Xylose is too rare and expensive to be used in commercial fermentation processes, and cobalt ions remaining in the spent fermentation medium constitute a serious environmental hazard. However, mutants that do not require xylose and cobalt have been obtained for commercial production of glucose isomerase.

Almost every known glucose isomerase is an intracellular enzyme, and can only be extracted from the cells in relatively low concentrations; so it is expensive to use this enzyme as a soluble and "once only" enzyme. Immobilized enzyme/cell technology is the key scientific advance that has made the use of glucose isomerase for HFCS production a commercial success. In many industrial cases, whole microbial cells are immobilized by physical means such as entrapment or encapsulation in polymeric materials or by chemical means such as intermolecular cross-linking with glutaraldehyde or covalent binding with diazotized diamino compounds. Commercially, soluble glucose isomerase is also immobilized on DEAE-cellulose. One industrial immobilization process goes as follows: The enzyme-containing cells are homogenized and mixed with glutaraldehyde and inert protein. The resulting gel is shaped into suitable granules, which are then washed and dried.

The immobilized glucose isomerase can usually be used for over 1000 hr at a temperature around 65°C. When the column enzyme activity decreases, the flow rate of the incoming glucose syrup can be adjusted so that the conversion of glucose to fructose is maintained constant.

Proteases

Proteolytic enzymes are by far the most important of the commercially available industrial enzymes. These enzymes, being essential parts of the metabolic system of most living organisms, can be isolated from innumerable sources.

Proteases are classified according to their catalytic mechanism. There are serine proteases, cysteine proteases, aspartic proteases, and metalloproteases. This classification is determined through reactivity toward inhibitors that act on particular amino acid residues in the active site region of the enzyme. The serine proteases are widely distributed among microbes. The enzymes have a reactive serine residue in the active site and are generally inhibited by DFP or PMSF. They are generally active at neutral and alkaline pH. The best-known serine protease is the alkaline protease Subtilisin. The occurrence of cysteine proteases has been reported in only a limited number of fungal species. The proteases are sensitive to sulfhydryl reagents, such as TLCK and iodoactic acid. Aspartic proteases are widely distributed among molds, including *Aspergillus, Penicillium,* and *Rhizopus,* but are rarely found in bacteria and protozoa. They have their maximal activity at low pH, 34. Many of the aspartic proteases are unstable above neutral pH and are not found in cultures growing at neutral or alkaline pH. Most aspartic proteases are sensitive to epoxy and diazoketone compounds. The pepsin-like aspartic proteases have been used commercially in processes such as soybean protein hydrolysis; the rennin-like aspartic proteases are capable of clotting milk in a manner similar to animal rennins. The rennin-like proteases from *Mucor* and *Endothia* species have commercial applications in cheese manufacture. The metalloproteases have pH optima between 5 and 9 and are sensitive to metal-chelating agents such as EDTA. Metalloproteases are widespread, and most of the bacterial and fungal metalloproteases contain zinc. Calcium is required to stabilize the protein structure. The best-known metalloprotease is the thermolysin produced by *Bacillus thermoproteolyticus*.

Microbial proteases with widely different properties are produced commercially. *Bacillus* protease, however, represents more than 95 percent of the sale of all proteases. The most important use of *Bacillus* proteases is in detergents. Proteinaceous dirt often precipitates on clothes, and it coagulates during the normal washing process. The addition of proteolytic enzymes to the detergent can easily dissolve proteinaceous stains, which are otherwise difficult to remove.

The alkaline serine protease of *Bacillus licheniformis*, also known as Subtilisin Carlsberg, is the preferred protease in most nonionic and anionic detergents. It attacks many peptide bonds and easily dissolves proteins. It may be used at temperatures up to 65°C, and its pH optimum is close to 9.0, the pH normally used in washing fluids.

A few other proteases are also used in detergents. The serine protease of *Bacillus amyloliquefaciens* has found some applications, presumably because of its substantial content of α-amylase. This may be an advantage for some applications.

Sales of proteases were small and relatively unimportant till about 1965. Then, the use of proteases in detergents created an explosion in the enzyme industry. But in 1971 the backlash came. Allergic symptoms were discovered in some workers handling enzymes in detergent factories. The public, particularly in the United States, was alarmed, and proteases were taken out of most detergents. It was later found that with proper precautions in handling every risk can be eliminated by using proteases in liquid form or by encapsulating the enzymes.

In the tanning industry, alkaline protease from *Bacillus amyloliquefaciens* is used in combination with sulfite for hide treatment and dewooling. A protease from an alkalophilic *Bacillus sp.* is successfully used for dehairing of ox hides in combination with lime, apparently because of the enzyme's stability at pH as high as 12.

In the brewing industry, there is a development toward substitution of malt with unmalted barley and amylase, glucanase, and protease of microbial origin. The neutral protease from *Bacillus amyloliquefaciens* and the therostable neutral protease *Bacillus subtilis var. thermo-proteolyticus* have been used by brewers successfully to hydrolyze barley proteins into amino acids and peptides.

Furthermore, *Bacillus* proteases are used in the preparation of protein hydrolyzates. The limited hydrolysis of soybean protein is an especially promising application.

The protease fermentation of the *Bacillus* bacteria takes place under strictly aseptic conditions in conventional equipment for submerged fermentations. The aeration rate is about 1 vvm (volume of air per volume of medium per minute). Vigorous agitation is used to improve air distribution and oxygen transfer. The fermentation temperature is around 37°C, and the time cycle is 24 days.

The composition of the fermentation medium is important to the yields of protease. Proteins of many different sources are used in commercial media. Carbohydrates are used as an energy source. The C/N ratio is important to the success of the process. Protein should be present in high concentration, and carbohydrate must not be in excess. A convenient way of obtaining this is to conduct fed-batch fermentation, feeding carbohydrate during the run and maintaining the carbohydrate concentration below 1 percent. Continuous fermentation of protease in commercial scale is not yet known.

The recovery and finishing of *Bacillus* protease involves the following steps:

1. Cooling to about 4°C to prevent microbial spoilage.
2. Precipitating undesirable salts using flocculants or filter aids.
3. Removing all particles by centrifugation or filtration.
4. Removing pigments and odors with activated carbon treatment.
5. Removing bacterial contaminants by filtration.
6. Concentrating at low temperature, either by reverse osmosis or by vacuum evaporation.
7. Recovering protease using precipitation by salts (ammonium or sodium sulfate) or solvents (acetone, ethanol, or isopropanol).
8. Recovering the precipitate by filtration.
9. Drying at low temperature.
10. Encapsulating granules of enzymes in a nonionic surfactant.

Proteases are also produced for special applications by many different organisms, such as *Streptomyces griseus, Aspergillus niger, Aspergillus flavusoryzae, Mucor miehei*, *Mucor pusillus*, and *Endothia parasitica*. The applications include protein hydrolysis to amino acids, milk coagulation, cheese manufacturing, and preparation of digestive aids.

Rennin, an aspartyl protease mentioned earlier, is found in the fourth stomach of nursing calves. Because of a decline in veal consumption and an increased demand for cheese, the dairy industry found it difficult to obtain sufficient quantitites of this enzyme from natural sources. The increasing market of rennin (around $100 million/year) provided the necessary incentive for commercializing a recombinant DNA-derived chymosin. Prochymosin, the self-processing precursor of chymosin, was expressed in *E. coli* as an insoluble, refractile protein aggregate. The cells were lysed, and the insoluble prochymosin mass was isolated by centrifugation. The prochymosin was solubilized by 8 M urea with high pH treatment, renatured, and activated, and then chymosin was purified by ion exchange chromatography.

Most, if not all, proteases are now produced on an industrial scale using genetically engineered organisms that over-express either the wild-type enzymes (through homologous expression or self-cloning) or variants that have been obtained through protein engineering. The major reason for using the genetically engineered organisms is to reduce production cost. The over-expressing Bacilli can secrete proteases into the fermentation broth in very high yields. The secreted protease becomes the dominant protein in the broth. This greatly facilitates and reduces the cost of downstream processing and purification. The other reason for using the genetically engineered organisms is to produce muteins with improved properties. As in the case of amylases, replacement of methionine with non-oxidizable amino acids has improved protease's storage stability in bleach-containing detergents. For example, Novo Nordisk markets a bleach-resistant protease, tradenamed Durazyme. In this case, the bleach-sensitive methionine residue is replaced by alanine.

24.8 SCP (Single Cell Protein)

The term "single cell protein" (SCP) was coined at MIT by Professor C.L. Wilson, in 1966, to represent the cells of algae, bacteria, yeast, and fungi grown for their protein contents. It should be noted that these microbial cells contain, in addition to proteins, carbohydrates, lipids, nucleic acids, vitamins, and minerals.

In the late 1960s and early 1970s, considerable research, development, and commercial interests from all over the world were directed toward SCP production against a background of an increasing deficit in protein supplies, an increasing output, at the time, of cheap petrochemicals, increasing interest in converting waste materials into profitable products, and relatively stable supplies and prices of agricultural products.

Most of this research and industrial effort did not achieve commercialization of SCP, largely because of political, sociological, and psychological factors. However, the SCP era witnessed several important advances in fermentation technology, such as continuous culture, improved bioreactor design, and computer control of fermentation processes.

Most of the SCP currently produced is for animal consumption. Minor quantitites of SCP are being produced for human use. This use has been controversial because of the content of nucleic acid, which can increase blood uric acid levels in a segment of the population. However, many traditional foods contain SCP, such as cheese, yogurt, and sour cream.

The following paragraphs summarize, by substrates, the status of major SCP projects.

Carbon Dioxide for SCP

The photosynthetic *Chlorella* (the green algae) has been grown on carbon dioxide mostly in the Oriental countries. Its market is largely in Asia, Japan in particular, as ingredients for health food and drinks.

Methane for SCP

Methane has been of interest as a substrate for SCP production because it is available in large quantity and high purity and, in many cases, is wasted at oil-producing wells. Many bacteria and some fungi utilize methane as a carbon and energy source for growth. A bacterial SCP continuous process was developed to the pilot scale by Shell in England using *Methanomonas capsulatas* and mixed cultures with good productivities and yields. The following factors forced the operation to be discontinued:

1. Limited transfer of methane and oxygen from the gas phase to the bacterial cells.
2. Explosion hazards.
3. High cooling demands because of excessive heat production.
4. Presence of inhibitory products.
5. High capital requirements because of problems stated above.

It is estimated that 300 million ft3 of methane is produced each day in the United States alone at a typical concentration of 0.20.6 percent in the ventilation air from coal mines. Purification of the methane from the vent gases is very expensive and impractical. Bioengineering Resources Inc. is reported to have demonstrated the feasibility of converting the dilute methane emissions to SCP, which commands a price of about $0.20/kg as an animal feed supplement.

Statoil of Trondheim, Norway, in a joint venture with Nycomed Amersham of Little Chalfone, U.K., has, since 1999, operated a 10,000 tons/year plant at its Tjeldbergodden facility to produce SCP from natural gas. The product, Bioprotein, is sold as animal feed to Norwegian farm cooperatives. It is produced in a patented loop fermentor, using a naturally occurring *Methylococcus capsulatus*.

n-Paraffins for SCP

The *Candida* yeast grown on *n*-paraffins was widely investigated for production of SCP. Continuous processes, using air-lift fermentors and operating nonaseptically in most cases, were developed by British Petroleum (using *Candida lipolytica*), Gulf (using *Candida tropicalis*), Kanegafuchi (using *Candida* sp.), Liquichemica Biosintesi (using *Candida maltosa*), and others.

Italproteine, a joint venture of BP and Anic (the petrochemical arm of the Italian state-owned Ente Nazionale Idrocarburi), constructed a 100,000 tons/year plant at Sarroch, Sardinia. Italproteine tradenamed its product "Troprina," made by BP-developed know-how. Liquichimica Biosintesi, owned by Liquichimica (a subsidiary of Italy's Liquigas), built its 100,000 tons/year plant at Saline di Montebello, Calabria. The product was tradenamed "Liquipron," made by the Kanegafuchi process. The operation of these two SCP plants has been blocked by the Italian government's health authorities on questions of levels of hydrocarbon residues.

Plans to construct large-scale yeast SCP plants (60,000120,000 tons/year each) in Japan by Kanegafuchi, Kyowa Hakko, Asahi, Mitsui Toatsu, and Dainippon were canceled because of public protests over questions on safety of the products and subsequent Japanese government regulatory action banning the use of these products.

However, yeast SCP plants are believed to be in operation in the former Soviet Union using purified *n*-alkanes as substrates.

Methanol for SCP

The events in Italy and Japan, as well as the dramatic increases in petroleum prices in the mid-1970s, stimulated much increased effort in selecting substrates other than *n*-paraffins for SCP production. Methanol has received perhaps the greatest attention. The use of methanol as a substrate for SCP production has a number of advantages, principally the following:

1. High solubility in water.
2. Low explosive hazards of methanol oxygen mixtures.
3. Moderate heat liberation (3.35.7 kcal/g of cells vs. 4.48.0 kcal/g of cells grown on *n*-paraffins).
4. Freedom from traces of aromatic hydrocarbons.
5. Ease of removal from the microbial cell product.

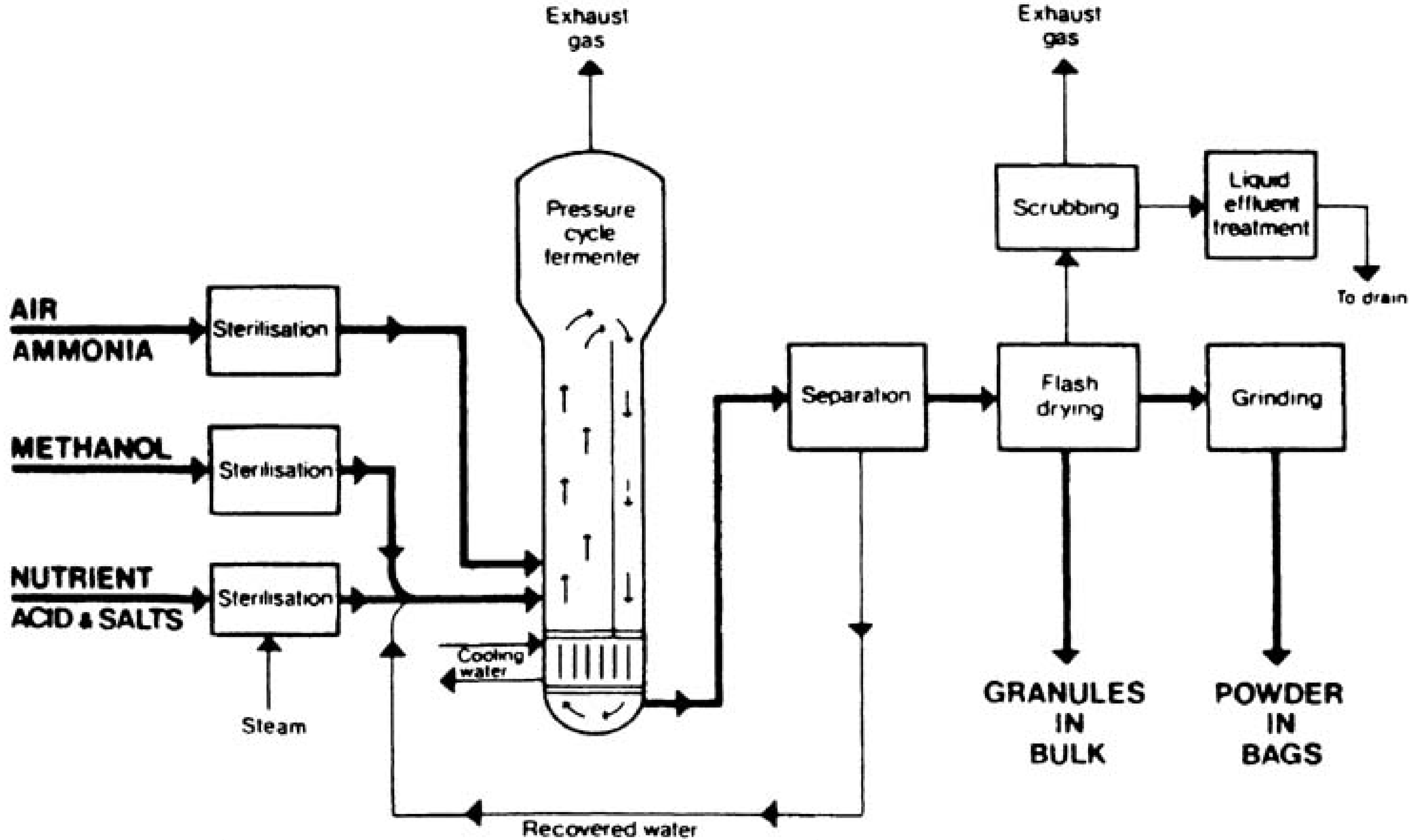
Organizations in many parts of the world have worked on methanol processes using both yeasts and bacteria, including Shell and ICI in England, Societa' Italiane Resine in Italy, Hoechst AG and Gelsenberg in Germany, Norsk Hydro in Norway, AB Marabou in Sweden, Yissum Research Development in Israel, Phillips Petroleum and Tenneco in the United States, and Mitsubishi in Japan. In addition, much work has been done in academic institutions. Among the industrial organizations, ICI and Hoechst are the ones that remain most active. Both use bacterial processes that have the advantages over yeast processes of higher growth rates, productivities, and yields.

Hoechst operates a 1000 tons/year pilot plant near Frankfurt. *Methylomonas clara* is grown on a mixture of air and methanol containing ammonia, water, and other essential mineral nutrients. The dried product, tradenamed Probion, contains some 70 percent protein, 10 percent nucleic acid, 8 percent fats, 7 percent mineral "ash," and 5 percent water. A purified version, with essentially all nucleic acids and fats removed, has a protein content of more than 90 percent, with minerals and water making up the balance. Hoechst intends to market this purified product, called Probion-S, for human consumption.

ICI has developed the most advanced process for SCP production from methanol. *Methylophilus methylotrophus* is the organism used. A novel type of air-lift bioreactor, called the pressure cycle fermentor, has been developed for use in the process. It is designed to maintain a high oxygen transfer rate and a homogeneous liquid phase without excessive shear, to remove the heat liberated during the growth phase at high productivities, and to avoid contamination problems encountered with conventional stirred fermentors because of leaks through drive shafts and mechanical seals. Air and fresh medium are introduced at the base of a tall vertical columnthe "riser." The riser is connected horizontally to another smaller diameter column, the "downcomer," at both the top and the base. A driving force for the movement of broth at high velocities is provided by the difference in air holdup between the riser and the downcomer. The high hydrostatic head and the level of turbulence in the riser provide excellent oxygen transfer because of the increased pressure, the small bubble size, and the mixing; up to 50 percent of the incoming oxygen is transferred. Spent air and CO2 (stripped by the nitrogen content of the air) are disengaged in the upper horizontal section. The broth is cooled in a simple heat exchanger set in the downcomer. ICI has also developed a proprietary agglomeration process for the initial separation of bacterial cells from the growth medium, which permits the final centrifugation of a much higher solid slurry than is possible otherwise.

A 70,000 tons/year plant came on stream in early 1980. A schematic diagram of the ICI protein process is shown in Fig. 24.30. The final product, tradenamed Pruteen, contains more than 70 percent protein and is sold as animal feed.

Fig. 24.30 Flowsheet of the ICI SCP process. (*Process Biochem.*, 12(1), 30, 1977. Reprinted with permission of the publisher.)



The ICI pressure cycle fermentor at Billingham, England, measuring 200 ft high and 25 ft in diameter, is claimed to be the largest such vessel in the world.

The naturally occurring *Methylophilus methylotrophus* used in the ICI SCP process is relatively inefficient in assimilating nitrogen. The inefficiency is caused by the absence of the gene that codes for glutamate dehydrogenase, the enzyme essential for incorporating hydrogen and ammonia into glutamic acid, the major pathway for the formation of α-amino groups directly from ammonia. Instead, *Methylophilus methylotrophus* possess the gene for producing glutamate synthase, which leads to a less efficient nitrogen assimilation pathway. It is reported that ICI has successfully overcome the problem by first inactivating the glutamate synthase enzyme using classical mutation techniques and then inserting into *Methylophilus methylotrophus* the glutamate dehydrogenase gene from *Escherichia coli*.

Ethanol for SCP

Ethanol can be used by certain bacteria and yeasts as a substrate for SCP production. It is completely water soluble, available in large quantity and high purity, and easy to store and transport. In addition, it has the advantage over methanol of low toxicity and good acceptability to regulatory agencies as a raw material for producing a food-grade end product. Organisms of interest that utilize ethanol in SCP production include:

1. *Candida utilis* (used by Pure Culture Products, formerly Amoco Foods).
2. *Acinetobacter calcoaceticus* (used by Exxon and Nestle).
3. *Candida acidothermophilum* and *Candida ethanothermophilum* (used by Mitsubishi Petrochemical).

The yeast-based SCP processes using ethanol as a substrate have been developed to a greater extent than the bacterial processes.

Pure Culture Products in the United States has taken the approach of using torula yeast, which is well known and has long been established as a food ingredient. As such, they have been able to obtain clearance for their ethanol-grown product as a food ingredient, and they are already marketing material from their plant in Hutchinson, Minnesota, which is ultimately designed to produce up to 15 million lb of product per year. The material, tradenamed Torutein, is being sold in the United States as a nutritional supplement and flavor enhancer for such processed foods as meat patties, pasta, baked goods, frozen pizzas, and sauces.

In the Pure Culture Product's continuous process, the ethanol concentration is maintained at approximately 200 ppm. The nitrogen source, either aqueous or anhydrous ammonia, is supplied continuously to keep the pH in the desired range. Other macro and micro elements are also supplied continuously. All liquid streams except ammonia are sterilized at 149°C. Ammonia is sterilized by filtration, and air is sterilized by compression and filtration. Typical operating conditions are 30°C, pH 4.6, and aeration and agitation to give oxygen absorption rates in the range of 100140 mmol of O2 L/hr. Cell concentrations of 67 g/L on a dry weight basis are obtained at a dilution rate of about 0.3 hr.

Torutein is now marketed by Burns Philip food and fermentation flavor ingredients, a unit of Ohly. Its web site, http://www.ohly.de/museum.htm, offers an interesting virtual tour of its Yeast Factory Museum.

Carbohydrates for SCP

The raw materials used for SCP production in this category are carbohydrate-containing wastes or by-products of various manufacturing processes. These carbohydrate-containing materials include molasses, sulfite liquor waste, whey, and cellulosic wastes.

Yeasts grown on molasses are produced for both food and feed uses by Standard Brands, Universal Food, and Yeast Products in the United States, and by many other processors abroad.

Sulfite liquor waste is used for feed yeast production by Boise Cascade, Rayonier (a subsidiary of ITT), and Wausau Paper Mills in the United States, by Environcon in Canada, by Attishola in Switzerland, and by G. A. Serlacins in Finland. Yeast plants utilizing sulfite liquor or wood hydrolyzate are also in operation in Austria, Rumania, Poland, Germany, Japan, the People's Republic of China, Saudi Arabia, and the former Soviet Union.

One well-publicized SCP process using spent sulfite liquor as a substrate is the Pekilo process developed by Tampella in Finland. The Pekilo process is said to be the first commercial continuously operating process in which filamentous fungi are cultivated for SCP production. A plant based on the Pekilo process was built jointly by United Paper Mills and Tampella at Jamsankoski in Finland with a capacity of 10,000 tons/year using a microfungus *Paecilomyces varioti*. Its product, the Pekilo protein, is sold for animal feed.

Whey, a waste of cheese manufacturing, is fermented with *Saccharomyces fragilis* by Amber Laboratories, Stauffer Chemical, and Kraft in the United States for both food and feed uses. For the production of food yeast, the yeast cream obtained by centrifugation is spray-dried. For feed yeast, the whole fermentor broth is concentrated by evaporation and then spray-dried.

Thermo Tech of Ontario, Canada, has since 1994, been operating plants of varying sizes to convert 2001200 tons/day of municipal wastewater sludge, manure, and pulp and paper plant sludge into feed or fertilizer products. At these "Thermo Master" plants, raw slurry undergoes a rapid, high-temperature fermentation using a thermophilic bacterium that transforms waste into SCP. The high fermentation temperature destroys pathogenic microbes present in the waste. Thermo Tech, in addition to selling its SCP to feed companies at about $200/ton, also receives portion of its income by taking away and processing the waste sludge generated by companies like Coca Cola, Kraft Foods, and Nestle. These companies would otherwise have to send its waste to landfill or wastewater treatment plants.

Cellulose and cellulosic wastes have received a great deal of attention in recent years as potential substrates for SCP production. Delignification, decrystallization, hydrolysis, and microbial growth are the important steps involved in SCP production from cellulosic materials. At the present time, no commercial production of SCP from cellulose is known. However, the utilization of cellulose, the most abundant renewable organic resource on earth, is of great importance from both technical and commercial points of view, not only for SCP production but also for chemicals and fuels.

24.9 Antibiotics

The science of microbiology was established by the French chemist Louis Pasteur. Among his many contributions, the discovery of the existence of microscopic forms of life should be especially pointed out: yeasts were established as the cause of the alcohol fermentation of wine. He later established pathogenic bacteria as the cause of many infectious diseases and even recognized viruses as causative agents of other infections.

Shortly before World War II, another significant role of microorganisms began to emerge. Alexander Fleming, an English bacteriologist, is credited with first suggesting that the product of one microorganism might be used to inhibit the growth of another. Fleming observed that a chance contaminant (a *Penicillium* mold) clearly prevented the growth of a pathogenic *Staphylococcus* he was culturing in a Petri dish. He succeeded in establishing some of the simple properties of the mold product penicillin and published his results in 1929. Nearly a decade later, another group of English biochemists undertook to further examine the phenomenon in the course of a broad research program for better chemotherapeutic agents. Florey, Chain, Heatley, and Abraham, of Oxford University, succeeded in isolating a small quantity of penicillin concentrate and by 1941 unequivocally demonstrated its potential usefulness in *Staphylococcus* septicemia. British government officials recognized its possible usefulness in military medicine, but could not further develop the discovery because of the serious war conditions then prevailing. The research was brought to the attention of the government and industry in the United States, with the result that an international governmentindustry research program was established to produce this remarkable chemotherapeutic agent. The success of this program established a new class of powerful therapeutic agents, the antibiotics, which have revolutionized medical practice. Thus, the product of one microorganism is used to combat an infection caused by another. To supply the huge amounts of antibiotics needed in modern medicine, the fermentation industry has undergone a virtual revolution.

Since the early 1940s, an intensive search for new and useful antibiotics has been in progress throughout the world. About 12,000 antibiotics had been discovered from microbial origins through 1994. Among them, about 6600, 2600, 1400, and 1300 were produced, respectively, by *Streptomyces* spp., filamentous fungi, non-actinomycete bacteria, and non-Streptomycete strains of actinomycetes. Of the thousands of antibiotics known, over 150 are in clinical use, with approximately one-half of them produced by microbial fermentation and one-half by a combination of microbial synthesis and chemical modification.

The antibiotic market was around $7 billion in 1997. Sales of cephalosporins, macrolides, penicillins, quinolones, and tetracyclines were about $2.3, $1.4, $1.4, $1.2 and $0.2 billions, respectively. These antibiotics act by inhibiting the synthesis of cell wall (cephalosporins and penicillins), protein (macrolides and tetracyclines), and nucleic acid (quinolones). Quinolones are commercially produced by chemical synthesis.

Penicillins

The original mold observed and preserved by Fleming was a strain of *Penicillium notatum*, a common laboratory contaminant. Later cultures of *Penicillium chrysogenum* were found to be better producers of penicillin, and the present industrial strains have been derived from this species. The original strains produced the antibiotic only by surface fermentation methods and in very low yields, a few ppm. Gradually, improved media and the eventual discovery of strains productive under submerged aerobic fermentation conditions led to dramatic yield increases, which made commercial production a reality. Subsequent improvements, principally in culture selection and mutation to productive strains, further improved yields, until today broths often contain 2030 g/L of penicillin. Dramatic price reduction has come with improved production. For instance, a million-unit vial of penicillin (1667 units = 1 mg of potassium penicillin G) had a wholesale price of $200 in 1943. In 1952, the same vial had a wholesale value of only $1.30. Today, a million units of potassium penicillin G is selling for as little as $0.50 or approximately $20/kg of free acid.

The original *P. chrysogenum* strains produced large amounts of unwanted yellow pigments that were difficult to remove from the recovered penicillin. Today, nonpigmented mutants, a strain known in the industry as Wisconsin 49133 (or progeny therefrom), are universally employed. The desired culture is propagated from a laboratory stock in small flasks and transferred to plant inoculum tanks. After 24 hr these are used to inoculate larger fermentors, which contain a typical production medium such as that shown below.

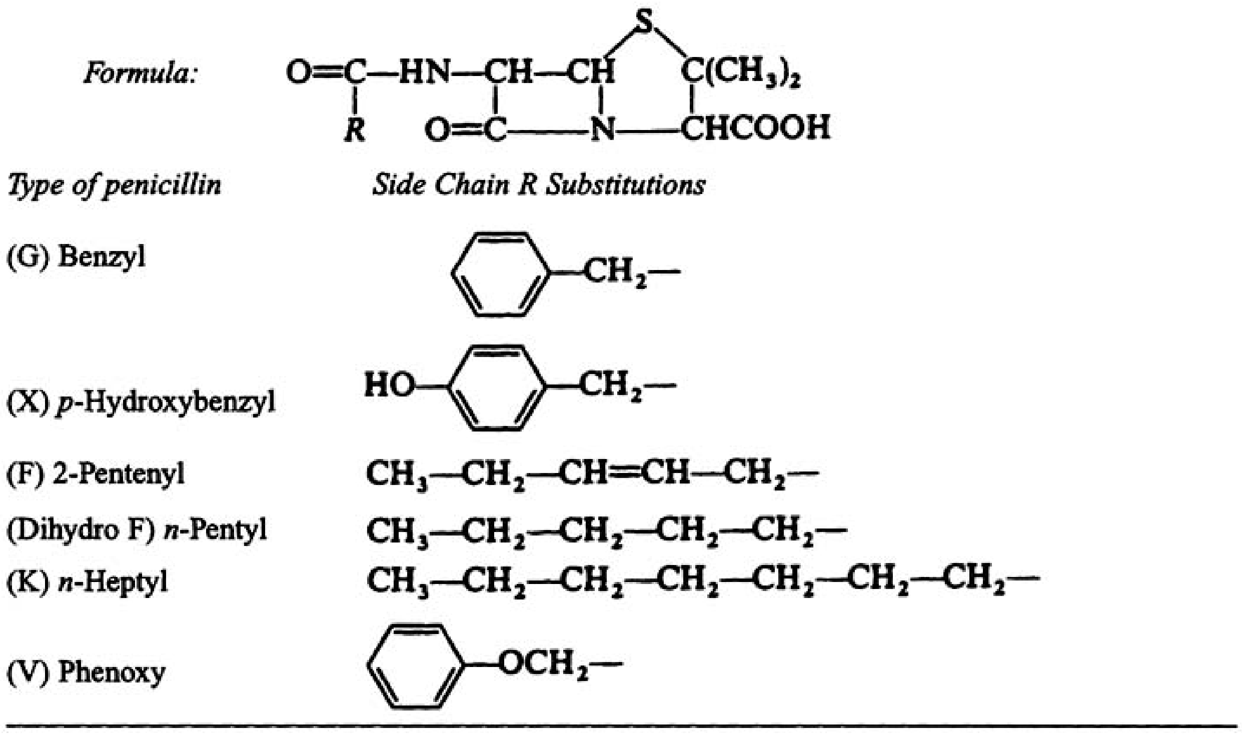
Components	Grams/Liter
Corn steep liquor	30
Lactose	30
Glucose	5.0
NaNO3	3.0
MgSO4	0.25
ZnSO4	0.044
Phenyl acetamide (precursor)	0.05
CaCO3	3.0

The medium is usually sterilized batch-wise, cooled to 24°C, and inoculated. The time of fermentation may vary from 60 to 200 hr. Sterile air is blown through the tank, usually at a rate of one volume per minute.

When penicillin concentration reaches its peak potency, as determined by microbiological or chemical assays, the broth is clarified by means of rotary vacuum filters. The penicillin, being acidic, is extracted from the aqueous phase into a solvent, such as methyl isobutyl ketone or amyl acetate, at a pH of 2.5 by means of a continuous countercurrent extractor, such as a Podbielniak. The penicillin extract is then reextracted with an aqueous alkaline solution or a buffer at a pH of 6.57.0. A 90 percent recovery is made at this step. The aqueous solution is chilled, acidified, and extracted again with a solvent, such as ether or chloroform. The solvent extract is then reextracted into water at a pH of 6.57.0 by titration with a solution of base. The base used depends on which salt of penicillin is desired. The popular forms are sodium or potassium salts. A typical flow sheet for antibiotic recovery is shown in Fig. 24.15.

Table 24.8 gives the structural formulae of the "natural penicillins," comprising several closely related structures with aliphatic and aromatic substitutions to the common nucleus. The early impure product contained mixtures of these types. For several reasons penicillin G became the preferred type and the crystalline product of commerce. Phenylacetic acid or its derivatives are used as precursors in the fermentation medium to enhance penicillin biosynthesis and suppress the production of the less desirable types.

TABLE 24.8 Structural Formulae of Natural Penicillins



Formula:

Type of penicillin	*Side Chain R Substitutions*
(G) Benzyl	
(X) *p*-Hydroxybenzyl	
(F) 2-Pentenyl	$CH_3—CH_2—CH{=}CH—CH_2—$
(Dihydro F) *n*-Pentyl	$CH_3—CH_2—CH_2—CH_2—CH_2—$
(K) *n*-Heptyl	$CH_3—CH_2—CH_2—CH_2—CH_2—CH_2—CH_2—$
(V) Phenoxy	

The fact that proper selection of precursors could lead to new variations in the penicillin side chain offered the first source of synthetic penicillins. Penicillin V, derived from a phenoxyacetic acid precursor, attracted clinical use because of its greater acid tolerance, which made it more useful in oral administration.

The widespread use of penicillin eventually led to a clinical problem of penicillin-resistant staphylococci and streptococci. Resistance for the most part involved the penicillin-destroying enzyme, penicillinase, which attacked the beta-lactam structure of the 6-aminopenicillanic acid nucleus (6-APA).

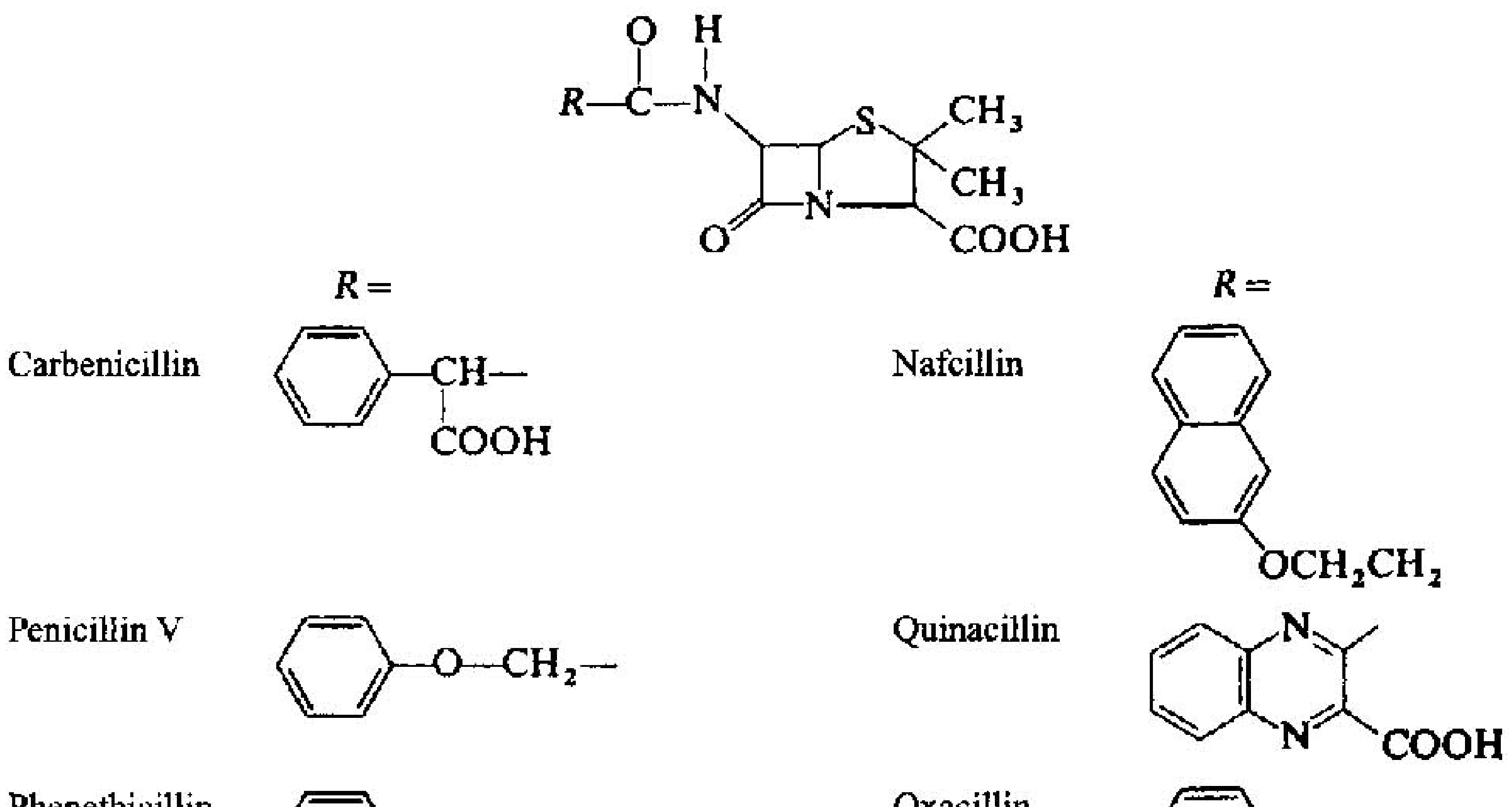
In 1959, Batchelor and coworkers in the Beecham Research Laboratories in England discovered that the penicillin nucleus, 6-APA, accumulated during fermentation when side chain precursors were omitted. This 6-APA could be used for the chemical synthesis of entirely new types of penicillin by coupling with new side chains. Shortly thereafter, several sources of penicillin amidase were found that would cleave the phenylacetyl side chain from penicillin G, thus producing a more economical source of 6-APA. A vast number of "synthetic penicillins" have been generated, and a few have achieved clinical importance. Several objectives were sought:

1. To broaden the inherent utility of penicillin to include gram-negative pathogens not inhibited by the natural penicillins.
2. To improve its stability and absorption.
3. To increase its resistance to penicillinase-producing pathogens.
4. To decrease allergenicity.
5. To improve other factors pertinent to clinical use.

The broad objectives have been achieved with varying degrees of success.

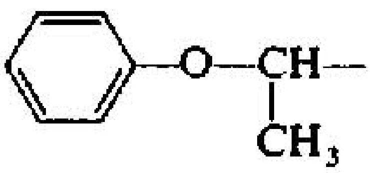
Table 24.9 shows the structures of some of the semisynthetic penicillin that have become important chemotherapeutics. Semisynthetic penicillins, on the average, sell for about $150/kg.

TABLE 24.9 Structural Formulae of Some Semisynthetic Penicillins

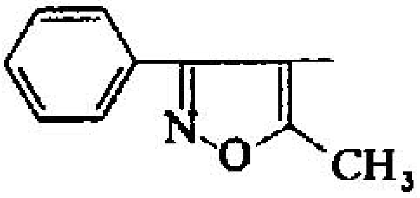


R =		*R* =	
Carbenicillin		Nafcillin	
Penicillin V		Quinacillin	
Phenethicillin		Oxacillin	

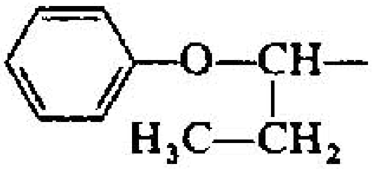
Phenethicillin



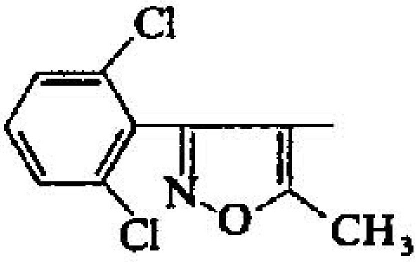
Oxacillin



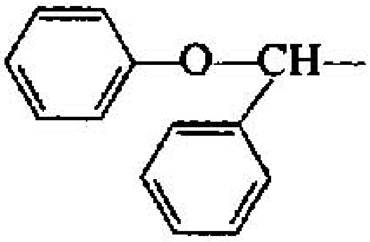
Propicillin



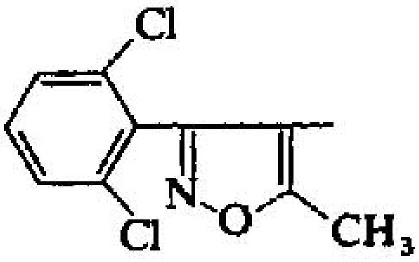
Cloxacillin

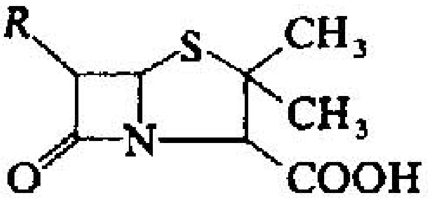


Phenbenicillin

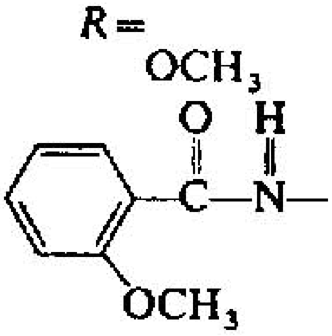


Dicloxacillin

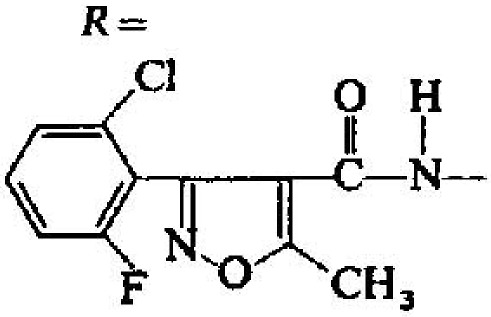




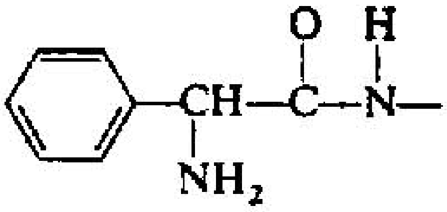
Methicillin



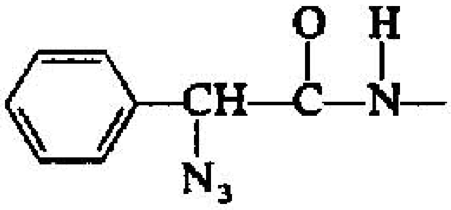
Flucoxacillin



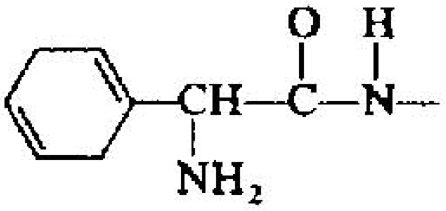
Ampicillin



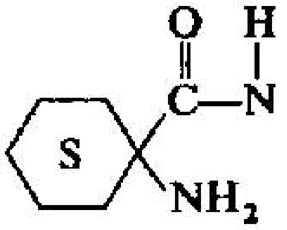
Azidocillin



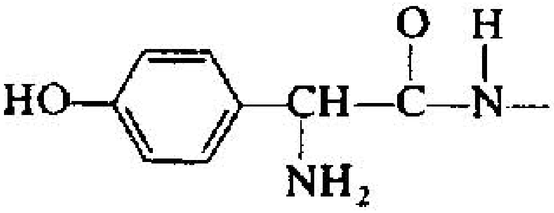
Epicillin



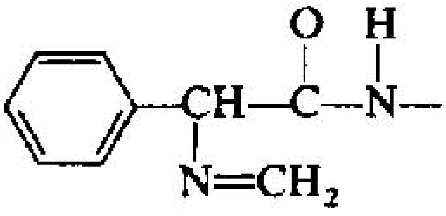
Cyclacillin



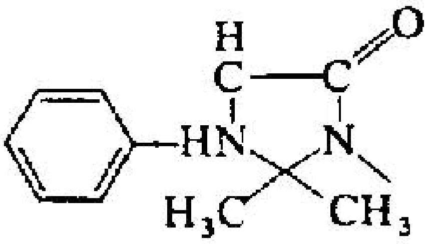
Amoxicillin



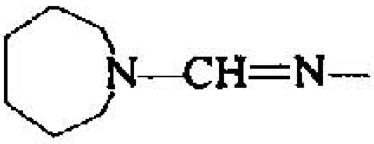
Metapilicillin



Hetacillin



Mecillinam



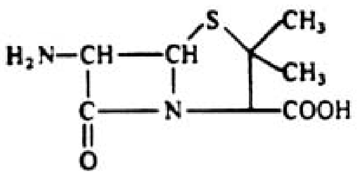
The natural penicillins, primarily G and V, have a relatively narrow spectrum. They act primarily on gram-negative organisms. Semisynthetic penicillins such as ampicillin, carbenicillin have a broader spectrum. Some, such as methicillin, oraficillin, and oxacillin, are resistant to penicillinase. In 1984, Beechem introduced Augmentin, which was the first combination formulation of a penicillin (amoxillin) and a penicillinase inhibitor (clavulanic acid). In 1985, Beechem introduced Timentin, a ticarcillin/clavulanic acid combination, and Pfizer, in 1987, introduced Unasyn, an ampicillin/sulbactam sodium combination.

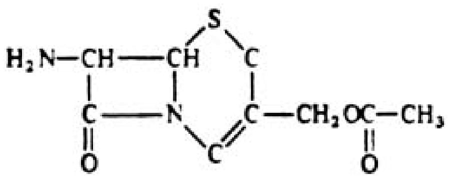
Cephalosporins

In 1948, Professor Guiseppe Brotzu isolated a *Cephalosporium* culture from seawater near the sewer discharge of Cagliarri, Sardinia. This culture produced a broth inhibiting both gram-positive and gram-negative bacteria. The team of Florey, Abraham, and Newton at Oxford University, isolated a compound identified as cephalosporin N; and during the same period a group in the Michigan Department of Health isolated synnematin B from another strain of *Cephalosporium*. Synnematin B and cephalosporin N proved to be identical. Structure studies eventually proved these antibiotics to be a new type of penicillin, α-aminoadipyl-6-APA (also called penicillin N), a naturally produced penicillin with gram-negative activity.

In the course of studies on the Brotzu strain of *Cephalosporium*, Abraham and Newton detected small quantities of a second antibiotic, cephalosporin C. Painstaking work proved it to be chemically similar to penicillin N, but not a penicillin. It had pronounced gram-negative activity, was more stable to acid, and was not destroyed by penicillinase. It possessed the same α-aminoadipyl side chain as a new penicillin, but the nucleus was 7-aminocephalo-sporanic acid (7-ACA). 7-ACA contains a six-membered 1,3-dihydrothiazine ring instead of the five-membered thiazole ring in 6-APA. The structures of 6-APA and 7-ACA are shown in Fig. 24.31.

Fig. 24.31 Structures of 6-APA and 7-ACA.





Microbiological processes for production of cephalosporin C resemble in many respects those used for penicillin production. Special strains of *Cephalosporium* have been selected that produce more cephalosporin C and less cephalosporin N than the parent culture. The growth of these strains in certain special fermentation media has resulted in higher antibiotic titers. Even with these improvements in processing, the antibiotic yields, averaging 1020 g/L, are much lower than those reported for the penicillins. Cephalosporins sell for around $250/kg.

As in the penicillin studies, the possibility of further improving the chemotherapeutic properties of cephalosporin C was apparent if the 7-ACA nucleus could be obtained. Enzymatic cleavage of the side chain failed, as did the use of precursors to generate new side chains; however, successful chemical methods have been found. Several semisynthetic cephalosporins have been produced and are used clinically.

The leading agents in this category include cefaclor (first-generation cephalosporin, developed by Lilly), cefoxitin (second-generation, by Merck), cefuroxime axetil (oral form, second-generation, by Glaxo), and ceftriaxone (third-generation, by Roche).

The enzymatic removal of the 6-acyl side chain of penicillins is widely used in the commercial production of 6-APA. Bacterial acylases have also been found that cleave the 7-acyl side chain of cephalosporin C to form 7-ACA. The discovery and the development of penicillin and cephalosporin acylases greatly accelerated the commercialization of many semisynthetic penicillins and cephalosporins. On the other hand, basic studies on biosynthesis of beta-lactam antibiotics have progressed rapidly with the recent applications of molecular biology techniques to many of the beta-lactam-producing organisms. Most of the biosynthetic reactions involved in the formation of penicillins and cephalosporins have been demonstrated in cell-free systems, and a few of the enzymes have been highly purified and cloned. Knowledge of these enzymes, their promoters, and efficient expression will have a positive influence on the industrial production of penicillins and cephalosporins.

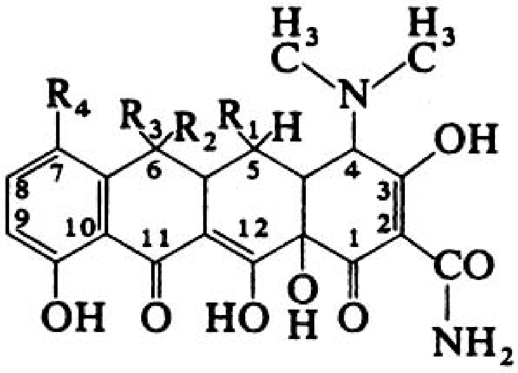
Tetracyclines

In 1948, a broad-spectrum antibiotic, chlortetracycline (Aureomycin), was announced from the Lederle Laboratories, Division of American Cyanamid Company. This antibiotic is produced by *Streptomyces aureofaciens* when grown under submerged aerobic conditions on media composed of sugar, cornsteep liquor, and mineral salts. The crystalline compound has a golden yellow color, which suggested the trade name.

The following year a second, related antibiotic, oxytetracycline (Terramycin), a product of *Streptomyces rimousus*, was announced by Pfizer Inc. It is also a yellow substance, chemically and biologically similar to chlortetracycline. Independent research by both companies eventually led to the disclosure of the structure of these two important chemotherapeutic agents; this has been regarded as one of the brilliant achievements of modern organic chemistry.

Both compounds may be regarded as derivatives of a nucleus known as tetracycline. Their structures along with those of other clinically important tetracyclines are shown in Table 24.10.

TABLE 24.10 Structure of Clinically Important Tetracyclines



	R_1	R_2	R_3	R_4
Tetracycline	H	OH	CH_3	H
7-Chlortetracycline (Aureomycin®)	H	OH	CH_3	Cl
5-Oxytetracycline (Terramycin®)	OH	OH	CH_3	H
6-Demethyl-7-chlortetracycline (Declomycin®)	H	OH	H	Cl
6-Deoxy-5-oxytetracycline (Vibramycin®)	OH	H	CH_3	H
6-Methylene-6-deoxyl-6-demethyl-5-oxytetracycline (Rondomycin®)	OH	H	CH_2	H

Tetracycline can also be produced by *Streptomyces aureofaciens* fermentations under special conditions, that is, chloride starvation or special strains of the organism that fail to halogenate efficiently. Tetracycline possesses many chemotherapeutic properties of chlortetracycline and oxytetracycline. It is an important broad-spectrum antibiotic.

Mutations of tetracycline-producing organisms have led to other tetracycline analogues, of which 6-demethyl-7-chlortetracycline (Declomycin) has clinical use. Chemical modifications of oxytetracycline have generated two other useful members of the family, known as Vibramycin and Rondomycin.

Tetracyclines are active in vivo against numerous gram-positive and gram-negative organisms, and some of the pathogenic rickettsiae and large viruses. The systemic administration of tetracyclines may be carried out by utilizing either oral or intravenous dosage forms. For veterinary use, tetracyclines are given by intravenous injection in bovines and equines, except that oral routes are used for young nonruminating calves.

In connection with work on animal protein factors, it was found that chlortetracycline fermentation mash containing some vitamin B12 gave growth responses well above those obtained with supraoptimal levels of vitamin B12 alone. At the present time, both chlortetracycline and oxytetracycline are used extensively for growth stimulation and improvement of feed efficiency in poultry and hogs, and for the reduction of losses from certain disease conditions. Purified antibiotics as well as dried fermentation residues with the mycelium of *Streptomyces aureofaciens* are used for these purposes. Tetracyclines, used as animal feed supplements, are selling for about $60/kg.

Other Antibiotics

Table 24.11 lists some antibiotics produced on a commercial scale. The importance of new antibiotics for use in clinical medicine is stressed in most discussions of antibiotics, but their importance in other areas should not be forgotten. Antibiotics such as streptomycin and tetracycline are being used against bacterial plant pathogens, and cycloheximide, blasticidin S, nystatin, and griseofulvin are being used against fungi. Antibiotics are also being used in livestock production, where they improve marketable weight and increase food utilization. Those that stimulate animal growth are bacitracin for poultry and swine; bambermycin for swine and calves; virginamycin for poultry and swine; avoparcin for poultry and swine; chlortetracycline for poultry, swine, calves, cattle, and sheep; erythromycin for chickens; nystin, oleandomycin, and procaine penicillin for poultry and swine; streptomycin for poultry; oxytetracycline for poultry, swine, cattle, and calves; tylosin for poultry and swine; lasolocid for poultry; and monensin for poultry and cattle. Combinations of antibiotics and their mixtures with sulfa drugs are also being used.

TABLE 24.11 Important Antibiotics and Bioactive Natural Products

Compound or Name	Classa
Aclacinomycin A	(NP) Anthracycline antitumor drug
Amikacin sulfate	(SS) Semisynthetic aminoglycoside antibiotic
Amoxicillin	(SS) Semisynthetic analog of ampicillin
Amoxicillin/clavulanate	(SS/NP) β-Lactam and β-lactamase inhibitor
Amphotericin B	(NP) Polyene
Amphotericin, liposomal	(NP) Liposome-entrapped polyene
Ampicillin	(SS) Semisynthetic β-lactam
Ampicillin/sulbactam	(SS/SS) β-Lactam and β-lactamase inhibitor
Arbekacin	(SS) Semisynthetic aminoglycoside
Avermectin	(NP) Macrolide
Azithromycin	(SS) Semisynthetic erythromycin
Bacampicillin·HCl	(SS) Semisynthetic β-lactam of penicillin class
Bacitracin	(NP) Peptide antibiotic complex
Bambermycins	(NP) Substituted aminoglycosides
Bialaphos	(NP) Phosphinoglutamic acid-ala-ala
Bleomycin sulfate	(NP) Glycopeptide antitumor antibiotic
Bradykinin antagonist	(NP) Antiviral protein
Carbenicillin indanyl	(SS) Semisynthetic β-lactam of penicillin class
Carminomycin	(NP) Anthracycline antitumor drug
Cefaclor	(SS) Second-generation cephalosporin
Cefadroxil	(SS) First-generation cephalosoprin
Cefamandole naftate	(SS) Third-generation cephalosporin
Cefazolin sodium	(SS) First-generation cephalosporin
Cefepime	(SS) New fourth-generation cephalosporin
Cefixime	(SS) Third-generation cephalosporin
Cefmetazole sodium	(SS) Second-generation cephalosporin
Cefonicid sodium	(SS) Second-generation cephalosporin
Cefoperazone sodium	(SS) Third-generation cephalosporin
Cefotaxime	(SS) Third-generation cephalosporin
Cefotetan, disodium	(SS) Second-generation cephalosporin
Cefoxitin sodium	(SS) Cepha antibiotic
Cefpodoxime Proxetil	(SS) Third-generation cephalosporin
Cefprozil	(SS) Third-generation cephalosporin
Ceftazidime	(SS) Third-generation cephalosporin
Ceftibuten	(SS) Third-generation cephalosporin
Ceftizoxime sodium	(SS) Third-generation cephalosporin
Ceftriaxone	(SS) Third-generation cephalosporin
Cefuroxime	(SS) Second-generation cephalosporin

Cephalexin	(SS) First-generation cephalosporin
Cephalothin sodium	(SS) First-generation cephalosporin
Cephapirin sodium	(SS) First-generation cephalosporin
Cephradine	(SS) First-generation cephalosporin
Chloramphenicol	(NP/C) *N*-Dichloroacyl phenylpropanoid
Chlortetracycline	(NP) Tetracycline.
Clavulanic acid	(NP) Clavam β-lactam inhibitor
Clindamycin	(SS) Chlorinated lincomycin analog
Cloxacillin sodium	(SS) β-Lactam of isoxazolyl penicillin series
Colistimethate sodium	(NP) Polymyxin E (cyclic peptide antibiotic)
Colistin sulfate	(NP) Polymyxin E (cyclic peptide antibiotic)
Cycloserine	(NP) Amino acid analog
Cyclosporin	(NP) Substituted cyclic peptide
Dactinomycin	(NP) Actinomycin D (acylpeptidolactone)
Daunorubicin·HCl	(NP) Anthracycline antitumor drug
Daunorubicin, liposomal	(NP) Liposome-entrapped anthracycline
Demeclocycline·HCl	(NP) Chlorinated, demethylated tetracycline analog
Dicloxacillin sodium	(SS) Dichlorinated β-lactam
Dirithromycin	(SS) Macrolide
Docetaxel	(SS) Semisynthetic taxol
Doxorubicin·HCl	(NP) Anthracycline antitumor drug
Doxorubicin, liposomal	(NP) Liposome-entrapped anthracycline
Doxycycline	(SS) Tetracycline analog
Efrotomycin	(NP) Natural product antibiotic
Epirubicin	(SS) Anthracycline antitumor drug
Erythromycin A	(NP) Macrolide antibiotic
Etoposide	(SS) Derivative of podophyllotoxin
Flomoxef	(SS) Fluorinated oxacephalosporin
Floxacillin	(SS) Fluorinated derivative of oxacillin
Flurithromycin	(SS) Macrolide
Fosfomycin	(NP) Phosphonic acid
Fusidic acid	(NP) Steroidal natural product
Gentamycin sulfate	(NP) Aminoglycoside antibiotic
Gramicidin	(NP) Peptide antibiotic
Hygromycin B	(NP) Substitute aminoglycoside
Imipenem	(SS) Carbapenem
Josamycin	(NP) Macrolide
Kanamycin sulfate	(NP) Aminocyclitol antibiotic
Laidlomycin propionate K	(NP) Polyether ionophore
Lamivudine (3TC)	(SS) Nucleoside analog antiviral drug
Lasolocid sodium	(NP) Polyether ionophore
Latamoxef	(SS) Oxycepham
Lincomycin·HCl	(NP) Lincomycin
Lovastatin	(NP) Hydroxymethylglutaryl-CoA reductase inhibitor
Maduramycin ammonium	(NP) Polyether ionophore
Methicillin	(SS) β-lactam of penicillin class\
Mezlocillin sodium	(SS) β-Lactam of penicillin class
Midecamycin	(NP) Marcolide
Milbemectin	(NP) Macrolide
Milbemycins	(NP) Macrolides
Minocycline	(SS) Tetracycline derivative
Miocamycin	(SS) Acetylated macrolide
Mitomycin C	(NP) Antitumor drug
Monensin sodium	(NP) Polyether ionophore
Mupirocin	(NP) Pseudomonic acid A
Nafcillin	(SS) β-lactam of penicillin class
Narasin	(NP) Polyether ionophore
Natamycin	(NP) Tetraene polyene
Neomycin sulfate	(NP) Aminoglycoside antibiotic
Netilmicin	(SS) Aminoglycoside antibiotic
Nisin	(NP) Lantibiotic
Novobiocin sodium	(NP) Coumarin
Nystatin	(NP) Polyene antifungal antibiotic
Oleandomycin	(NP) Macrolide
Oxacillin sodium	(SS) Isoxazolyl penicillin series
Oxytetracycline	(NP) Tetracycline
Paclitaxel	(NP) Taxol
Penicillamine	(SS) 3-Mercapto-D-valine
Penicillin G	(NP) Penicillin
Penicillin V	(SS) Phenoxymethyl analog of penicillin
Pentostatin	(NP) Nucleoside analog antineoplastic agent
Phosphinothricin	(NP) Phosphinoglutamic acid
Piperacillin sodium	(SS) β-Lactam related to penicillin
Piperacillin/tazobactam	(SS) β-Lactam and β-lactamase inhibitor
Plicamycin	(NP) Natural product
Polymyxin B sulfate	(NP) Cyclic peptide antibiotic
Polyoxin A	(NP) Nucleoside antibiotic
Pravastatin	(NP) Hydroxymethylglutanyl-CoA reductase inhibitor
Pristinamycin	(NP) Antibiotic mixture
Quinine sulfate	(NP) Cinchona alkaloid
Quinupristin/dalfopristin	(SS) Streptogramin mixture
Rifabutin	(SS) Derivative of rifamycin S
Rifampin	(SS) Derivative of rifamycin B
Rokitamycin	(SS) Esterified macrolide
Roxithromycin	(SS) Macrolide
Salinomycin	(NP) Polyether ionophore
Semduramycin	(SS) Polyether ionophore
Simvastatin	(SS) Hydroxymethylglutaryl-CoA reductase inhibitor
Spectinomycin·HCl	(NP) Aminocyclitol
Spiramycin	(NP) Macrolide
Streptomycin sulfate	(NP) Aminoglycoside antibiotic
Streptozocin	(NP/C) Glucose analog
Sulbactam	(SS) Penicillanic acid sulfone
Sultamicillin	(SS) Double ester of ampicillin and sulbactam
Tacrolimus (FK506)	(NP) Macrolide
Teichoplanin	(NP) Glycopeptide
Tenipocide (VM-26)	(SS) Derivative of podophyllotoxin
Tetracycline·HCl	(NP) Tetracycline
Ticarcillin disodium	(SS) β-Lactam of penicillin class
Ticarcillin/clavulanate	(SS/NP) β-Lactam and β-lactamase inhibitor
Tobramycin sulfate	(NP) Aminoglycoside
Topotecan	(SS) Semisynthetic analog of camptothecin
Troleandomycin	(SS) Macrolide antibiotic
Tylosin phosphate	(NP) Macrolide
Vancomycin·HCl	(NP) Glycopeptide antibiotic
Vidarabene (Ara-A)	(NP) Purine nucleoside antiviral drug
Vinblastine sulfate	(NP) Vinca alkaloid
Vincristine sulfate	(NP) Vinca alkaloid
Vinorelbine tartrate	(SS) Semisynthetic vinca alkaloid
Virginiamycin	(NP) Mixture of antibiotics

[a] C, chemically synthesized; NP, natural product; SS, semisynthetic compound.

Source: Adapted from "Industrial Antibiotics: Today and the Future" by W. R. Strohl in *Biotechnology of Antibiotics*, edited by W. R. Strohl, Marcel Dekker, 1997.

Further research on new agents to treat both human and animal diseases is certainly in order, particularly to treat those diseases that are not successfully controlled at present.

24.10 Bioactive Natural Products

In the last six decades that followed the first clinical application of penicillin, the study of microorganisms as sources of therapeutically useful compounds has focused on those with antibiotic activity against bacterial and fungal infections. Indeed, antibiotics discovered in this period have been the basis of a major advance in the practice of medicine. In the last twenty years, however it has become evident that the contribution of microorganisms does not have to be limited to the realm of infectious diseases. Metabolites of microorganisms have been found to have many other interesting therapeutic applications. Particularly as our understanding of disease processes at the biochemical and genetic levels becomes more advanced, enzymes or receptors have increasingly served as specific targets for therapeutic intervention. Recombinant DNA technology, in particular, has helped to produce reagents as tools for the development of practical and high throughput screening methodologies based on mechanism of action. Automation and miniaturization have also allowed a rapid increase in the throughput of the screening process. Several of these bioactive natural products have been successfully developed as clinically useful therapeutics. They are listed in Table 24.11 together with the antibiotics.

Cyclosporin A is a fungal decapeptide. It is the principal drug used to suppress immune rejection in organ transplant patients.

Cyclosporin A binds to a cytosolic protein, cyclophilin, inhibits its peptidyl-prolyl isomerase activity, and suppresses T cell activation. Cell-based screens for fermentation products with similar immuno-modulating activity have resulted in compounds with biological activities similar to cyclosporin A. Fujisawa's FK5O6, the most interesting among them, also binds to a cytosolic protein, FKBP (FK5O6 binding protein). It, too, inhibits peptidyl-prolyl isomerase activity and suppresses T cell activation. FK5O6 has been proven as an effective immunosuppressant and has fewer side effects than cyclosporin A.

Mevinolin and compatin were independently discovered by Merck and Sankyo. Both control cholesterol synthesis in human. The Merck group screened directly for inhibition of HMG-CoA reductase, a key enzyme involved in cholesterol synthesis, while the Sankyo group sought agents that would interfere with the incorporation of radiolabeled acetate into cholesterol in a cell-free enzyme system. The search for microbial products that interfere with cholesterol metabolism continues, given the important roles serum cholesterol plays in coronary heart disease. Compatin and mevinolin, being HMG-CoA reductase inhibitors, also prevent maturation (a post-translational modification near the carboxy terminal) of *ras* proteins. Thus they may be of value in the treatment of ras-dependent tumors.

Monobactam, a new class of natural beta-lactam antibiotics, was found from a beta-lactamase screen. Aztreonam, a chemically synthesized monobactam, is in clinical use. Thienamycin was also discovered in an assay based on beta-lactamase susceptibility. Mutastein, a fungal glycoprotein that inhibits the synthesis of insoluble glucan by the bacteria causing the formation of dental plaque, has been used commercially as a prophylactic agent for tooth decay.

The search for enzyme- or receptor-targeted microbial products does not limit itself to medical use. Several commercially important "antibiotics" of agricultural applications have been found in recent years. Nikkomycin, a chitin synthase inhibitor, is an antifungal agent. Avermectin, acting upon the synaptic gamma-aminobutyric acid (GABA) receptor system, is an effective antiparasitic and antihelmintic agent.

In the past two decades, as the human genome has been mapped and as therapeutic screening becomes increasingly mechanism-based, production of target enzymes or receptors using recombinant DNA techniques has drastically expanded the number of targets that can feasibly be screened. The following table lists some of the targets and their applications.

Targets	Applications
HIV reverse transcriptase	AIDS therapy
HIV protease	AIDS therapy
IL-I convertase inhibition	Antiarthritic
IL-I inhibition	Antiarthritic
IL-I receptor antagonist	Antiarthritic
MHC antagonist	Antiarthritic
Collagenase inhibitor	Antiarthritic
Anticonvulsant Monoamine oxidase A inhibitor	Antidepressant
5-HT receptor agonist	Antidepressant
5-HT reuptake inhibitor	Antidepressant
	Antiobesity
Lipase inhibitor	Antiobesity
H2-receptor antagonist	Antiulcerant
Enkephalinase inhibitor	Antiulcerant
Opioid receptor kappa agonist	Analgesics
Alpha glucosidase inhibitor	Antidiabetes
H I-receptor antagonist	Antihistamine
5-Lipoxygenase inhibitor	Antiasthmatics
Phosphodiesterase inhibitor	Antiasthmatics
P AF Antagonist	Antiasthmatics
IL-4 Antagonist	Antiasthmatics
Leukotriene synthase inhibitor	Antiasthmatics
Atrial peptidase inhibitor	Antihypertensive
Renin inhibitor	Antihypertensive
ACE inhibitor	Antihypertensive
Cholesterol acyltransferase	Cholesterol control
HMG-CoA reductse inhibitor	Cholesterol control
LDL receptor antagonist	Cholesterol control
Elastase inhibitor	Cystic fibrosis

While screening against targets at the protein level for inhibitors of enzymes and for molecules (agonists and antagonists) that interact with receptors is expanding, the search for compounds that modulate gene expression, at the DNA level, is rapidly emerging.

24.11 Biopharmaceuticals

The term "biopharmaceuticals" has been generally accepted for distinguishing the large molecule pharmaceuticals (mainly proteins) that have emerged as a result of the modern "biotechnology" from the traditional small molecule drugs. The so-called biotechnology is largely based on two scientific breakthroughs of the late 1970s: recombinant DNA and hybridoma methodologies. Over the past two decades, genetic engineering or recombinant DNA technology has affected all facets of pure and applied biological research. The main thrust of the technology lies with the ability to introduce a manipulated genetic material into a cell in such a way as to permit its replication in and passage to progeny cells. The primary, commercial application of genetic engineering has been to produce a protein of interest in large quantity by inserting its gene into a cell in which the protein can be efficiently expressed. The major steps involved in commercializing such a protein include,

1. gene cloning using homologous probes, by expression screening, or through gene synthesis
2. construction of expression vectors
3. transforming host cells
4. large scale propagation of the transformed cell, and
5. recovery and purification of the expressed product.

The commonly used host cells for expressing heterologous genes include *E. coli*, yeast, and mammalian cells. The most commonly used *E. coli* promoters are those from the *lac* or *trp* operons and the late promoter, *pL*, from phage lambda. These promoters can be "turned on" by the addition of an inducer (derepressor) such as IPTG (isopropyl thiogalactoside) or lactose for *lac*, IAA (indoleacrylic acid) for *trp*, or by temperature shift from 3037°C to 42°C for *pL*. Following process optimization and scale-up, an expression level of 15 g/L is achievable with *E. coli* expression cultures.

For yeast, there are in general two modes of transformation, integrative and autonomous. The integrative vectors are simple plasmids and achieve clonal expression through integration into the host genome by homologous recombination. The autonomous vectors possess a sequence known as ARS (autonomously replicating sequence). It, like an origin of replication, allows plasmids to be passed to progeny cells. The yeast 2-µm circle, a naturally occurring high copy number plasmid, is a commonly used ARS vector. The autonomous transformants are in general less stable than the integrative ones. The inclusion on an ARS plasmid of a yeast centromeric sequence can overcome the instability problem. To achieve efficient expression, several yeast promoters are used, PGK promoter from phosphoglycerate kinase, GAL1 from galactose kinase, PHO5 from acid phosphatase, and HSP90 from a heat shock protein. Certain "supersecreting" hosts have been sought. An optimized host/vector yeast system can produce a heterologous protein over 1 g/L level, while 1100 mg/L is more of the norm.

Mammalian plasmid-based vectors, in addition to a prokaryotic replicon and a selection marker to permit DNA replication in *E. coli*, commonly have a eukaryotic repilcon and a eukaryotic selection marker. The eukaryotic replicon usually comes from viruses such as Simian virus 40 (SV 40), bovine papilloma virus (BPV), or Epstein Barr virus (EBV). The commonly used promoters are the SV 40 early promoter (including its upstream enhancer region) and the metallothionine promoter (including its enhancer and metal regulatory elements). Some plasmid-based vectors also contain a splicing site to promote RNA stability and a poly-adenylation sequence to ensure transcription termination. The most commonly used mammalian hosts are CHO (Chinese hamster ovary) and BHK (baby hamster kidney).

The selection of an expression system for a cloned gene depends on many factors: host/vector stability, expression level, ease of scale-up, post-translational modification (glycosylation, secretion, amidation, folding, proteolysis, phosphorylation, etc.), and product immunogenicity and heterogenicity.

The biopharmaceuticals that are industrially produced by microbial (bacterial or yeast) fermentation include insulin, human growth hormone, hepatitis B surface antigen vaccine, alpha interferon, beta interferon, gamma interferon, granulocyte colony stimulating factor, and interleukin-2. A more complete list of industrial biopharmaceuticals produced by microbial fermentation and mammalian cell culture as well, together with their therapeutic applications and sales, is available in the next chapter.

Insulin via Recombinant DNA Technology

Insulin is the first human biopharmaceutical that was commercialized using the recombinant DNA technology. Its fermentative expression in *E. coli*, downstream recovery, and purification serves as a good example for the large-scale production of recombinant proteins. Insulin, given its rather small size, is expressed as N-terminal extended proinsulin using the *trp* promoter. The fermentation is carried out at about 37°C and near neutral pH. It is a fast, aerobic fermentation. So it is important to prevent oxygen starvation. The *trp* operon is turned on when the fermentation runs out of its natural repressor, tryptophan. The chimeric protein as it is expressed accumulates very rapidly inside the cells as insoluble aggregates (inclusion or reflectile bodies). The formation of inclusion bodies prevents proteolysis and facilitates product recovery. The recombinant *E. coli* fermentation usually runs for 1824 hr. At the end of fermentation, the inclusion bodies account for about 1020 percent of total dry cell mass.

After the cells are killed by heat or acid treatment to prevent release of recombinant organisms into the environment, the treated fermentation broth is concentrated by passing through a continuous disk stack centrifuge. It is then diluted in a buffer solution usually containing Tris-Base and EDTA before being sent through a high pressure homogenizer, two or three times, in which the cells are broken down and the inclusion bodies released. The inclusion bodies are then separated from the cell debris and recovered in another continuous disk stack centrifuge. The inclusion body slurry is then washed with a buffer solution usually containing a detergent like Triton-X-100 and recentrifuged. The entire recovery process is carried out at refrigerated temperature typically around 10°C.

The inclusion bodies are then dissolved in a 5 M urea chaotropic solution containing beta-mercaptoethanol, the latter used to reduce disulfide bonds. Urea and beta-mercaptoethanol are then replaced with deionized water by diafiltration. The solution is then clarified using a dead-end filter. The N-terminal extension separated from the proinsulin sequence by a methionine residue is then removed by treatment with CNBr (cyanogen bromide). The reaction is carried out in a formic acid solution. When the reaction is completed, formic acid, excess CNBr, and cyanide gas, a by-product of the cleavage reaction, are removed in a rotary vacuum evaporator.

The denatured mature proinsulin is then sulfonated in the presence of a denaturing agent such as urea of guanidine. Following diafiltration to reduce the concentration of urea or guanidine and to exchange it with a proper chromatographical buffer, the sulfonated proinsulin is purified using an ion-exchange column. Refolding of the purified proinsulin then takes place in the presence of beta-mercaptoethanol and at a low concentration of proinsulin, typically below 1 g/L. After the refolding reaction is completed in 1224 hr, the proinsulin solution is diafiltered to reduce volume and to exchange into a proper chromatographical buffer before it is further purified in a hydrophobic interaction column. The C-peptide is removed from proinsulin using trypsin and carboxypeptidase. After the enzymatic reaction is completed, the solution is diafiltered to reduce volume and to exchange into a proper chromatographical buffer before an ion-exchange column is employed to purify now properly folded and cleaved insulin. Fermentatively or microbially produced recombinant proteins can also be expressed as soluble proteins inside the cells or secreted outside the cells. Their purification differs from the scheme outlined above for insulin. As essentially all mammalianly produced recombinant proteins are expressed in the soluble form, details for recovery and purification of this type of proteins can be found in the next chapter.

24.12 Bibliography

Aiba, S., Humphrey, A. E., and Milis, N., Biochemical Engineering, 2nd ed., University of Tokyo Press, Tokyo, 1973.
Aida, K., Ichiba, I., Nakayama, K., Takinami, K., and Yamada, H. (Eds.), Biotechnology of Amino Acid Production, Elsevier, Amsterdam, 1986.
Anderson, S., Marks, C. B., Lazarus, R., Miller, J., Stafford, K., Seymour, J., Light, D., Rastetter, W., and Estell, D., Science, 230, 144 (1985).
Aunstrup, K., "Enzymes of Industrial InterestTraditional Products," Ann. Report of Fermentation Processes, 6, 175 (1983).
Armiger, W. B. (Ed.), "Computer Applications in Fermentation Technology," Biotechnology and Bioengineering Symposium Series No. 9, John Wiley and Sons, New York, 1979.
Atkinson, B., and Mavituna, F., Biochemical Engineering and Biotechnology Handbook, The Nature Press, New York, 1983.
Bailey, J. E., and Ollis, D. F., Biochemical Engineering Fundamentals, McGraw-Hill Book Company, New York, 1977.
Baumberg, S., Hunter, I., and Rhodes, M. (Eds.), Microbial Products: New Approaches, Cambridge University Press, Cambridge, England, 1989.
Baker, S. A., "Industrial Enzyme Process," Sci. Prog. Oxf., 65, 477 (1980).
Boze, H., Moulin, G., and Galzy, P., "Production of Microbial Biomass," in Biotechnology, Vol. 9, G. Reed and T. W. Nagodawithana (Eds.), VCH Publishers Inc., New York, 1995.
Bull, A. T., Ellwood, D. C., and Ratledge, C. (Eds.), Microbial Technology: Current State, Future Prospects, Cambridge University Press, Cambridge, England, 1979.
Chater, K., and Bibb, M., "Regulation of Bacterial Antibiotic Production," in Biotechnology, Vol. 8, Kleinkaufand, H., and von Dohren, H. (Eds.), VCH Publishers Inc., New York, 1997.
Cheremisinoff, N. P., Gasohol for Energy Production, Ann Arbor Science Publishers Inc., Ann Arbor, Michigan, 1979.
deBoel', L., and Dijkhuizen, L., "Microbial and Enzymatic Processes for L-Phenylalanine Production," Adv. Biochem. Engineering/Biotechnol., 41, (1990).
Demain, A. L., "Riboflavin Oversynthesis," Ann. Rev. Microbiol., 26, 369, 1972.
DiMarco, A., and Pennella, P., "The Fermentation of the Tetracyclines," Prog. Ind. Microbiol., 1, 45 (1959).
Ebner, H., Follmann, H., and Seller, S., "Vinegar," in Biotechnology, Vol. 9, G. Reed and T. W. Nagodawithana (Eds.), VCH Publishers Inc., New York, 1995.
Esaka, N., Soda, K., Kumagai, H., and Yamada, H., "Recent Advances in Enzymatic Synthesis of Amino Acids," Biotech. Bioeng., 22, Suppl. 1, 127 (1980).
Foster, K. A., Frackman, S., and Jolly, J. F. "Production of Enzymes as Fine Chemicals," in Biotechnology, Vol. 9, G. Reed and T. W. Nagodawithana (Eds.), VCH Publishers Inc., New York, 1995.
Friendmann, H. C., and Cagen, L. M., "Microbial Biosynthesis of B12-like Compounds," Ann. Rev. Microbiol., 24, 159 (1970).
Godfrey, T., and Reichelt, J., Industrial Enzymology, The Nature Press, New York, 1983.
Hatch, R., "Computer Applications for Analysis and Control of Fermentation," Ann. Report Fermentation Processes, 5, 291 (1982).
Hirose, Y., and Shibai, M., "Amino Acid Fermentation," Biotech. Bioeng., 22, Suppl. 1, 111 (1980).
Ho, N., "Yeast Alcohol Tolerance and Recombinant DNA for Improved Alcohol Production," Ann. Report Fermentation Processes, 4, 235 (1980).
Hopps, H. E., and Petricciani, J. C. (Eds.), Abnormal Cells, New Products and Risk, Tissue Culture Association, Gaithersburg, MD, 1985.
Humphrey, A. E., "Biochemical Reaction Engineering," Chapter 8 in Chemical Reaction Engineering Reviews, ACS Symposium Series 72, 1979.
International Technical Information Institute, "Japan's Most Advanced Industrial Fermentation: Technology and Industry," International Technical Information Institute, Tokyo, 1977.
Kalisz, H. M., "Microbial Proteinases," Adv. Biochem. Engineering/Biotechnol., 36, 1 (1988).
Kleinkauf, H., and von Dohren, H., "Peptide Antibiotics," in Biotechnology, Vol. 8, H. Kleinkauf and H. von Dohren (Eds.), VCH Publishers Inc., New York, 1997.
Kosaric, N., "EthanolPotential Source of Energy and Chemical Products," in Biotechnology, Vol. 6, M. Roehr (Ed.), VCH Publishers Inc., New York, 1996.
Leuchtenberger, W., "Amino AcidsTechnical Production and Use," in Biotechnology, Vol. 6, M. Roehr (Ed.), VCH Publishers Inc., New York, 1996.
Lowenhem F. A., and Moran, M. K. (Eds.), Faith, Keyes and Clark's Industrial Chemicals, 4th ed., John Wiley and Sons, New York, 1975.
Maddox, I. S., "Microbial Production of 2,3-Butanediol," in Biotechnology, Vol. 6, M. Roehr (Ed.), VCH Publishers Inc., New York, 1996.
Magee, R. J. and Kosaric, N., "The Microbial Production of 2,3-Butanediol," Adv. Appl. Microbiol., 32, 89 (1987).
Martin, J. F. and Demain, A. L., "Control of Antibiotic Biosynthesis," Microbial Rev., 44, No. 2, 230 (1980).
Martin, J. F., and Liras, P., "Enzymes Involved in Penicillin, Cephalosporin and Cephamycin Biosynthesis," Adv. Biochem. Engineering/Biotechnol., 39, 153 (1989).
Miller, D. N., A.I.Ch.E.J., 20, 445 (1975).
Miller, L., "Baculoviruses as Gene Expression Vectors," Ann. Rev. Microbiol., 42, 177 (1988).
Morris, C. E. "America's Gold," Food Engineering, 52, No. 5, 95 (1980).
Nakayama, K., "The Production of Amino Acids," Process Biochemistry, 11, No. 2, 4 (1976).
Nuesch, J., Heim, J., and Treichler, H. J., "The Biosynthesis of Sulfur-Containing Beta-Lactam Antibiotics," Ann. Rev. Microbiol., 41, 25 (1987).
Peppler, H. J., and Perlman, D. (Eds.), Microbial Technology, Vol. 1 and 2, 2nd ed., Academic Press, New York, 1979.
Parker, W. L., O'Sullivan, J., and Sykes, R., "Naturally Occurring Monobactams," Adv. Appl. Microbiol., 31, 181 (1986).
Perlman, D., "Influence of Penicillin Fermentation Technology to Processes for Production of Other Antibiotics," Process Biochemistry, 10, No. 9, 23 (1975).
Perlman, D., and Tsao, G. T. (Eds.), Annual Reports on Fermentation Processes, Vol. 13, Academic Press, New York, 19771979.
Raju, G. K., and Cooney, C. L., "Media and Air Sterilization," in Biotechnology, Vol. 3, G. Stephanopoulos (Ed.), VCH Publishers Inc., New York, 1993.
Reed, G., and Peppler, H. J., Yeast Technology, AVI Publishing Company, Westport, Conn., 1973.
Roehr, M., Kubicek, C. P., and Kominek, J., "CitRic Acid," in Biotechnology, Vol. 6, M. Roehr (Ed.), VCH Publishers Inc., New York, 1996.
Rushton, J. H. et al., Chem. Engineering Progress, 46, 467 (1959).
Schierholt, J., "Fermentation Processes for the Production of Citric Acid," Process Biochem., 12, No. 9, 20 (1977).
Scrimshaw, N. S., and Murray, E. B., "Nutritional Value and Safety of Single Cell Protein," in Biotechnology, Vol. 9, G. Reed and T. W. Nagodawithana (Eds.), VCH Publishers Inc., New York, 1995.
Strohl, W. R., "Industrial Antibiotics: Today and the Future" in Biotechnology of Antibiotics, W. R. Strohl (Ed.), Marcel Dekker, New York, 1997.
Tabor, J. (Ed.), Genetic Engineering Technology in Industrial Pharmacy, Marcel Dekker, New York, 1989.
Tanaka, H., and Omura, S., "Screening of Novel Receptor-Active Compounds of Microbial Origin," in Biotechnology, Vol. 8, H. Kleinkauf and H. von Dohren (Eds.), VCH Publishers Inc., New York, 1997.
Tannenbaum, S., and Wang, D. I. C. (Ed.), Single Cell Protein Il, M.I.T. Press, Cambridge, MA, 1975.
Tsao, G. T. (Eds.), Annual Reports on Fermentation Processes, Vol. 48, Academic Press, New York, 19801985.
Weinstein, M. J., and Wagman, G. H. (Eds.), "Antibiotics: Isolation, Separation and Purification," Chromatography, 15 (1978).
U.S. International Trade Commission, Synthetic Organic Chemicals: U.S. Production and Sales, U.S. Government Printing Office, Washington, D.C., 1989.
Woodruff, H. B., and Burg, R. W., "The Antibiotic Explosion," in Pharmacological Methods, Receptors and Chemotherapy, Elsevier, Amsterdam, 1986.

25.1 Introduction

Tissue culture has been practiced since the beginning of the last century. It passed through its general exploratory and expansion phases in the 1950s and now is in its specialization phase concerned with control mechanisms and differentiated functions. An increasing number of cell types have been grown in culture, including connective tissue elements (fibroblasts), skeletal tissue (bone and cartilage), cardiac, smooth muscle, and epithelial tissues (liver, lung, breast, bladder, and kidney), neural cells, endocrine cells, melanocytes, and many other tumor cell types. Cell culture technique has been used widely in many different disciplines from the basic sciences of cell, developmental, and molecular biology to the applied fields of biotechnology. In the past decade in particular, the biotech industry has successfully brought this technology forward towards many important medical applications. In this chapter we will focus on industrial cell culture used for the commercial production of biological products, mainly therapeutic proteins or biopharmaceuticals.

Fermentation, meaning specifically "microbial fermentation," has been used to produce many important therapeutic proteins. And its principles, processes, and products are covered in the previous chapter. Cell culture, meaning specifically "mammalian cell culture" in this context, offers important advantages for the expression of heterologous proteins. Mammalian proteins produced in cell culture are more native and functional because of cell culture's mammalianlike transcription, translation, and post-translational modification processes. When it comes to engineering principles, they are basically the same for microbial fermentation and mammalian cell culture. However, the physiological characteristics are quite different. In this chapter, we will therefore focus more on the physiological aspects of cell culture than the engineering ones.

25.2 Cell Culture Fundamentals

Introduction

A good manufacturing process using cell culture to produce therapeutic proteins requires:

1. the generation and selection of a suitable cell type/line which produces the product of interest in a stable and consistent manner
2. the development of a simple medium and process in which the producing cells can be cultured with desirable growth kinetics and high productivity
3. the consistent operation of both cell culture and recovery/purification equipment to achieve the required yield, purity, and quality (physical and functional) of the product

Cell Lines/Expression System

Only relatively few mammalian cell types, Chinese Hamster Ovary (CHO), baby hamster kidney (BHK), human lymphoblastoid (Namalva), myoloma, and hybridoma, have been used for the commercial production of biopharmaceuticals. Those that have been approved by the Food and Drug Administration (FDA) are given in Table 25.1. The specific cell lines include CHO-K1, DG44, DUKX-B11, BHK21. NS0, and SP2/0. Among all cell lines, CHO and NS0 are most often used for the production of recombinant proteins and monoclonal antibodies, respectively. Although insect cell culture has gained popularity for protein expression in recent years because it has some advantages over mammalian cell culture,1,46 there are as yet no biologicals derived from insect cells that have been commercialized today. For insect cells, Spodoptera frugiperda ovarian (Sf-9, Sf-21, and TN368), Drosophila schneider S2 (S2) and High Five (BTI-TN-5B1-4, derived from Trichoplusia ni egg homogenates) cell lines are most commonly used.1

TABLE 25.1 Biopharmaceuticals Approved by FDA

					Application		
Product	Trade Name	Companies	Appr. Date	Sales $M	Process	(Prevention or Treatment of)	Type
Anti-CD3 antibody	Orthoclone (OKT3)	Ortho/Johnson J	1986	na		Graft rejection	mAb
Dornase alpha	Pulmozyme	Genentech	2/24/98	122	CHO/stir.tk suspension	Cystic fibrosis	R. protein/enzyme
Abciximab	ReoPro	Centocor/Lilly	12/22/94	418		Acute cardiac complication	mAb (CD3)
Aldesleukin (IL2)	Proleukin	Chiron/Roche	5/5/92	na	*E. coli*	Metastatic renal cell carcinoma	Recomb. protein
Alemtuzumab	Campath	Millenn./scheringAG	5/7/2001	na	CHO	B-cell chronic lymphocytic	mAb (CD52)
Alteplase (t-PA)	Activase	Genentech	1987;6/18/96	206	CHO	Myocardual infarction/cute ischemic stroke	R. glycoprotein
(2nd generation t-PA)	TNKase	Genentech	6/2/00	incl. activas	CHO	Myocardual infarction/cute ischemic stroke	R. glycoprotein
(2nd generation t-PA)	Retavase	BoehringM/Centocor.	10/30/96	na	CHO	Myocardial infarction	Nonglycoprotein
Anihemophilic factor (recomb)	ReFacto	GI/Pharmacia	3/6/00	90	CHO/cont. perfusion	Hemorrhagic episodes, hemophiliaA	R. glycoprotein
Basiliximab	Simulect	Novartis	5/12/98		M myeloma, recomb./stir.tk	Acute organ rejection	mAb (IL2 recet)
Becaplermin	Regranex	OMJ	12/16/97		CHO/stir.tk, susp., fed-batch	Lower extremity diabetic neuropathic ulcers	R. glycoprotein
Coagulation factor IX (recomb)	BeneFix	GI	2/11/97	273		Hemorrhagic episodes, hemophilia B	R. glycoprotein
Coagulation factor VII (recomb)	Novoseven	Novo	3/25/99	na	BHK	Hemorrhagic episodes, hemophilia A or B	R. glycoprotein
Coagulation factor VIII (recomb)	Recombinate	GI/AHP				Hemophilia A	R. glycoprotein
Daclizumab	Zenapax	Roche	12/10/97	36	SP2/O	Acute organ rejection	mAb
Denileukin diftitox	Ontak	Seragen	2/5/99	na		T-cell lymphoma	
Epoetin alpha	Epogen\Procrit\NeoRecorm	Amgen\Johnson J \Roche	1988;7/26/99	3605	CHO/roller bottle	Anemia	R. glycoprotein
				380	CHO	Anemia	R. glycoprotein
Etanercept	Enbrel	Immunex	11/2/98	690	CHO/stir.tk, suspension	Rheumatoid arthritis	IgGl/TNFrecep.
Filgrastin(G-CSF)	Neupogen	Amgen	2/20/91	1223	*E. coli*	Infection as manifested by febrile neutropenia	Recomb. protein
	Neupogen	GI/Roche	Jun-94	169	*E. coli*	In patients receiving anticancer drugs; Acute myelid leukemia	Recomb. protein
Hepatitis B vaccine (recomb)	Recombivax\Engerix-B	Merck\Smithkline	7/7/981	~800	*S. cerevisiae*, CHO	Hepatitis B	Vaccine
		Chiron	1987			Hepatitis C	Vaccine
Human growth hormone	Humatrope	Lilly	1985	350	*E. coli*	Abnormal growth, Turner's syndrome	Recomb. protein
	Norditropin	Novo		253	*E. coli*, CHO	Abnormal growth. Turner's syndrome	Recomb. protein
	Protropin/Nutropin	Genentech		272	*E. coli*	Abnormal growth, Turner's syndrome	Recomb. protein
Human insulin	Humulin	Lilly	10/20/83	1,114	*E. coli*, *S. cerevisiae*	Diabetic	Recomb. Protein
	Novolin	Novo		1,428	*E. coli*, *S. cerevisiae*	Diabetic	Recomb. Protein
Human insulin lispro	Huma;og	Lilly	2/1/96	225	*E. coli*	Diabetic	Recomb. Protein
Infliximab	Remicade	Centocor	8/24/98	370	NSO/stir.tk susp., perfusion	Rheumatoid arthritis	mAb
Interferon alfacon 1	Infergen	Amgen	10/6/97	14.5	*E. coli*	Hepatitis C	Recomb. Protein
Interferon alpha 2a	Roferon-A	Roche	11/1/96	169	*E. coli*	Hepatitis C	Recomb. Protein
Interferon alpha 2b	Intron-A	Schering Plough	11/6/97	1,360	*E. coli*	Conjunction/chemotherapy for lymphoma	Recomb. Protein
Interferon alph-n1	Wellferon	Chiron/Wellcome	3/25/99	na	human lymphoblastoid cell	Chronic hepatitis C	End.glycoprotein
Interferon alph-n1/beta 1b	Betaferon/Betaseron	Chiron/schering AG	7/93,11/30/95	530		Multiple sclerosis	R. glycoprotein
Interferon beta 1a	Avonex	Biogen	5/17/96	433?	CHO/stir.tk suspension	Multiple sclerosis	R. glycoprotein
Interferon gamma 1b	Actimune	Intermune	2/10/00			Malignant osteropetrosis	R. glycoprotein
Lyme disease vaccine (recomb)	LYMErix	Smithkline	12/21/98	na		Vaccine for lyme disease	Vaccine
Opervekin (IL11)	Neumega	GI	11/25/97			Thrombocytopenia	
Palivizumab	Synagis	Medimmune	6/19/98		NSO/stir.tk susp., fed-batch	Lower respiratory tract disease	mAb
Pegaspaaaargase	Oncaspar	Enzon	2/1/94			Conjunction/chemotherapy for leukemia	
PEGinterferon alfa 2b	PEG-Intron	Schering P/Enzon	5/25/00,1/19/01	na		Hepatitis C	
Rituximab	MabTheraRituxan	Genentech/IDEC	11/26/97	848	CHO/stir.tk, suspension	Lymphoma	mAb (CD20)
Sargramostin	Leukine, Prokine	Immunex/AHP	11/7/96	88	yeast	Acute myelogenous leukaemia	R. glycoprotein
Trastuzumab	Herceptin	Genentech	9/25/98	252	CHO/stir.tk suspension	Metastatic breast cancer	mAb

Source: (1) FDA Web site; (2) Pharmaceutical Companies Analysis, Espicom Ltd (2001, September).

In addition to choosing proper cell lines as hosts, also required for high levels of protein expression is the construction of expression vectors with selectable markers, strong promoters, and favorable RNA binding, and processing sequences. The commonly used expression systems in the biotechnology industry include:

1. *Dihydrofolate reductase (DHFR) expression system*.2,3 Provision of the DHFR gene in an expression vector enables the transformed cells to overcome the toxicity of MTX (methotrexate, a folate analog and an inhibitor of DHFR) and allows the transformed cells to survive. A possible DHFR expression system would employ a dhfr-cell line as host, a weak promoter like SV40 to regulate DHFR expression, and a strong promoter like CMV to drive the expression of the protein of interest. MTX is used as the selection and amplification pressure. DHFR is among of the most commonly used expression systems in mammalian cells.
2. *Glutamine synthetase (GS) expression system*.3 A variation of the DHFR system in which the DHFR gene is replaced by the GS gene can also result in high level of heterologous protein expression. With the GS system, amplification is often not necessary. Amplification using high levels of selecting agent such as MTX can be detrimental to the general cell growth characteristics. GS synthesizes glutamine, an essential amino acid, and allows cells to grow in a glutamine-free medium. MSX (methionine sulphoxamine, a GS inhibitor) can be used for selection and amplification if needed. GS is another commonly used expression system in mammalian cells.
3. *Ecdysone-indacible expression system*.4 This system consists of two vectors. The first is pVgRXR, which constitutively expresses the heterodimeric ecdysone receptor of Drosophilia. The receptor when bound with Ecdysone or Muristerone A (an analog of Ecdysone) activates the second vector pIND. pIND drives the expression of the gene of interest. This allows for a rather stringent regulation of gene expression in mammalian cells using Ecdysone or Muristerone A. Special cell lines including EcR-293, EcR-CHO, and EcR-3T3 have been developed to take advantage of this expression system.
4. *Baculovirus expression vector system (BEVS)*.1,4 The BEVS uses the baculovirus, Autographa californica nuclear polyhedrosis virus (AcNPV), as a vehicle to express heterologous proteins in insect cell lines such as Sf9, Sf21, and High Five. The BEVS takes advantage of the strong viral polyhedrin promoter and replaces the viral polyhedrin gene with the gene of interest. The system offers efficient infection and high yield of expressed protein.

Cell Line Preservation (Cell Bank)

Cell banking5 is the first important step of a cell culture based manufacturing process. The purpose of cell banking is to ensure a consistent source of seed stocks for the anticipated commercial life of an approved product. Production cell lines are usually banked using a two-tiered system mat consists of a master cell bank (MCB) and a working cell bank (WCB). Each production batch starts with one or more vials from the WCB. When a WCB is depleted, a new WCB can be created from the MCB. Cell banks are usually kept in -80 degree freezers or in liquid nitrogen. The cell banks need to be tested to ensure they are free of microbial, viral, and mycoplasma contaminants.

Cell Line Stability

To ensure consistency in the production process and in product quality requires not only a continuous supply of seed stocks from qualified cell banks, but also demonstrated stability of the producing cells. The production cell line needs to show stability in product yield, product quantity, product formation rate, and other parameters such as cell concentration and growth rate. The stability needs to be tested and documented over a period that is a reasonable number of generations above and beyond the number of generations required for the cell line to be propagated from the initiation of production process (i.e., WCB) to its normal end. Such stability studies are an important part of a manufacture's regulatory submission for a product license. For details, please refer to the FDA guideline, CBER (1993)6 "Points to Consider in Characterization of Cell Lines Used to Produce Biologicals" and ICH (1996)7 "Quality of Biotechnological Products: Analysis of the Expression Construct in Cell Lines Used for Production of r-DNA Derived Protein Products."

Medium

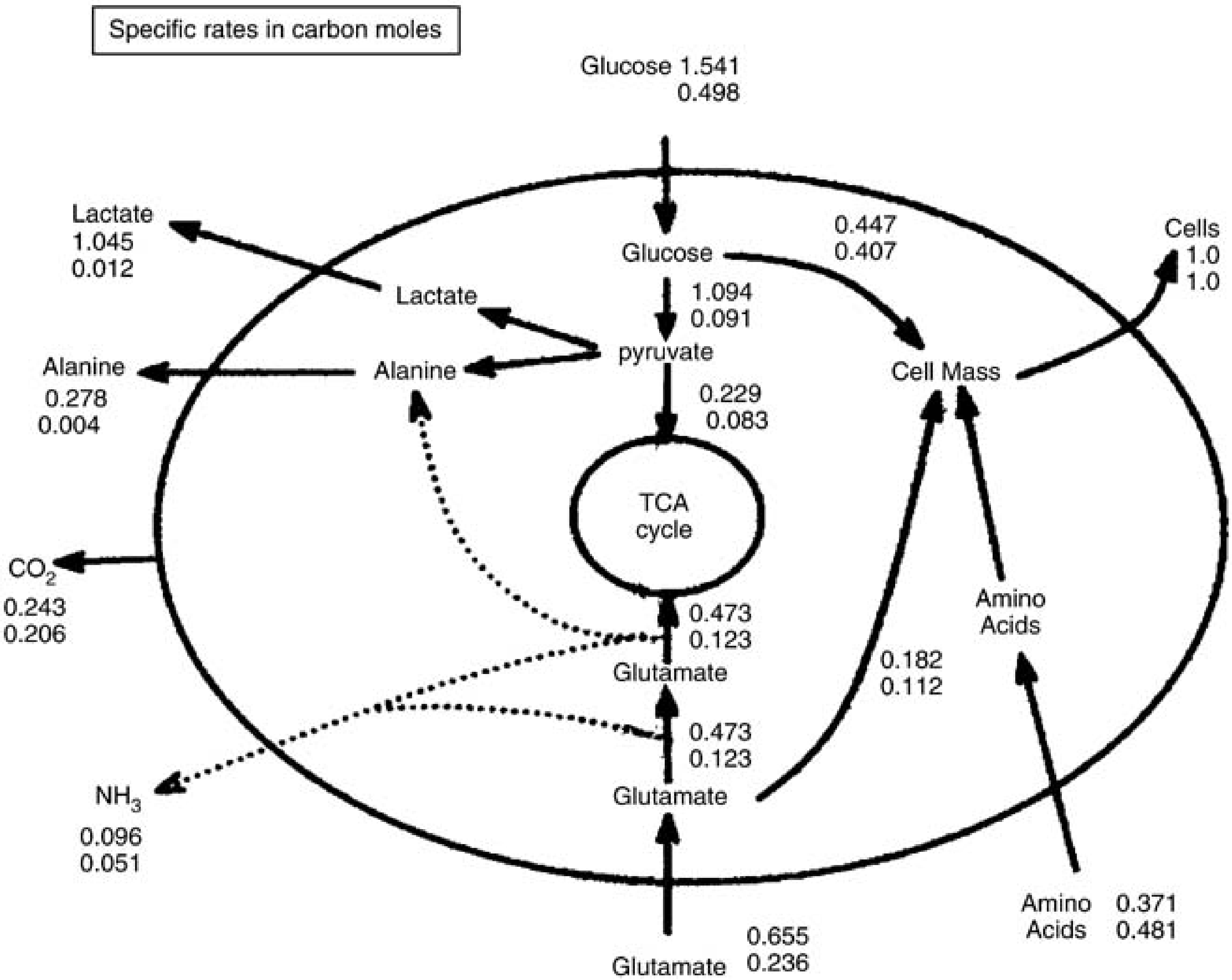
Serum is an extremely complex mixture of small and large molecules. It has been commonly used in culture media as a source of growth and attachment factors, hormones, binding proteins, lipids, minerals, and so on.8 However, since it is of animal origin, it raises concern of containing adventitious agents, particularly viral contaminants. And, because of its complexity, it also causes concern regarding batch-to-batch variation in composition. Its high cost and high protein content (which makes product recovery more difficult) have also fueled the industry-wide interest in the development of serum-free (SF), protein-free (PF), and chemically defined (CD) media for the manufacture of biopharmaceuticals in recent years.9 However, SF media tend to be highly specific to one cell type and sometimes even to one particular cell line. So it is not uncommon that a different optimal medium is required for a particular cell line. To develop SF or animal-component-free media requires considerable experience and expertise and it can be a time-consuming process. One possible approach is to start with a commercially available SF medium for a particular cell type. Several SF, PR, and CD media are available commercially, for examples, Gibco's Hybridoma-SFM, Sf-900 II SFM, Express Five SFM, CHO-S-SFM II; Hyclone's HyQ SFX-CHO, HyQ PF-CHO, HyQ SFX-Mab, HyQ PFX-Mab, HyQ SFX-insect. One drawback of using commercial media is that one is simply replacing one black box, serum, with another. To gain full knowledge about the growth and production characteristics of one's own cell line, there is no substitute to developing one's own media.

Metabolic Control and Culture Parameters

Since the production of a biopharmaceutical usually involves the over-expression of a heterologous protein in a naïve host, the cellular energy consumed for such production can amount to a substantial part of the total energy of the cell. Therefore, conservation and optimization of cellular energy consumption by providing a well-balanced physical, chemical, and physiological environment becomes rather important in developing a cell culture process for biopharmaceutical production. Wagner10 gave an excellent overview on this subject by analyzing the biochemical constituents of production cells, cellular metabolisms, energy control, and maintenance.

Glucose, galactose, and fructose are possible carbon sources for the cultivation of animal cell cultures. However, glucose metabolized through the glycolysis pathway is consumed at a much faster rate than the other sugars. It was measured to be around 0.13 × 10-6 nmol/sec/cell for mammalian cells and 0.07 × 10-6 nmol/sec/cell for insect cells. Glutamine is actually often one of the main energy sources in cultivation of mammalian cells. The metabolic flux of glucose and glutamine are shown in Fig. 25.1.11 CO2 is an end product of aerobic metabolism, but is also used by cells as a substrate for the carboxylation of pyruvate into oxalacetic acid catalyzed by pyruvate carboxylase. Oxalacetic acid feeds into the TCA cycle. CO2 is also used in a physical mechanism to control the pH of the medium containing the bicarbonate buffer system. Lactate and ammonia are two major end metabolites of mammalian cell culture. A lactate concentration of >60 mmol/L or ammonium concentration of >5 mmol/L can be toxic to cell growth and inhibitory to product formation.12,13 A lactate concentration of <20 mmol/L or an ammonium concentration of <2 mmol/L should be maintained to avoid any inhibitory effect.1316

Fig. 25.1 Metabolic flux analysis. (Hu, W-S, Lecture notes, Cell and Tissue Reactor Engineering, University of Minnesota, 2000. By permission.



In addition to the concentrations of nutrients such as glucose, glutamine, and other essential amino acids and the concentrations of waste products such as lactate and ammonia, the following parameters constitute the basic parameters used for monitoring and controlling cell growth and product synthesis.

1. *Cell number and viability*. Total cell numbers can be measured using haemocytometer or flow cytometer by counting whole cells or stained nuclei. Cell viability is most commonly measured by a dye (0.4% Trypan blue) exclusion method. They can also be determined using MTT colorimetric assay or by indirect measurements of the metabolic rates of oxygen, carbon dioxide, glucose, or lactate.
2. *pH*. Optimal pH is around 7.07.5 for mammalian cells and 6.26.6 for insect cells. CO2 and NaHCO3 are commonly used for pH control.
3. *Temperature*. Optimal temperature is around 37°C for mammalian cells and 2530°C for insect cells.
4. *Osmolarity*. Optimal osmolarity is around 280320 mOsm/kg for mammalian cells and 360400 mOsm/kg for insect cells.
5. *Growth rate*. Doubling times for mammalian and insect cells are similar and are in the range of 1248 hr.
6. *Oxygen requirement*. Oxygen demand is a key design consideration for any aerobic biological process. The oxygen requirement for cell culture is relatively small compared to microbial culture.7 The oxygen demand is 5100 mmol/L/hr for microorganisms (bacteria, yeast) against 0.52 mmol/L/hr for mammalian and insect cells. However, because of its sensitivity to shears which can result from aeration and agitation, it is not exactly straightforward to supply oxygen to mammalian cells, especially when cultured in large bioreactors. Surface aeration, membrane diffusion, medium perfusion, oxygen enrichment, and increasing the partial pressure of oxygen are often required to meet the oxygen demand in cell culture.
7. *Shear sensitivity*. Because of the lack of cell wall and its relatively large size, mammalian or insect cells are quite sensitive to shear. High shear due to agitation and aeration as discussed above will damage the cells. Specially designed spargers and agitators are used to minimize shear. A number of polymers including methylcellulose, dextran, Pluronic F-68, and various reverse Pluronic polyols have also been used to protect cells from shear.
8. *Other process related parameters*. Parameters specific to cell types, expression systems, and operation modes can also be important. For instance, the number of viruses added per cell (MOI, multiplicity of infection), the timing of infection, the age of the virus, and the post-infection medium compositions are critical for BEVS.

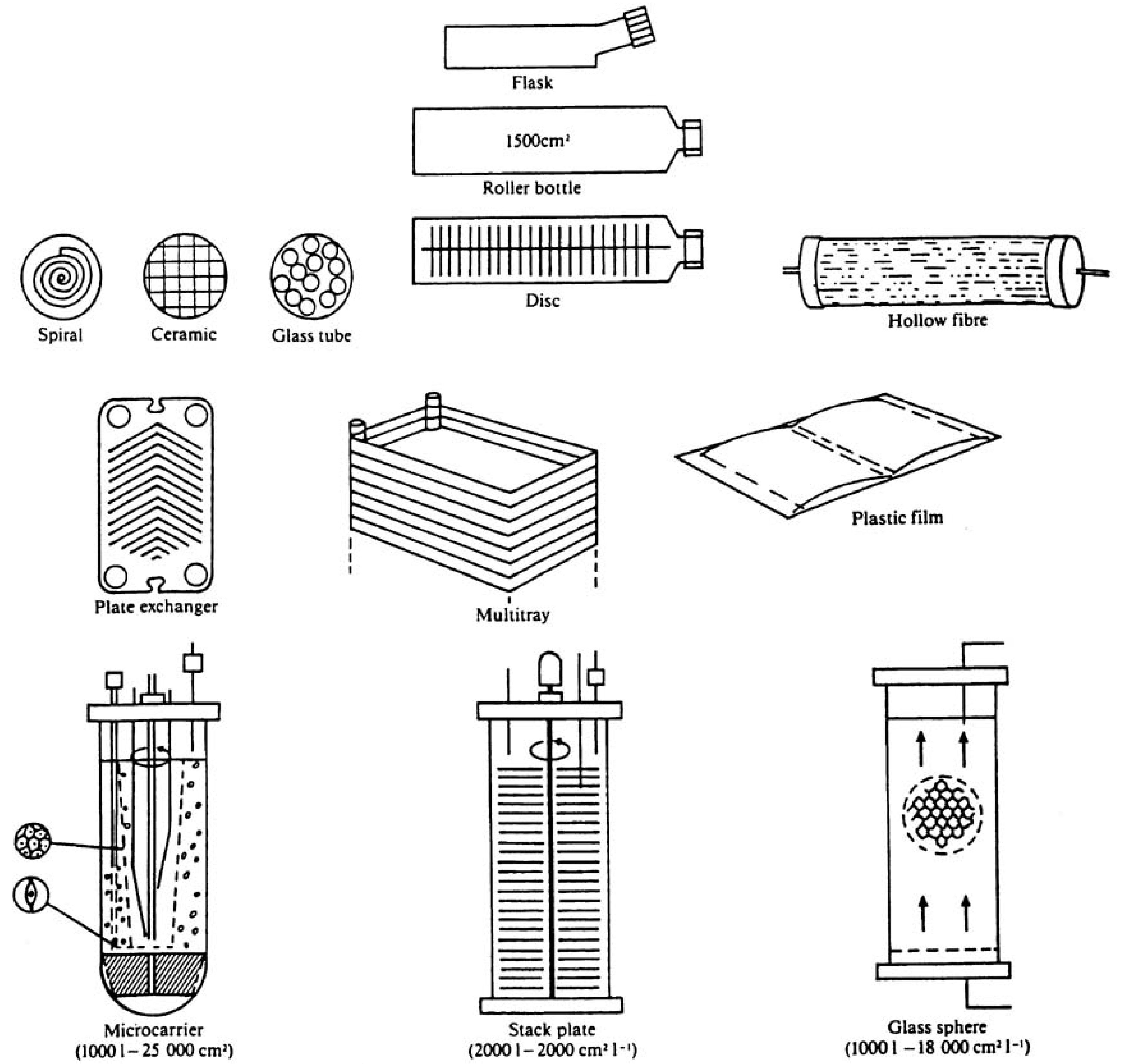
Scale-up/Bioreactors1719: Monolayer and Suspension Cultures

Depending upon whether a cell line is anchorage-dependent or independent, bioreactor configurations and scale-up criteria can be quite different. In many cases anchorage-dependent cells can be adapted to anchorage-independent growth and scaled up as a suspension culture. Suspension culture is generally the preferred mode for scale-up because of controllability, resource savings, and ease in process operation and validation. The advantages of monolayer culture include ease in medium replacement and perfusion. However, anchorage-independent or suspension cultures can also be immobilized on or entrapped in solid matrices to facilitate medium replacement and perfusion if needed.

1. *Monolayer culture*. The objective in scaling up a monolayer (anchorage-dependent) culture is to maximize the surface area for cell attachment and growth and to minimize the overall volume (medium and headspace) while maximizing the productivity.

Figure 25.2.17 shows the scale-up of anchorage-dependent cells from T-flasks to roller bottles, modified roller bottles with Spiral-Cel (by Sterilin, UK), glass tubes (by Bellco), and extended surface roller bottles (ESRB by Becton Dickinson). Large-capacity stationary culture system also includes multi-tray units (by Nunc). Also available are hollow fiber systems (by A/G Technology), Opticell (ceramic cartridges by Charles River), plastic film bag (by Du Pont), Heli-Cel (by Sterilin, UK), plate heat exchangers (by APV), and Fibra-Cel (by New Brunswick/ Sterilin, UK) for culturing anchorage dependent cells.

Fig. 25.2 Bioreactor system for anchorage-dependent cultures. (© Oxford University Press, 1992. Reprinted from Animal Cell Culture: A Practical Approach, edited by R.I. Freshney (2-edition, 1992), page 68, by permission of Oxford University Press.)



As a unit operation, these bioreactor systems can be classified into three categories:

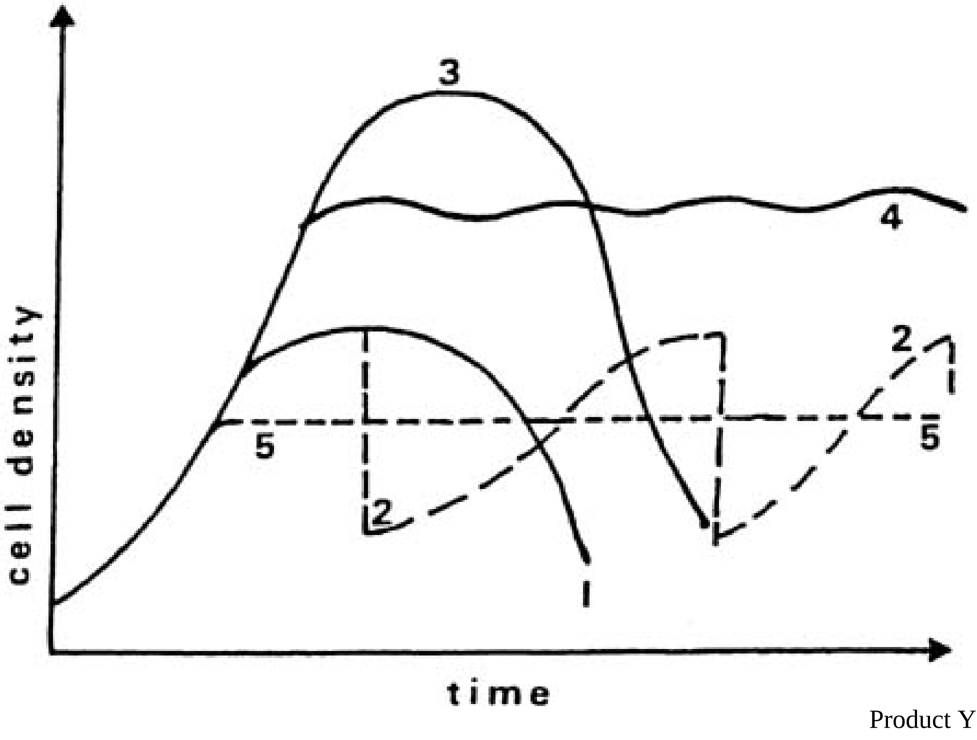
1. cells in stationary phase and medium in moving phase: such as glass bead bed bioreactors and Celligen Plus (by NBS);
2. heterogeneous mixing: such as stack plate reactors;
3. homogeneous mixing: such as microcarrier reactors.

2. *Suspension culture*. Anchorage-independent cell culture can be considered almost as a microbial culture. Bioreactor scale-up criteria are generally similar, except that for cell culture it is important to minimize shears by eliminating baffles, using curved bottom and water-jacket, and using impellers that cause minimal shears (such as marine propellers or helical-ribbon impellers) for mechanically agitated bioreactors or using draft-tubes for air-lift bioreactors.

Cell Culture Processes

Similar to the microbial fermentation processes discussed in the previous chapter, there are batch, fed-batch, semi-continuous batch (repeated draw-off), continuous perfusion, and continuous processes for cell culture. With mathematical equations similar to those in the previous chapter, these cell culture processes can be illustrated in Fig. 25.3.17 There are pros and cons for each process type. All bioreactors described in the previous section can be operated in one of these process modes.

Fig. 25.3 Comparison of culture processes. The numbers 15 are explained in the table above. (© Oxford University Press, 1992. Reprinted from Animal Cell Culture: A Practical Approach edited by R.I. Freshney (2-edition, 1992), page 60, by permission of Oxford University Press.)



Culture process	Cell Count (10E6)	Product Yield per Liter	
		mg/weeka	mg/montha

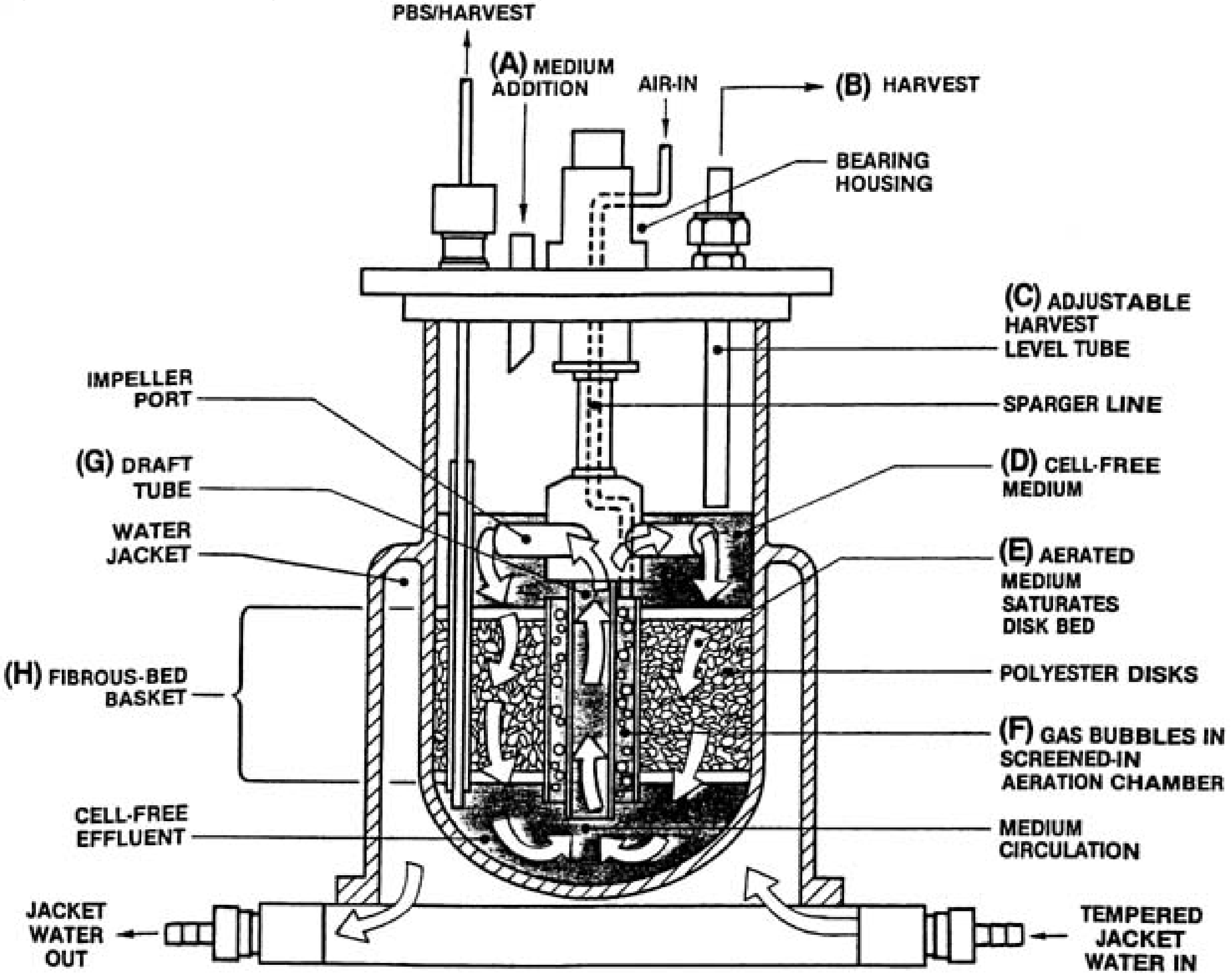
1. Bath	3	6	12
2. Semi-continuous batch	3	15	60
3. Fed-batch	30	60	120
4. Perfusion	10	70	280
5. Continuous	2	15	60

[a]Including turn-around time.

Batch, fed-batch, and continuous perfusion processes are most commonly used in the industry. Two different systems are used in the perfusion mode:

1. *Circulation inside of the bioreactor*. Figure 25.4 shows a typical perfusion system using a bioreactor like Celligen Plus (by NBS) where cells are retained inside of the bioreactor and so is the medium circulated and the environment controlled,

Fig. 25.4 Schematic of Celligen Plus bioreactor. Courtesy of New Brunswick Scientific, Edison, NJ, USA.



2. *Circulation outside of the bioreactor*. Figures 25.5 and 25.6 show a typical perfusion of a bioreactor incorporating glass beads, microcarriers, and hollow fibers where cells are retained inside the reactor but medium is circulated and the environment is controlled outside of the bioreactor.

Fig. 25.5 A glass bead bioreactor. A: glass beads; B: reservoir; C: pump; D: inoculation and harvest line; E: temperature-controlled water jacket. (© Oxford University Press, 1992. Reprinted from Freshney, R., ed., Animal Cell Culture: A Practical Approach, page 69, by permission of Oxford University Press.)

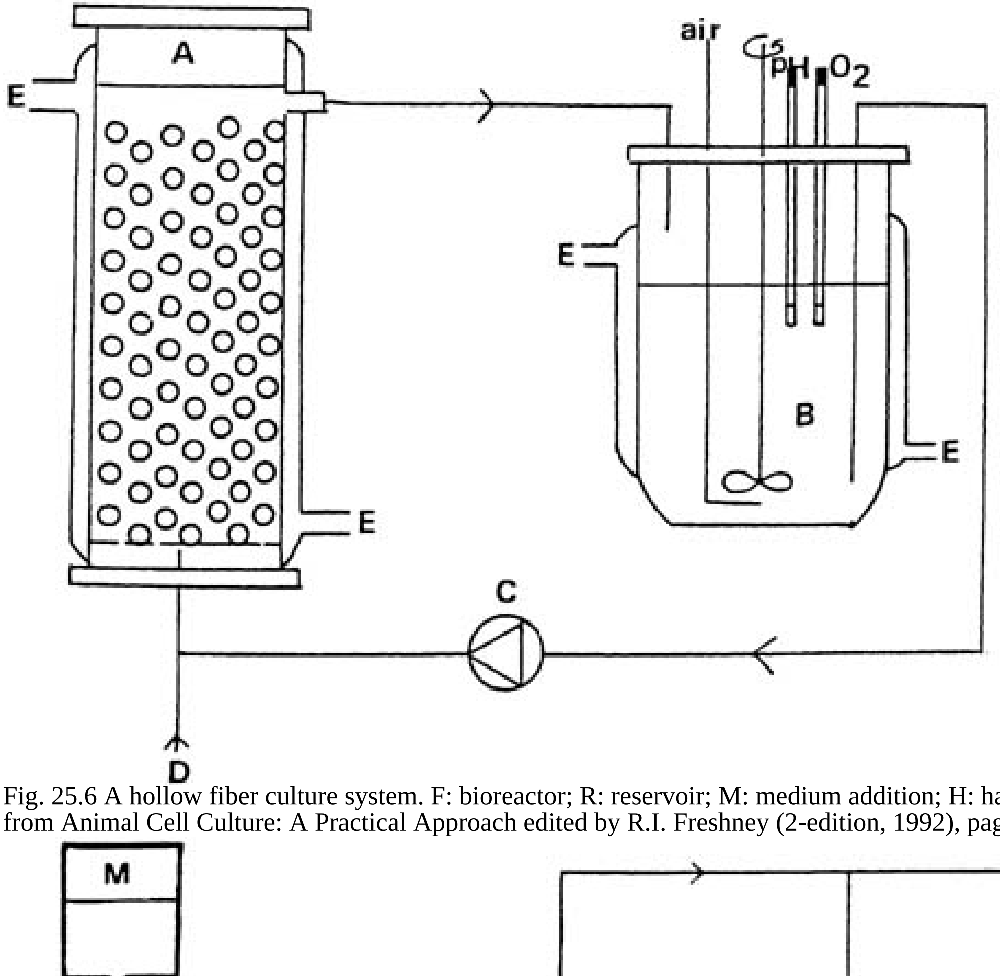
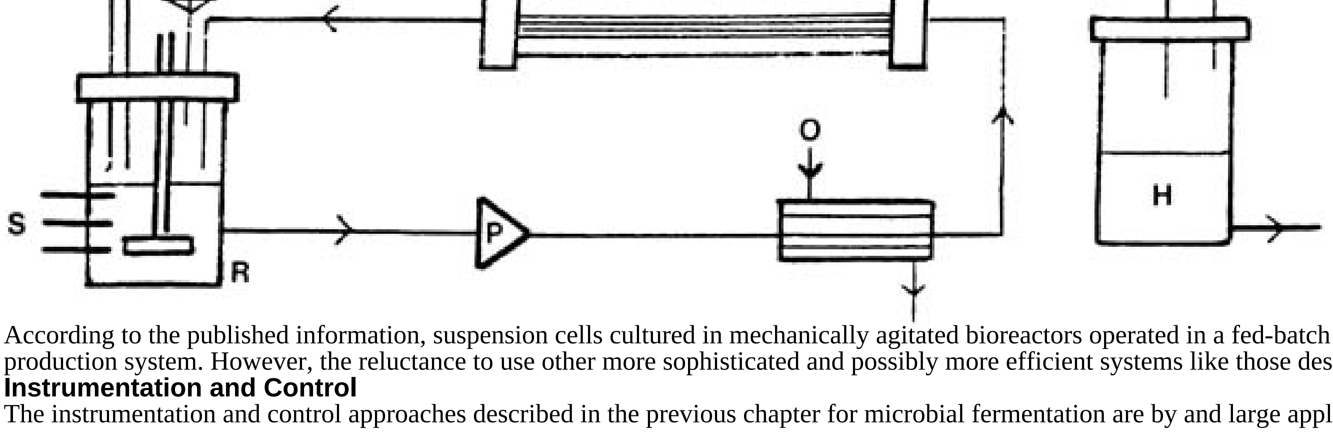


Fig. 25.6 A hollow fiber culture system. F: bioreactor; R: reservoir; M: medium addition; H: harvest vessel; S: sensors; O: oxygenation module; P: pump. (© Oxford University Press. Reprinted from Animal Cell Culture: A Practical Approach edited by R.I. Freshney (2-edition, 1992), page 88, by permission of Oxford University Press.)



According to the published information, suspension cells cultured in mechanically agitated bioreactors operated in a fed-batch mode are the most commonly practiced biopharmaceutical production system. However, the reluctance to use other more sophisticated and possibly more efficient systems like those described above seems to have gradually diminished.[20]

Instrumentation and Control

The instrumentation and control approaches described in the previous chapter for microbial fermentation are by and large applicable here to cell cultures.

Table 25.221 lists biosensors and on-line systems used for monitoring and/or controlling low molecular weight (e.g., glucose, lactate, and glutamine) and high molecular weight (e.g., IgG, antithrombin, tPA) components in cell culture medium. An affordable flow cytometer recently developed by Guava Technologies22 provides another powerful tool for the cell population analysis in a culture system either on-line or off-line. Other techniques used for monitoring microbial cultures such as on-line HPLC,23 on-line mass spectroscopy,24 and NMR25 can also be applied to cell cultures.

TABLE 25.2 Application of Biosensor Systems for On-line Monitoring of Cell Culture Processes

Component	Bioprocess	Sensor System	Duration of On-line Application
Low molecular weight			
Glucose	BHK cells	FIA, optrode	220 hr
		FIA, luminometer	220 hr
		FIA, oxygen electrode	100 hr
	Hybridoma cells	FIA, YSI analyzer	200 hr
Lactate	BHK cells	FIA, optrode	220 hr
		FIA, oxygen electrode	100 hr
	Hybridoma cells	FIA, HO electrode	270 hr
Glutamine	Hybridoma cells	FIA, HO electrode	270 hr
High molecular weight			
Mouse IgG	Hybridoma cells	FIA, affinity chromatography streaming potential	160 hr
		FIA, immunoassay	500 hr
Antithrombin III	Hybridoma cells	FIA, immunoassay	35 hr
tPA	Hybridoma cells	FIA, immunoassay	35 hr

Source: Adapted from Mammalian Cell Biotechnology in Protein Production, deGruyter, Berlin, 1997.

The following indirect approaches have also been used to estimate the metabolic states and fluxes of cells. Such information can then be used to develop on-line control strategies.18,26,27

1. Use oxygen uptake rates (obtained by material balancing) coupled with lactate production rates to estimate cell concentrations and then to adjust glucose feeding rates.
2. Use glucose and glutamine feeding strategies to reduce ammonia and lactate productions.
3. Use a fluorescence probe to determine cellular NAD(H) and to control nutrient feeding for the maintenance of desired redox status (ratio of reduced and oxidized coenzyme).

Recovery/Purification Processes

What is generally called "downstream processing" that includes recovery and purification of the protein is often actually the more critical part of a commercial cell culture process. The cost of downstream processing can account for 8090 percent of the total manufacturing cost. The purity requirement for products is usually very high, particularly those for therapeutic applications. As of 1990, the U.S. guidelines suggest that the bulk material, used to formulate intravenous or subcutaneous agents, must contain less than 0.1 percent protein impurities and less than 100 pg nucleic acid per dose, and be free of viral contaminants. The strategy to select an overall recovery scheme becomes critically important in order to minimize production costs and achieve purity objectives. Blanch and Clark28 and Jungbaue29 gave excellent overviews on this matter.

The order of the steps generally used to recover cell culture products is as follows:

1. *Separation of insolubles*. The insolubles include whole cells, cell debris, aggregated proteins, and insoluble nutrients. The commonly employed unit operations to achieve such separation include sedimentation, centrifugation, and filtration. Table 25.3 is a summary of various classes of membrane filters and their properties. If the desired product is intracellular or membrane-bound, cell disruption is required prior to separation of the insolubles. Unlike microbial cells, animal cells are relatively easy to rupture. Usually, detergents or blending will do.

TABLE 25.3 General Classes of Membrane Filters and their Properties

Membrane Filters	Pore Size (microns)	Retains	Materials of Construction	Applications
Microporous	0.110	Bacteria, cell debris	Cellulose esters, polyvinylidene fluoride, polycarbonate	Sterile filtration, cell processing
Ultrafiltration	0.0010.1	Macrosolutes (e.g., viruses, pyrogens, proteins, peptides)	Polysulfone, cellulosics	Cell & macrosolute processing, pyrogen removal
Reverse osmosis	0.00050.001	Ions	Cellulose acetate, polyamides, polysulfone	Water purification, concentration of small organics

Source: Adapted from Biochemical Engineering, Blanch & Clark, © Marcel Dekker, New York, 1997. By Permission.

2. *Isolation and concentration*. This refers to the separation of the desired product from impurities and the minimization of its volume. In this early downstream processing stage, significant concentration often can be achieved during isolation. The commonly used unit operations for this category of downstream processing include ultrafiltration (Table 25.3), extraction (aqueous two phase system), and ion exchange chromatography. Table 25.4 lists some of the more commonly used chromatographic methods for protein purification. Alternatively, expanded bed chromatographic columns, such as Pharmacia's Streamline,30 can be used to achieve both cell/debris removal and early isolation and concentration.

TABLE 25.4 Chromatographic Methods for Protein Purification

Basis of Separation	Chromatography Type	Characteristics	Applications
Charge	Ion-exchange	Resolution can be very high.	Most widely used method. Is most effective at early stage where large volumes are handled. Can also be used in batch mode
		Capacity is high.	
		Speed can be high, depending on matrix	
van der Waals interactions	Hydrophobic interaction	Resolution is good.	Can be applied at any stage, but is most usefully applied when the ionic strength is high (after ion exchange or salt precipitation). Possibility of some denaturation. May involve organics during elution.
		Capacity is good.	Speed is high.
Size and shape	Gel filtration	Resolution moderate for fractionation.	Fractionation is best for later stages of purification where loads are lower.
		Good for buffer exchange and desalting.	Buffer can be exchanged at any time.
		Relatively low capacity for loaded protein, but high flow rates possible with some packings	
Biological affinity	Affinity	Selectivity can be very high.	Can be used at any stage, but the materials are expensive; best to use when protein loads and fouling substances have been reduced by less-costly methods.
		Capacity can be high, depending on the ligand.	
		Speed is high.	

Source: Adapted from Biochemical Engineering, Blanch & Clark, © Marcel Dekker, 1997. By Permission.

3. *Primary purification*. This is meant for separation steps that are more discriminative and selective than isolation. Some of the primary purification steps take advantage of the dissimilarity in physical and chemical properties between the desired product and contaminating species. The commonly used unit operations here include precipitation (by changing the solvent environment such as salt concentration, dielectric constant, and pH) and chromatography (Table 25.4).
4. *Final purification*. Final purification steps often include those for viral removal or inactivation. See the U.S. and European regulatory guidelines (CBER, 1994; CPMP, 1996).6,31 Crystallization and drying are also often used for final purification. In addition to helping maximize purity, they also help to extend the bulk material's shelf life. However, many therapeutic protein bulks remain in liquid form.

Based upon the above sequence of operations, a generalized recovery and purification scheme is shown in Fig. 25.7. Figure 25.832 gives a typical recovery and purification scheme for monoclonal antibodies. The guidelines proposed by Wheelright33 for devising a recovery scheme are a useful starting point for the design of a downstream process.

Fig. 25.7 A generalized flow chart for purification of proteins. (Adapted from Blanch, H. and Clark, D., Biochemical Engineering, 1997, Fig. 6.42, *Generalized flow chart for the purification of recombinant proteins*, p. 546. Copyright © Marcel Dekker, 1997. By permission of the copyright owner.

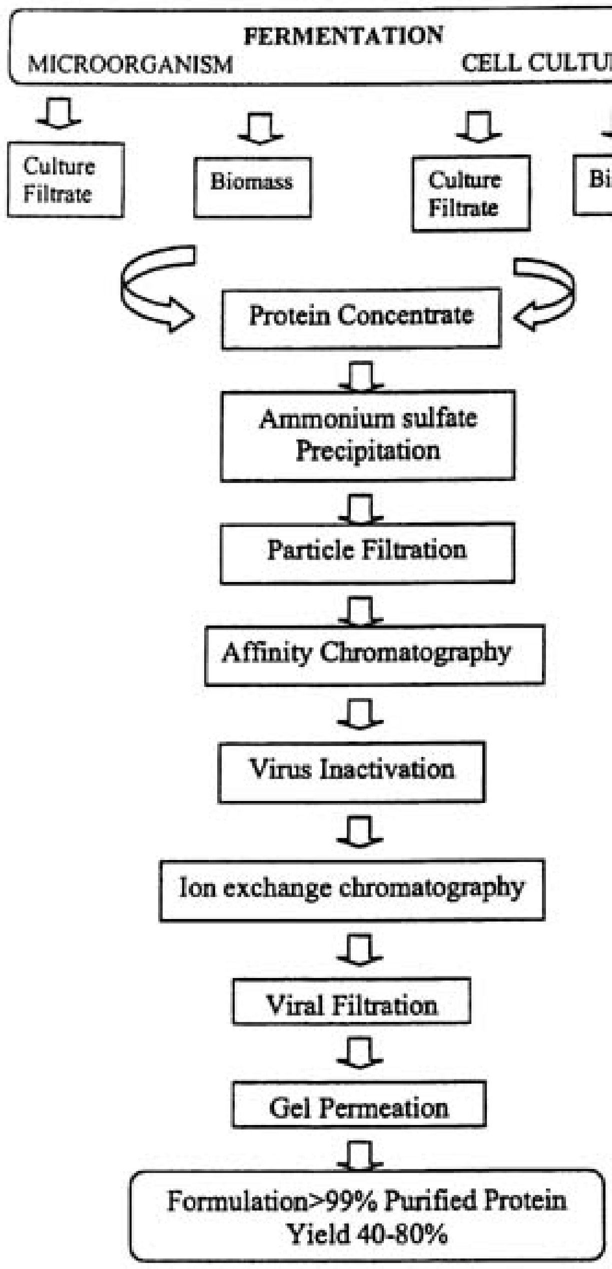
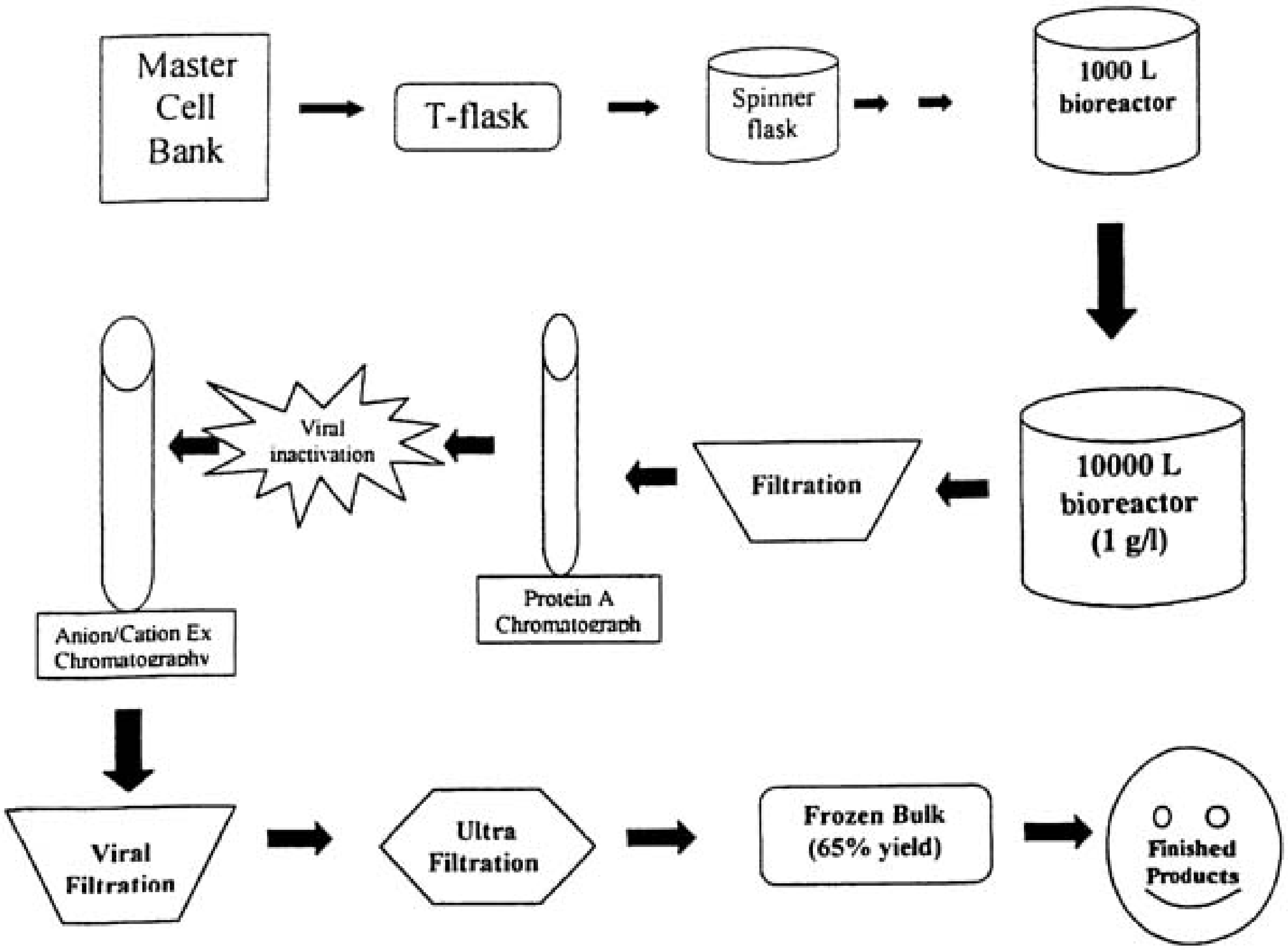


Fig. 25.8 A generalized flow chart for monoclonal antibody production. (*Source:* The biopharm guide and cell culture and industry sources.)32



1. Choose separation unit operations based on differentiating physical properties, for example, charge or size.
2. Choose isolation unit operations that exploit the greatest difference in the physical properties of the desired product and impurities.
3. Choose the step that will result in the largest reduction of mass or volume first.
4. The best process in the laboratory may not be the most economical process in the production plant.
5. The fewer steps, the higher the yields; the simpler the process, the easier the manufacturing.

25.3 Products of Cell Cultures

Table 25.2 lists the major recombinant biopharmaceuticals approved by the FDA up to the year 2000. It includes proteins produced by either microbial fermentation or cell culture. Wetzel34 categorized the proteins used for medical applications as follows:

1. Proteins for non-immune based diseases including cardiovascular and cerebrovascular thrombembolic disorders; stroke, restenosis, heart failure, fibrosis, osteoporosis, angiogenesis, insulin resistance, and non-insulin-dependent diabetes mellitus and sepsis.
2. Proteins for immuno-enhancement including tumor therapy and vaccines.
3. Proteins for immuno-suppression including co-stimulation, antigen presentation, immune deviation, multiple sclerosis, psoriasis, arthritis, insulin-dependent diabetes mellitus, inflammatory bowel disease, allergy/asthma, allo-transplantation, and xeno-transplantation.
4. Proteins for other specific medical applications

1. Cytokines including interferons, interleukins, growth and colony stimulating factors, nerve growth factors, and other growth factors.
2. Hormones including insulin, growth hormones, insulin-like growth factors, and other hormones such as follicle stimulating hormone.
3. Anti-inflammatory proteins such as IL-1ra and interleukin receptor antagonists.
4. Enzymes and enzyme inhibitors: such as Pulmozyme for cystic fibrosis, endopeptidase for inflammation, and glucocerebrosidase for Gaucher's disease.
5. Clotting factors and related proteins such as Factor VIII and IX for clotting deficiency, Factor VIIa for hemophilia A, tissue plasminogen activator for thrombosis, and prourokinase for stroke.
6. Monoclonal antibodies for therapeutic use such as Zenapax for organ transplant, Ramicade and Enbrel for arthritis, Rituxan for lymphoma, and Herceptin for breast cancer.

Not much exact process detail is available from the manufacturers of the commercialized biopharmaceuticals. However, a few published processes presented below may provide some insights into how biopharmaceuticals are processed in large scale in general.

Tissue Plasminogen Activator (tPA)

Recombinant tPA is used clinically as a thrombolytic agent for the treatment of myocardial thrombosis. It has a specific affinity for fibrin and its enzymatic activity is markedly increased in the presence of fibrin. Genentech has marketed the product under the trade name of Activase since 1987.

Production of tPA by Recombinant CHO Cells at Genentech.35

CHO cells, transfected with the expression vector pETPFR (expressing tPA), was grown in a medium consisting of 1:1 mixture of Ham F12 and Dulbescco modified Eagle medium without hypoxanthine or thymidine but supplemented with diafiltered FBS (7% v/v) and methotrexate (up to 500 nM) in the seed stages. It is followed by further growth in a bioreactor with a medium containing 2 percent FBS, hypoxanthine, and thymidine but without methotroxate. When the cell density increased from 0.25 × 106 to 1 × 106 cells/ml, the culture broth in the reactor was pumped and circulated through a hollow fiber filter to remove the permeate containing waste and residual serum while replenished with fresh medium containing no serum. The process was continued until serum concentration was reduced to less than 0.0001 by volume. Then the batch culture for the production phase was run at 37°C for 90 hr.

Large Scale Purification of tPA.36

TPA produced by a human melanoma cell line (Bowes) was purified from 30 L supernatant on immunosorbent chromatographic column using monoclonal antibodies (AF-10 and AF-54), followed by chromatography on Arginine-Sepharose 4B and gel filtration on Sephadex G-150. The step yield and purity were shown in the following table. The over yield was about 50 percent. The final purity of the product gave a specific activity of about 200,000 IU/mg protein.

Purification Steps	Total Amount of Protein by Amino Acid Analysis		Total Amount of Protein by Clotlysis		Total Amount of activator by RIA		Specific Activity (IU/mg)
	mg	%	IU × 106	%	mg	%	
Cell supernantant	3790	100	3.85	100	20	100	1,000
IgG Sepharose	23	0.6	3.5	91	19	96	155,000
Arginine-Sepharose	13	0.35	2.45	64	13	64	187,000
Sephadex G-150	9	0.25	1.95	51	9	47	208,000

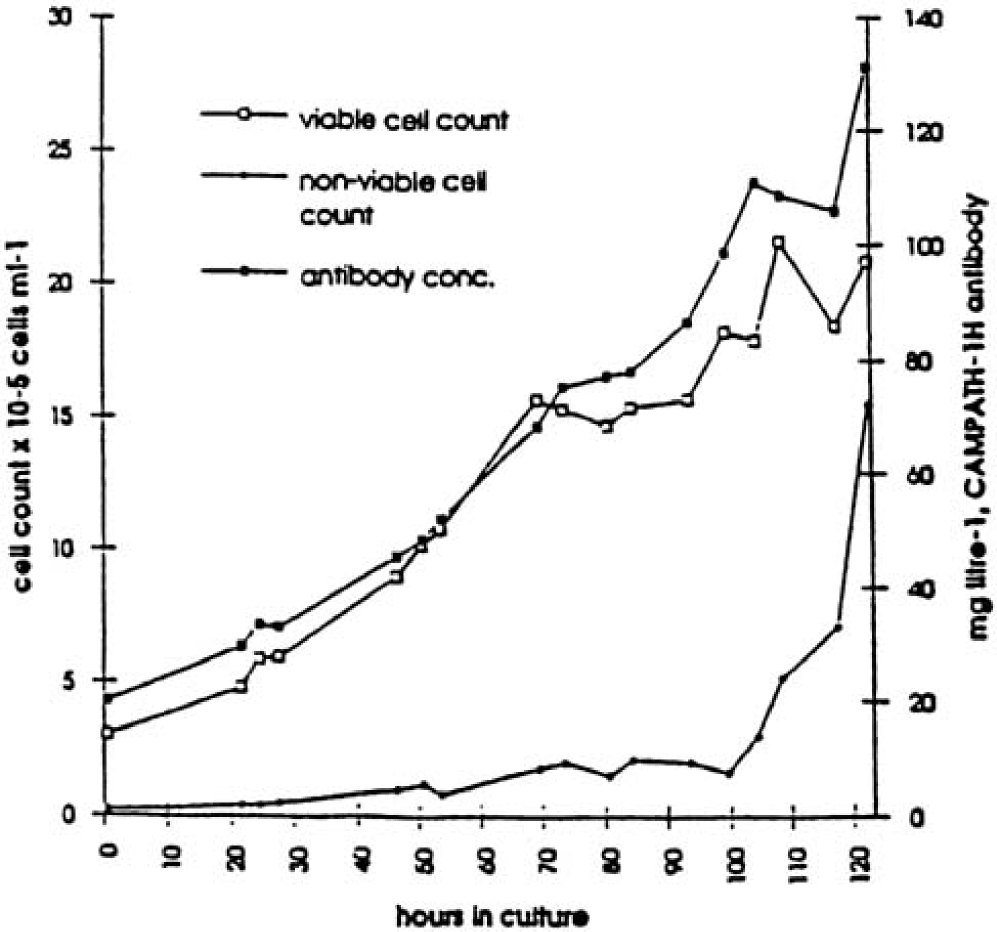
Monoclonal Antibodies

Monoclonal antibodies, for their high affinity and selectivity for the target antigens, are among the most commercially developed cell culture products. They have been used for transplant rejection, rheumatoid arthritis, non-Hodgkins lymphoma, and so on. Commercial products include ReoPro, Campath, Simulect, Zenapax, Remicade, Synagis, Enbrel, Rituxan and Herceptin, and the like.

Production of Humanized IgG1 Monoclonal Antibody (Campath1) at Welcome.37

CHO 3D11 cells, derived from CHO DUK B11 dhfr-cells, were engineered to secrete a humanized antibody Campath-1H using dhfr selection and MTX amplification. The cells were grown in Iscoves medium containing 10 percent dialyzed FBS, NEAA, free of 10-6 M MTX in static culture. It was then washed, grown, and adapted in a SF medium designated WCM5, an Iscoves medium without lecithin, albumin, or transferrin, but supplemented with selected components of F12 medium and 5 mg/L of human recombinant insulin. The adapted cells were renamed as CHO 3D44. The batch cell culture was carried out in a 8000-L fermentor fitted with Rushton turbine type stainless steel impellers run at 70 rpm through a magnetically coupled rotor. Temperature was controlled at 37°C, dissolved oxygen at 15 percent of saturation, and pH at 7.1. The fermentor was seeded with 2 × 105 cells/ml and reached a maximal viable cell density of 2.16 × 106 cells/ml after 108 hr in culture and a maximum antibody concentration of 131 mg/L after 122 hr in culture. The cell culture profiles are shown in Fig. 25.9.

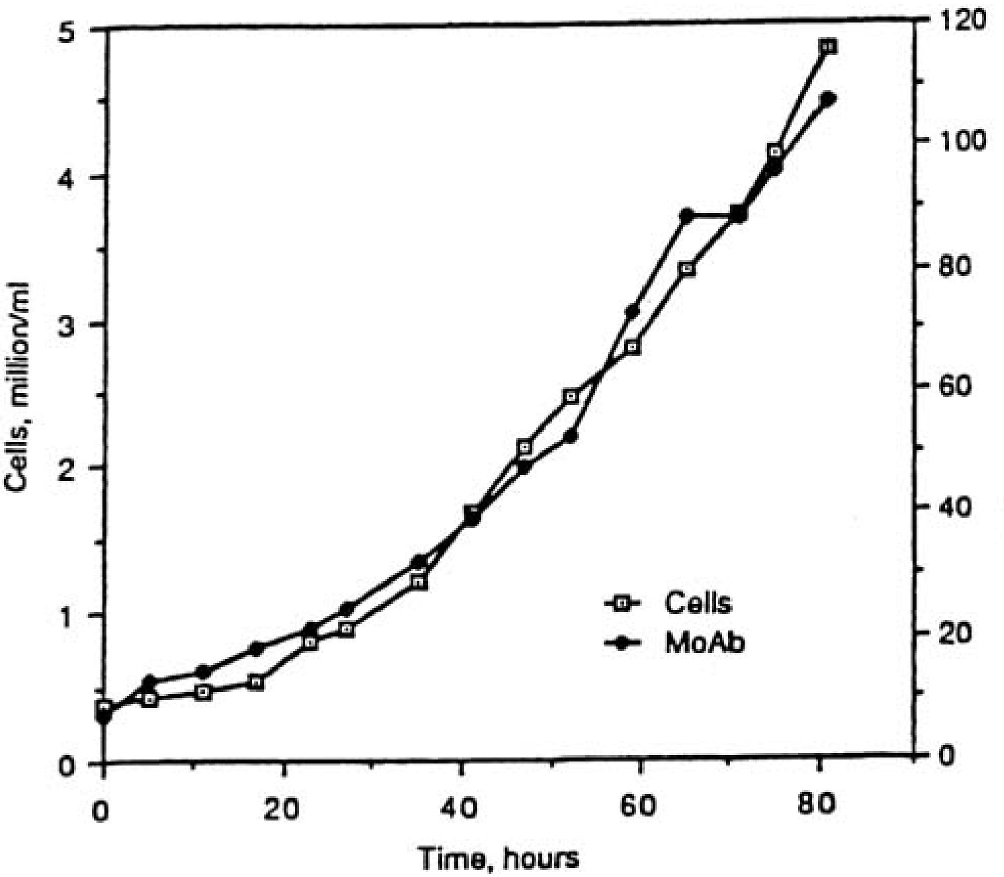
Fig. 25.9 CHO 3D44 cell growth and antibody production in an 8000-L bioreactor using WCM5 serum-free medium. (Reproduced with permission, from Keen, M., and Rapson, N., *Cytotechnology*, 17, 153163, 1995. © Kluwer Academic/Plenum Publisher, New York.)



Production of KS1/4 Antibodies in Suspension Cell Culture at Eli Lilly.38

Hybridoma cell line KS1/4 was used to produce monoclonal antibody that has an affinity for the cell surface antigens found on lung and colorectal adenocarcinomas. The hybridoma culture, seeded from a 150-L initial culture, was conducted in a 1300-L fermenter and controlled at 37°C, pH 7.1, and dissolved oxygen (DO) of 20 percent. The set DO was maintained by sparging with air once the culture's oxygen consumption rate exceeded the rate of oxygen transfer through both surface. Agitation rate was increased from 30 rpm at the beginning to a maximum of 140 rpm in the 150-L tank and 75 rpm in the 1300-L tank. Typical cell concentrations ranged from 34 × 105 cells/ml at inoculation to 4 × 106 cells/ml at harvest in 84 hr of culture. Cell viability was generally 95 percent throughout most of the cell culture but might drop to 80 percent in last few hours before harvest and re-inoculation to start the next run (in a semi-continuous, drew-and-feed mode). The antibody concentration exceeded 100 mg/L. Figure 25.10 shows the cell growth and antibody production profiles. The 150-L and the 1300-L computer-controlled stirred fermentors were retrofitted for mammalian cell culture by modifying the agitation and aeration systems. The reactors were 304-L s/s vessels with standard ASME dished heads and four baffles. Rushton turbines were replaced with a 9″ diameter marine propeller in the 150-L tank and 22″ propeller in the 1300-L tank. Agitator drives were modified to run between 25 and 250 rpm. Incoming air was filtered through two 0.1-µm sterile filters and supplied to the culture either as a surface overlay or through a sintered stainless sparger with an average 1-µm pore size. The overlay air was maintained at a rate sufficient to give a steady 4 psi back pressure. The pH was maintained at 7.1 using carbon dioxide and sodium bicarbonate.

Fig. 25.10 Cell and antibody concentrations produced by a KS 1/4 hybridoma in a semi-continuous culture. (Baker, M., Metzger, L., Slaber, K., Nevitt, K., and Boder, G., *Biotechnol. Bioeng.*, 32, 9931000, John Wiley & Sons, Inc., 1988. Used by permission of copyright owner.)



Production of Monoclonal Antibodies for Use in Human at Celltech.39

At Celltech, a 2000-L air-lift cell culture system as shown in Fig. 25.11 was used to produce monoclonal antibodies. By varying the composition of the gas mixture, dissolved oxygen and pH of the culture were controlled. Normally a batch or fed-batch culture was the preferred mode of production for monoclonal antibodies. A typical profile is shown in Fig. 25.12.

Fig. 25.11 Celltech's air-lift process system. (Reproduced with permission, from Hill, C., *Biochem. Soc. Transc.*, 18, 245247, 1990, © the Biochemical Society.)

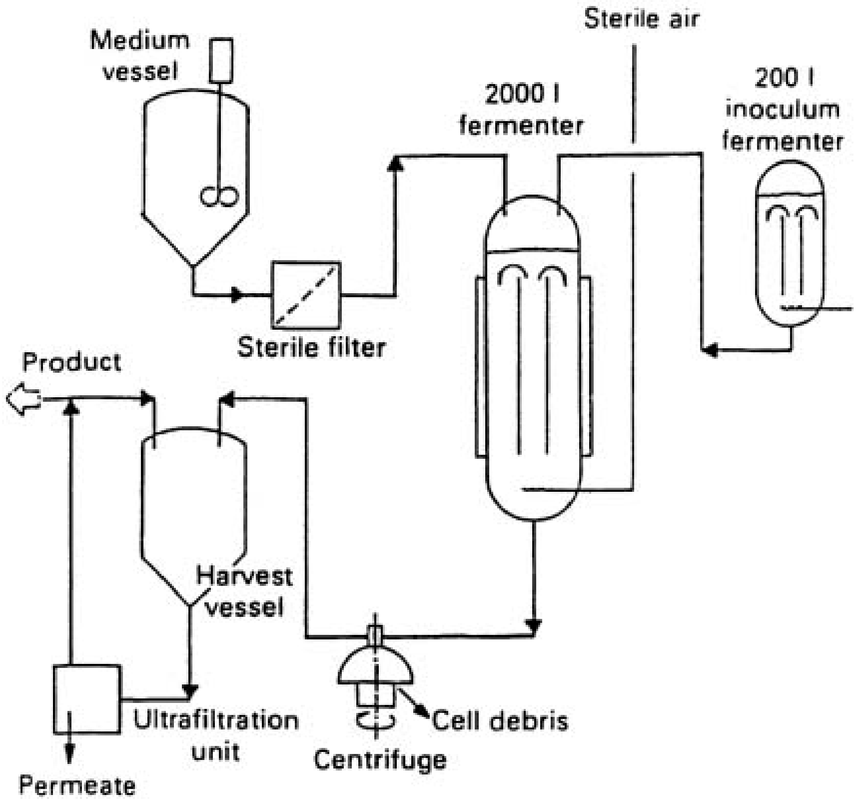
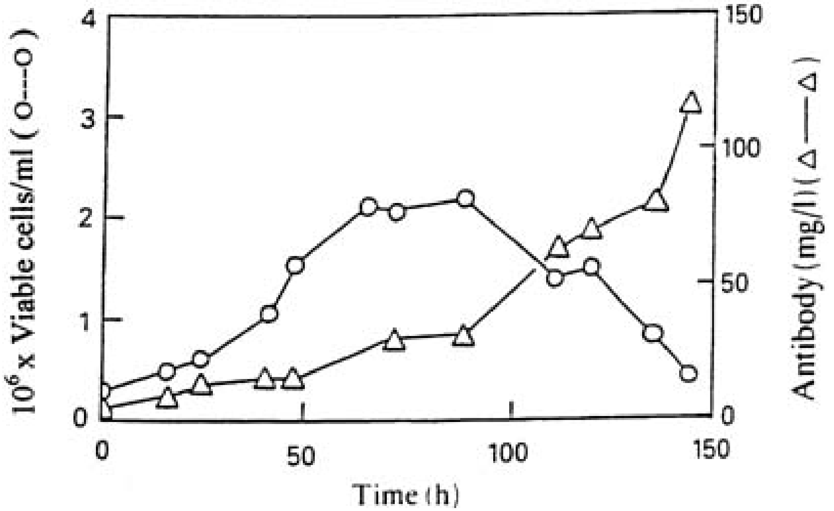


Fig. 25.12 Fed-batch airlift culture of murine hybridoma producing monoclonal antibody. (Reproduced with permission, from Hill, C., *Biochem. Soc. Transc.*, 18, 245247, 1990, © the Biochemical Society.)



For large scale purification of monoclonal antibodies, Celltech used ultrafiltration to concentrate the culture supernatant. It is then followed by a combination of ion-exchange chromatography, affinity chromatography, and gel-filtration chromatography. Using these techniques, purity of >95 percent was routinely achieved.

Batch Production of Monoclonal Antibody by Large Scale Suspension Culture at NCI, NIH, Maryland.40

A batch culture of hybridoma cells (9.2.27) for the production of an IgG2a was carried out in a 50-L fermenter at the National Cancer Institute. The basic fermentor setup, operating parameters, and culture medium are shown in Fig. 25.13. The production flow chart is given in Fig. 25.14. The cell culture was conducted with pH and DO controls (pH 7.0 ± 0.3 and DO >10%) by sparging or overlaying the culture with air and/or CO2. Agitation rates were maintained at a tip speed of approximately 4000 cm/min with a marine impeller. Once the stationary phase was reached the culture was held at the operating parameters through a log death phase of the culture. The culture was chilled when its viability dropped to <20 percent. Typical cell growth and antibody production curves are shown in Fig. 25.15. The cells were then removed by passing through a continuous flow centrifuge. The supernatant was then concentrated on a Millipore Pellicon unit with a 100 kDa polysulfone membrane and polished using a 0.2-µm positive charged Pall filter. The antibody was concentrated by 1020× at a yield of >95 percent. The concentrate was further purified by a Protein A-Sepharose CL-4B column chromatography. The purity was increased by 60120 fold with a yield of >70 percent. The purified solution was further concentrated by the Pellicon unit using 100 kDa polysulfone membrane to an antibody concentration of 20 mg protein/mL. It was then formulated by addition of sodium chloride until a proper chloride level was achieved. Following QC analysis and validation, the bulk material was processed through a 0.22-µm sterile filter into glass serum bottles.

Fig. 25.13 NCI's basic 50-L fermentor design, operating parameters, and culture medium. (Reprinted from Lebherz, W., "Batch Production of Monoclonal Antibodies by Large Scale Suspension Culture," *Bioprocess Technology*, Seaver, S. S. (Ed.), Vol. 2, pp. 93118, Copyright Marcel Dekker, 1987. By permission of Marcel Dekker.)

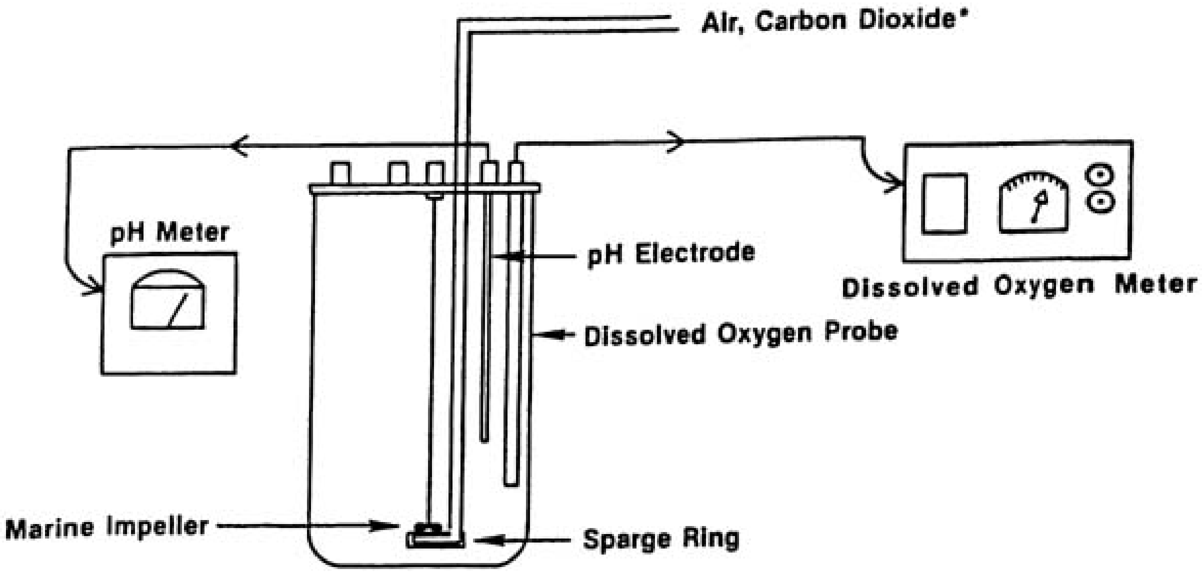


Fig. 25.14 NCI's production process of 9.2.27 monoclonal antibodies. (Reprinted from Lebherz, W., "Batch Production of Monoclonal Antibodies by Large Scale Suspension Culture," *Bioprocess Technology*, Seaver S. S. (Ed.), Vol. 2, pp. 93118, Copyright Marcel Dekker, 1987. By permission of Marcel Dekker.)

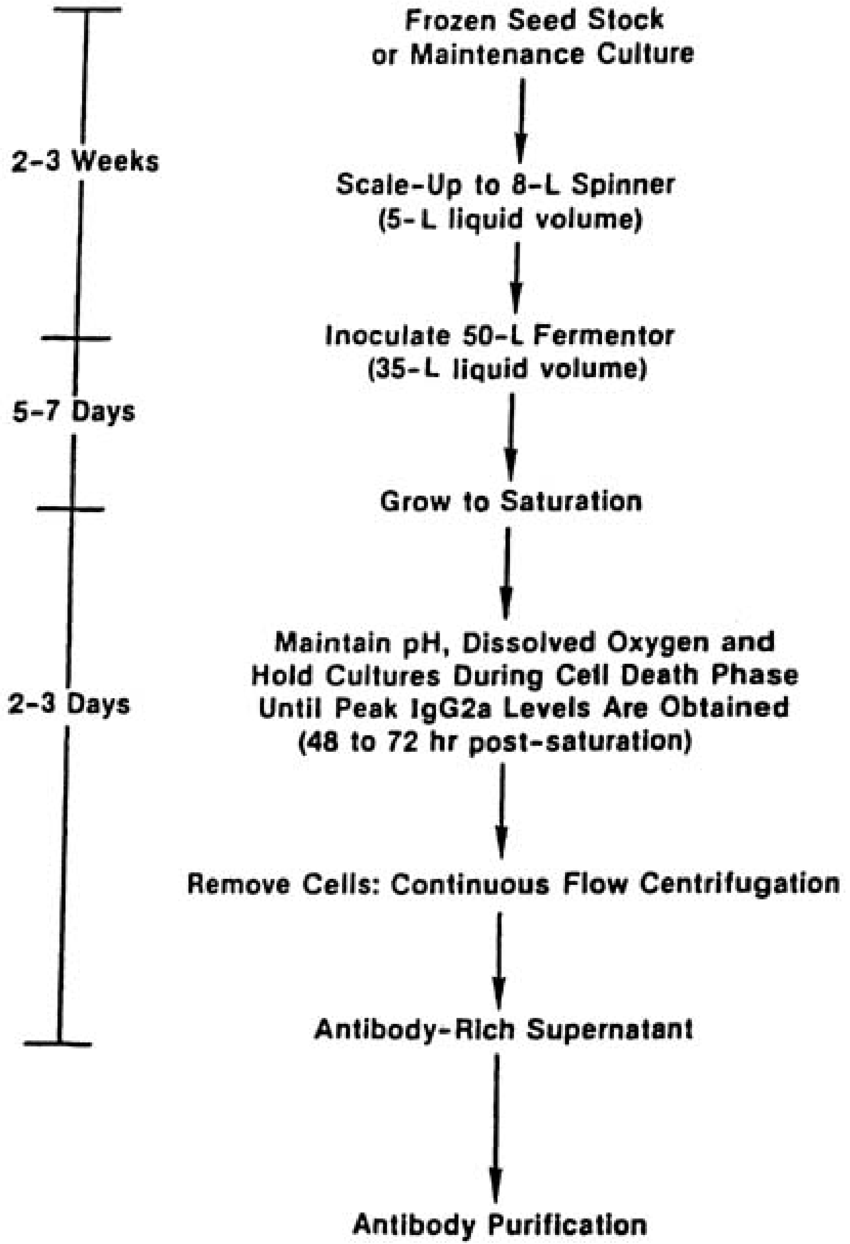
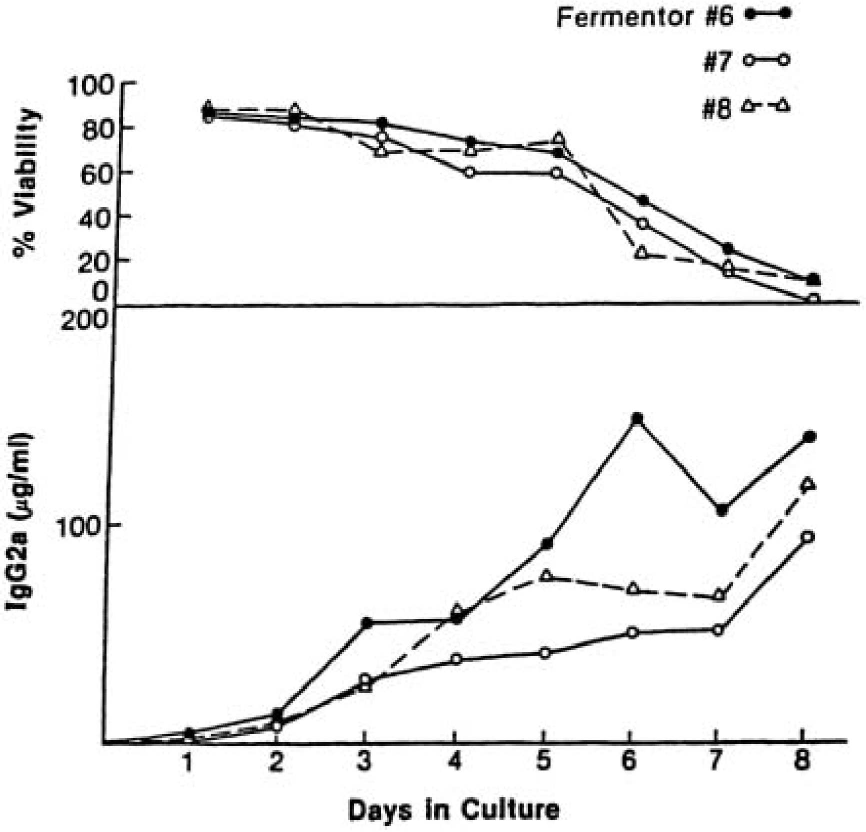


Fig. 25.15 NCI's 9.2.27 monoclonal antibody production profiles from three different runs. (Reprinted from Lebherz, W., "Batch Production of Monoclonal Antibodies by Large Scale Suspension Culture," *Bioprocess Technology*, Seaver, S. S. (Ed.), Vol. 2, pp. 93118, 1987. Copyright © Marcel Dekker, 1987. By permission of Marcel Dekker.)



B. Operating Parameter

Temperature	37 ± 0.58C
Agitation rate (tip speed)	100 rpm (4,000 cm/min)
pH Range	7.0 ± 0.3
Dissolved oxygen	>10% saturation
Pressure (head-space)	5 ± 3 psig

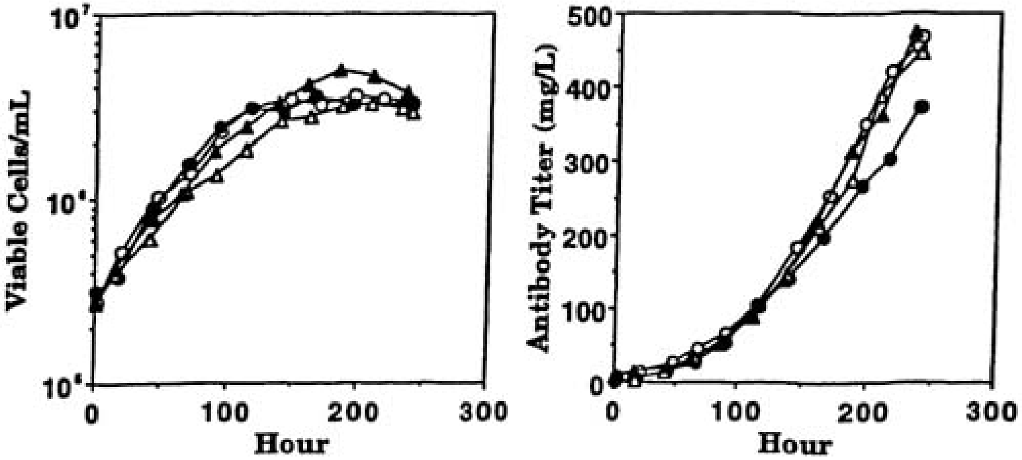
Working volume/total volume	35/50 L
C. Culture medium	
DMEM (powder) with 4,500 mg D-glucose/L	13.37 g/L
Fetal bovine serum	7% (v/v)
Sodium pyruvate	100 mg/L
HEPES buffer	10 mM
L-Glutamine	4 mM
Sodium bicarbonate	3.7 g/L

*Venting mechanisms, pressure regulating system, heating/cooling systems not shown.

Large Scale Production of Humanized Monoclonal Antibody Expressed in a GS-NSO Cell Line at Roche.41

A GS-NSO cell line was engineered for the production of a humanized monoclonal antibody in a 2000-L bioreactor. The culture was kept in suspension and well mixed by agitation. Microprocessor based automation was used to control temperature, pressure, dissolved oxygen, pH, agitation, and gas flow rates. The cell growth and antibody production profiles are shown in Fig. 25.16, and the performance of the production batches is summarized in the following.

Fig. 25.16 Consistency of a cell culture process producing a humanized monoclonal antibody expressed in a GS-NSO cell line in a 2000-L bioreactor. (Reproduced with permission from Ray, N., Rivera, R., Gupta, R., and Mueller, D., *Animal Cell Technology*, pp. 235241, 1997. © Kluwer Academic/Plenum Publisher, New York.)



Batch	Culture Time (hr)	Mab Titer (mg/L)	Integral Viable Cell (10E6 cell × hr/ml)	Sp. Productivity (µg/10E6 cell/hr)	Viability at Harvest (%)
1	240	468	576	0.81	76
2	240	372	563	0.66	70
3	240	444	472	0.94	80
4	235	473	616	0.77	80
Average		439 ± 11%	557 ± 11%	0.80 ± 14%	77 ± 6%

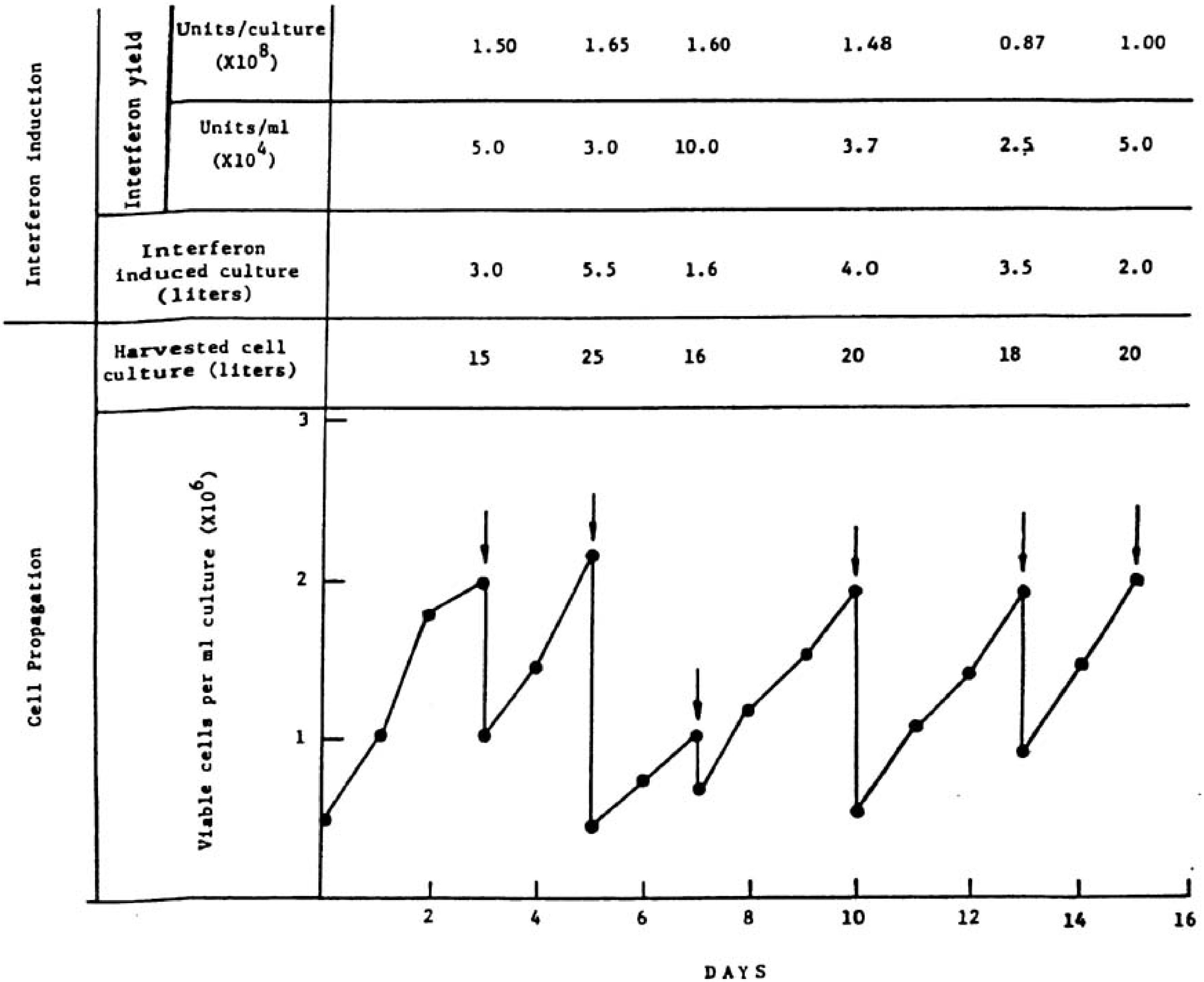
Interferon

Human interferons (hIFNs) are a family of proteins initially discovered for their ability to inhibit viral infection and their immunomodulating activity. They have been grouped into hIFN-(leukocyte), hIFN-(fibroblast), and hIFN-(immune). Their therapeutic applications include treatments for hepatitis C, multiple sclerosis, and lymphoma.

Large Scale Production of Human Lymphoblastoid Interferon α.42,43

Cells from the Namalva human lymphoblastoid cell line were cultured in a 130-L bioreactor, in a semi-continuous mode, and in a low cost medium based on RPMI-1640 and containing heat-inactivated and PEG-treated bovine serum, Primatone, and Pluronic F-68. Every 4872 hr, when the cell density reached 2×106 viable cells/ml, a portion of the cell culture was withdrawn and an equal volume of fresh medium was added. The harvested cells were collected by centrifugation and resuspended in a SF RPMI-1640 medium at concentration of 107 cells/ml and primed by the addition of 100 units/ml of IFN. After 2 hr of incubation at 36°C, the culture was then infected with 100 HAU/ml (hemagglutination units/ml) of Sendai virus as an IFN inducer. After 20 hr, the cells were removed by centrifugation and the supernatant was ready for further processing. Fig. 25.17 summarized cell growth and IFN production during the two-week production period. Culture supernatant (at 1 × 106 µ/mg protein) was incubated in pH 2.0 to inactivate the inducer virus. Then the pH was adjusted to 7.0 prior to concentration by ultrafiltration (with a 20,000 Da MW cut-off). The concentrated IFN solution was centrifuged before gel filtration was carried out on a Sephadex G-75 column. The collected IFN fractions were then concentrated by trichloroacetic acid precipitation, followed by solubilization in a 0.1 M potassium phosphate buffer and centrifugation. The partially purified IFN (at 12 × 106 µ/mg protein) was then purified by a monoclonal anti-Namalva INF-α chromatography (NK2 Sepharose, Celltech, UK). A highly purified INF (at 12 × 108 µ/mg protein) preparation was finally obtained. The overall yield was 5055 percent.

Fig. 25.17 Namalva cell growth and IFN production in a 130-L bioreactor using semi-continuous culture. (Reuveny, S. et al., *Ann. Virol.*, 133E, 191199, 1982. Reprinted with permission from Publications Scientifiques, Institut Pasteur, Paris.)



Interferon induction	Interferon yield	Units/culture ($X10^8$)	1.50	1.65	1.60	1.48	0.87	1.00
		Units/ml ($X10^4$)	5.0	3.0	10.0	3.7	2.5	5.0
		Interferon induced culture (liters)	3.0	5.5	1.6	4.0	3.5	2.0
Cell Propagation		Harvested cell culture (liters)	15	25	16	20	18	20

Interleukin

Production of Interleukin-3 (IL-3) in a Suspension Culture.44

At the NCI-Frederick cell culture facility, IL-3 was constitutively produced using WEHI-3, a murine cell line that is reported to be a myelomonocyte. Because of potential biohazard concerns, a contained system as shown in Fig. 25.18 was designed and used. The medium consisted of the RPMI-1640 basal medium supplemented with glutamine, HEPES buffer, gentamicin, and FBS. The culture process started with a frozen seed stock. After six stages of spinner flask cultures, an inoculation density of 1.25 × 105 cells/mL was reached for use in the final 50-L production fermenter. The culture doubling time was 79 hr and the inoculum cell density was above 5 × 104 viable cell/mL with 95 percent viability throughout the inoculum build-up. The agitation speed was maintained at 16 rpm during inoculation, increased to, and maintained at 50 rpm until cell density reached 5 × 105 cells/mL, and then increased to 75 rpm and further to 100 rpm as cell density reached 1 × 106 cells/mL about 68 hr. The culture reached a saturation density between 75 and 80 hr and then entered the log death phase evidenced by a rapid drop in both cell viability and cell density. Prior to reaching the saturation density, IL-3 titers remained below 100 units/mL. After reaching saturation density, IL-3 titers more than doubled in one day as shown in Fig. 25.19. The culture was then chilled (410°C) prior to transfer to a continuous-flow centrifuge (Alfa-Laval). The supernatant was collected at a flow rate of 1 L/min (RCF of 8000*g*) in a chilled 200-L stainless steel holding tank. About 560 g/L of ammonium sulfate was used to precipitate IL-3 by mixing for 1 hr, followed by recirculation for 1215 hr. The precipitate was recovered using a continuous-flow centrifuge (Cepa) at a flow rate of 0.5 L/min and at 17,000*g*. The precipitate was then re-suspended and dialyzed prior to column purification. In a second batch of recovery, 123 L of supernatant was concentrated using an ultra-filtration unit equipped with 10 k MW membrane (Millipore Pellicon) to 7.14 L at overall recovery yield of 81 percent. The concentrate then went on to the first purification column.

Fig. 25.18 Schematic of an interleukin-3 production system. (Klein, F. et al., "Interleukin 2 and Interleukin 3: Suspension Cultures of Constitutive Producer Cell Lines," *Advances in Biotechnological Processes*, Mizrahi (Ed.), 1984, Copyright Wiley-Liss, Inc., 1984. This material is used by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.)

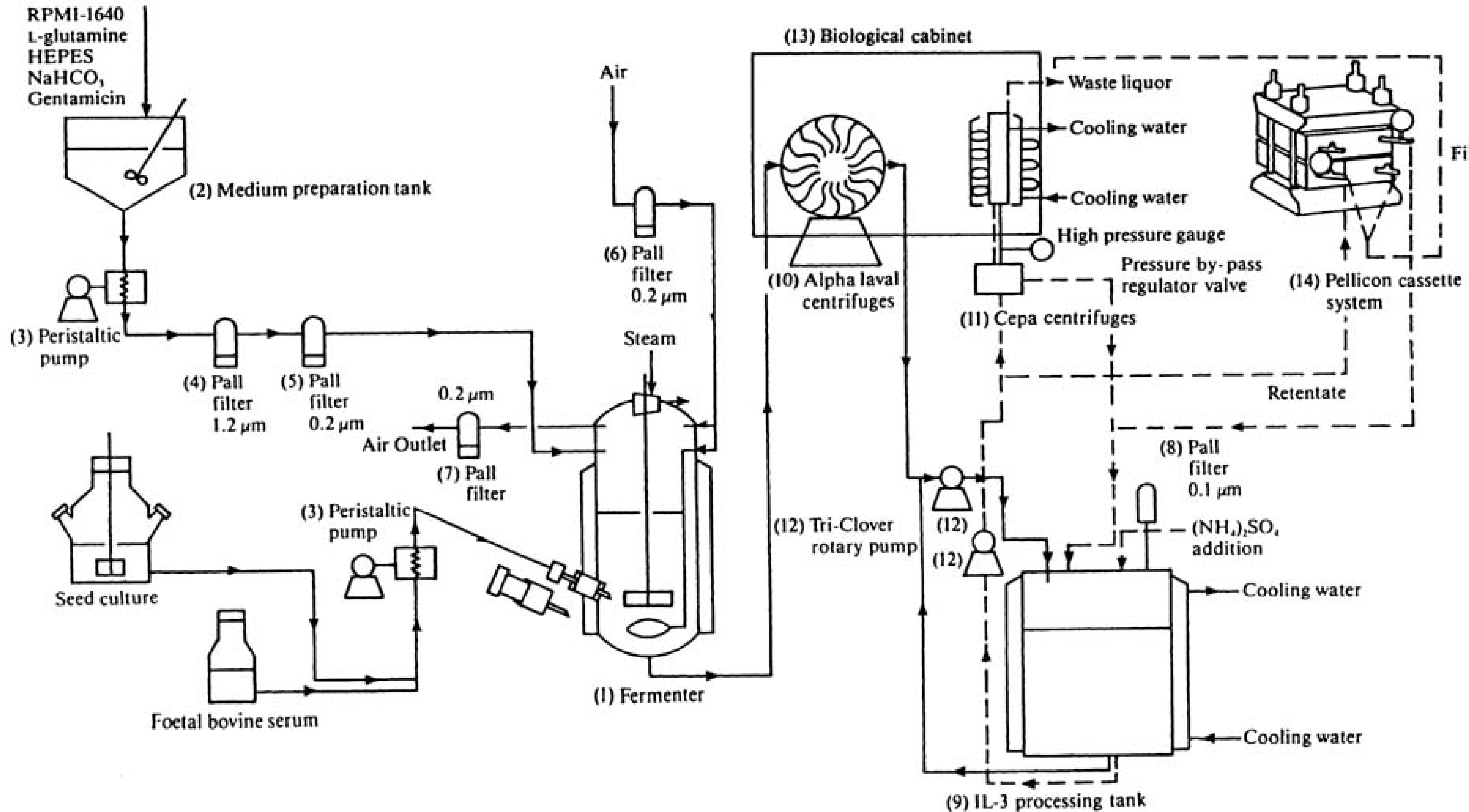
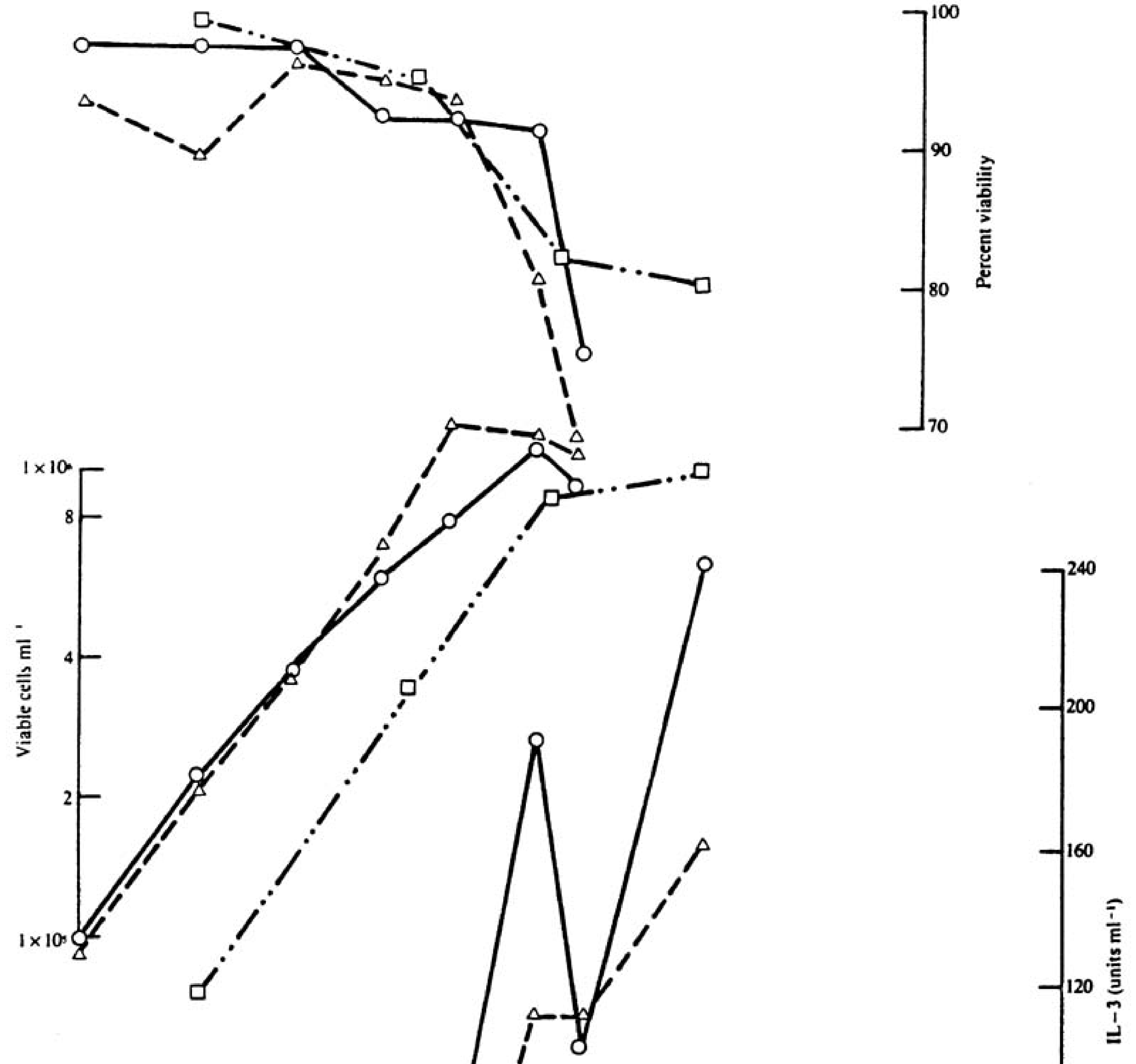
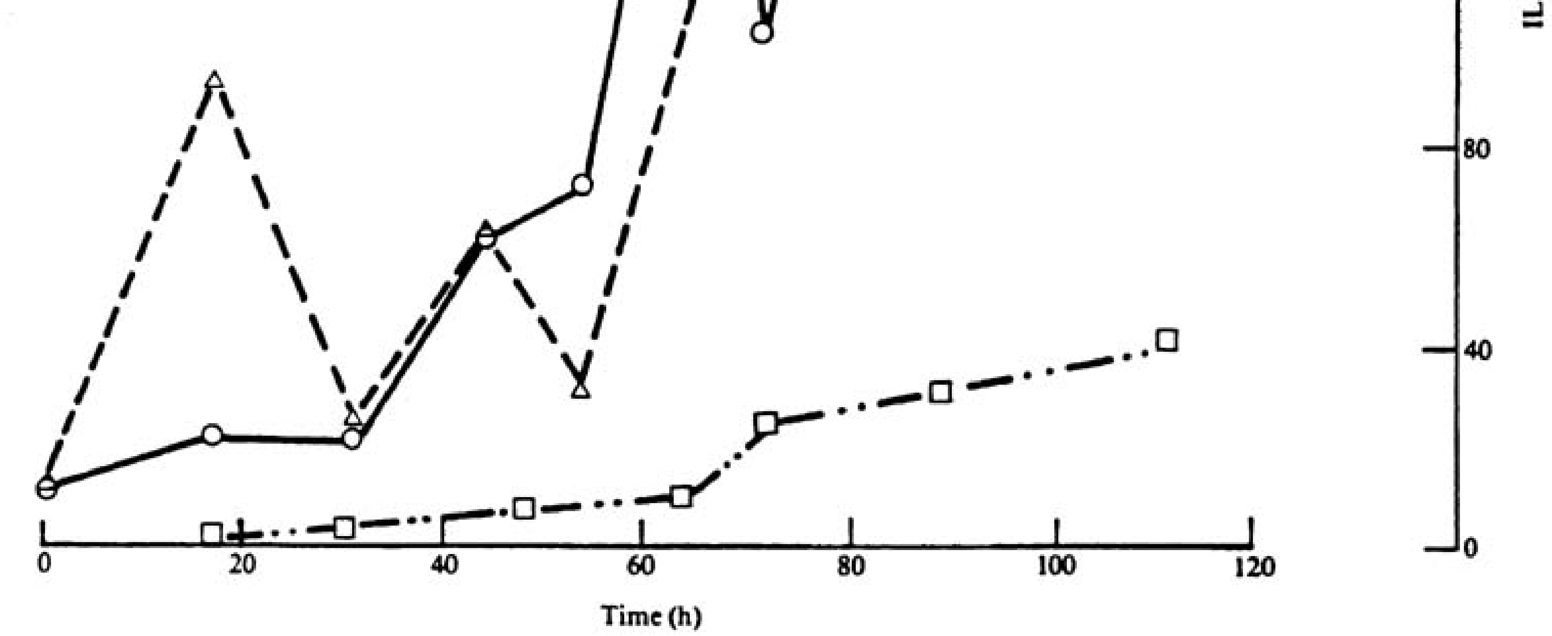


Fig. 25.19 Cell growth and IL-3 production profiles from three different runs. (Klein, F. et al., "Interleukin 2 and Interleukin 3: Suspension Cultures of Constitutive Producer Cell Lines," *Advances in Biotechnological Processes*, Mizrahi (Ed.), Copyright © 1884, Wiley-Liss, Inc. This material is used by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.)




IL
80
40
0
0
20
40
60
80
100
120
Time (h)

25.4 Regulatory Considerations

The biopharmaceuticals resulting from modern biotechnology (most importantly, recombinant DNA and hybridoma technologies) are large proteins (with a molecular weight easily over 10,000 Da) while the conventional synthetic drugs are small molecules with a molecular weight usually less than 1000. A major difference between these two classes of products is that it is much more difficult to determine the purity and structure of the former than the latter. So, the quality assurance of biopharmaceuticals has to rely largely on the stringent control of production cultures and processes, in addition to identity, purity, and potency tests of the products. Several guidelines have been published by CPMP31 since 1989 and by CBER6 since 1985. Foulkes and Traynor45 outlined the development of the regulatory environments in the United States and in Europe. They also discussed the current regulatory processes and the latest initiatives from these governing bodies.

25.5 Glossary

1. *Anchorage-dependent (monolayer) cells*: cells that need to anchor or attach to a solid surface for growth.
2. *Expression*: conversion of genetic (DNA) information into protein.
3. *Expression vector*: a DNA vehicle which carries a particular gene into cells in which the gene is expressed.
4. *Hybridoma*: a hybridized cell resulted from the fusion of an antibody producing cell and a continuous, immortal cell such as a myeloma.
5. *Osmolarity*: a measurement of osmotic pressure.
6. *Perfusion*: a mode of operation in which cells are completely or partially retained inside the bioreactor.
7. *Tissue culture*: cultivation of cells derived and isolated from mammalian, insect, or plant tissue.
8. *Transcription*: a step of the expression process in which RNA polymerase catalyzes the synthesis of mRNA (messenger RNA) using DNA as a template.
9. *Translation*: a step of the expression process in which the information contained in mRNA is used to direct the synthesis of protein, a polymer of peptides.

25.6 References

O'Reilly, D., Miller, L., and Luckow, V., Baculovirus Expression Vectors: A Laboratory Manual, 1994.
Wurm, F., Aspects of Gene Transfer and Gene Amplification in Recombinant Mammalian Cells, Mammalian Cell Biotechnology in Protein Production, pp. 87120, Walter de Gruyter, New York, 1997.
Racher, A., Tong, J., and Bonnerjea, J., Biotechnology, p. 247, Vol. 5a, Wiley-Vch, New York, 1999.
Invitrogen, Carlsbad, CA, 92008 USA.
Wiebe, C., and Boss, M., "Cell Banking," Bioprocess. Technol., 10, 147160 (1990).
Center for Biologics Evaluation and Research, FDA, Bethesda, MD, USA.
International Conference on Harmonization established in 1990.
Barnes, D., and Sato, G., Cell, 22, 649 (1980).
Maurer, H., Toward Chemically-defined, Serum-free Media for Mammalian Cell Culture, Animal Cell Culture, IRL Press, Oxford, 1994.
Wagner, R., Metabolic Control of Animal Cell Culture Process, Mammalian Cell Biotechnology in Protein Production, pp. 193232, Walter de Gruyter, New York, 1997.
Fu, P-C., Europa, A., Gambhir, A., and Hu, W. S, "Metabolic Flux Distributions in Hybridoma Cells at Different Metabolic States, Animal Cell Technology: Challenges for the 21st Century," Proceedings of JAACT and ESACT, pp. 5155, 1999.
Glacken, M., Biotechnology, 6, 1941 (1988).
Glacken, M. et al., Biotech. Bioeng., 32, 491 (1988).
Wagner, R. et al., Citotechnology, 1, 145 (1998).
Miller, W. et al., Biotech. Bioeng., 32, 947 (1998).
McQueen, A., Biotech. Bioeng., 35, 897 (1990).
Griffiths, B., Scaling-up of Animal Cell Cultures, Animal Cell Culture, IRL Press, Oxford, 1995.
Prokop, A., and Rosenberg, M., "Bioreactor for Mammalian Cell Culture," Advances in Biochemistry Engineering/Biotechnology, Vol. 39, Springer, Berlin, 1989.
Tokashiki, M., and Yokoyama, S., "Bioreactor Designed for Animal Cells," Mammalian Cell Biotechnology in Protein Production, pp. 87120, Walter de Gruyter, New York, 1997.
Lubiniecki, A., Advances in Animal Cell Biology and Technology for Bioprocesses, Butterworths, Guildford, 1989.
Scheper, T., Freitag, R., and Scrienc, F., "On-line and Off-line Process Analysis," Mammalian Cell Biotechnology in Protein Production, pp. 372410, Walter de Gruyter, New York, 1997.
Phi-Wilson, J. et al., American Biotechnology Laboratory, 2001, May.
Compton, B., "On-line Monitoring of Bioprocesses using HPLC, Sensors in Bioprocess Control," Bioprocess Technology, Marcel Dekker, New York, 1990.
Kotiaho, T. et al., "Membrane Introduction Mass Spectroscopy," Anal. Chem., 63(18), 875 (1991).
Fernandez, E., Manchuso, A., and Clark, D., "NMR Spectroscopy Studies of Hybridoman Metabolisms in a Simple Membrane Reactor," Biotech. Bioeng., 22, 457 (1980).
Glacken, M., Fleischaker, R., and Sinskey, A., Biotech. Bioeng., 28, 1376 (1986).
Zou, W., and Hu, W., "On-line Characterization of a Hybridoma Cell Culture Process," Biotechnol. Bioeng., 44, 170 (1994).
Blanch, H., and Clark, D., Biochemical Engineering, Marcel Dekker, New York, 1997.
Jungbauer, A., "Principles of Products Extraction from Cell Culture and Purification for Pharmaceutical Proteins," Mammalian Cell Biotechnology in Protein Production, pp. 413451, Walter de Gruyter, New York, 1997.
Amersham Pharmacia Biotech, Process Products Catalog, 2000.
Committee for Proprietary Medicinal Products, European Union.
Molowa, D., and Shenouda, M., "The State of Biologics Manufacturing," Industry Analysis, JPMorgan, 2001.
Wheelwright, S., "Designing Downstream Processes for Large-scale Protein Purification," Biotechnol. Bioeng., 5, 789 (1987).
Wetzel, G., "Medical Applications of Recombinant Proteins in Human and Animals," Biotechnology, pp. 125188, Vol. 5a, Wiley-Vch, New York, 1999.
Arathoon, W., and Birch, J., US Patent No. 5053334 (1991).
Einarsson, M., Brandt, J., and Kaplan, L., Biochim. Biophys. Acta, 830, 110 (1985).
Keen, M., and Rapson, N., Cytotechnology, 17, 153163 (1995).
Backer, M., Metzger, L., Slaber, K., Nevitt, K., and Boder, G., Biotechnol. Bioeng., 32, 9931000 (1988).
Hill, C., Biochem. Soc. Transc., 18, 245247 (1990).
Lebherz, W., "Batch Production of Monoclonal Antibody by Large Scale Suspension Culture", Bioprocess. Technol., 2, 93118 (1987).
Ray, N., Rivera, R., Gupta, R., and Mueller, D., Animal Cell Technology, Carrondo, Griffiths and Moreira (Eds.), pp. 235241, Kluwer Academic, Dordrecht, 1997.
Reuveny, S. et al., Ann. Virol., 133E, 191199 (1982).
Marcus, D., Lasar, A., Reuveny, S., and Traib, A., Biotechnol. Lett., 5(7), 453456 (1983).
Klein, F. et al., "Interleukin 2 and Interleukin 3: Suspension Cultures of Constitutive Producer Cell Lines," Advances in Biotechnological Processes, Alan Liss, New York, 1983.
Foulkes, L., and Traynor, G., "Regulation of Antibodies and Recombinant Proteins," Biotechnology, pp. 495516, Vol. 5a, Wiley-Vch, New York, 1999.
Shuler, M., Wood, H., Granados, R., and Hammer, D., Baculovirus Expression Systems and Biopesticides, Wiley, New York, 1995.

26.1 Introduction

This chapter will discuss the role of chemistry within the pharmaceutical industry. Although the focus will be upon the industry within the United States, much of the discussion is equally relevant to pharmaceutical companies based in other first world nations such as Japan and those in Europe. The major objective of the pharmaceutical industry is the discovery, development, and marketing of efficacious and safe drugs for the treatment of human disease. Of course drug companies do not exist as altruistic, charitable organizations but like other share-holder owned corporations within our capitalistic society must achieve profits in order to remain viable and competitive. Thus, there exists a conundrum between the dual goals of enhancing the quality and duration of human life and that of increasing stock-holder equity. Much has been written and spoken in the lay media about the high prices of prescription drugs and the hardships this places upon the elderly and others of limited income. Consequently, some consumer advocate groups support governmental imposition of price controls, such as those that exist in a number of other countries, on ethical pharmaceuticals in the United States.

However the out-of-pocket dollars spent by patients on prescription drugs must be weighed against the more costly and unpleasant alternatives of surgery and hospitalization, which are often obviated by drug therapy. Consideration must also be given to the enormous expense associated with the development of new drugs. It can take 10 years or more from the laboratory inception of a drug to its registrational approval and marketing at an overall cost which is now $600800 million dollars and increasing. Only 1 out of 10 to 20,000 compounds prepared as drug candidates ever reach clinical testing in man and the attrition rate of those that do is >80 percent. The expense of developing a promising drug grows steadily the further through the pipeline it progresses; clinical trials can be several orders of magnitude more costly than the preclinical evaluation of a compound. While the sales of successful drugs that run the gauntlet and reach the shelves of pharmacies can eventually recoup their developmental expenses many times over, the cost of the drugs that fail is never recovered.

To a large extent, the difficulties associated with bringing a drug to market have arisen from the increasingly stringent but appropriate criteria that have been imposed by the Food and Drug Administration (FDA) in the United States and analogous regulatory agencies in other countries. It is unlikely that an occurrence like that of the thalidomide disaster, which resulted in horrible birth defects several decades ago, would happen again today. Furthermore the era of easy approval of "me-too" drugs is long past. During this era, which prevailed until the final two decades of the past century, it was possible to gain approval for drugs which, although they fell outside the scope of the patents covering a particular marketed drug, offered little advantage over the marketed agent. It is now necessary for a company to demonstrate that a drug, for which a New Drug Application (NDA) is submitted to the FDA, affords significant benefits in terms of efficacy and/or safety relative to the existing drug therapy. The approvability bar may be lowered for agents aimed at the treatment of life-threatening maladies such as cancer and AIDS or for those such as Alzheimer's disease where no effective therapy currently exists; but even in these cases it is incumbent upon the sponsoring company to provide compelling empirical evidence that their drug is safe and effective. The restrictions imposed by Health Mainten-ance Organizations (HMOs) can also have significant impact on the sales of any given drug. Most HMOs list only a select few drugs, for which they will cover costs, within any given category, such as antidepressants, antihypertensives, or cholesterol-lowering agents.

A major consequence of the financial and logistical impediments to the successful introduction of new drugs has been the high incidence of mergers and acquisitions among U.S.-based pharmaceutical companies in the recent past. These events have not occurred because bigger is necessarily better but because the critical mass of internal resources required to bring a drug from the test tube to the pharmacy continues to grow. In contrast to this trend among the major drug companies (often dubbed "big pharma") there has been a proliferation of start-up companies often founded by entrepreneurial scientists with "big pharma" or academic experience and financed by venture capital investment. While many such start ups are strictly bio-techs, others function as mini drug companies and are staffed by both chemists and biologists. Unlike their much larger brethren, the small companies cannot attempt to cover the breath of drug research but instead focus upon a particular therapeutic area and perhaps even a particular disease. Their mission is to discover drug candidates, which a large company may be interested in licensing and developing. The "big pharma" companies do not rely exclusively upon filling their developmental pipelines with drug candidates that have been discovered in-house but often enter into collaborations and licensing agreements to acquire the rights to promising agents from the labs of smaller companies or academic researchers.

26.2 Medicinal Chemistry

Chemistry has long been an integral part of the pharmaceutical industry and its importance should not diminish. Many currently marketed drugs such as the antineoplastic agent, paclitaxel, and the antibiotic, vancomycin, are natural products. The extracts of plants and marine organisms and the products of soil bacteria fermentation will continue to be investigated as potential sources of powerful new drug substances. Chemists are certainly involved in this arena of drug discovery as they conduct the painstaking isolation, purification, and structural characterization of pharmacologically active components which most often are present in minute amounts in the natural source and which have extremely complex chemical structures. The enormous advances in molecular biology have resulted in the successful development of bioengineered therapeutic agents, for example, human insulin, Herceptin (Genentech drug for breast cancer), and Enbrel (Immunex drug for rheumatoid arthritis). It is anticipated that many other biomolecules may be forthcoming for the treatment of human disease.

However the great majority of existing drugs are small organic molecules (MW •200600) that have been synthesized by medicinal chemists. There is no reason to doubt that most drugs of the future will also fall in this category. It is thus important to define what is meant by "medicinal chemist" and what role is played by the practitioners of this sub-discipline in the pharmaceutical industry. A traditional and perhaps somewhat narrow definition of medicinal chemist is that of a researcher engaged in the design and synthesis of bioactive molecules. As part of their academic training many medicinal chemists carried out doctoral and postdoctoral work that involved the total synthesis of natural products and/or the development of synthetic methodology. They are hired by pharmaceutical companies because of the skills they have gained in planning and conducting the synthesis of organic compounds. While such skills can remain important throughout chemists' careers, they alone are insufficient for the challenging task of drug discovery in which, unlike the academic environment, synthetic chemistry is just a means to an end rather than an end in itself. Thus, the enterprising young chemical researcher who enters the industry must be able and willing to undergo an evolution from that of pure synthetic chemist who knows how to make compounds to that of medicinal chemist who also has an insight into what to make and why.

Such insight is gained by acquiring an expanded knowledge base. It is important for the medicinal chemist to know what structural components act as pharmacophores in existing drugs. Pharmacophores, which can be of varying complexity, comprise the essential structural elements of a drug molecule that enable it to interact on the molecular level with a biological macromolecule such as a receptor or enzyme and thus impart a pharmacological effect. The medicinal chemist must become skilled at analyzing the structure activity relationships (SAR) that pertain to the series of compounds on which he/she is working. That is, how does the activity in a biological test of analogs within the series change depending on the introduction of substituents of various size, polarity, and lipophilicity at various domains of the parent drug molecule? Elucidation of the SAR within a series of active compounds is the key to optimizing the potency and other desirable biological properties in order to identify a new chemical entity (NCE) as a bona fide drug candidate. Quantitative structure activity relationships (QSAR) are often employed in this effort; analyses employing linear free energy relationships, linear regression, and other techniques can be utilized to correlate biological activity with the electronic, steric, polarizability, and other physical/chemical parameters of the substituent groups on members of a series of structurally related compounds.

The synthesis and isolation of pure enantiomers has become increasingly important. In the past chiral drugs were most often marketed as racemic mixtures since it was not deemed cost-effective to provide them in enantiomercially pure form. However, in many cases one or the other enantiomers of an optically active drug may have a significantly greater level of the desired biological activity and/or less side effect liability than its antipode. Regulatory agencies such as the FDA now routinely require that each enantiomer of a chiral drug be isolated and evaluated in tests of efficacy, side effects, and toxicity. If one of the enantiomers is shown to be clearly superior then it is likely that it is the form that will be developed as the drug candidate. Thus enantioselective chemical reactions which can afford a high enantiomeric excess(ee) of one or the other of a pair of enantiomers are valuable components of the medicinal chemist's synthetic tools. Enzyme chemistry plays a prominent role in drug R&D since isolated enzymes or microorganisms can often achieve an enantiospecific chemical transformation much more efficiently and economically than conventional synthetic methods. Many "big pharma" companies now have dedicated groups that exclusively study enzymatic reactions.

Research Strategies

The discovery of new drugs may occur by luck or serendipity or as the result of some brilliant insight. However pharmaceutical companies cannot depend on chance occurrences as a research strategy. The aforementioned "me-too" approach has hardly been abandoned and it is likely that the marketing of a novel drug will soon be followed by a number of competitors' agents but with the caveat that the latter offer some therapeutic advantage over the prototype.

The most scientifically sound approach is that of rational drug design, which is based on an understanding of the biochemical mechanisms underlying a particular disease. If, for example, overactivity or underactivity of a certain neurotransmitter system is believed to be responsible for a central nervous system (CNS) disorder such as depression, then medicinal chemists can endeavor to design agents capable of normalizing neurotransmission by their action upon the receptor proteins through which interneuronal communication is mediated. Cloning and expression of human genes to afford functional receptors and enzymes that can be studied in cell culture has been a tremendous advance in the ability to evaluate drug action at the molecular level. Likewise, molecular biology has afforded macromolecules that are essential to the life cycle of pathogens such as bacteria and viruses, thus enabling novel mechanistic strategies for the treatment of infectious disease. In many cases, X-ray crystallography has provided a detailed three-dimensional structure of a macromolecule such as an enzyme with and/or without a bound substrate. Researchers having expertise in computer assisted drug design (CADD) can depict the determined structure on silicon graphics terminals and in collaboration with medicinal chemists can propose drug molecules to fit the active site. Such detailed analysis of protein structure was instrumental in the design of a number of drugs that inhibit HIV protease, an enzyme essential to the integrity of the AIDS virus.

Up until now there have been approximately 1000 human proteins identified as potential targets for drug intervention in various diseases. It is estimated that the determination of the human genome will increase this number by at least tenfold. Therefore, it seems safe to predict that the rational approach to drug discovery will grow accordingly and with it the role of synthetic/medicinal chemistry. There will be intense competition within the pharmaceutical industry to determine the functional relevance of this multitude of new targets in the absence and presence of disease and a close nexus to this quest will be the search for compounds that can impart selective pharmacological effects upon the target proteins. But it is not likely that these goals can be met by employing only the classical iterative approach which entails one-compound-at-a-time synthesis and low volume testing. Instead the challenges of this exciting new era of research must be met by methodologies that can synthesize and test large numbers of compounds in a short period of timethat is, combinatorial chemistry and high-throughput screening (HTS). In the context of its application within pharmaceutical research, combinatorial chemistry should not be regarded as a separate discipline but instead as a technologically specialized part of medicinal chemistry. This topic will be discussed in detail in a later section of the chapter.

Another important interface occurs with chemists in process research and development. In most cases medicinal chemists are not overly concerned with the cost, toxicity, or environmental impact of the starting materials, reagents and solvents they employ to synthesize target compounds since they are dealing with relatively small quantities of materials. Neither are reaction conditions employing very low or elevated temperatures and pressures problematic on the discovery scale. However these and other pragmatic considerations must be taken into account for the bulk scale preparation of experimental drugs. Process chemists must very often modify the synthetic procedures of their medicinal chemistry colleagues and in many cases devise an entirely new synthetic pathway. Process chemistry will also be discussed in an ensuing section.

Pharmacodynamics

Medicinal chemists must be generally knowledgeable about pharmacodynamics, that is, the effect of drugs upon biological systems. In addition to being aware of the state-of-the-art understanding of the biological mechanisms that underlie the particular diseases for which they are endeavoring to discover drug therapy they should know the basis of the various in vitro and in vivo tests that the biologists employ to evaluate both the potential efficacy and side-effect liability of the synthesized compounds. Because drug research covers a plethora of human diseases, each with its own unique combination of etiology and biochemical mechanisms, the number and diversity of biological tests are far too great to discuss in this chapter. Suffice it to say that in a general sense the primary and often even the secondary biological tests of drugs for a particular disease target are in vitro tests that can be run rather quickly, inexpensively, and on small amounts of compound. For example, these can be receptor binding assays for CNS drugs, enzyme assays for antihypertensive agents, inhibition of bacterial colony growth by antibiotics, and the killing of cultured cancer cells by oncolytic drugs. Encouraging in vitro results lead to in vivo testing in some appropriate animal model. In vivo tests are more laborious and costly but are necessary to establish that a drug is effective in an intact living organism; they can range from complex behavioral paradigms for CNS drugs to enhancement of survival time of tumor-implanted mice by experimental cancer drugs. Evaluation of a NCE's propensity to cause side effects is as important as efficacy testing. Even if a compound shows an encouraging level of the desired activity, a lack of selectivity can cause it to induce a number of undesirable pharmacological effects thus precluding its further development. The medicinal chemist must be able to interpret the results of the tests run on his/her compounds and use this information as a guide to further synthetic work.

Pharmacokinetics and Toxicity

It is also necessary that chemists are attuned to various aspects of pharmacokinetics (PK), that is, the effects of biological systems upon drugs. These aspectsabsorption, distribution, metabolism and excretion (ADME)are as critical as biological activity in determining whether a NCE is a viable drug candidate. A compound may exhibit high affinity for a biological receptor or potent inhibition of an enzyme in an *in vitro* assay but if it is poorly absorbed or rapidly metabolized to inactive species then it will be ineffective as a drug. For example, the empirically based Lipinsky's rules of five (Table 26.1) define the limits of such physical/chemical parameters as molecular weight, lipophilicity, and hydrogen bond forming moieties that must be considered for the absorption of orally administered drugs. A compound with potent intrinsic activity can be rendered ineffective in vivo by its rapid conversion to inactive metabolites. The susceptibility of compounds to metabolic conversion can be assessed by incubating them with liver homogenates from various species including rodent, dog, monkey, and man or with cloned, expressed human hepatic enzymes. Analysis of the incubates by liquid chromatography/mass spectometry (LC/MS) can quantify the extent of metabolism and even identify some specific metabolites. In vivo adminstration of a NCE to one or several animal species is required to determine its oral bioavailability, half-life, and other PK properties such as distribution and elimination. If an unsatisfactory PK profile threatens to be the demise of an otherwise promising drug candidate, it falls upon the medicinal chemist to make structural permutations aimed at correcting the problem. If poor absorption is the problem this may entail modifying the lipophilicity of the drug molecule to render it more membrane permeable. A metabolic liability might be rectified by blocking the site of biotransformation with a metabolically inert atom or group.

TABLE 26.1 Lipinski's Rules for Drug Absorption

Absorption of a drug following oral administration is favored by:
1. Molecular weight is <500
2. The drug molecule has <5 hydrogen bond donors
3. The drug molecule has <10 hydrogen bond acceptors
4. The distribution coefficient, log P, is < 5

Source: Lipinski, C. A., Lombardo, F., Dominy, D. W., and Feeny, P. J., *Adv. Drug Delivery Rev.*, 23, 325 (1997).

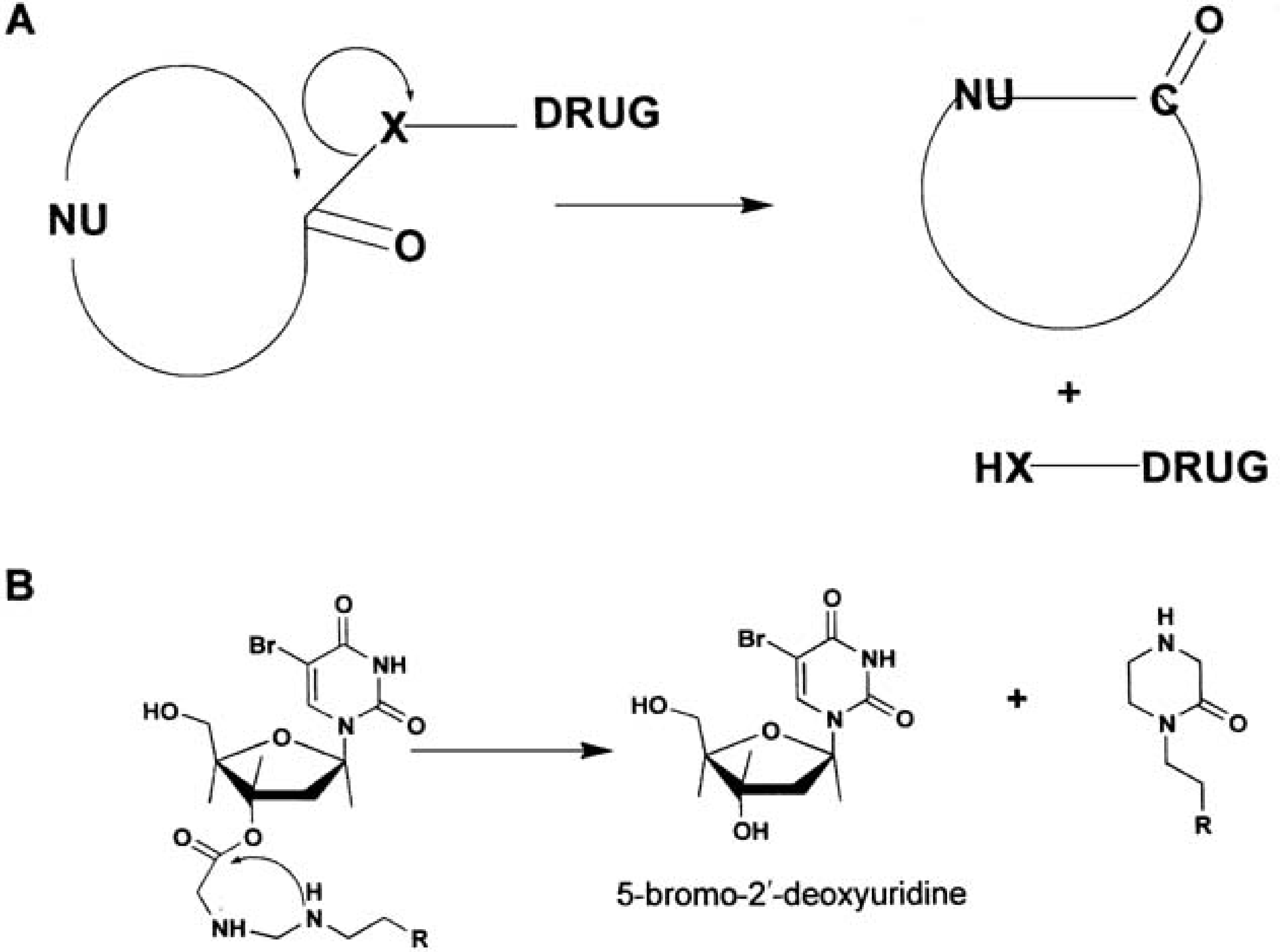
Toxic effects upon blood or organs or the potential to cause gene aberrations will red flag a compound regardless of its having both excellent biological activity and PK properties. Promising lead compounds are screened in in vitro tests in bacteria and mammalian cells to determine whether they cause gene mutations and DNA damage. If they pass this hurdle the compounds are dosed on a daily basis for several weeks to several months in both a rodent and non-rodent (usually dog or monkey) species and the animals are observed for any adverse effects; the test animals are necropsied following conclusion of the study to ascertain whether any organ or tissue damage occurred. Unacceptable toxicological findings will invariably kill a drug candidate and again it is the medicinal chemist who will be called upon to save the day by devising and implementing structural modifications to eliminate the toxicity. This may be a more daunting task than overcoming a side effect or metabolic issue, especially if the toxicity is mechanism-based.

Drug Delivery

Drugs can be administered to patients in many ways. The most common and preferred route is oral administration and oral drugs are generally formulated as tablets or capsules in which a specific dose of the drug substance is homogeneously mixed with some inert filler or excipient. Some oral medications, such as pediatric formulations of antibiotics, are in solution form, as are injectable drugs. Obviously this requires satisfactory solubilization of the drug, preferably in aqueous medium. Compounds bearing some ionizable group such as a basic amine or an acidic function can usually be converted into water-soluble salts but neutral molecules present greater difficulties. In some cases the results of clinical trials will indicate that an experimental injectable drug shows promise of efficacy but does not elicit a robust response because its poor solubility limits the amount that can be administered and thus does not allow adequate plasma levels to be attained. Inadequate membrane permeability can restrict the absorption and bioavailability of an orally administered drug.

Medicinal chemists can respond to such findings by investigating the feasibility of preparing a suitable prodrug. A prodrug is a derivative in which a cleavable solubilizing group is covalently appended to the parent drug molecule, most often via a hetero atom such as oxygen or nitrogen. An effective prodrug is one which has much higher solubility than the parent drug and which following its administration is rapidly cleaved in vivo to achieve a therapeutically beneficial plasma concentration of the parent drug. Figure 26.1 shows a generic and a specific example of intramolecular activation of prodrugs by cyclization-elimination.

Fig. 26.1 A. General reaction scheme for intramolecular activation of prodrug by cyclization-elimination. B. Activation of basic ester prodrug of the radiation sensitizer 5-bromo-2-deoxyuridine by cyclization-elimination. (From Testa, B., and Mayer, J.M., *Drug Metab. Rev.*, 30(4), 792793, Marcel Dekker, Inc., New York, 1998. Reprinted by permission of Marcel Dekker, Inc.)



Patents

Patent protection on both its approved and experimental drugs is of critical importance to a pharmaceutical company. Issued patents provide the company with exclusivity for the manufacture, use, and sale of its drug products and it is highly unlikely that a company would undertake the risks and costs of developing an agent for which it had no patent protection. There are several types of patents of which the "composition of matter" (COM) or "product" patent may be deemed to have the greatest value. An approved COM patent covers specifically claimed compounds of a certain structural chemotype and provides empirical evidence that the claimed compounds have been prepared, characterized, and found to have some utility. In order to be patentable the compounds must have structural novelty and cannot have been publicly disclosed either in the scientific or patent literature or by a presentation. But structural novelty alone is not sufficient grounds for a patent; it must be demonstrated that the compounds are useful and in the context of a drug patent the proposed utility is for the treatment of some disease. The basis of such utility is activity in appropriate and relevant biological tests. Clinical data may also be used in support of a patent application although in the great majority of cases the applications are filed well before any compound within the application reaches clinical trials. Medicinal chemists are closely affiliated with the patent process and are most commonly the inventors listed on COM patents covering drug substances. The chemists and other researchers with whom they collaborate must provide the chemical and biological data for the patent and the chemists will also provide input as to the scope and claims of the patent. Since patents are legal documents that provide the assignee exclusive proprietary rights to the covered subject matter for 20 years from the date of the patent's issue, it is essential that all supportive data be accurate and instructive. If a patent is ever challenged by another party and is found to contain erroneous information then it could be invalidated. Moreover, in the United States, patents are granted on a "firstto-invent" basis. Thus if two or more parties submit applications on identical subject matter to the U.S. Patent Office then the patent will be awarded to the party that can prove that it had the earliest conception and reduction to practice of the subject matter. Therefore it is imperative that chemists maintain accurate records of all experimental work in a bound notebook and that such records are dated, signed, and witnessed.

Other types of drug related patents include process, use, and formulation patents. Chemists are responsible for process patents, which describe an improved method of preparation of some drug substance but are minimally involved with the others. Use patents are based on the discovery of some unobvious utility of a compound that is either part of the public domain or covered by an existing patent; such discoveries are most likely to be made by biologists. Formulation patents disclose a preferred means of drug delivery of a known drug substance.

Clinical Trials

Even though there is no involvement of chemistry in the clinical evaluation of drugs, any discussion of the pharmaceutical industry must include clinical trials for the results of such trials determine whether or not an experimental drug has the combination of efficacy, safety, and tolerability which will allow it to achieve registrational approval and reach the market. If a drug candidate survives the hurdles of pharmacological, pharmacokinetic, and toxicological testing, the next customary step in the United States is the sponsoring firm's filing of an Investigational New Drug (IND) application with the FDA. This is a formal request to initiate clinical investigation in man and is accompanied by a detailed description of the planned studies and clinical protocols. Upon approval of the IND, Phase I clinical studies are initiated.

Phase I studies are conducted in healthy volunteers in order to establish the drug's safety and to determine appropriate dosage levels. If the drug is found to have an acceptable human pharmacokinetic profile and to be free of untoward side effect liabilities, it is advanced into Phase II trials, which are typically carried out in several hundred patients and may last from 6 months to 2 years. Phase II trials are designed to ascertain the appropriate dosing regimen for the drug and whether it is effective in treating the target disease. Only about one third of drugs pass Phase II trials, most failing because of the lack of efficacy. Those that pass are advanced into Phase III trials which may involve from several hundred to several thousand patients and which can last from one to three years or even longer depending on the type of drug under study and the complexities of the study design. Phase III trials provide the ultimate test of an experimental drug since they are designed to verify the drug's effectiveness against the target disease as well as its safety. For agents that are intended for chronic use, studies also monitor adverse reactions that may develop only after long-term use and the development of tolerance. Clinical studies of many drug classes will commonly employ several patient groups of approximately equal size with one group receiving the experimental drug, another placebo (non-drug), and another a positive control, that is, a marketed drug used to treat the same disease for which the experimental agent is being evaluated. In order to minimize the possibility of bias in favor of the test drug, such studies are most often run in a double-blinded manner with neither patients nor clinical investigators knowing which group is receiving which treatment until the conclusion of the trial.

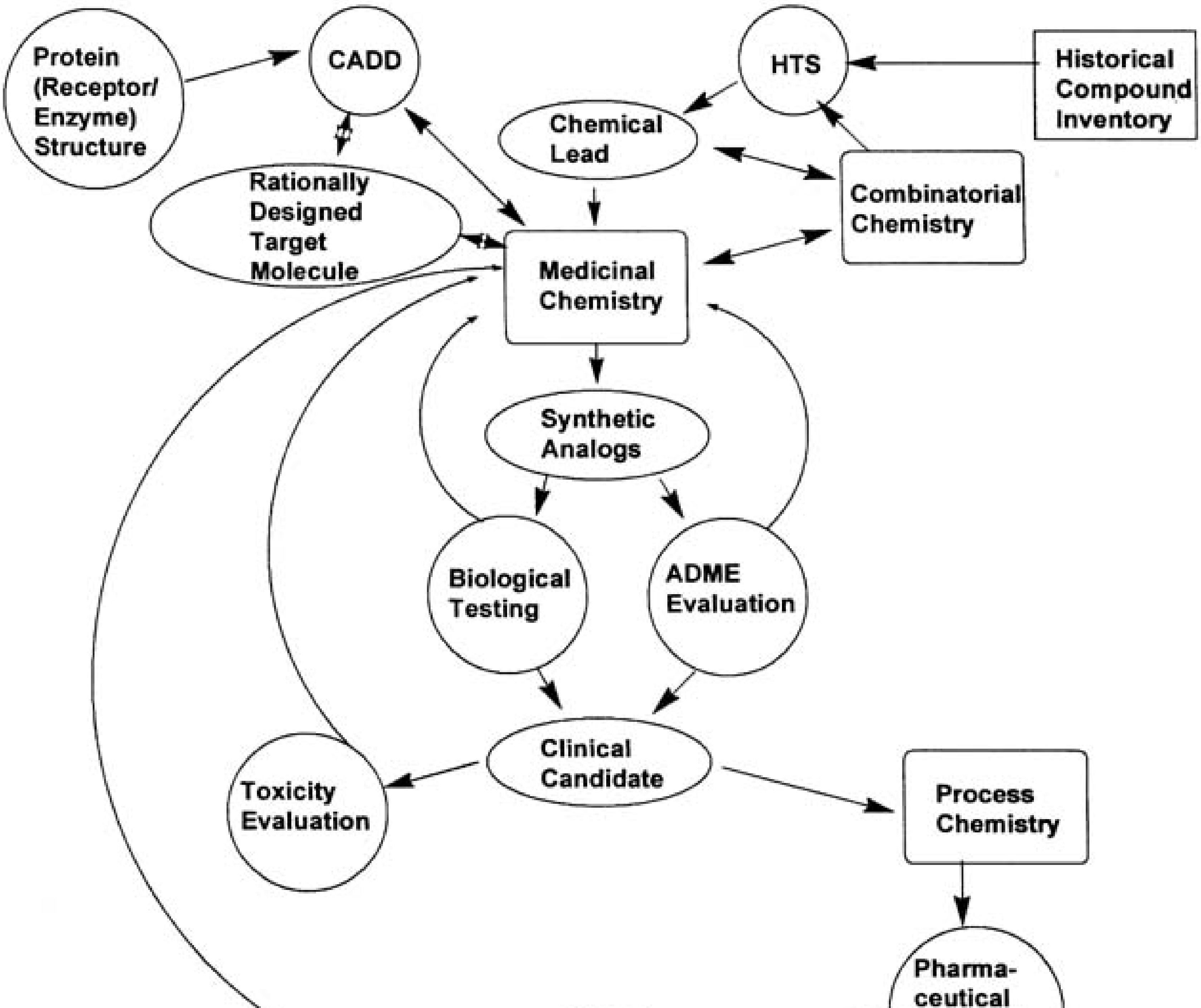
If a drug candidate is among the one in four to five that gets through Phases IIII and if statistical analysis of the clinical data supports its efficacy then the sponsoring firm will assemble the voluminous data into the NDA which is submitted to the FDA. Review of the NDA can take 12 years and often the FDA may request that additional information be provided or even that some additional studies be done. When approval is granted the company is then free to market the drug.

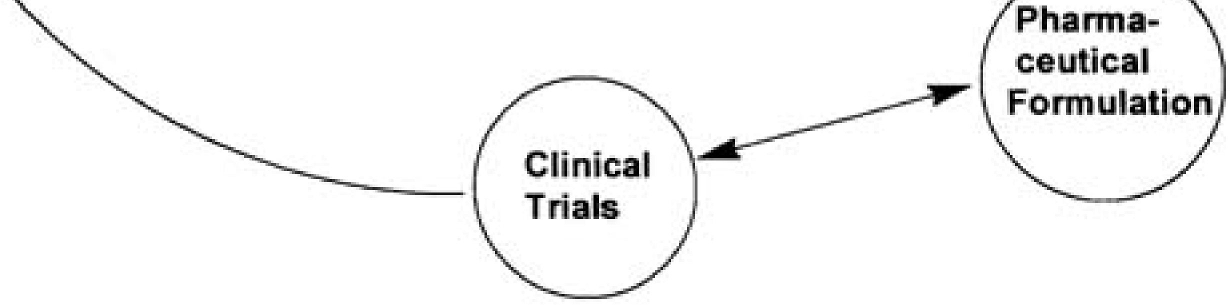
The results of clinical evaluation of an experimental drug can feedback into medicinal chemistry. For example, if a drug is found to fail because of poor bioavailability in humans then medicinal chemists will endeavor to design and prepare an analog with improved pharmacokinetic properties.

Summary

The preceding sections present what is an admittedly superficial overview of the very extensive and complex topic of medicinal chemistry, its role in the pharmaceutical industry, and its interface with other disciplines. An acquired understanding of relevant biology, pharmacology, toxicity, and so on is not just of heuristic value but is necessary for the chemist to engage in meaningful dialog with their colleagues who work in these specialties. Successful drug discovery and development cannot be done by individuals working in isolation but requires the interactive collaboration of many researchers representing a multiplicity of scientific disciplines as depicted in Fig. 26.2. It may be argued that medicinal chemists are the most versatile generalists among these researchers in that they must have primary expertise in chemistry along with extensive knowledge of numerous other areas.

Fig. 26.2 Drug discovery and development is a complicated process that involves the interaction of researchers in various disciplines. Medicinal chemists may synthesize analogs based on chemical leads arising from the high throughput screening of combinatorial libraries or historical compound inventories. Alternatively, analog synthesis can be based on the collaboration between medicinal chemistry and computer assisted drug design to rationally design small molecules capable of interacting with a macromolecular biological target (receptor or enzyme). Subsequent biological, pharmacokinetic, and toxicological evaluations lead to identification of a drug candidate that, following development of a suitable bulk scale synthesis by process chemistry and pharmaceutical formulation, is advanced into clinical trials. Feedback to medicinal chemistry from any of these developmental steps can give rise to further synthetic modifications and refinements.



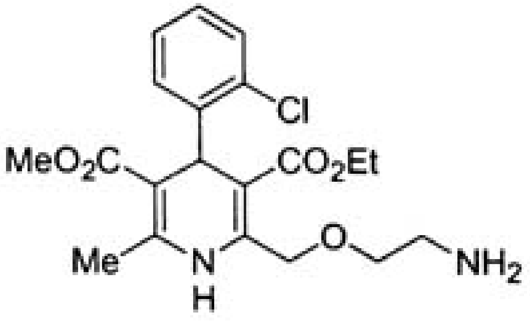


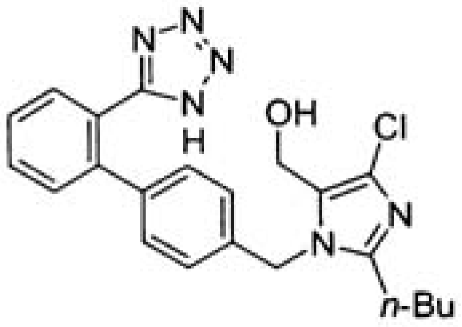
The following section presents examples of marketed drugs in a number of different therapeutic categories.

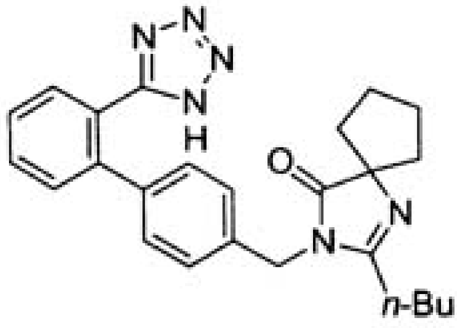
26.3 Cardiovascular Agents

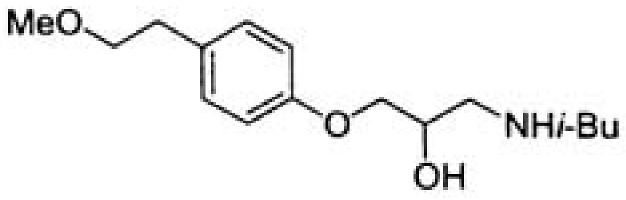
Hypertension

A variety of agents of several mechanistic types are currently available for the treatment of hypertension (elevated blood pressure). The dihydropyridine derivative amlodipine (Norvasc®/Pfizer) is a receptor operated, calcium entry blocker which prevents Ca++ entry into vascular smooth muscle cells. Amlodipine is also useful for the treatment of angina. Losartan (Cozaar®/Merck) and irbesartan (Avapro®/Bristol-Myers Squibb) are angiotensin receptor antagonists which inhibit the action of angiotensin II on the AT1 receptor. Metoprolol (Toprol®/AstraZeneca) is a cardioselective, β1-adrenergic receptor blocking agent and is also useful in the treatment of angina.


amlodipine

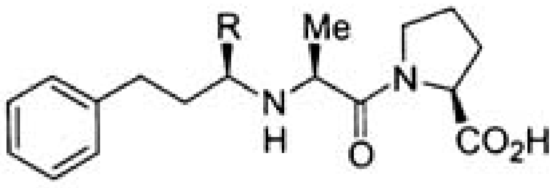

losartan

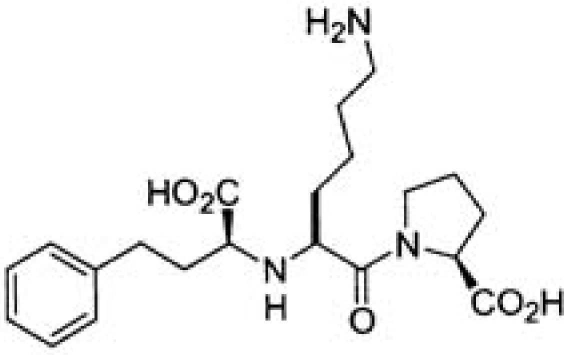

irbesartan

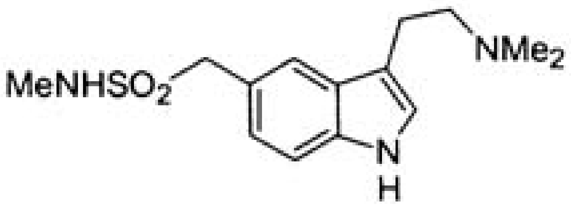

metoprolol

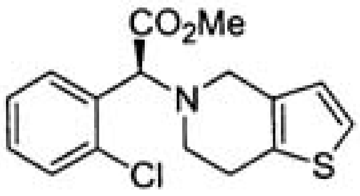
Congestive Heart Failure, Migraine, and Thrombolytic Agents

Enalapril (Vasotek®/Merck) and lisinopril (Zestril®/AstraZeneca and Prinvil®/Merck) are angiotensen-converting enzyme (ACE) inhibitors and useful in the treatment of congestive heart failure and hypertension by suppression of the reninangiotensenaldosterone system. Enalapril is an ethyl ester prodrug, which is hydrolyzed in the liver to the active carboxylic acid enalaprilat. Sumatriptan (Imitrex®/Pfizer) is a selective agonist of serotonin (5hydroxytryptamine) type-1 receptors (most likely the 5-HT1B and 5-HT1D subtypes) in the vasculature. It is thought to exert its beneficial effects on migraine headaches by selective constriction of certain large cranial blood vessels and/or possibly through suppression of neurogenic inflammatory processes in the central nervous system. Clopidogrel (Plavix®/Bristol-Myers Squibb, Sanofi-Synthelabo) is an inhibitor of ADP-induced platelet aggregation and is useful in the treatment of various thrombolytic events such as stroke and myocardial infarction.


enalapril, R = CO_2Et
enalaprilat, R = CO_2H


lisinopril

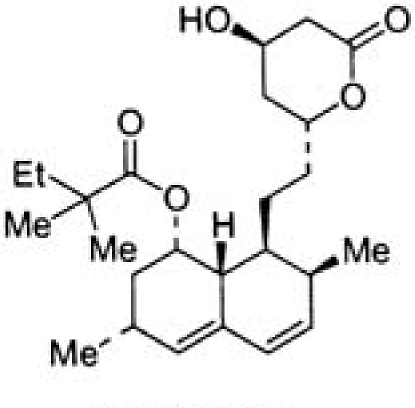

sumatriptan


clopidogrel

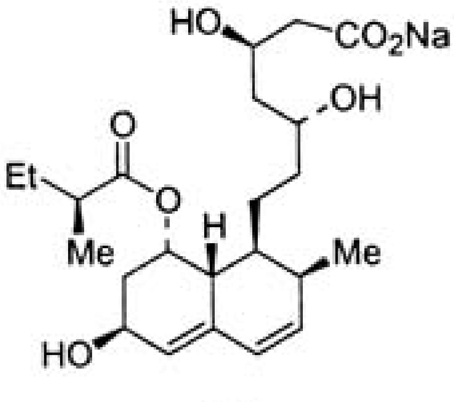
26.4 Metabolic Agents

Hyperlipidemia

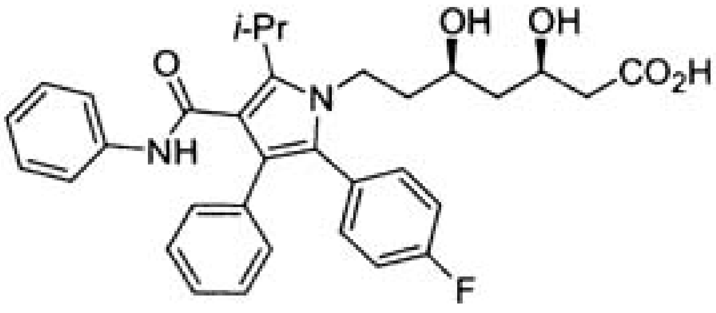
Simvastatin (Zocor®/Merck), pravastatin (Pravachol®/Bristol-Myers Squibb), and atorvastatin (Lipitor®/Pfizer) are hydroxymethylglutaryl-CoA (HMG-CoA) reductase inhibitors (statins) which lower serum lipid levels by inhibiting cholesterol biosynthesis. Simvastatin and pravastatin are semi-synthetic, mevalonic acid-derived antilipidemic agents while atorvastatin is a wholly synthetic, pentasubstituted pyrrolo-heptanoic acid. Unlike pravastatin and atorvastatin, simvastatin is a lactone prodrug which must be converted to the corresponding, ring-opened δ-hydroxy acid in vivo.



simvastatin



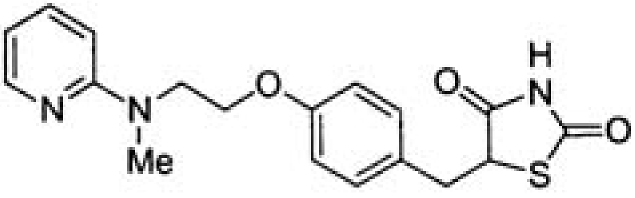
pravastatin



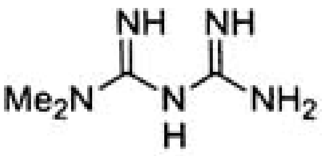
atorvastatin

Diabetes

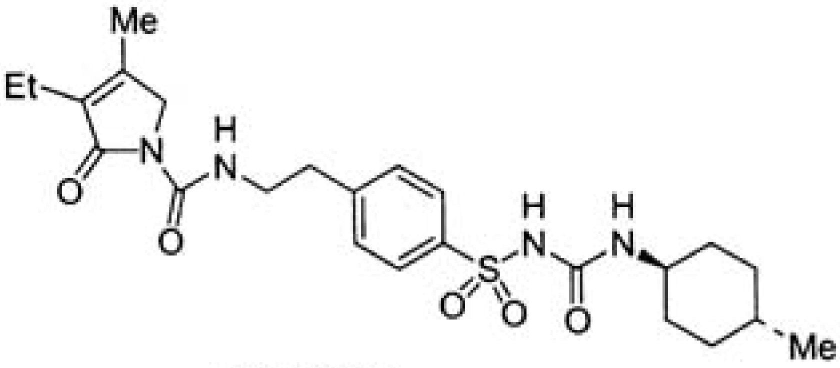
A variety of mechanistic agents are currently available for the treatment type 2 (noninsulin-dependent) diabetes mellitus (NIDDM). Rosiglitazone (Avandia®/GlaxoSmithKline) is a thiazolidinedione (glitazone) antidiabetic agent and an agonist at the peroxisome proliferator-activated receptorgamma (PPARgamma). Activation of this receptor enhances insulin sensitivity in target tissues by increasing insulin-responsive gene transcription. Metformin (Glucophage®/Bristol-Myers Squibb) is an antihyperglycemic agent, which improves glucose tolerance in patients with type 2 diabetes. The compound acts by decreasing both hepatic glucose production and intestinal absorption of glucose, and improves insulin sensitivity by increasing peripheral glucose uptake and utilization. Glimepiride (Amaryl®/Aventis) is in the sulfonylurea class of antidiabetic agents. Glimepiride is thought to lower blood glucose concentration by stimulating insulin secretion in pancreatic beta cells.



rosiglitazone



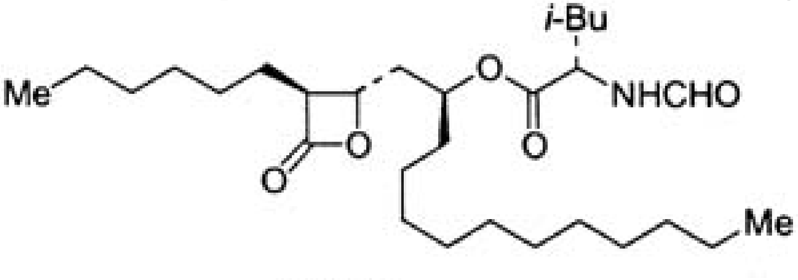
metformin



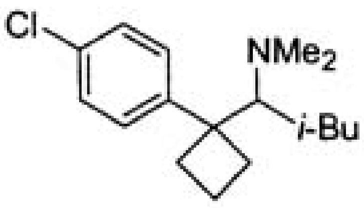
glimepiride

Obesity

Orlistat (Xenical®/Roche) is a reversible gastric and pancreatic lipase inhibitor. The compound has no effect on appetite suppression but rather acts by inhibiting dietary fat absorption from the GI tract. Sibutramine (Meridia®/Abbott) and its major active metabolites are reuptake inhibitors of norepinephrine, serotonin, and dopamine and exerts its beneficial effect through appetite suppression.



orlistat

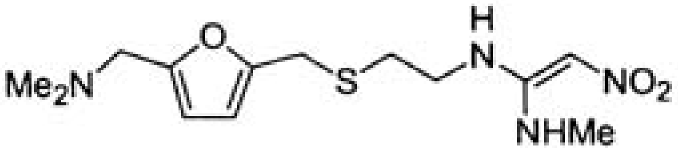


sibutramine

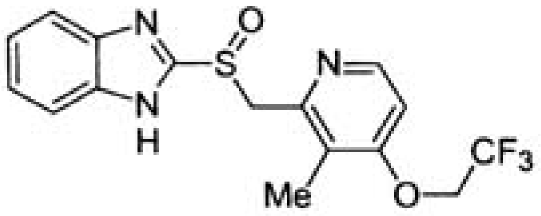
26.5 Gastrointestinal and Genitourinary Agents

Antisecretory

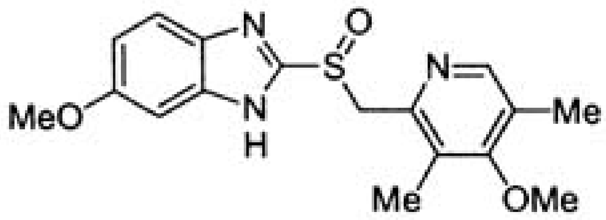
Ranitidine (Zantac®/GlaxoSmithKline) is a histamine H2-receptor antagonist which inhibits the release of gastric acid and is useful in the treatment of a variety of hypersecretory conditions [dyspepsia, heartburn, duodenal and gastric ulcers, and gastroesophageal reflux (GERD)]. Lansoprazole (Prevacid®/TAP) and omeprazole (Prilosec®/AstraZeneca) are benzimidazole gastric antisecretory agents and unrelated both chemically and pharmacologically to the H2-receptor antagonists. These agents are known as proton pump inhibitors due to their ability to inhibit the H+K+-ATPase (the proton pump) in gastric parietal cells and block the secretion of hydrochloric acid. Lansoprazole and omeprazole are also useful in the management of duodenal and gastric ulcers, and GERD.



ranitidine



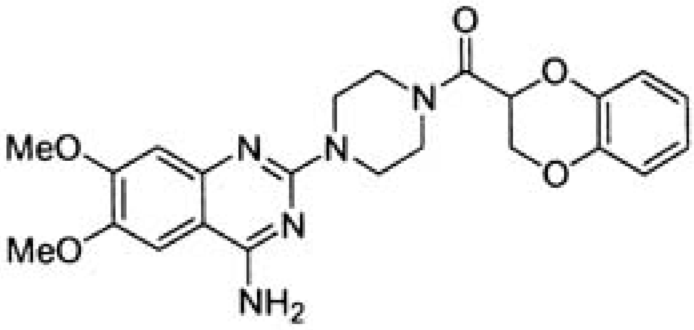
lansoprazole



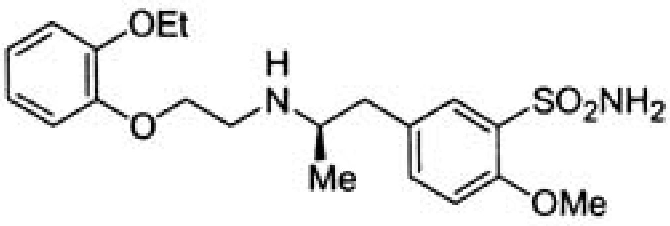
omeprazole

Benign Prostatic Hyperplasia and Urinary Urge Incontinence

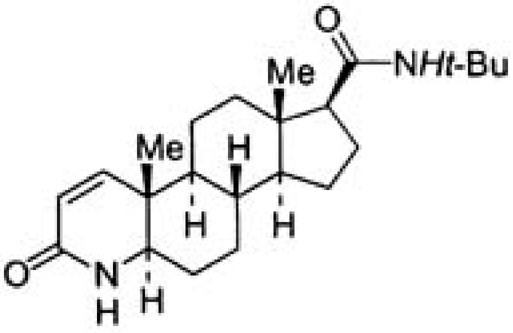
Doxazosin (Cardura®/Pfizer) and tamsulosin (Flomax®/Boehringer Ingelheim) are used in the management of benign prostatic hyperplasia (BPH). Both compounds are postsynaptic, α1-adrenergic blocking agents that relax the prostatic tissue and increase urinary outflow in men. Because tamsulosin demonstrates selectivity for the α1A-adrenergic receptor subtype located in prostate over that of α1B-subtype located in vascular tissue, there is a reduced incidence of cardiovascular side effects (hypotension, dizziness, and syncope). Because doxazosin is not selective for the α1A-subtype, it is also useful in the treatment of hypertension. Finasteride (Proscar®/Merck) is a 5α-reductase inhibitor, which blocks the conversion of testosterone to 5α-dihydrotestosterone (DHT). Because DHT is an androgen responsible for prostatic growth, inhibition of the 5α-reductase enzyme is beneficial in reducing prostatic size associated with BPH. Tolterodine (Detrol®/Pharmacia) is a nonselective muscarinic receptor antagonist which decreases contraction of the detrusor muscle in the overactive urinary bladder. It is useful for relieving the symptoms associated with urinary urge incontinence and urinary frequency.



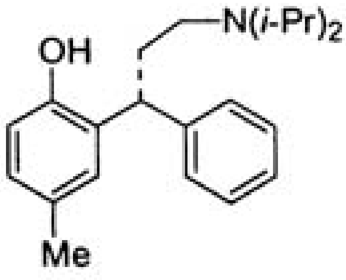
doxazosin



tamsulosin



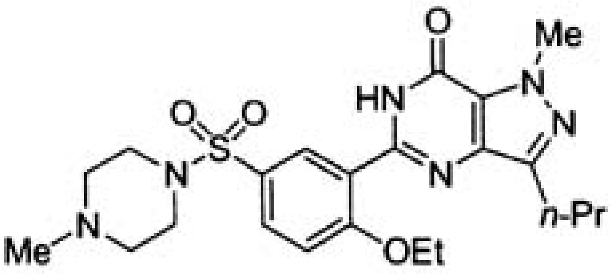
finasteride



tolterodine

Erectile Dysfunction

The pyrazolopyrimidinone derivative sildenafil citrate (Viagra®/Pfizer) is a selective phosphodiesterase (PDE) type 5 enzyme inhibitor. The drug acts by selectively blocking the PDE type 5 isoenzyme ultimately causing vascular vasodilation in the corpus cavernosal tissue which, in turn, leads to penile tumescence and rigidity.

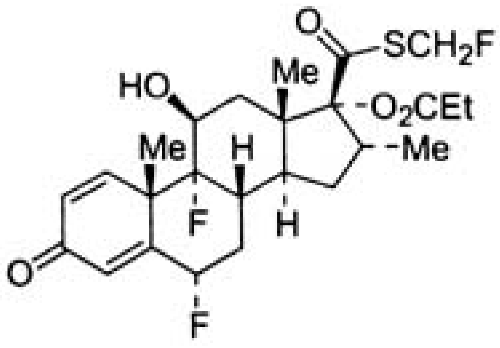


sildenafil

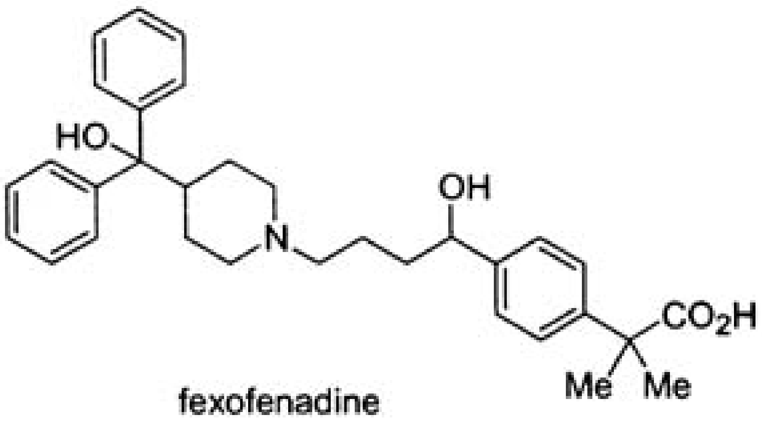
26.6 Pulmonary Agents

Asthma and Allergic Rhinitis

Fluticasone (Flovent®/GlaxoSmithKline) is a synthetic corticosteroid derivative, which is a selective agonist at the human glucocorticoid receptor and useful in the treatment of asthma. Although the precise mechanism of fluticasone in asthma is unknown, it is believed its anti-inflammatory property contributes to its beneficial effect. The butyrophenone derivative fexofenadine (Allegra®/Aventis) is an anti-histamine and used in the treatment of seasonal allergic rhinitis. Because fexofenadine does not readily cross the bloodbrain barrier, it is considered a "nonsedating" anti-histamine.



fluticasone

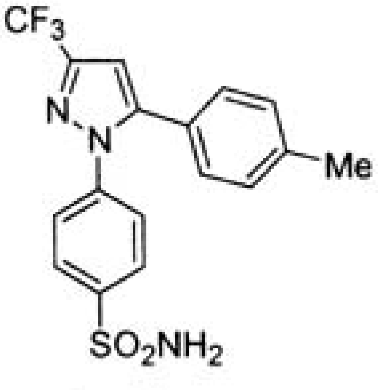


fexofenadine

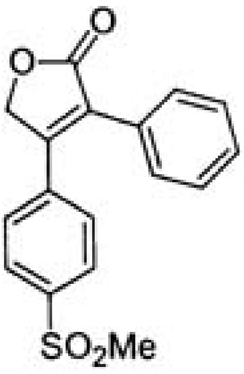
26.7 Inflammation and Osteoporosis

Arthritis

The diaryl pyrazole derivative celecoxib (Celebrex®/Pharmacia, Pfizer) and the furanone rofecoxib (Vioxx®/Merck) are selective cyclooxygenase type 2 (COX-2) inhibitors and are useful in the treatment of arthritis. Both compounds exert their pharmacological effect by selectively blocking the COX-2 enzyme to produce an anti-inflammatory effect without adverse gastrointestinal side effects.



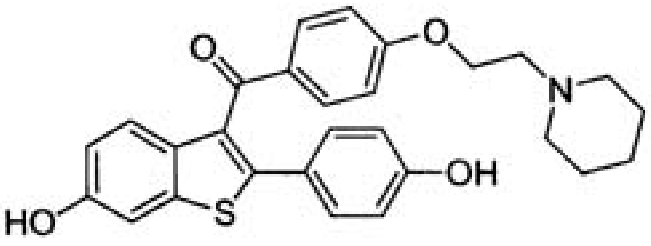
celecoxib



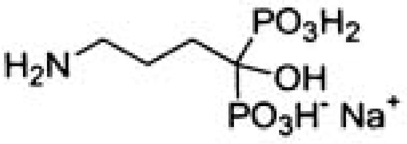
vioxx

Osteoporosis

The benzothiophene derivative raloxifene (Evista®/Lilly) is a selective estrogen receptor modulator (SERM). Raloxifene produces its biological actions via modulation (both activation and blockade) of estrogen receptors, which ultimately results in decreased resorption of bone. The bisphosphonate derivative alendronate (Fosamax®/Merck) is an inhibitor of osteoclast-mediated bone resorption and is also useful in the treatment of osteoporosis. Both raloxifene and alendronate are useful in the treatment of osteoporosis in postmenopausal women.



raloxifene

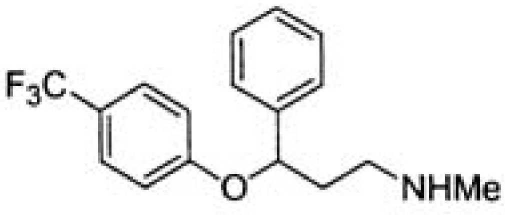


alendronate

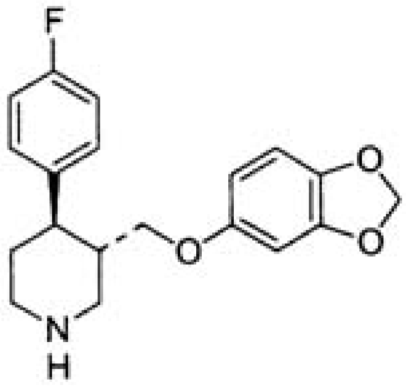
26.8 Central Nervous System Agents

Antidepressants

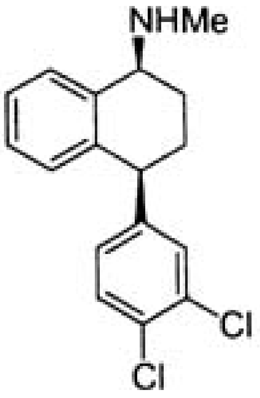
Fluoxetine (Prozac®/Lilly), paroxetine (Paxil®/GlaxoSmithKilne), and sertraline (Zoloft®/Pfizer) are selective serotonin reuptake inhibitors (SSRI's) and are useful in the treatment of depression. These agents potentiate the pharmacologic actions of the neurotransmitter serotonin by preventing its reuptake at presynaptic neuronal membranes. In addition to its SSRI properties, venlafaxine (Effexor®/Wyeth-Ayerst) also appears to be a potent inhibitor of neuronal norepinephrine reuptake and a weak inhibitor of dopamine reuptake therefore enhancing the actions of these neurotransmitters as well. Venlafaxine is indicated for use in anxiety and depression.



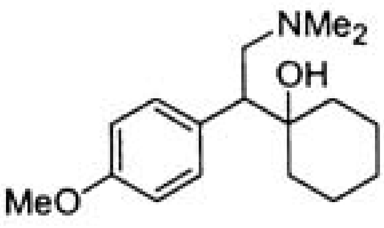
fluoxetine



paroxetine



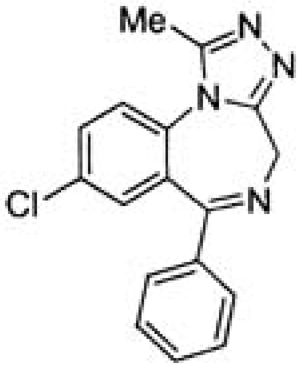
sertraline



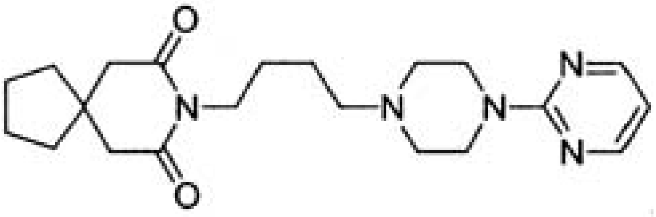
venlafaxine

Anxiolytics

Alprazolam (Xanax®/Pharmacia), a benzodiazepine derivative is used for the treatment of both anxiety and panic disorder while buspirone (Buspar®/Bristol-Myers Squibb) is indicated for the treatment of anxiety disorders. The mechanism of action of buspirone is distinct from that of the benzodiazepines and is believed to be mediated mainly through modulation of serotonergic neurotransmission via interaction with the 5-HT1A serotonin receptor subtype.



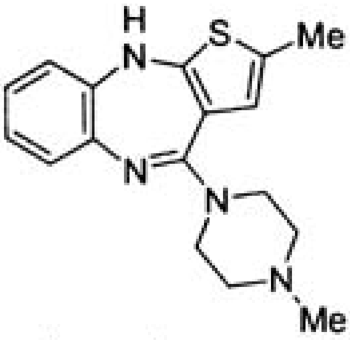
alprazolam

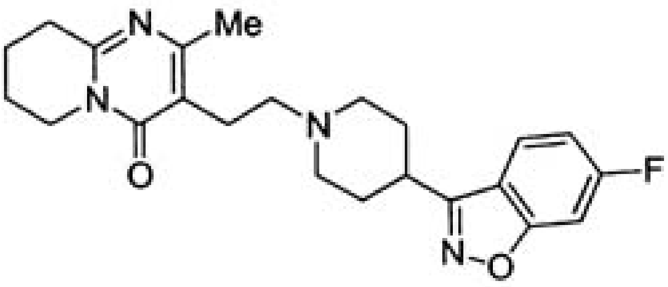


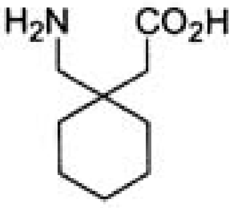
buspirone

Bipolar Disorders, Schizophrenia, and Epilepsy

The thienobenzodiazepine derivative olanzapine (Zyprexa®/Lilly) and benzisoxazole compound risperidone (Risperidal®/Janssen) are atypical antipsychotic agents. Olanzapine is used in the treatment of bipolar disorder while risperidone is useful in the management of schizophrenia. It is believed both compounds exert their beneficial effects through antagonism of serotonergic and dopaminergic receptors. Gabapentin (Neurontin®/Pfizer) is a γ-aminobutyric acid (GABA) analog and useful in the treatment of epilepsy. Although structurally related to GABA, it has no GABA-minergic activity. The mechanism for its anticonvulsive action is currently unknown.

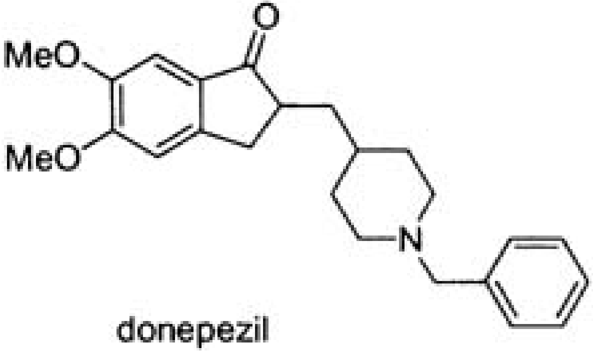

olanzapine


risperidone


gabapentin

Alzheimer's Disease

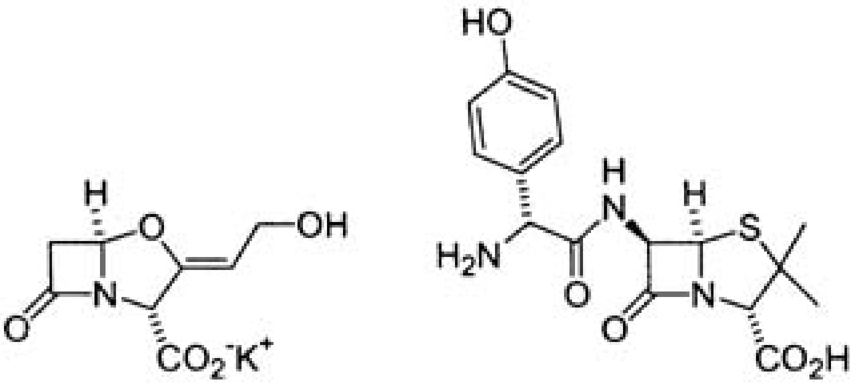
The indanone derivative donepezil (Aricept®/Pfizer, Eisai) is an acetycholinesterase inhibitor and is structurally unrelated to other cholinesterase inhibitors. Because its mechanism increases the concentration of the neurotransmitter acetycholine at cholinergic sites, it is useful in the treatment of Alzheimer's disease (dementia).


donepezil

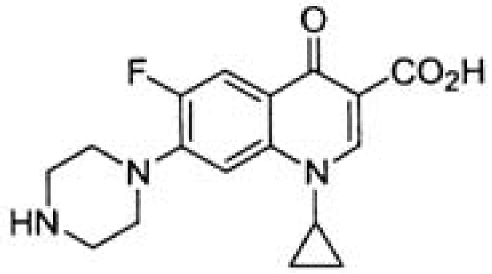
26.9 Infectious Diseases

Antibacterials

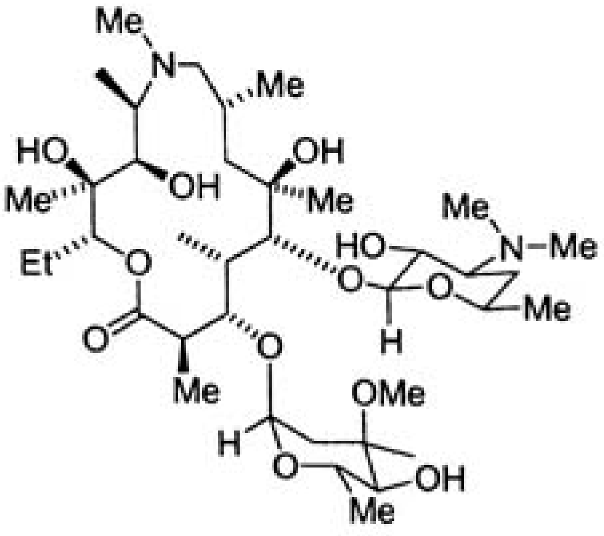
The primary driver for research in the antibacterial area over the past decade has been the emergence of resistant organisms. Important members of the ever-growing armamentarium of antibacterials include azithromycin (Zithromax®/Pfizer), linezolid (Zyvox®/Pharmacia), amoxicillin®/clavulanate potassium (Augmentin/Glaxo-SmithKline), and ciprofloxacin (Cipro®/ Bayer). Azithromycin is a semisynthetic 9a-azalide analog of erythromycin possessing improved resistance to acid mediated degradation, increased activity against gram negative organisms, and improved pharmacokinetics. Its indications include the treatment of mild to moderate upper and lower respiratory tract infections and otitis media in pediatrics. Interestingly, azithromycin tends to concentrate in the lung tissue which is the site of many bacterial infections. Amoxicillin/clavulanate is one of the few approved drug mixtures and is the drug of choice for the treatment of otitis media. It is also an alternative treatment for anthrax exposure in pediatrics. Ciprofloxacin is a totally synthetic antibacterial that acts as a DNA gyrase inhibitor. It is active against a broad range of pathogens including both gram positive and gram negative aerobic bacteria and is effective against urinary tract and lower respiratory tract infections. Linezolid is the first new antibacterial chemotype to be put on the marketplace in over 20 years. It is a totally synthetic oxazolidinone derivative, which has a unique mechanism of action resulting in a low potential for cross resistance to other antibacterials. Linezolid is indicated for the treatment of community acquired pneumonia, MRSA, and VRE infections and has the unique characteristic of being 100 percent orally bioavailable.



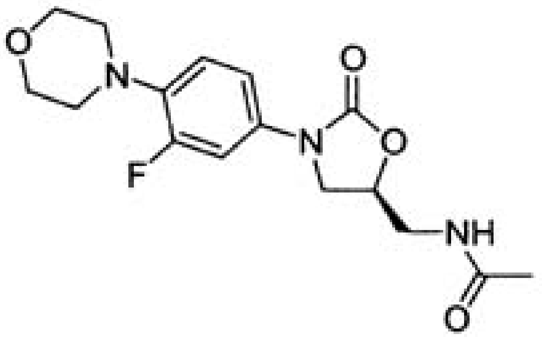
amoxicillin / clavulanic Acid



ciprofloxacin



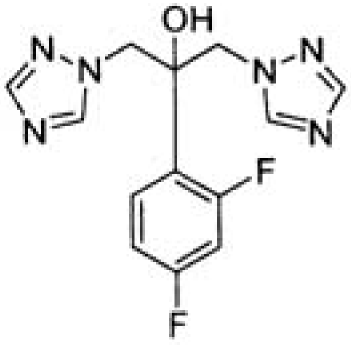
azithromycin



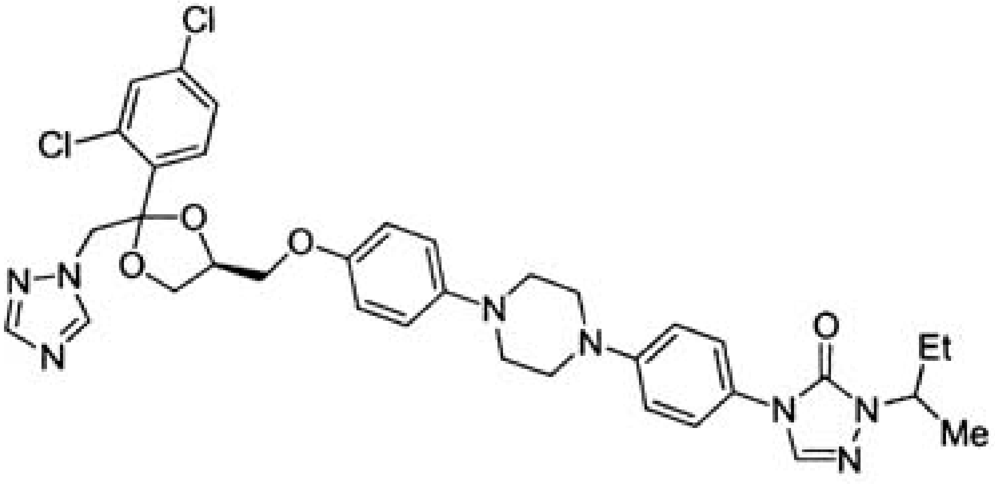
linezolid

Antifungals

The increasing immunocompromised patient population has exacerbated the need for effective antifungal agents to combat opportunistic fungal infections which arise in these patients. Fluconazole (Diflucan®/Pfizer), an achiral triazole derivative, is indicated for the treatment of systemic candidiasis as well as meningitis caused by *Cryptococcus neoformans*. Itraconazole (Sporanox®/Janssen, Ortho Biotech), a mixture of four diastereomers, is used to treat aspergillosis, oral candidiasis, and histoplasmosis. These agents are structurally related to other imidazole based antifungals such as ketoconazole and miconazole but have better antifungal activity and broader coverage.



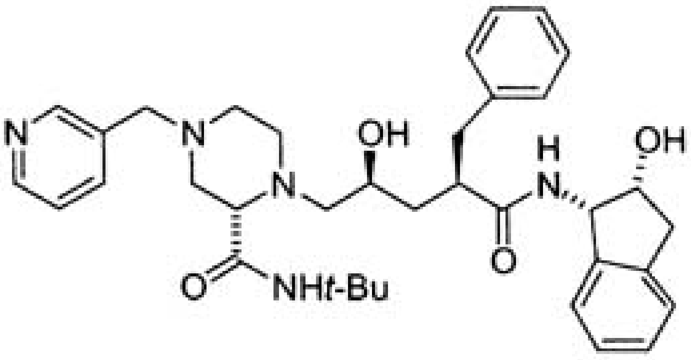
fluconazole



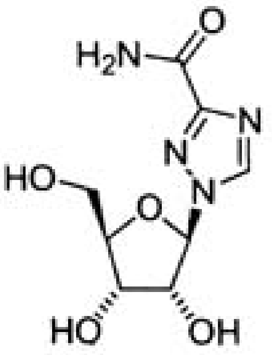
itraconazole

Antivirals

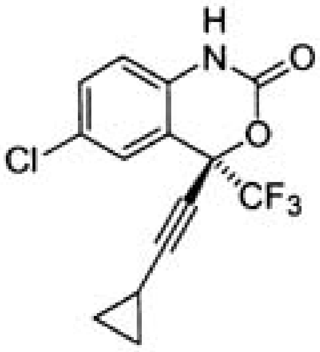
Antiviral research has become a major focus in the pharmaceutical industry over the past decade as evidenced by the marketing of a plethora of agents active against HIV/AIDS, Hepatitis B and C, and respiratory syncytial virus (RSV). Indinavir (Crixivan®/Merck) is one of a group of HIV protease inhibitors which are used in conjunction with other antiretroviral chemotherapeutic agents for the treatment of AIDS in adults and adolescents. It is a Phe-Pro scissile bond peptidomimetic with an hydroxyindane moiety which was optimized for selectivity and potency. Ribavirin (Rebetron®/Schering and Virazole®/ICN) is a synthetic nucleoside used to treat RSV in hospitalized infants and is also used in combination therapy with interferon for the treatment of chronic hepatitis C. Efavirenz (Sustiva®/Bristol-Myers Squibb) is a synthetic non-nucleoside reverse transcriptase inhibitor (NNRTI) used in conjunction with other antiretroviral agents for the treatment of HIV.



indinavir



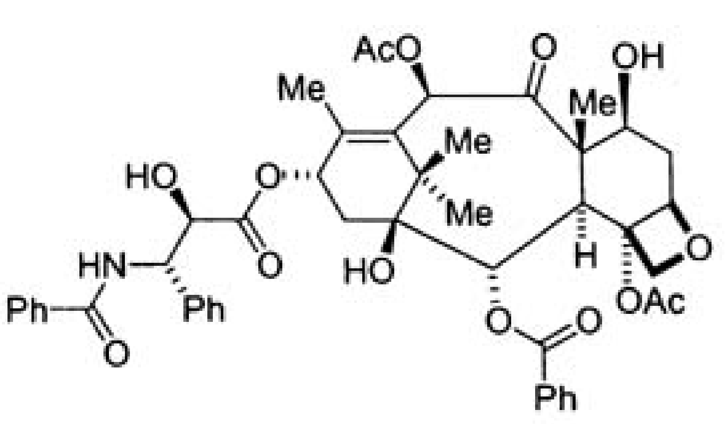
ribavirin



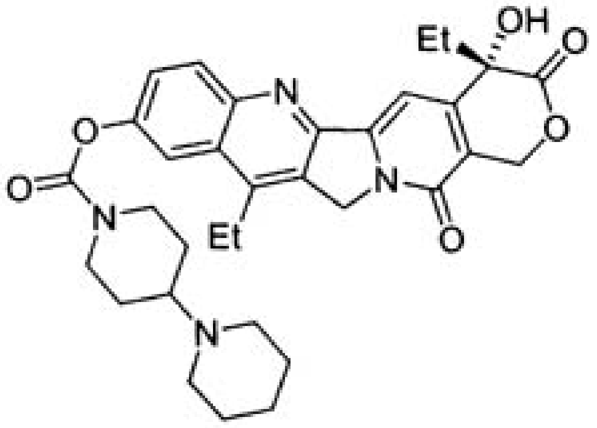
efavirenz

26.10 Antineoplastics

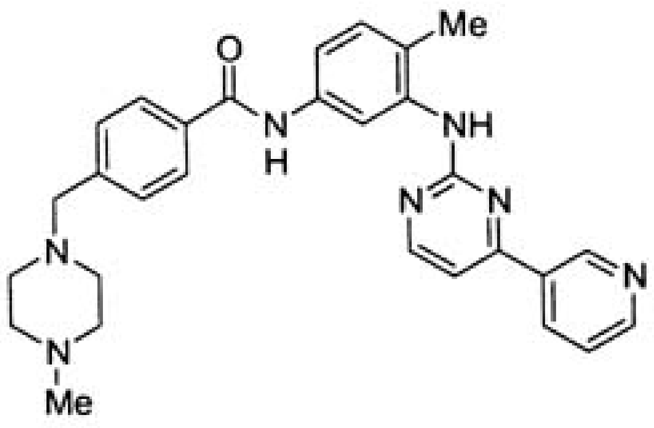
Paclitaxel (Taxol®/Bristol-Myers Squibb) and irinotecan (Camptosar®/Pharmacia) were discovered as a result of natural product extract screening done at the NIH in the late 1960s by Monroe Wall and Mankush Wani. Paclitaxel is a naturally occurring diterpene which exerts its antineoplastic effect via stabilization of the mitotic spindle during cell replication. It is used for the treatment of non-small cell lung, breast, ovarian, and esophageal carcinomas as well as Kaposi's sarcoma. Irinotecan is a prodrug, which upon release of the piperidinylpiperidine carbamate moiety reveals the pharmacologically active parent SN-38, which is itself a derivative of the naturally occurring camptothecin. Irinotecan exerts it's antineoplastic activity via the inhibition of Type I DNA topoisomerase and stabilization of the transiently formed Topoisomerase I/DNA cleavable complex. Tamoxifen (Nolvadex® AstraZeneca), a non-steroidal anti-estrogen chemotherapeutic possessing both agonistic and antagonistic properties, is used for the treatment and prevention of breast cancer. Imitanib (Gleevac®/Novartis), an inhibitor of Bcr-Abl tyrosine kinase, recently received FDA approval for the treatment of chronic myelogenous leukemia.



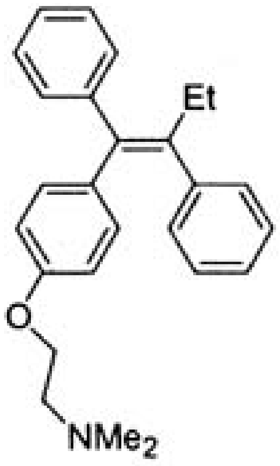
paclitaxel



irinotecan



Imatinib

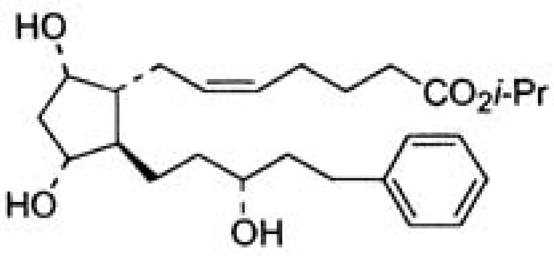


Tamoxifen

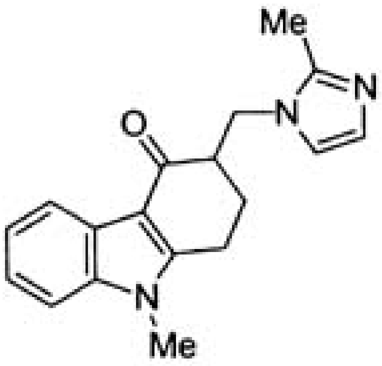
26.11 Miscellaneous Agents

Glaucoma and Nausea

Latanoprost (Xalatan®/Pharmacia & Upjohn) is topical, ocular hypotensive agent used to treat glaucoma. The compound is a synthetic analog of the naturally occurring pros-taglandin PGF2a and is thought to reduce intraocular pressure by increased outflow of the aqueous humor. Odansetron (Zofran®/ GlaxoSmithKline) is a selective, serotonergic, 5-HT3 receptor antagonist and is used to ameliorate nausea and vomiting associated with chemotherapy-induced emesis.



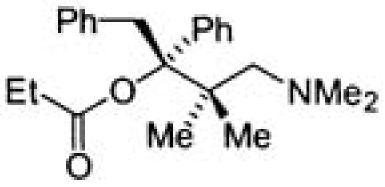
latanoprost



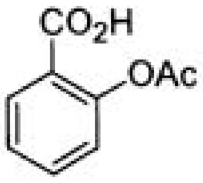
odansetron

Analgesics

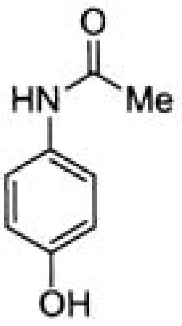
Propoxyphene, along with aspirin, acetominophen and ibuprofen are among the most widely used agents for the treatment of mild to moderate pain.



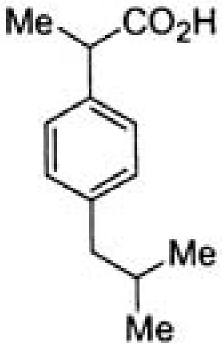
Propoxyphene



Aspirin



Acetominophen



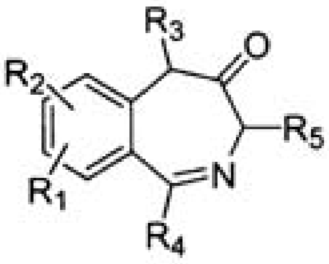
Ibuprofen

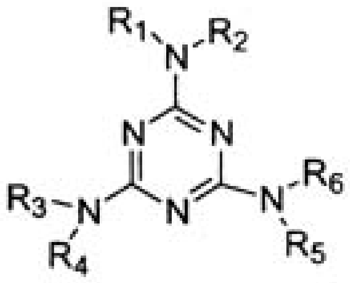
26.12 Small Molecule High Throughput Synthesis

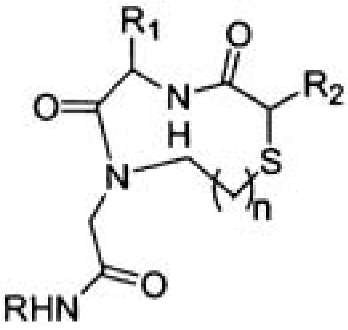
The field of non-oligomeric, small molecule high throughput synthesis came into existence in 1992. Since then, small molecule library synthesis has impacted drug discovery efforts in lead identification, as well as lead optimization. In a recent review, R. Dolle has categorized synthetic libraries as follows: (1) discovery libraries: libraries synthesized with no preconceived notion about which molecular target it may be active against. These libraries tend to be large in size, typically >5000 compounds. (2) targeted libraries: these libraries are biased in their design and contain a pharmacophore known to interact with a specific target, or a family of targets. (3) optimization libraries: libraries are constructed around an existing lead with the intent to improve potency, selectivity, pharmacokinetic profile, and so on. These libraries tend to be smaller in size, usually ranging from tens up to a few hundred compounds.

Discovery Libraries

Researchers have employed several different strategies to create populations of molecules that are used for broad-based screening. One strategy is to synthesize libraries of "privileged pharmacophores" such as benzodiazepines (1), triazines (2), and so on. A second strategy is to design scaffolds or templates for library synthesis that are based on important molecular recognition. Libraries of β-turn mimetics (3) synthesized by Ellman et al. are examples of templates for molecular recognition.


1

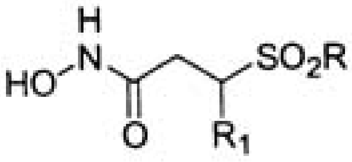

2

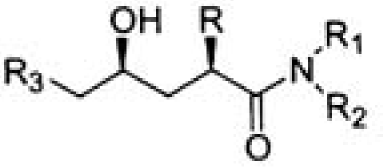

3

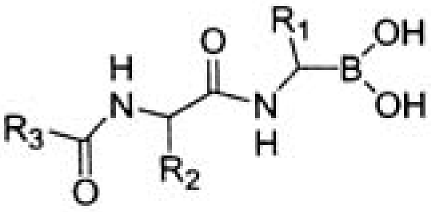
The discovery of chemical encoding technologies and radio-frequency (Rf) encoded synthesis have had a major impact on synthesis of lead discovery libraries. Both technologies utilize resin-based split-pool synthesis to prepare large arrays of compounds. Libraries of >50 K members were prepared by using chemically encoded beads. Chemical encryption, in the form of unique chemical markers (tags), is associated with synthetic identity of the library member tethered to the resin bead. The technology for chemical encoding was pioneered by W. Clark Still and subsequently commercialized by Pharmacopeia, Inc. Restricted amount (200300 µg), lack of analytical characterization of library members, and the requirement of a specialized screening format for chemically encoded libraries have limited the utility of this technology. Radio-frequency encoded synthesis, developed, and commercialized by IRORI, Inc. overcomes the aforementioned limitations while retaining the efficiency of split-pool synthesis. Libraries of 1015 K members can be prepared, with individual members quantitated and characterized by LC/MS. Most pharmaceutical companies have utilized Rf-encoded synthesis in their lead identification efforts.

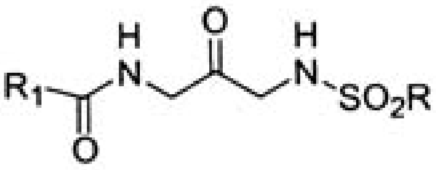
Targeted Libraries

Libraries targeted towards proteolytic enzymes, non-proteolytic enzymes, G-protein coupled receptors (GPCRs), and ion-channels have been very successful in lead identification. Libraries of hydroxamates (4), hydroxy ethylenes (5), boronic acids (6) and α-keto sulfonamides (7) have been prepared as inhibitors of metallo-, aspartyl, serine, and cysteine proteases respectively, using either solid phase or solution phase synthesis.

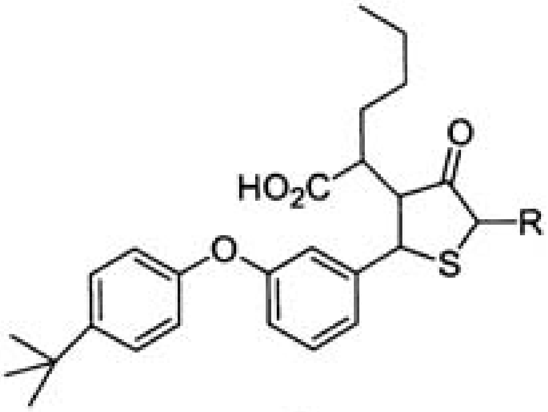

4


5

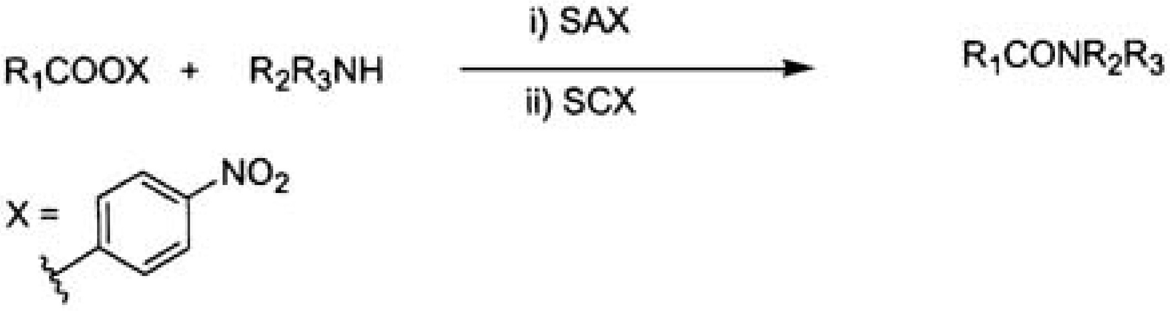

6


7

Structure-based design has been effectively utilized in synthesis of inhibitors of non-proteolytic enzymes. Inhibitors of MurB, an essential bacterial enzyme required for biosynthesis of peptidoglycan, were identified using the X-ray structure of the enzyme for library design. Thiazolidinone inhibitors (8) thus identified are the first examples of small molecule inhibitors of MurB.

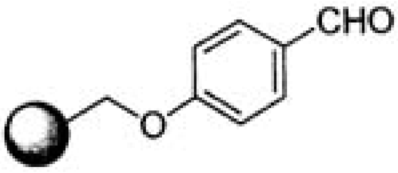

8

Substituted indoles (5HT2a; D4 and α2a receptor antagonists) and piperazines (δ opioid antagonists) are representative chemotypes targeted towards GPCRs. Advances made in solid phase extraction (SPE) and in development of resin-based scavengers have increased the versatility of chemistries implemented for synthesis of targeted libraries. A combination of cation exchange (SCX) and anion exchange (SAX) resin was effectively utilized to prepare libraries of highly substituted amides.

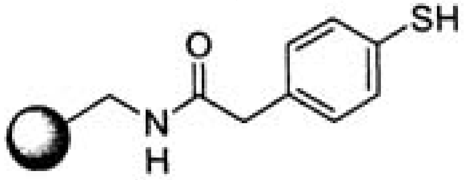

$$R_1COOX + R_2R_3NH \xrightarrow[\text{ii) SCX}]{\text{i) SAX}} R_1CONR_2R_3$$

Scavenger resins and polymer-bound reagents are routinely used to prepare medium-sized (5001000 member) libraries. Polymer-bound isocyanates (9) and aldehydes (10) are used to remove amines from reaction mixtures, while polymer-bound thiols (11) are used to scavenge halides.

9



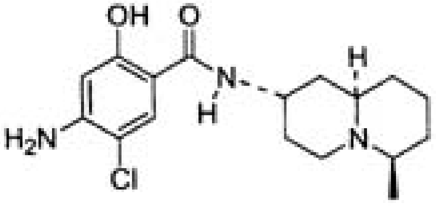
10



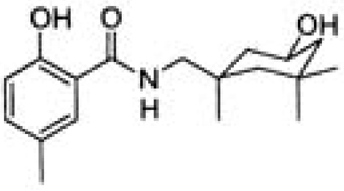
11

Optimization Libraries

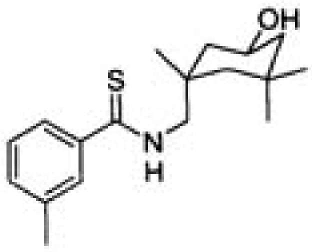
Starting with a lead structure, researchers have demonstrated that parallel synthesis can be effectively utilized to optimize activity, as well as reduce timelines for optimization. Parallel synthesis strategy was implemented to identify more potent analogs of influenza hemagglutinin inhibitor (12) (IC50 4 µg/ml). Solid phase extraction was used to automate preparation of >400 analogs resulting in identification of compounds (13) (IC50 = 20 ng/ml) and (14) (IC50 = 20 ng/ml).



12

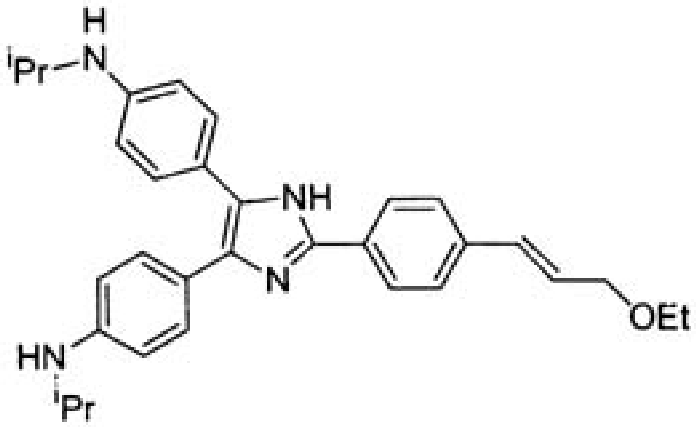


13

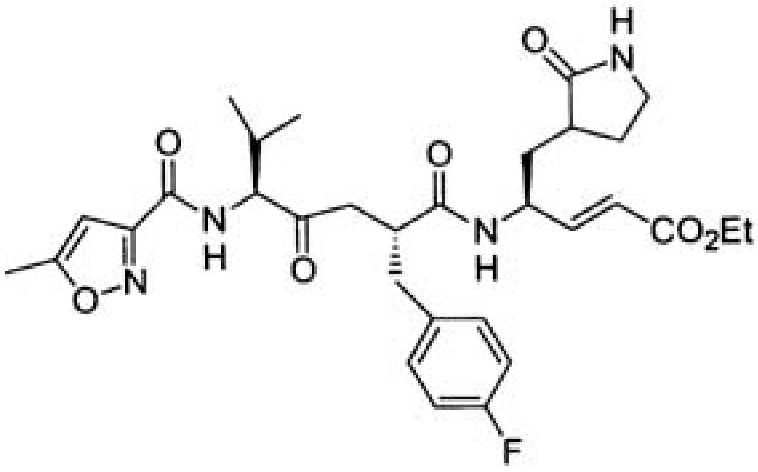


14

During the past 10 years, the pharmaceutical industry has expended significant resources in developing and assimilating technologies to increase synthesis throughput and decrease pre-clinical time lines. There are numerous examples in the literature demonstrating effective use of high throughput synthesis for lead discovery and optimization. There are two publicly known examples of clinical candidates that have emerged directly from optimization libraries. Ontogen Corporation identified OC144-093, (15) (IC50 = 50 nm) as a P-glycoprotein modulator and Agouron Pharmaceuticals reported identification of AG-7088, (16) (kobs/I = 1,470,000 M-1S-1), a clinical candidate for treatment of rhinovirus infection.



15



16

< previous page

next page >

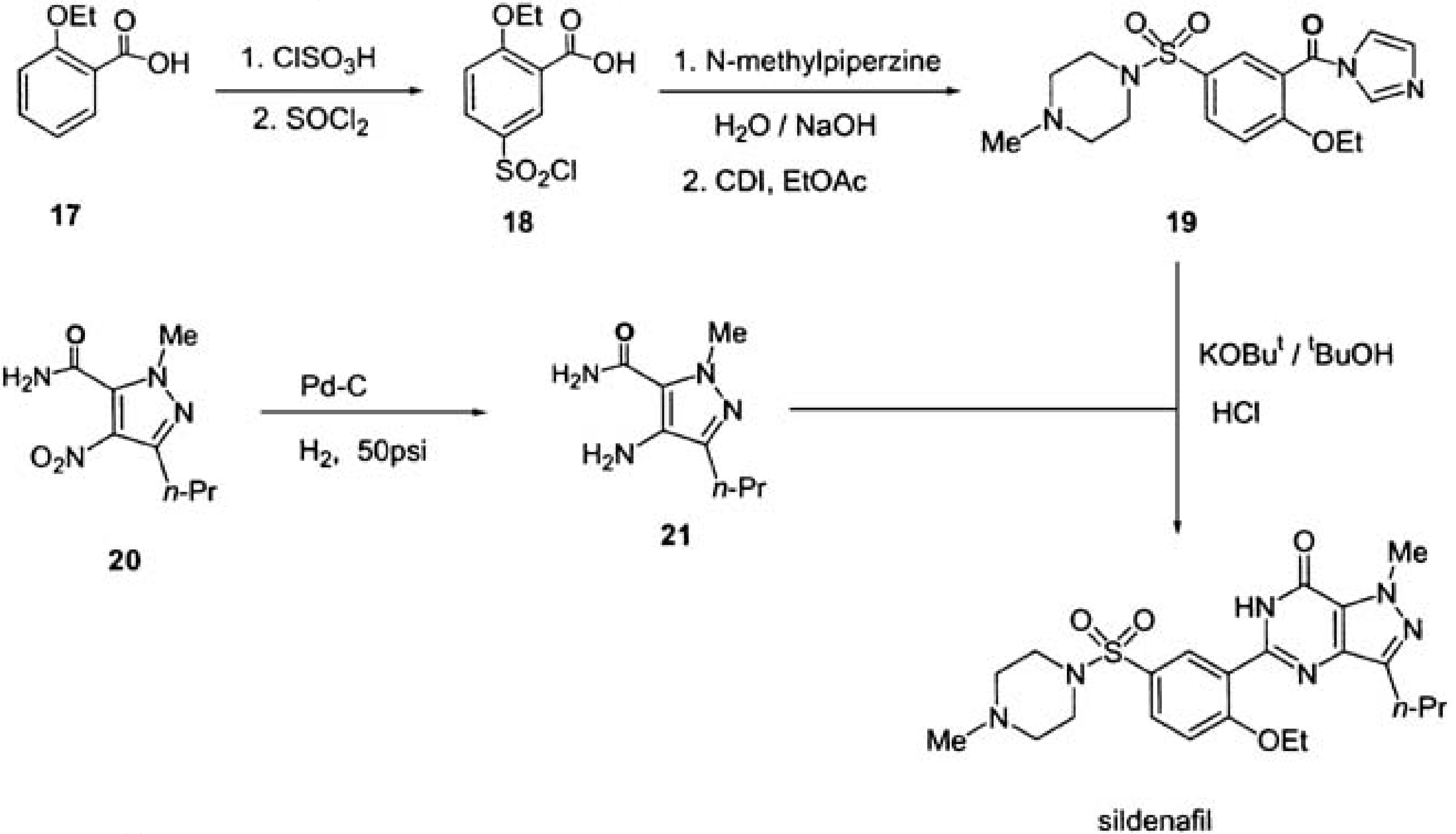
26.13 Chemical Process R&D in the Pharmaceutical Industry

Most of the active pharmaceutical ingredients (API's) of commercially available pharmaceuticals are manufactured either by chemical syntheses or microbial fermentations. However, some of the active ingredients are directly obtained from natural sources. This section addresses the development and manufacture of APIs. Recent trend shows that >75 percent of the drug candidates in development are chiral and of complex structure. Incessant demand to shorten the timelines for the discovery, development, and launch of NCEs coupled with environmental concerns has necessitated the development of higher yielding, more robust, and environmentally friendly processes in shorter times. The success of a pharmaceutical company greatly depends not only on discovering blockbuster NCEs but also on its ability to design, optimize, and scale-up a chemical process to commercial manufacturing with increasing rapidity. The chemical manufacturing process must be a robust procedure capable of operating routinely in a manufacturing environment. Considerable attention has to be given to various parameters in developing a manufacturing process for an API, including for example: efficiency of the synthesis, availability and cost of starting materials, toxicity of the reagents, stability and toxicity profiles of intermediates, formation of byproducts, and safe disposal of waste materials. Data from various aspects of chemical process development, including process structure and flowsheet, operational guidelines, optimization, process management, process control, fault diagnostics and equipment management need to be in place in order to support a smooth transition from laboratory to manufacturing plant. Safety is another critical factor requiring consideration for large scale manufacture. All reactions should undergo a process hazard analysis for incident-free and successful plant implementation before scale-up. The use of automation in accelerating the design of cost-effective and well-understood synthetic processes has been demonstrated over the past few years by pharmaceutical companies. Automation concepts and tools such as statistical design of experiments and parallel experimentation using in-house built reactor blocks or commercially available systems such as Zymark robots, ReactArray, Bohdan, Argonaut's Surveyor or Mettler Toledo MultiMax will play a major role in increasing the productivity of Process R&D with respect to speed and economics, as well as obtaining process knowledge. The application of microreaction technology (micro piloting) is another area that is growing rapidly to understand the chemical engineering aspects of process development. Some beneficial features of microreaction technology include mixing efficiency, enhanced heat transfer, and more uniform residence time distributions.

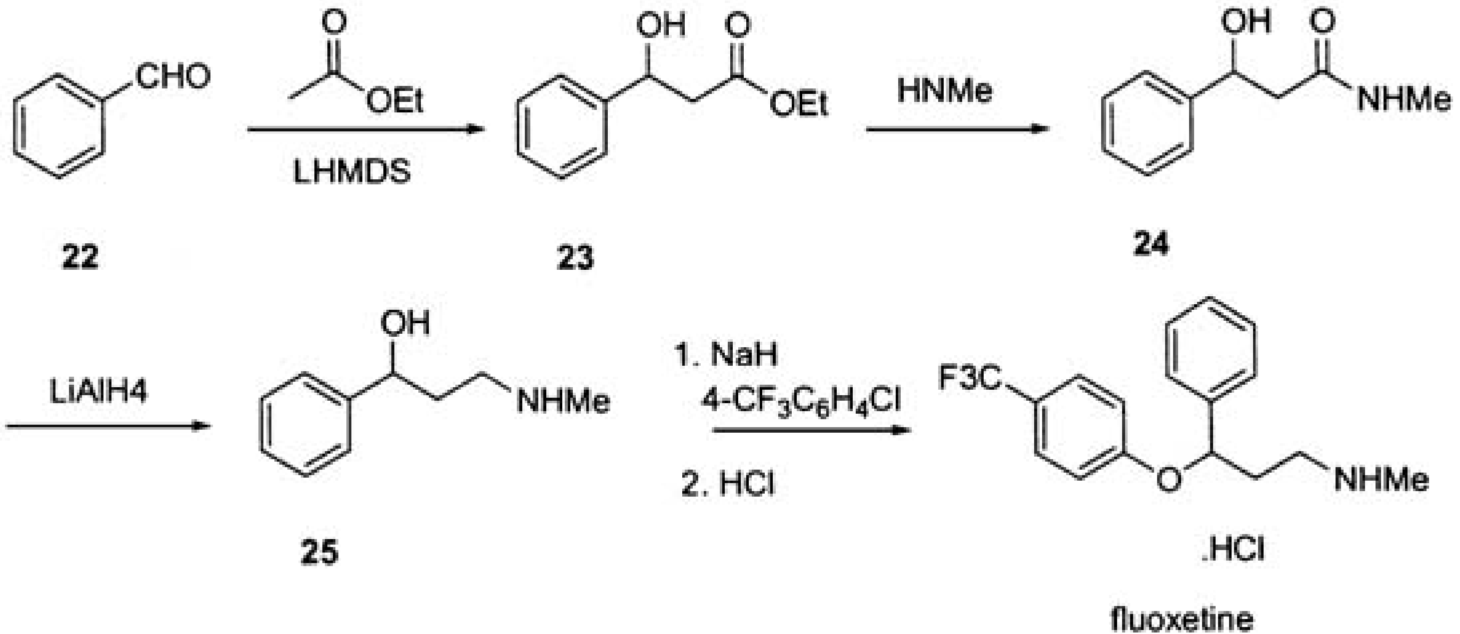
Production and logistical processes are becoming more complex due to an increasing number of products and smaller batch sizes. To manage this, supply chain optimization and production planning activities need to be addressed. Production simulation can be used for performance measurements and capacity assessments of manufacturing as well as material and information flow processes. Some applications of production simulation include bottleneck analysis, examination of process alternatives, assessment of investment decisions, and solution of sequencing problems. Batch process development is a fairly complex series of engineering tasks. In the pharmaceutical industry, the production of a majority of APIs is based on a batch concept. This concept offers many advantages with respect to quality assurance as an individual batch can be accepted or rejected. However, the scale-up of the batch size without proper controls may lead to problems. The variety of the equipment involved often does not facilitate the scale-up process. In order to avoid scale-up problems, continuous or semi-continuous processes need to be adopted as alternatives to a batch production.

Crystallization, filtration, drying, and milling (if required) are other important factors that need to be defined well before a process to manufacture solid APIs is finalized. Physicochemical properties of APIs play a vital role in providing the pharmaceutical drug products with desired bioavailability, manufacturing properties, and good final product quality. Particle size, density, flowability, polymorphism, hygroscopicity, and stability are critical properties for solid APIs in the formulation development. Polymorphism is very important in determining the physical properties of various crystal forms of a drug for optimal chemical and formulation processing, as well as for satisfying regulatory and patent issues for producing consistent solid forms of a drug.

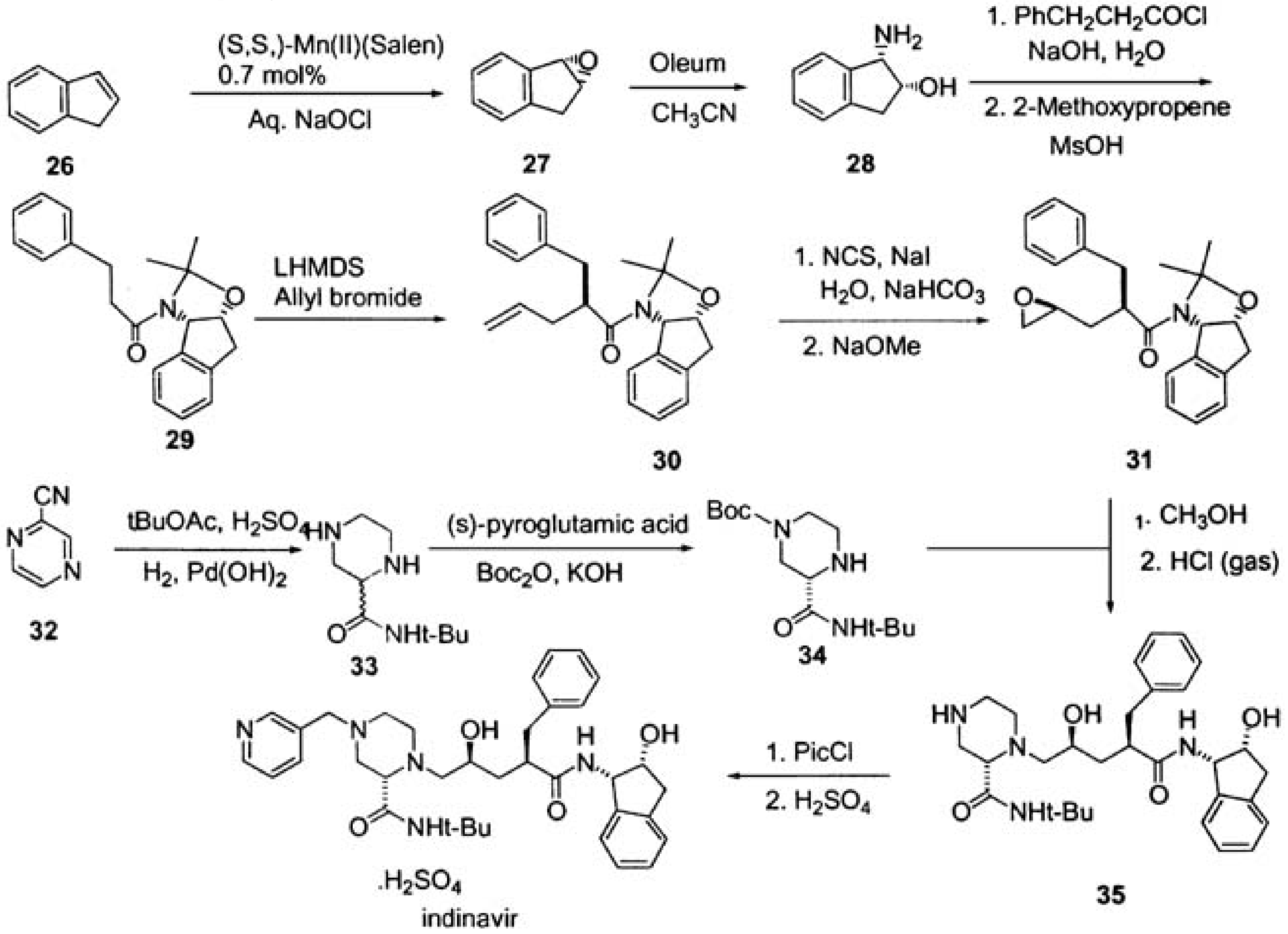
The following flow diagrams shows the preparation of APIs of some widely used pharmaceutical drugs in today's market. Scheme-1 shows the preparation of sildenafil. This route has a greater synthetic convergency than other published routes.



Synthesis of fluoxetine as a racemic mixture is shown in Scheme-2. Recently several patents and publications have appeared in the literature describing the synthesis of (S)- and (R) enantiomers.



The single enantiomer of indinavir has five stereogenic centers, four of which are derived either directly or indirectly from epoxide (27). Synthesis of indinavir sulfate developed by Merck is shown in Scheme-3.



26.14 Conclusion

The discovery and development of novel therapeutic agents by the pharmaceutical industry has afforded physicians with an extensive armamentarium of weapons to fight a wide range of human disease. Of course there remains the opportunity for even more effective drugs with greater benefit to risk ratios than those currently available. The elucidation of the human genome will eventually lead to the identification of many new macromolecular targets for drug intervention. Chemistry has been and will likely continue to remain at the forefront of pharmaceutical research which will afford the drugs of the future.

26.15 Reading References

Krogsgaard-Larsen, P., Liljefors, T., and Madsen, U. (Eds.), A Textbook of Drug Design and Development, 2nd ed., Harwood Academic Publishers, Amsterdam, 1996.
Spilker, B., Multinational Drug Companies; Issues in Drug Discovery and Development, Raven Press, New York, 1989.
Wermuth, C. G. (Ed.), The Practice of Medicinal Chemistry, Academic Press, San Diego, 1996.
Lipinski, C. A., Lombardo, F., Dominy, D. W., and Feeny, P. J., Adv. Drug Delivery Rev., 23, 325 (1997).
Testa, B., and Mayer, J. M., Drug Metab. Rev., 30, 787807 (1998).
Wess, G., Urmann, M., and Sickenberger, B., Angew. Chem. Int. Ed., 40, 33413350 (2001).
Miertus, S., and Fassina, G. (Eds.), Combinatorial Chemistry and Technology; Principles, Methods and Applications, Marcel Dekker, Inc., New York, 1999.
Dolle, R., J. Combinatorial Chem., 3, 477518 (2001).
Anderson, N. G., Practical Process Research and Development, Academic Press, New York, 2000.

27.1 Soap and Fatty Acids

Introduction

The origin of the word "soap" is traced to sacrificial Mount Sapo of ancient Roman legend. The mixture of fat and wood ashes that reacted to form soap was carried by rain to the banks of the Tiber River and was found as a clay deposit useful for cleaning clothes.1

The boiling of fats with ashes was recorded as early as 2500 B.C. Commercial soap-making was a widespread art in the Middle Ages in Europe. The invention of the soda ash process by LeBlanc in 1791, and the discovery by Chevreul in 1811 that soap was composed of a mixture of fatty acids paved the way to modern soap-manufacturing processess.1,2

Chemistry

The modern definition of soap relates to cleansing agents derived from fats, oils, and other fatty derivatives: the soaps are alkali and ammonium salts of fatty acids containing from 8 to 22 carbon atoms. These fatty acids are generally a mixture of saturated, unsaturated, and polyunsaturated moieties:

$$CH_3(CH_2)_{10}COOM\text{: saturated soap}$$

$$\begin{gathered}CH_3(CH_2)_6CH_2CH{=}CHCH_2(CH_2)_6COOM\text{:}\\ \text{mono-unsaturated soap}\end{gathered}$$

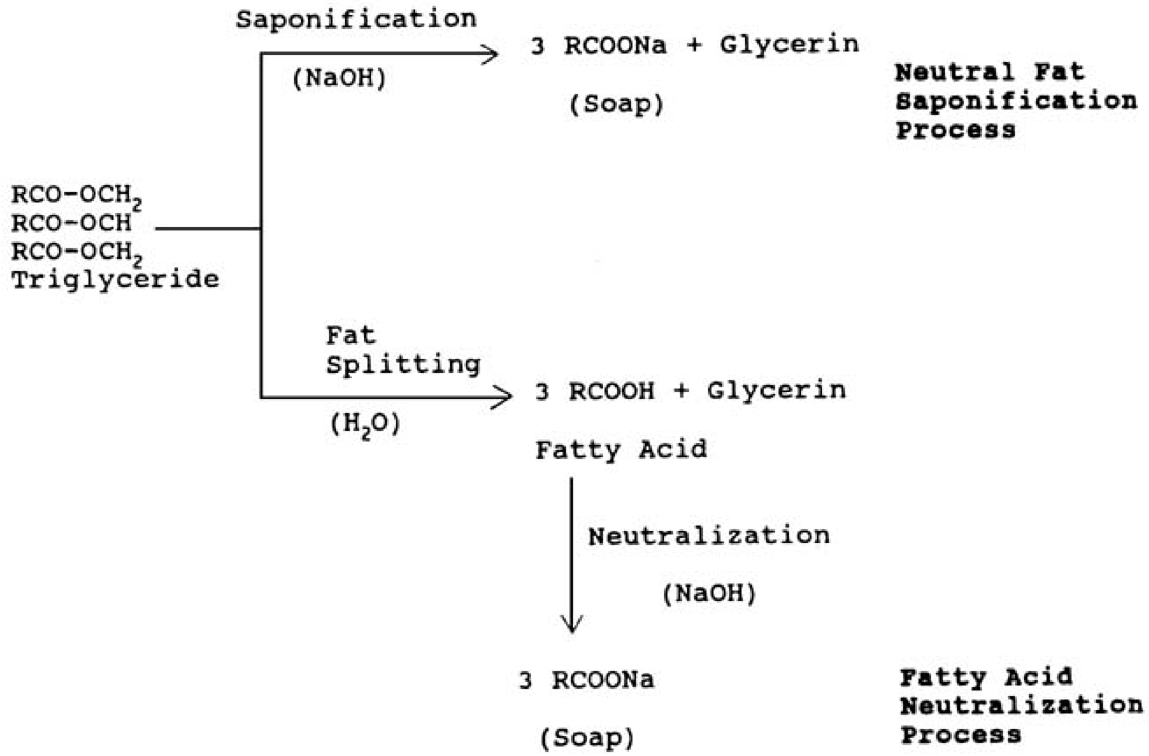
$$\begin{gathered}CH_3(CH_2)_3(CH_2CH{=}CH)_2CH_2(CH_2)_6COOM\text{:}\\ \text{di-unsaturated soap}\end{gathered}$$

$$\begin{gathered}CH_3(CH_2CH{=}CH)_3CH_2(CH_2)_6COOM\text{:}\\ \text{poly-unsaturated soap}\end{gathered}$$

$$[M = Na, K, (HOCH_2CH_2)_3NH^+]$$

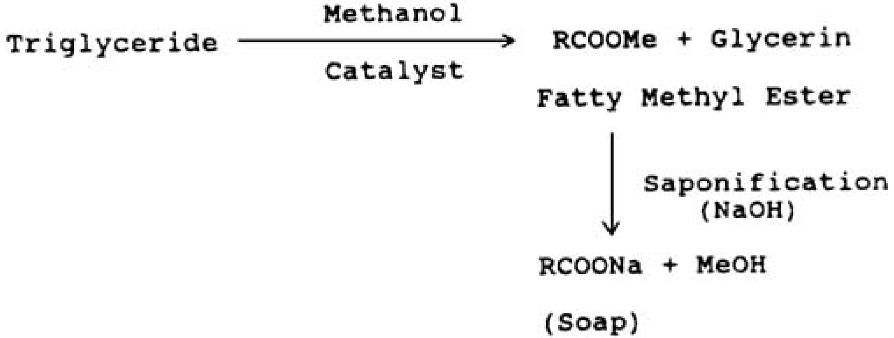
In general, two types of chemical reactions are utilized in the manufacture of soap: the saponification of triglycerides (fats and oils) and the neutralization of fatty acids (which themselves are produced from the triglycerides by a variety of methods, most notably splitting or hydrolysis of fats and oils with steam under high pressure). Sodium hydroxide (the predominant alkali employed in the manufacture of soaps), potassium hydroxide, sodium carbonate, and triethanolamine are the most commonly utilized alkaline moieties in these processes (Fig. 27.1).3,4

Fig. 27.1 The chemistry of soap manufacture.



In recent years, soap manufacture by an alternate route, the saponification of fatty methyl esters, has been under development, most notably in Japan (Lion Corporation) and Italy (Ballestra). The fatty methyl esters are obtained from the methanolysis of triglycerides; inorganic alkali, quaternary ammonium salts, and enzymes (lipase) have been used as catalysts for methanolysis in commercially practiced processes (Fig. 27.2). The methyl ester process for soap manufacture is typically more capital and cost intensive vs. the more traditional fats/oils saponification and fatty acid neutralization processes.

Fig. 27.2 Fatty methyl ester process for soap manufacture.



Manufacturing Technology

Several designs of commercial equipment, based on the two processes described in Fig. 27.1, are available for the manufacture of soap.

Neutral Fat Saponification	Fatty Acid Neutralization
Kettle Processes	Mazzoni SC and SCC
Semi-boiled	Alfa Laval
Boiled	Sato
Modern Processes	
Mazzoni SCNC-N, SCNV-N	Dial Saponiflex
Binacchi CSWE, BV	
Alfa Laval Centipure	
Hitachi	

Dial Saponiflex

A critical review of these processes is beyond the scope of this chapter. However, the processes themselves are described in a subsequent section.

Raw Materials

The triglycerides (triesters of fatty acids with glycerin) are the basic raw materials for the manufacture of soap utilizing a variety of processes. These triglycerides, as a renewable resource, occur widely throughout the plant and animal kingdoms.

Soap products in the United States are produced almost exclusively from tallow (beef) fat and coconut oil or their fatty acid derivatives. The utilization of palm oils, palm kernel oil, and their derivatives for soap manufacture is more commonly practiced in many other parts of the world, especially in Asia. More recent trends in lauric oil pricing have driven the purchase and use of palm kernel oil and palm kernel olein as cost-effective alternatives to coconut oil in the West. The frequent substitution of coconut oil with palm kernel oil is done primarily for cost optimization of soap-manufacturing feedstocks. Among other triglycerides, cottonseed oil, ricebran oil, castor oil, neem oil, sunflowerseed oil, fish oil, and olive oil have also been used as indigenous feedstocks in many developing nations. The use of palm stearin, the by-product of palm oil fractionation, is increasing in use as a substitute for tallow outside of tallow-producing markets. Depressed pricing in 2000 and the recent fears of Bovine Spongiform Encephalopathy (BSE) or Mad Cow disease have driven this use trend in Europe. Fatty acid composition and analytical characterization data of some of the key fats and oils utilized for soap production are given in Table 27.1.

TABLE 27.1 Fatty Acid Distribution and Analytical Characterization of Soap Base Fatty Acids

Fatty Acid Distribution	Coconut	Stripped Coconut	Palm Kernel	Palm	Tallow
Caprylic ($C_8H_{16}O_2$)	7	1	4		
Capric ($C_{10}H_{20}O_2$)	6	1	5		
Lauric ($C_{12}H_{24}O_2$)	48	55	50		
Myristic ($C_{14}H_{28}O_2$)	19	22	15	2	2.5
Palmitic ($C_{16}H_{32}O_2$)	9	11	7	42	27
Stearic ($C_{18}H_{36}O_2$)	2	2	2	5	20
Oleic ($C_{18}H_{34}O_2$)	8	7	15	41	42
Linoleic ($C_{18}H_{32}O_2$)	1	1	1	10	5
Linoleic ($C_{18}H_{30}O_2$					1
C20/C22 isomers			1		2.5
Characterization Data:					
Titer, C	26	29	25	40	40
Iodine value (IV)	10	5	18	55	55
Color, Gardner, max.	5	1			5
Acid value (AV)	270	255			205
Saponification value (SV)	270	255			205
Color, Lovibond, 5-¼″	10Y/1R	4.5Y 0.6R			

Most of the fats and oils utilized for soap manufacture require a combination of steps: degumming, deodorization, refining, and bleaching, which are further discussed in the section on fatty acids.

For both saponification and neutralization reactions, sodium hydroxide is the most commonly used base. In some special situations, such as the manufacture of transparent and translucent soaps, potassium hydroxide and triethanolamine are also employed.

Functional Properties of Soap

Almost as a rule, the solubility of soap in water increases with an increase in the size of the monovalent cation (base); however, an increase in the size of a divalent cation (Mg^{+2}, Ca^{+2}) results in a decrease in the solubility. An increase in the chain length of the fatty acid moiety of soap results in a decrease in the solubility of soap, whereas the presence of unsaturation results in an increase in its solubility. In general, an increase in the water solubility is accompanied by an increase in the softness and use-up properties of soap. Adequate solubility results in good performance in soft water, although in hard water, lather decreases due to the formation of insoluble Ca^{+2} and Mg^{+2} salts. Cleansing bars formulated with various levels of synthetic detergents (combars or syndets) provide improved performance in hard water conditions.

Consumer-acceptable lather requires the consideration of a variety of lathering attributes, such as the speed with which lather is generated, volume, quality (i.e., loose or creamy), and the stability of the lather. In general, the use of fatty acids of C10C12 chain length provides a voluminous, lacey foam while fatty acids of longer chain length, from C16 to C18, contribute to a richer, creamier foam.

Thus, a combination of coconut (C10C12) and tallow (C16C18) fatty acids is most suitable from the viewpoint of providing a balance in lathering attributes to deliver consumer-acceptable lather. Commonly accepted ratios of these soap feedstocks are 8085 percent tallow and 1520 percent coconut oil for standard purpose bars and 6070 percent tallow, 3040 percent coconut oil for the higher lathering bar-soap products. Cleansing bars formulated with free fatty acids (superfatting) improves the quickness of lather generation and provides an open, bubbly foam. Commonly accepted levels of superfatting range from 17 percent for various types of bar soaps. Various methods are employed to evaluate bar lather attributes.5

Manufacturing Processes

The following describes neutral fat saponification and fatty acid neutralization, the two principal processes used for manufacture of soap in the world markets today.

Neutral Fat Saponification

In the saponification of triglycerides with an alkali, the two reactants are rather immiscible. Because saponification is a bimolecular nucleophilic substitution (SN2), the kinetic rate enhancement is achievable by both a high-shear mixing and an increase in the reaction temperature during the processing. Increasing the pressure of the reaction assures that vaporization does not occur at the high reaction temperatures and helps to accelerate conversion of the fatty feedstocks and of alkali to soap as well. The saponification reaction is self-catalyzed to some extent: the formation of soap product affects the emulsification of the two immiscible reactants, causing an acceleration of the reaction rate by providing increased surface area for the saponification reaction to occur.

The amount of alkali required to saponify a triglyceride blend is calculated by the following equations:

$$\begin{aligned} & \text{Triglyceride} + 3\text{NaOH} \\ & \quad \rightarrow\ 3\text{ROONa} + \text{Glycerin} \\ \text{NaOH} \quad = \quad & \text{weight oil} \times [\text{SV} \times 0.000713] \\ & \times 100/\text{NaOH}(\%) \text{ in solution} \\ = \quad & \text{weight oil} \times \text{SV}(1\,\text{g}/1000\,\text{mg}) \\ & (\text{MW NaOH}/\text{MW KOH}) \\ & (100\%/\%\text{NaOH}) \qquad\qquad (27\text{-}1) \end{aligned}$$

where SV = saponification value (mg KOH/g triglyceride) and MW = molecular weight. In general, the most commonly used soap feedstocks have these saponification values:

	g NaOH/100 g Triglyceride)
Beef tallow	14.09
Coconut oil	18.32
Palm kernel oil	17.81
Palm oil	14.19
Palm stearin	14.26

The use of these values simplifies the calculation of the stoichiometric alkali equivalent, for example:

$$\begin{aligned} & 100\,\text{g beef tallow} \times 14.09\,\text{g NaOH}/ \\ & 100\,\text{g tallow} \times 1\,\text{g}\,50\%\,\text{NaOH caustic}/0.5\,\text{g NaOH} \\ & \quad = 28.2\,\text{g}\,50\%\,\text{NaOH caustic required} \end{aligned}$$

The process of a typical triglyceride saponification plant is given as a flow diagram in Fig. 27.3.6 Feed streams to the plant are the fats/oils, caustic, brine, and water both to convert the triglycerides to soap and to provide the necessary processing environment. The two distinct product streams of saponification are wet, or neat soap and spent lye that contains the bulk of the by-product glycerin from this reaction. The essential components of this plant include proportioning pumps (mass flow meters are replacing proportioning pumps for reasons of reduced maintenance, higher accuracy, and reliability) for introducing correct quantities of reaction components into a reactor autoclave, which operates at temperature and pressure conditions appropriate for a rapid and complete reaction. The saponification mixture is recirculated within the autoclave. Since the temperature of the saponified soap/lye mixture is elevated, effective separation is made possible by passing this mass through a cooler and then to a static separator for separation of the unwashed soap from the spent lye solution. The soap, then, is contacted with washing lye countercurrently in an extraction column to separate glycerin (as spent lye) from the soap mass. Centrifugal separators remove entrained lye from the soap directly after extraction. This centrifuging step reduces moisture and electrolyte content from the soap that is required in downstream processing stages. Excess alkalinity in the neat soap (6063% total fatty matter) is neutralized, or trimmed with lauric oil (coconut or palm kernel typically). This finished base soap is dried (typically on vacuum spray-type dryers) to produce soap as pellets (7882% total fatty matter), ready for processing into finished product on finishing lines (Fig. 27.4).

Fig. 27.3 Continuous neutral fat oil saponification plant. (*Courtesy G. Mazzoni S.p.A., Busto Arsizio, Italy*)

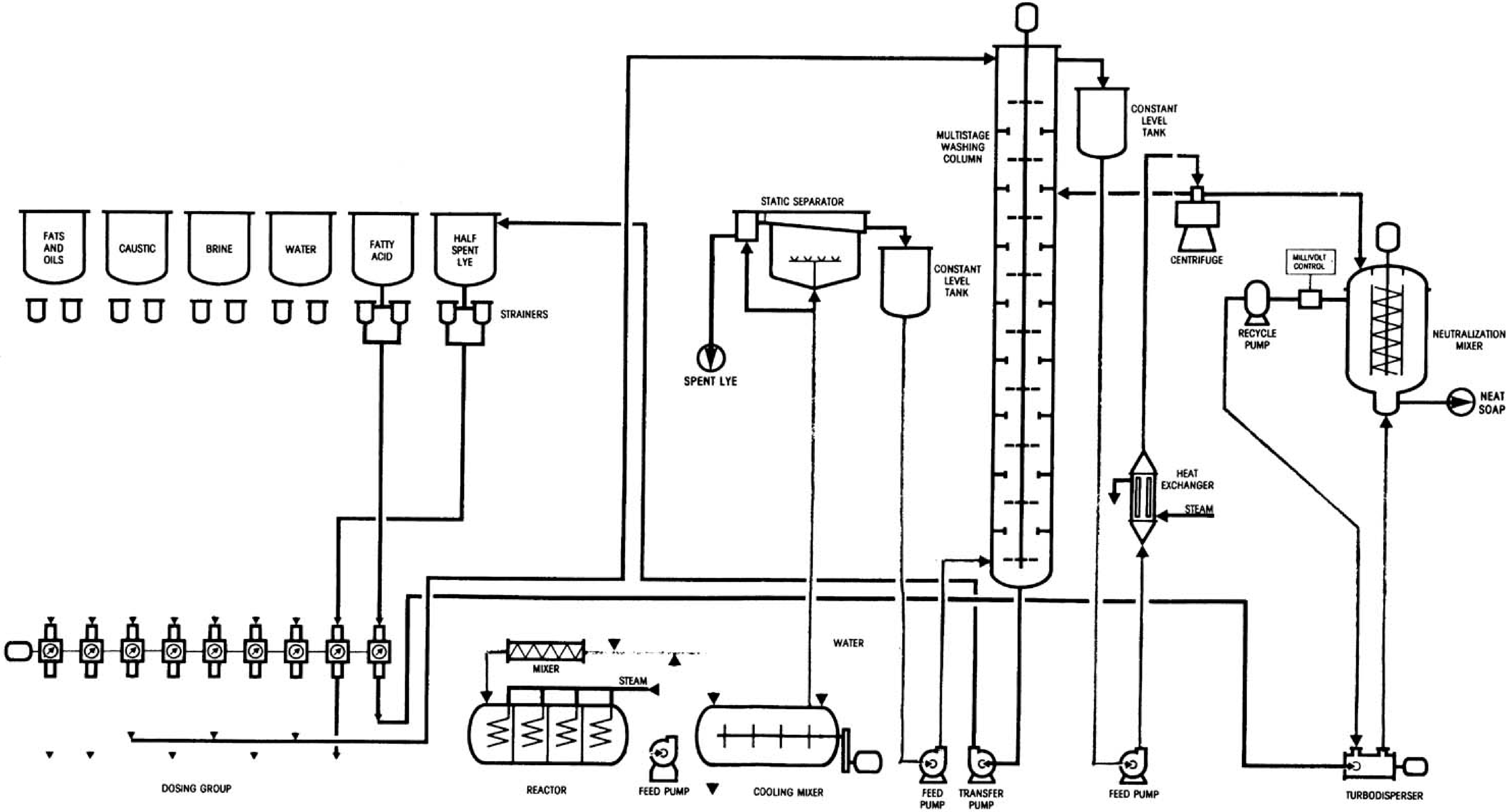
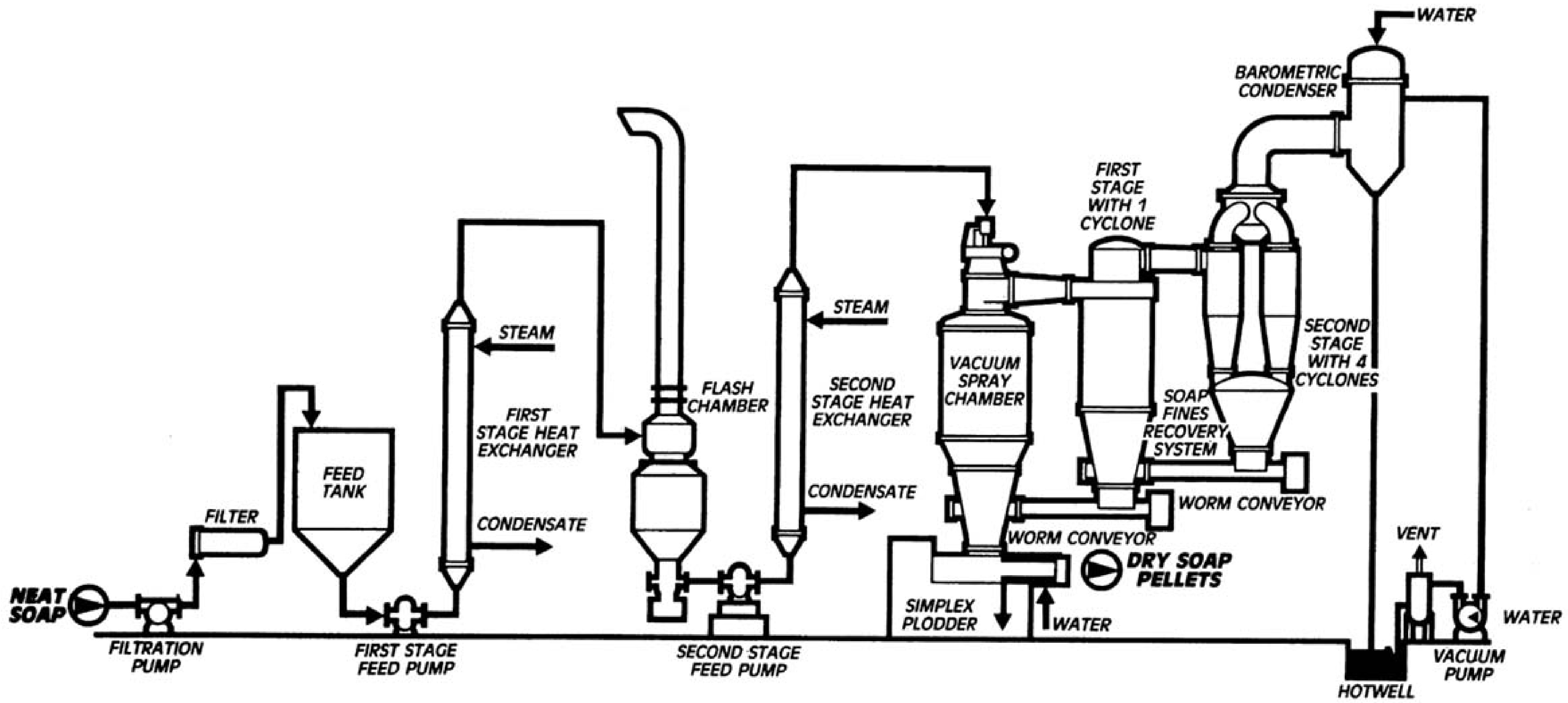


Fig. 27.4 Two-stage toilet-soap spray dryer. (*Courtesy G. Mazzoni S.p.A., Busto Arsizio, Italy*)



Soap Drying

The soap mass obtained after the completion of saponification (neat soap) is most commonly dried nowadays on vacuum spray dryers. The moisture content of soap is thus reduced from 3035 percent in neat soap to 818 percent in soap pellets (soap flakes or chips are produced on APV-type dryers, after milling). A variety of vacuum spray-dryers, from single-stage to multi-stage designs, are available from several manufacturers. The operation of a single-stage vacuum spray-dryer involves the pumping of neat soap at 8085°C through a shell-and-tube heat exchanger where the soap is heated by high-pressure (610 bar) steam passing through the outside tube in a countercurrent manner. The soap is preheated to 135155°C and sprayed onto the walls of a cylindrical vacuum chamber through a revolving nozzle. The thin layer of dried and partially cooled soap deposited on the walls of the vacuum chamber is removed by rotating scrapers and falls to the bottom of the chamber onto a discharge plodder. This plodder extrudes the dried soap mass in the form of noodles or pellets.

The excess moisture is flashed under 4560 mm Hg vacuum and removed as vapor after scrubbing to recover any soap fines carried-over with the exiting vapor stream. Proper control of moisture, temperature, vacuum, flow rate, nozzle pressure, and scraper function ensure the drying of acceptable quality soap pellets. Typically these soap pellets are conveyed pneumatically to storage bins/silos that act as surge for the soap finishing process. Soap storage silos are constructed of stainless steel and have angled bottoms that direct soap pellets toward centrally located twin discharge screw conveyers for efficient transfer to the finishing lines.

Fatty Acid Neutralization

The acidbase reaction of fatty acids with alkali to produce soap (below) is much faster than the corresponding reaction of triglycerides with alkalis.

$$\mathrm{RCOOH} + \mathrm{NaOH} \rightarrow \mathrm{RCOONa} + \mathrm{H_2O} \qquad (27\text{-}2)$$

The amount of alkali (NaOH) required to neutralize a fatty acid blend can be calculated as follows:

$$\begin{aligned}\mathrm{NaOH} \quad &= \quad [\text{weight fatty acid} \\ &\quad \times 40]/[\text{MW fatty acid}] \qquad (27\text{-}3) \\ \text{or NaOH} \quad &= \quad \text{weight of fatty acid} \times \mathrm{AV} \\ &\quad \times 0.000713\end{aligned}$$

where AV = acid value (mg KOH/g of fatty acid), and the remaining portion of the equation being derived as in (27-1).

The average molecular weight of a fatty acid blend is calculated from the following equation:

$$\begin{aligned}&\text{MW fatty acid} \\ &\quad = 56.1 \times 1000/\mathrm{AV} \\ &\quad = (\text{MW KOH}/\mathrm{AV}) \times (1000\,\mathrm{mg/g}) \\ &\qquad \times (1\text{ mole KOH}/1\text{ mole fatty acid})\end{aligned}$$

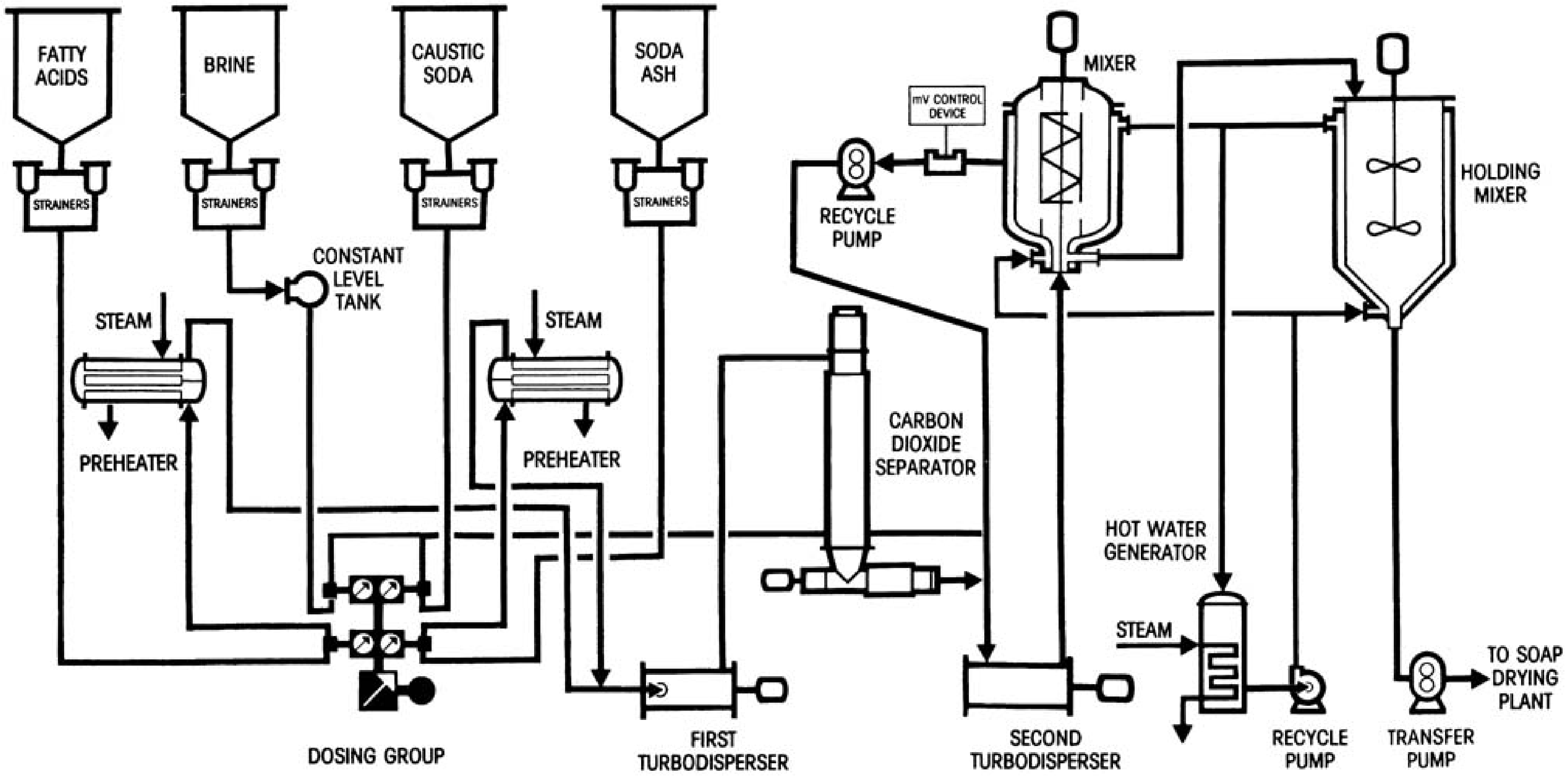
The average molecular weight of a fatty acid blend is calculated from the following equation:

$$\text{MW fatty acid} = 56.1 \times 1000/\mathrm{AV} \qquad (27\text{-}4)$$

where AV (acid value of fatty acid blend) = mg of KOH required to neutralize 1 g of fatty acid.

A typical Mazzoni fatty acid neutralization soap plant is illustrated in Fig. 27.5.6 The operation of this plant involves the pumping of the reactants through preheaters to a turbodisperser, or high-shear mixer, where the contact of the reactants with each other initiates the formation of neat soap. The soap mass, which is partially reacted at this stage, then proceeds to the mixer where it is recirculated until the neutralization is complete. The completion of the neutralization reaction is monitored continuously by an electric potential (millivolt, mV) measurement for alkalinity. The neat soap is then dried by vacuum spray-dryers, as described earlier, to produce soap pellets ready for finishing into soap bars.

Fig. 27.5 Continuous fatty acid neutralization plant. (*Courtesy G. Mazzoni S.p.A., Busto Arsizio, Italy*)

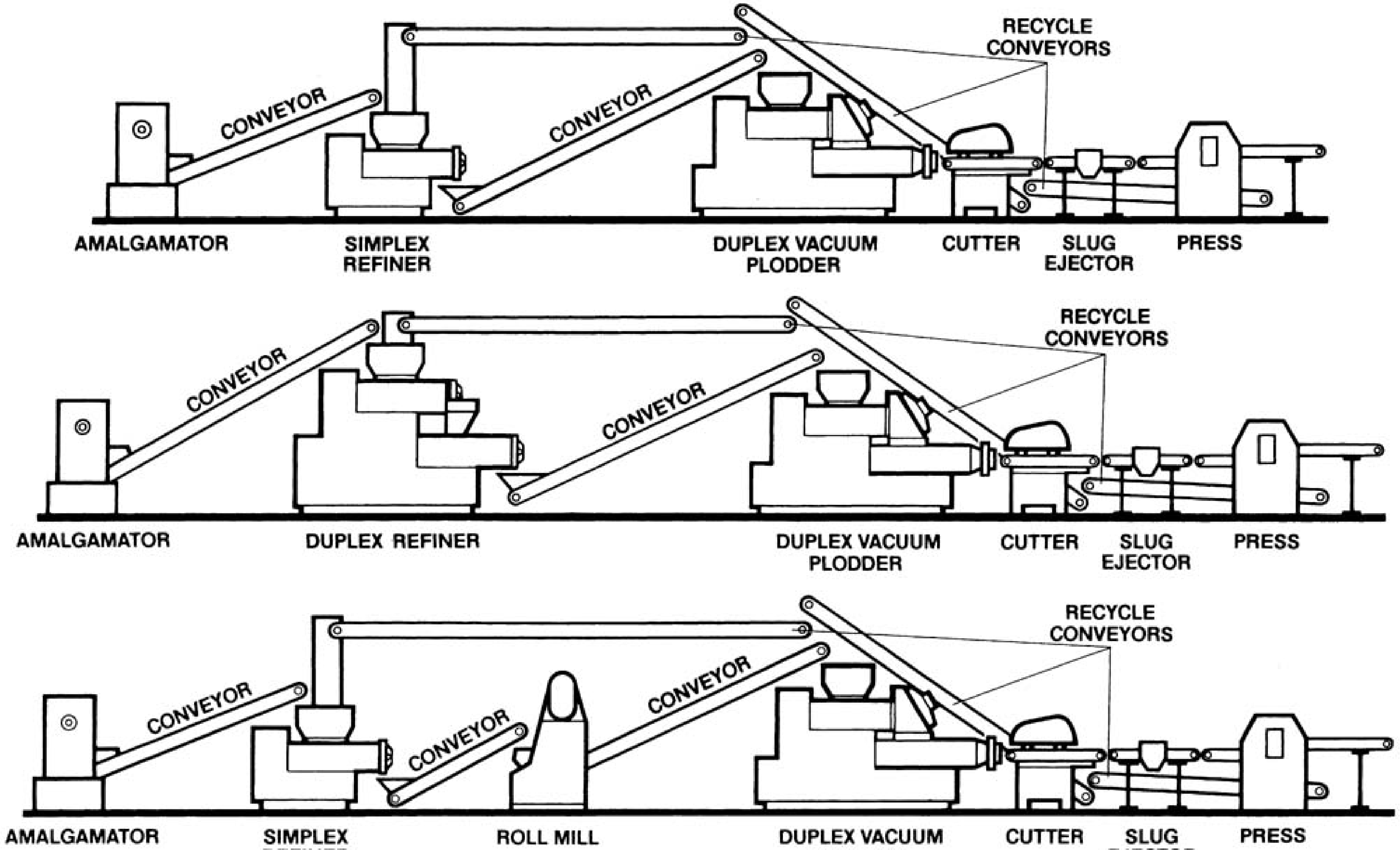


Soap Finishing

The starting point for soap finishing is usually the amalgamator, a horizontal mixer that coarsely blends soap pellets with minor ingredients such as color and fragrance. The soap mix from amalgamation is conveyed to subsequent processing stages like milling and refining to provide satisfactory homogenization and working of the soap before extrusion. A mixer/refiner process is also available for blending of soap pellets with minor ingredients on a more continuous basis. Sufficient working of the soap through fine-mesh screens in the refining stages provides uniformity and frictional heat necessary to produce acceptable-quality billets for pressing into bars. The finished product is extruded from the last plodder as a continuous slug. Vacuum is provided at the final extrusion chamber to evacuate entrained air. This step ensures the extrusion of a uniform slug that will not separate or exhibit surface bubbles/deformities. A multi-blade chain cutter or uni-blade cutter with blades separates the extruded slug into individual "billets" that are stamped with a press into soap bars of the desired size and shape.

The wrapping, cartoning, and end packaging machinery complete the bar finishing lines. The schematic diagrams of a variety of soap finishing configurations are illustrated in Fig. 27.6.

Fig. 27.6 Toilet-soap finishing line. (*Courtesy G. Mazzoni S.p.A., Busto Arsizio, Italy*)



Fatty Acids

The fatty acids are produced industrially from triglycerides. This process involves the steps shown in Fig. 27.7.7a, b

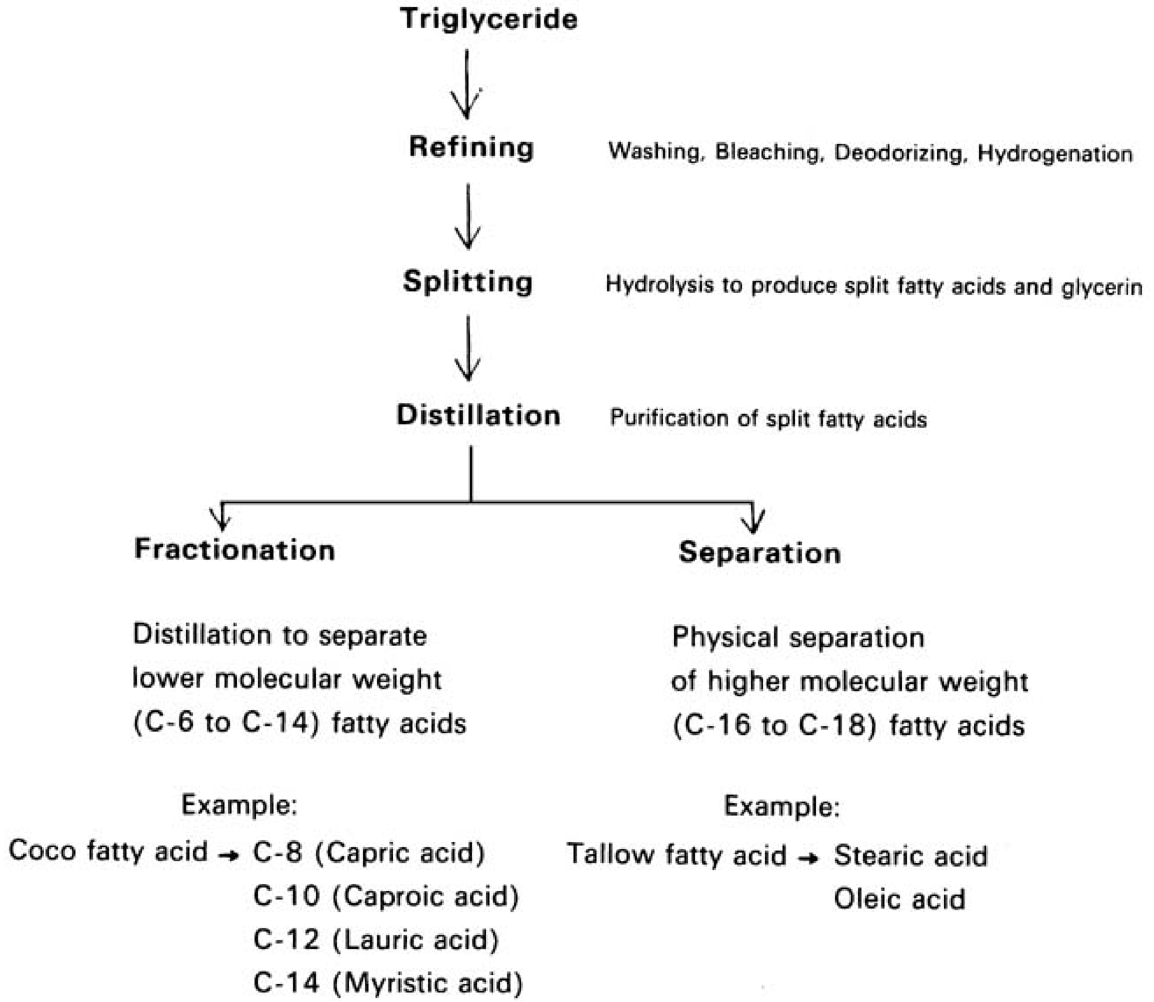


Fig. 27.7 Production of fatty acids from triglycerides.

Refining involves the purification of triglyceride to remove impurities (phosphatides, polyethylene, chlorophyll, heavy metals, off-odors, color bodies) by a combination of acid/alkali washing, clay/activated silica bleaching, deodorization, and hydrogenation steps.

Splitting, or the hydrolysis of triglycerides is usually performed with high-pressure steam, resulting in the formation of split crude fatty acids and glycerin. The production of fatty acids by more sophisticated "splitting" processes, such as hydrolysis of fatty methyl esters, ozonolysis of unsaturated fatty acids, and chemical oxidation is practiced in special situations.

Distillation of split fatty acids results in the improvement of chemical, color, and odor quality of the finished product.

Fractionation and Physical Separation

Distilled fatty acid streams are separated to obtain individual components of those fatty acid streams in a higher state of chemical purity. Fractionation is performed to separate the more volatile lower molecular weight fatty acids (C8C14 chain length). Physical separation is practiced to obtain higher molecular weight, less volatile, or thermally unstable fatty acids (C16C18 chain length).

Hydrogenation is utilized to reduce the unsaturation and poly-unsaturation of fatty acids. Hydrogenated fatty acids can also be produced by the hydrogenation of triglycerides followed by splitting of the resulting fat/oil blend. The selective reduction of di- and poly-unsaturation is accomplished frequently by the use of nickel-based specialized catalysts in a process often called "touch" hydrogenation.

In the partial hydrogenation of the unsaturated moities the isomerization of the double bonds of mono-unsaturated components is a common occurrence; it results in the transformation of the *cis* to the corresponding *trans* stereoisomer. As the physical properties of these isomers may vary significantly, their impact on the properties of the final product should be evaluated rather carefully, for little information is available on this subject at present.

Fat Refining

A typical fat refining plant (Alfa-Laval process) for the acid washing of fats and oils is illustrated in Fig. 27.8.8 The triglyceride is degummed to remove phosphatides and other impurities like mucilage, proteinaceous matter, and trace metals by acid washing with citric or phosphoric acids, and then is sent to a bleaching plant where it is dehydrated and treated with an activated clay for the removal of color bodies, heavy metals, chlorophyll, and polyethylene (Fig. 27.9).6

Fig. 27.8 Acid degumming plant for acid washing offats and oils. (*Courtesy Alfa-Laval, Tumba, Sweden*)

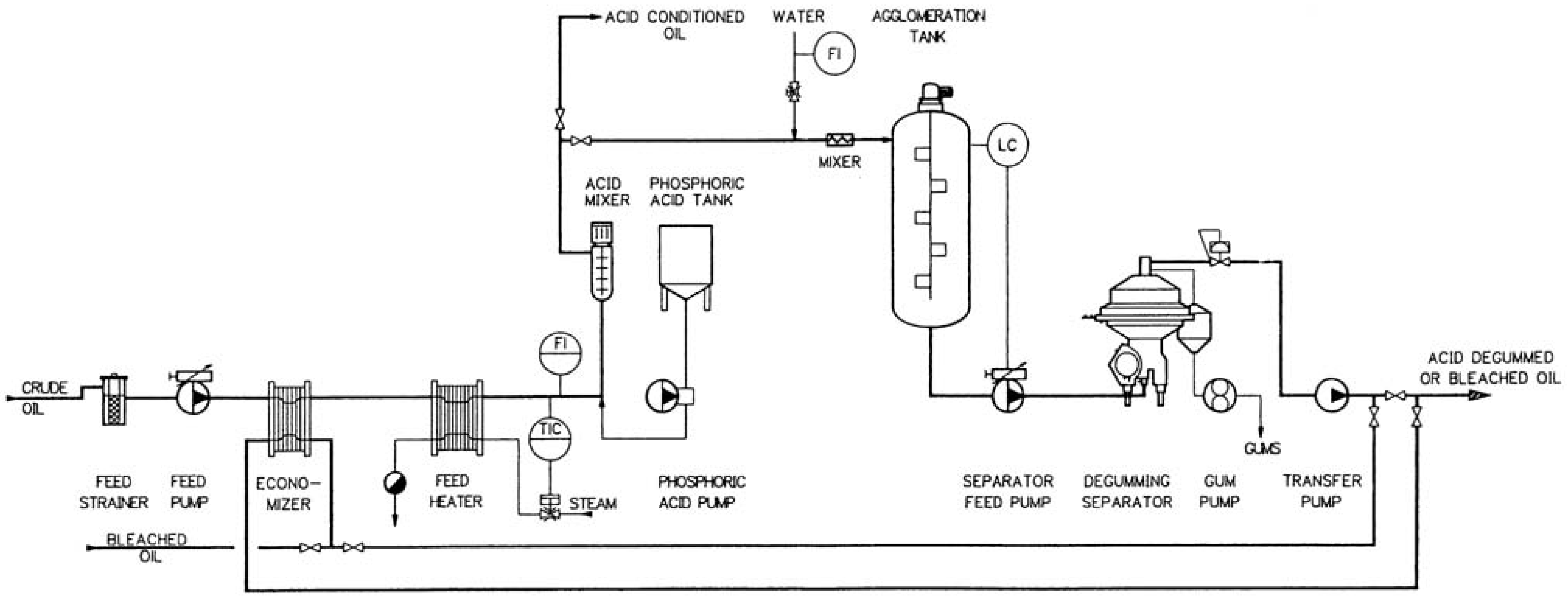
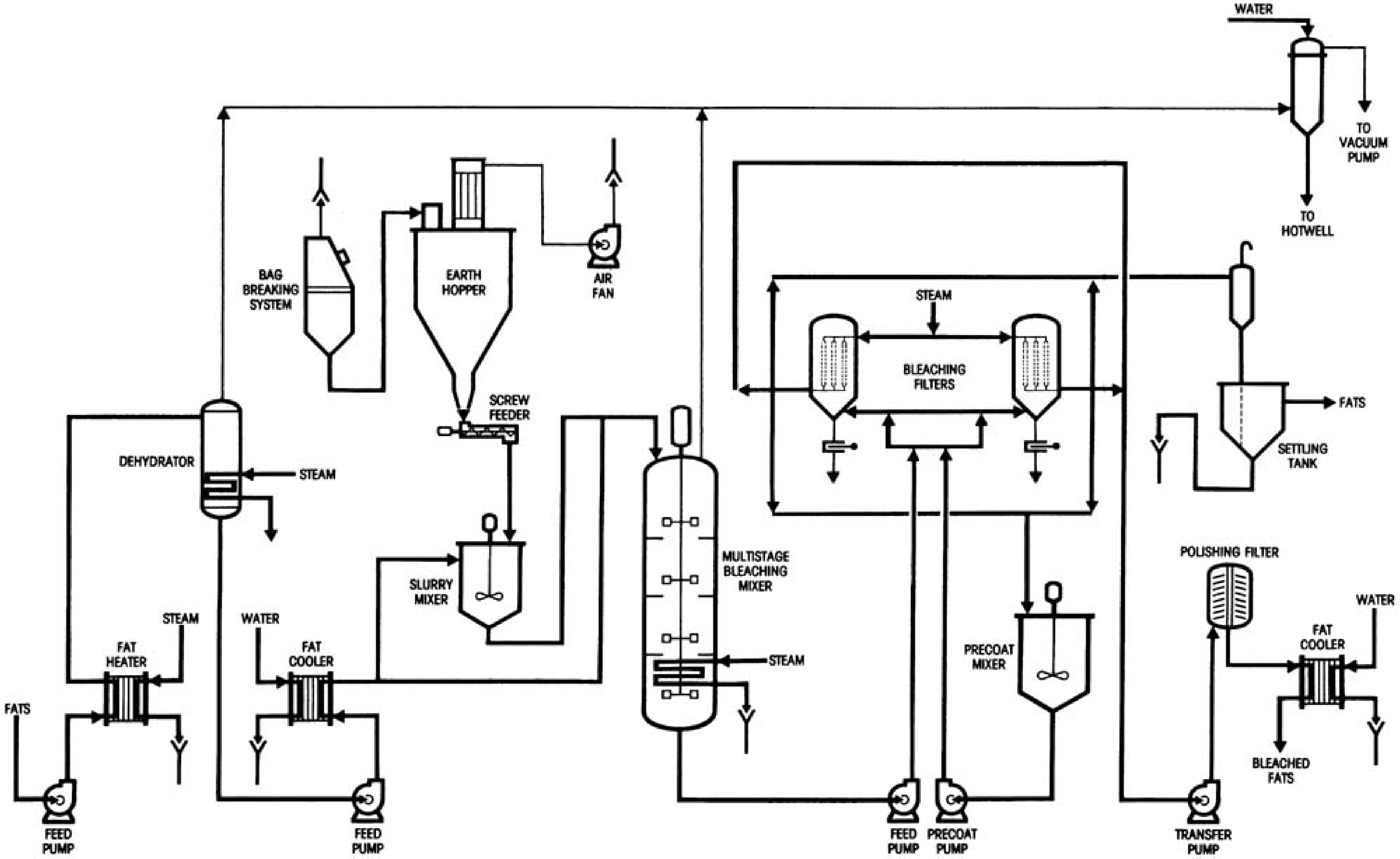


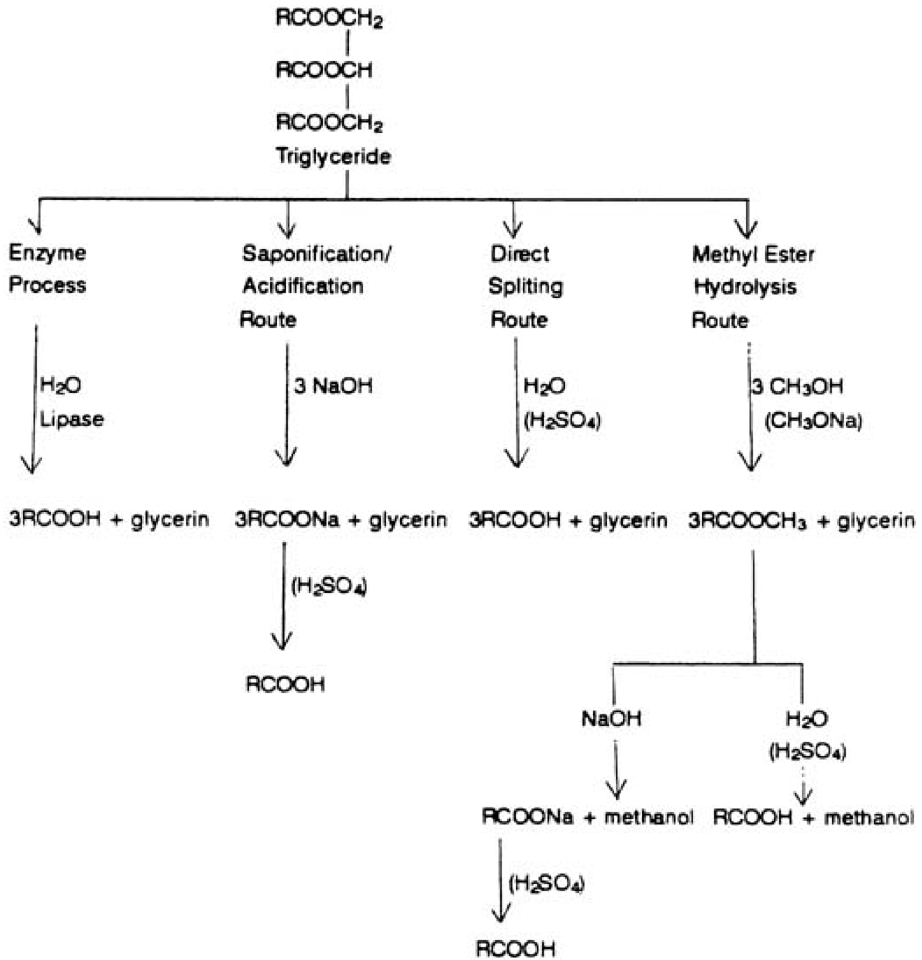
Fig. 27.9 Continuous fatty matter bleaching plant. (*Courtesy G. Mazzoni S.p.A., Busto Arsizio, Italy*)



Fat Splitting Processes

The direct reaction of triglycerides with alkali followed by acidification or the treatment of triglyceride with water (with or without an acid catalyst) results in the splitting of fats/oils into fatty acids and glycerin. The splitting, or hydrolysis of fats/oils with water can also be performed by using various lipase enzymes as catalysts. The enzyme process is still in the developmental stages. The inter-esterification of triglycerides with methanol followed by acid or alkali cleavage of the resulting fatty methyl ester is used for the production of certain molecular weight (C8C10) fatty acids; those methyl esters are generally the by-products of fatty alcohol production (Fig. 27.10).

Fig. 27.10 Fat splitting processes.



The splitting of fats/oils can be done either in a batch (Twitchell process) or in a continuous (Colgate-Emery process) mode; the enzyme-based processes are still under development for commercial applications. A comparison of these processes follows.9a, b

Process	Hydrolysis Time (hr)	Conversion (%)
Twitchell batch	824	95
Continuous	13	9899
Enzyme	72	98

The manufacture of fatty acids from triglycerides via their reaction with alkali followed by acidification is not used commercially. However, the recovery of fatty acids from alkali refining of fats/oils (to produce acid soaps) and soap reboiling operations (the recovery of fatty acids from the acidification of soap waste streams) are practiced still in special situations, for example, in edible oils refining where soapstock that is generated from chemical (alkali) refining requires acidulation to produce acid oils with commercial value.

Industrially, the fatty acids are produced predominantly by the high-pressure, high-temperature splitting of triglycerides. Both batch and continuous processes are available, the continuous process being in greater worldwide use today.

Batch Process

The triglyceride is heated with water at 150250°C and 1025 bar pressure. The splitting is achieved in 610 hr, depending on the nature of the triglyceride feed, for example, the higher the molecular weight of the triglyceride, the slower the splitting reaction. The reaction can be accelerated by increasing the temperature; under high pressure and at 225260°C, no catalyst is required.

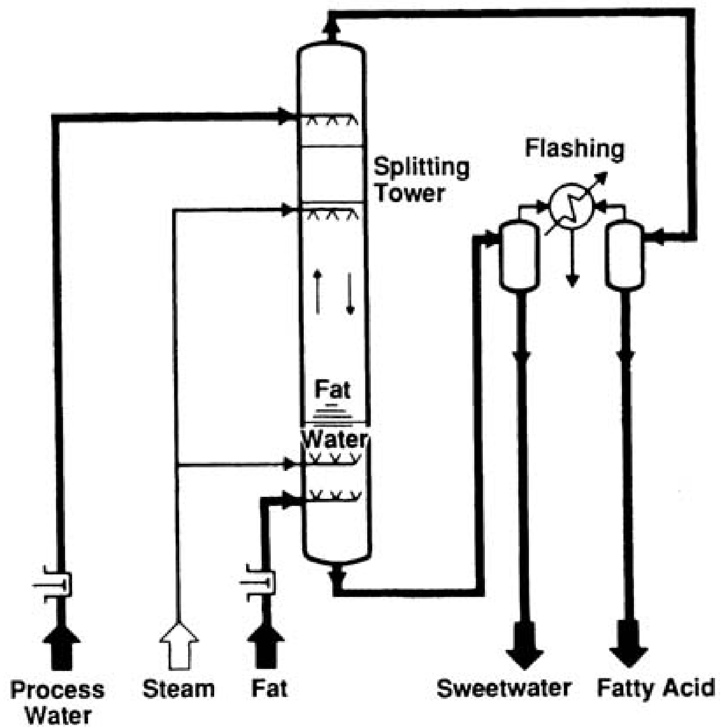
A typical commercial batch splitting plant (Lurgi) operates at 230°C/400 psig with over 96 percent split yields after two changes of water to remove the glycerin by-product. The sweetwater (glycerin/water mixture) is circulated continuously from the base to the top of the autoclave in this process.10

A number of other batch fat splitting plants is also available.

Continuous Fat Splitting Process

Currently, continuous fat splitting is the most widely used process in the industry. In the single-stage splitting process, fat and water are fed countercurrently into a continuous flow column at up to 260°C/60 bar pressure conditions. The glycerin by-product generated in the splitting process is removed by water (sweetwater), thus accelerating the reaction equilibrium toward completion (Fig. 27.11).10 The high splitting temperature ensures adequate dissolution of the water phase in the fat to the extent that mechanical forces for contacting the two reactants are not required. The fat phase passes from bottom to top through the tower, with the void volume of the tower acting as a reaction chamber. The water phase travels downward through the mixture of fat and fatty acid, concentrating the glycerin by-product in the sweetwater phase. The top and bottom sections of the tower are also designed to effect the transfer of heat from the products to the incoming reactant streams. Split yields of up to 99 percent can be achieved by the use of pre-purified fat/oil feedstocks.

Fig. 27.11 Single-stage countercurrent fat splitting plant. (*Courtesy Lurgi GmbH, Frankfurt am Main, Germany*)



In summary, the continuous high-pressure splitting of fats and oils involves:

1. The use of excess water to drive the reaction to completion and for removal of glycerin from the reaction zone.
2. The use of a proper combination of temperature and pressure conditions to increase the solubility of water in the fat phase while maintaining the boundaries of the fat/water interface. These conditions also result in an increase in the rate of splitting.
3. The possible use of an acid catalyst to accelerate the reaction rate.

Saponification/Acidification Process

The recovery of fatty acids from soapstocks and soap reboil operations is of commercial value:

$$\underset{\text{(unrefined)}}{\text{Triglyceride}} \xrightarrow{\text{caustic refining}}$$

$$\text{Triglyceride} + \underset{\text{(soapstock)}}{\text{RCOONa}} \xrightarrow{(H_2SO_4)} \text{RCOOH}$$

$$\underset{\text{(impure soap)}}{\text{RCOONa}} \xrightarrow{\text{reboiling}}$$

(impure soap)

$$\underset{\text{(purified neat soap)}}{RCOONa} \xrightarrow{(H_2SO_4)} \underset{\text{(purified fatty acid)}}{RCOOH}$$

The recovery of fatty acids from soapstocks by a continuous process has been described: soapstocks obtained from degumming and alkali refining operations are subjected to a saponification step followed by controlled acidification for cost efficiency and pollution control.11a, b

The saponification/acidification route is used for the manufacture of fatty acids that are sensitive to excessive heat. The splitting of castor oil with NaOH at 100220°C followed by acidification has been used to produce castor fatty acids.12

Enzymatic Processes

The enzyme hydrolysis of triglycerides is catalyzed by lipases:

$$\text{Triglyceride} + 3H_2O \xrightarrow{\text{enzyme}} 3RCOOH + \text{Glycerin}$$

Enzyme-based splitting of triglycerides is under active development and offers the following advantages over modern steam hydrolysis plants:

1. the hydrolysis occurs under mild conditions (3537°C, ambient pressures)
2. the color quality of the products is superior
3. the capital investment may be lower
4. the split yields are equivalent

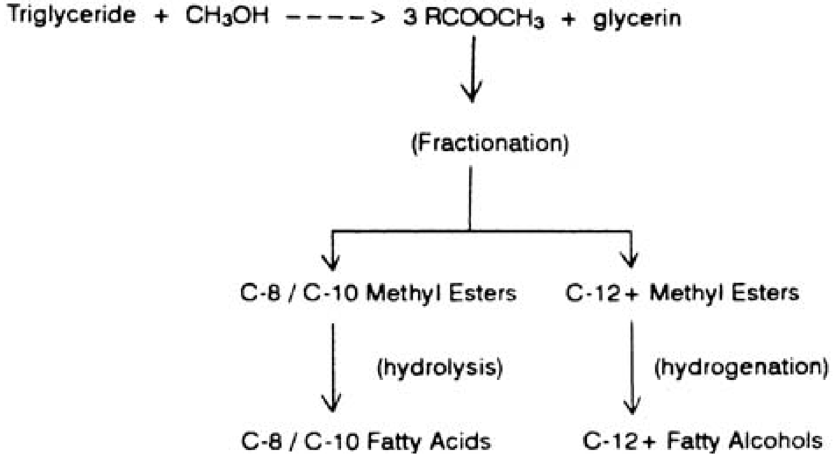
The selection of a suitable lipase for a commercially viable process is based on the following criteria:

1. it should be nonspecific and nonselective as to the structure of the triglyceride
2. it should be stable at above ambient processing temperatures, typically 4565°C
3. it should be economically viable (This aspect has initiated a number of ongoing studies on the recycling and immobilization issues of lipase utilization for triglyceride hydrolysis.)14a, b

Methyl Ester-Based Processes

The fatty methyl esters are produced predominantly by the transesterification of fats and oils with methanol in the presence of an alkaline catalyst under very mild reaction conditions.15a, b They are used in the production of lauric-type (C12) alcohols. The short-chain fatty methyl esters (C8C10), produced as by-products via the fractional distillation of crude lauric-type (coconut, palm kernel) methyl esters, are converted to fatty acids via acidic or alkaline hydrolysis (Fig. 27.12). The hydrolysis of short-chain fatty methyl esters by stream splitting or Twitchell-type processes is not very efficient because of unfavorable equilibrium constants.16a, b

Fig. 27.12 Fatty acid production via methyl esters.



Fatty Acid Distillation

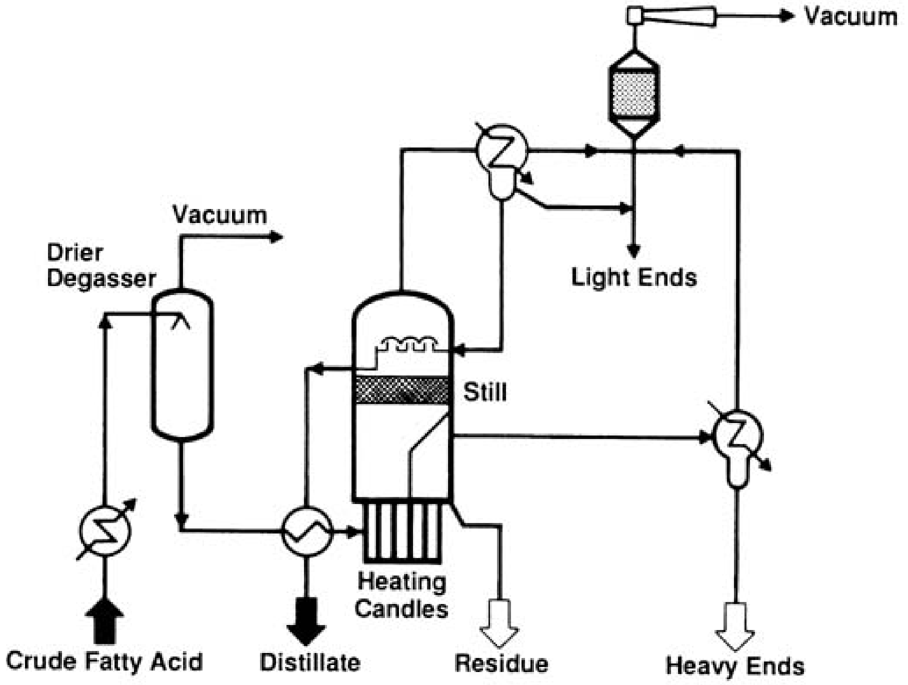
The removal of contaminants (unreacted triglycerides, odor and color bodies, polymerized matter, and decomposition products) from split fatty acids is achieved by simple distillation.17a, b The products obtained from such straight distillations are called "whole cut" fatty acids (e.g., whole coconut fatty acid).

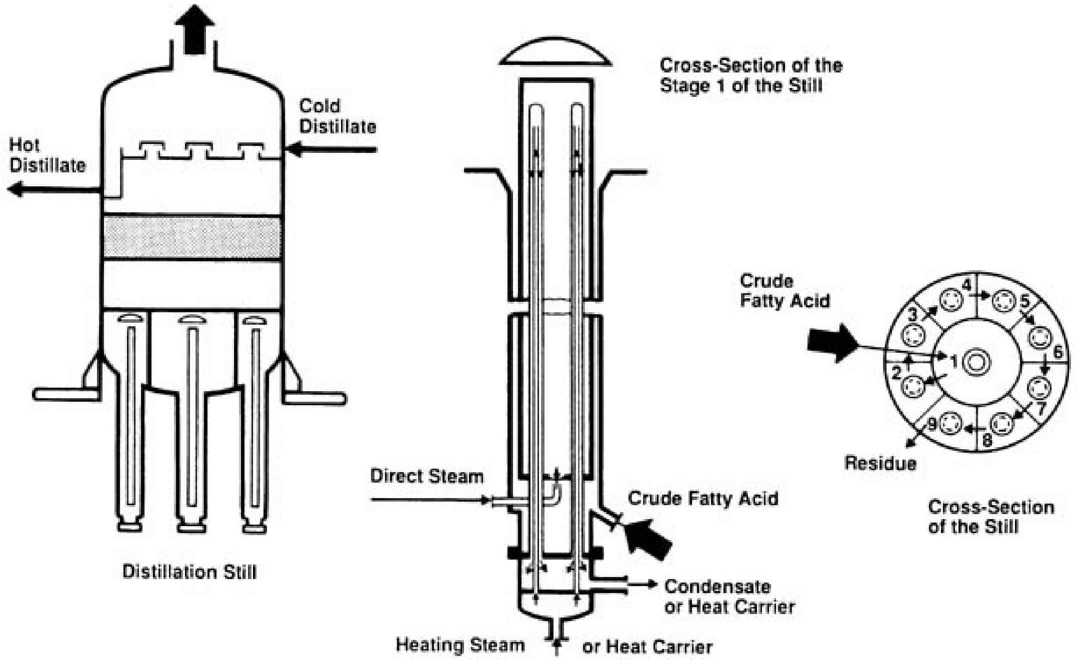
The separation of individual fatty acid components from whole fatty acid distillates is achieved by fractional distillation.18a, b However, fractional distillation cannot efficiently separate C16C18 saturated from the corresponding unsaturated fatty acids. In such cases, the melting points vary considerably with the degree of unsaturation and this property is used for their separation via a number of industrial processes.19a, b

Distillation

Straight distillation is used for the removal of low-boiling (top cuts) and high-boiling (pitch or bottoms) impurities from the crude split fatty acid stock. The distillation can be done either in a batch or a continuous manner. Generally batch distillation is employed for small-capacity manufacturing; continuous distillation is utilized for large-scale manufacturing and is highly automated (Fig. 27.13)10

Fig. 27.13 Continuous fatty acid distillation plant. (*Courtesy Lurgi GmbH, Frankfurt am Main, Germany*)





The split fatty acid (deaerated and preheated) is sent to a distillation column (Fig. 27.13) where it is vaporized rapidly by flowing over heated trays under high vacuum (10 mbar) with the addition of live steam (for fatty acid circulation and reduction of the partial pressure). The high boilers are removed by two stages of entrapment: the high boilers and bottoms from the distillation column pass through additional heated trays and a reboiler section to vaporize residual fatty acids from the bottoms; the light end condenser traps the low boilers and the distilled product. The heavy "bottoms" distillate is collected at the bottom section of the still.

The distillation process is not used for the narrow-cut fractionation. However, by passing a second time through the distillation tower the main-cut distillates may be separated into a light and heavy fraction; the light end is collected at the top products section and the heavy end is collected at the bottom residue section of the still. The separation of light and heavy fractions in the first pass is possible for high lauric-type split fatty acids, for example, coconut and palm kernel fatty acids where the lower molecular weight C6C10 fraction is "topped" from the heavier C12C18 fraction.

The operation of a Lurgi fatty acid distillation system10 proceeds as follows. The crude fatty acid is dried and degassed under vacuum. This dried and deaerated crude fatty acid is pumped into the distillation chamber (vacuum of 1012 mbar) that is divided into 510 chambers by vertical plates. Each chamber is equipped with a heating element and a baffle. Live steam is passed into the system to effect recycling of the liquid between the heating elements and the baffles by the principle of an air-lift pump. The fatty acid is sprayed on the baffles as a fine mist to allow it to evaporate easily; non-vaporized liquid is recycled from the bottom. The fatty acid passes through each chamber, from which it is discharged continuously. The fatty acid vapor is condensed in the surface condensers. Typical yields from the fatty acid distillation process are:

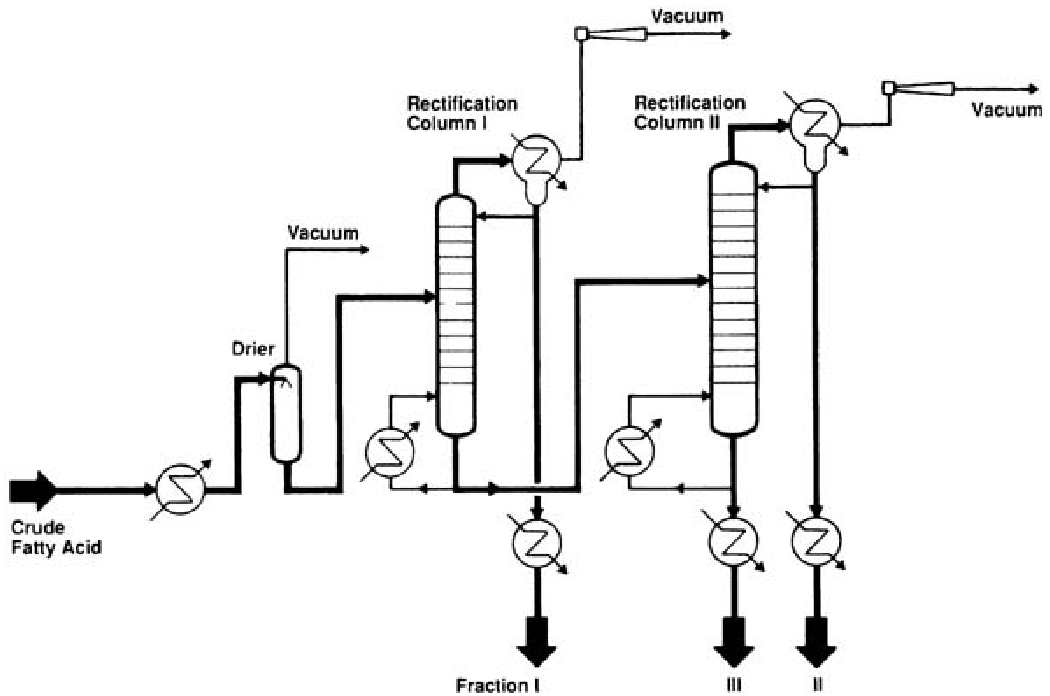
Tops or head cut	15%
Refined product	8591%
Residue	711%

Fractionation

The separation of individual components of a fatty acid mixture into high purity fractions (98% +) is achieved by fractional distillation. In a continuous fractionation still the fractionation column can be equipped with trays, in the "Thormann" system, or with the traditional packing to effect contact between the upward-moving vapors and the downward-flowing condensing liquid states.

A typical high-efficiency fractionation system (Fig. 27.14) consists of one continuous straight distillation and one, or preferably two fractionation columns.10,17a, b This system is very efficient for the separation of low-boiling fatty acid components,6,10 a C12 fraction of 99.5 percent purity from coconut fatty acid or a C22 fraction of 95 percent purity from rapeseed fatty acid is readily obtained.

Fig. 27.14 Fatty acid fractionation. (*Courtesy Lurgi GmbH, Frankfurt am Main, Germany*)



From an operational perspective, in rectification column 1 of Fig. 27.14, crude fatty acid is dried and vaporized with the addition of live steam under a vacuum of 840 mbar. A mixed fraction is withdrawn from the column top. The product from the bottom of column 1 is vaporized in rectification column 2 with live steam under a pressure of 68 mbar. The low-boiling fraction is collected in downstream scrubbers after the vapor has been condensed in surface condensers.

The separation scheme for coconut fatty acid is shown in Fig. 27.7. A variety of alternate fractionation schemes can readily be devised.

Fatty acid distillation has been done for over 100 years.17a, b However, most of the currently practiced art has originated in the past several years.18a, b The majority of the recent work falls in the following areas:

1. designs with gentle heating
2. improvements in the column design to reduce residence time
3. enhancements of system design to minimize pressure drops
4. demister designs for better air/effluent water pollution control

Special Separation Methods

The separation of long-chain saturated fatty acids (C16C18) from the corresponding unsaturated fatty acids by distillation is not practical because of the proximity of their boiling points. However, both the melting points and the solubility (in organic solvents) of these two types of fatty acids are vastly different.20 These properties are used to advantage for separating mixtures of saturated and unsaturated fatty acids into fractions that are rich in either saturated (stearin) or unsaturated (olein) components.19a, b

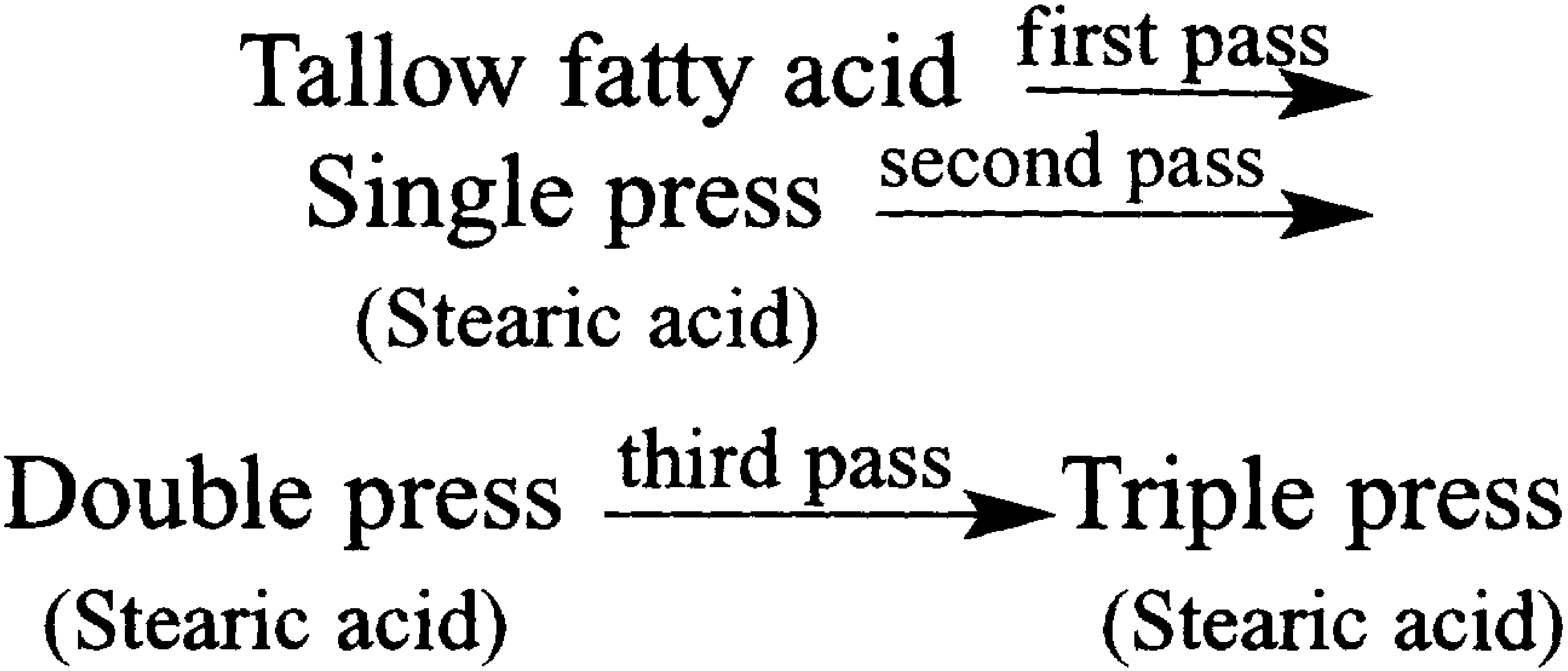
$$\text{Tallow fatty acid} \xrightarrow{\text{separation}} \underset{\text{(Stearin)}}{\text{Stearic acid}} + \underset{\text{(Olein)}}{\text{Oleic acid}}$$

The following processes are used commercially for the separation of such fatty acids: mechanical pressing, solvent crystallization, and hydrophilization.

Mechanical Pressing

This method, also called panning or pressing, is the simplest and oldest method, and is still practiced in many developing countries. In this method, fatty acids are melted in shallow pans and then allowed to cool. The material transforms into a crystalline solid with a liquid phase trapped between the crystals of the solid phase. The liquid portion (olein) is pressed, or squeezed out by the application of mechanical force, separating it from the solid (stearin) component.

The pressing can be done by hydraulic presses in either a batch or a continuous mode. A multiple number of melting/crystallization/pressing cycles is utilized for obtaining fractions of increased purity:



The unsaturation, as determined by the iodine value (IV), decreases from an IV of 10.0 for a single press to an IV of 1.0 for a triple press stearic acid grade.

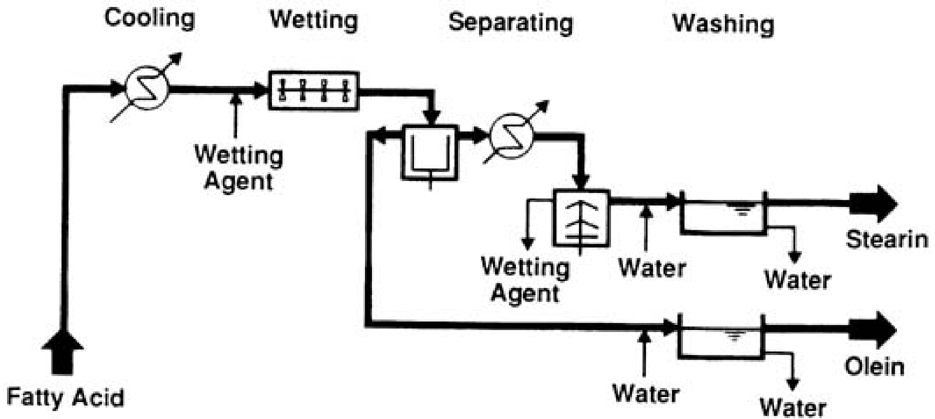
Solvent Crystallization

Two processes, one utilizing acetone (Armour) and the other employing methanol (Emery), are well known. Using the latter, fatty acid is dissolved in 90 percent aqueous methanol in a 1:2 acid/solvent ratio by the application of heat. The resulting solution is then cooled to -15°C in a multi-tubular crystallization chamber equipped with scrapers for efficient heat transfer. The crystallized fatty acids are removed by filtration. The filter cake is melted and stripped of any residual solvent to yield the refined stearin fraction, and then the liquid stearin is converted to flakes or powder by a variety of processes, for example, chill roller, and the like. The mother liquor from the filtration is stripped to obtain the olein fraction. The separated stearin and olein fractions have a variety of commercial applications in both the chemical and food processing industries.

Hydrophilization Process

In this process, olein and stearin are separated by forming a suspension of the stearin moiety in water with the aid of a surfactant followed by filtration to remove the stearin (Fig. 27.15). The mixture of fatty acids is melted and then partially crystallized by cooling. The resulting slurry, which is a mixture of liquid olein and crystalline solid stearin, is mixed with water containing a wetting agent (sodium dodecylbenzene sulfonate, sodium lauryl sulfate). After mixing for a specified period, the slurry is centrifuged; the light fraction provides the liquid olein moiety, the middle fraction is a suspension of solid stearin in water and the lower fraction is a water/surfactant solution.10

Fig. 27.15 Hydrophilization process for separation of olein and stearin. (*Courtesy Lurgi GmbH, Frankfurt am Main, Germany*)



Thermodynamically the fatty acid crystals, after being wetted by the surfactant solution, transfer from the liquid oil into the water solution. The oil droplets then coalesce to give the lighter olein phase. The separated stearin is remelted to remove residual water/surfactant and then reprocessed to the desired level of purity.

Although the quality of the final product from these three processes is fairly comparable, currently hydrophilization is the most widely used process for the separation of stearic and oleic acids.

27.2 Synthetic Surfactants Detergents

Synthetic detergents were initially developed to replace soap in an economy that was running short of edible fats and oils. It was soon discovered that the synthetic compounds were more resistant to hard water ions and provided improved wetting and detergency over soap. Over the years, the term "synthetic detergent" has been shortened to "syndets" to describe detergent compositions comprising synthetic detergents, hard water control agents (builders), and other detergent additives. The term "surface active agent" has been shortened to "surfactant" to describe the surface active ingredients. Since their introduction in the 1940s, synthetic detergents have continued to grow in both total volume and the percentage of detergent use as shown in Table 27.2.

TABLE 27.2 U.S. Consumption of Soap and Synthetic Detergent

Year	Soap (MM lbs)	Synthetic Detergent (MM lbs)	Total
1940	2306	30	2336
1950	2882	1443	4325
1960	1230	3940	5170
1970	1050	5650	6700
1980	1300	6400	7700
1990	1625	9000	10625

Source: SRIC's Chemical Economics Handbook, based on a compilation from various sources. A personal communication (08/14/02) from SRI states that because of the variety of forms and concentrations in which detergent products now appear, such data are no longer available. Also, while soap production was relatively flat for many years, it began to decline again in the late 1990s, as the liquid hand "soaps" and "body washes" that are gaining market share are largely based on surfactants. In contrast, the consumption of synthetic detergents has continued to grow at a rate somewhere between population and GDP growth.

Whether in liquid or powdered forms, modern detergent products are complex mixtures of many different ingredients. Typical formulations consist of surfactants, builders, and other additives designed to maximize performance for the consumer while maintaining reasonable raw material and manufacturing costs. Recent trends in both liquids and powders show increased use of multiple builder and surfactant systems.21,22 Performance additives such as bleaches, bleach activators, enzymes, enzyme stabilizers, anti-redeposition polymers, fluorescent whitening agents, dye transfer inhibitors, and fabric softeners are commonly used to increase overall performance. Regardless of the product format or formulation type, surfactants play a critical role in the cleaning process.

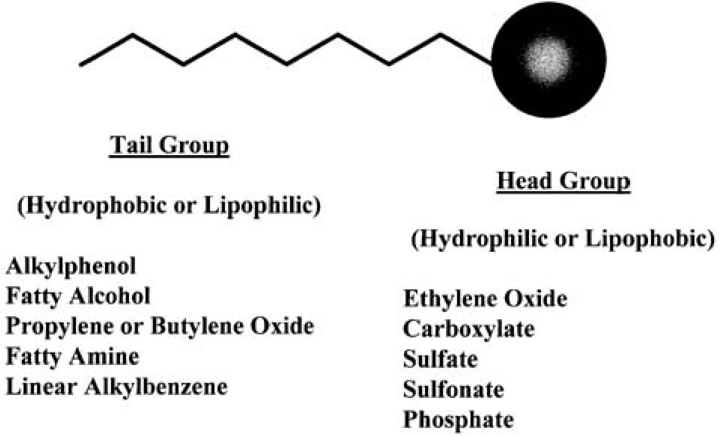
Characteristic Features of Surfactants

Surfactants have a characteristic molecular structure consisting of one structural unit that has very little attraction for the solvent (lyophobic group), combined with a group that has a strong attraction for the solvent (lyophilic group). This duality is known as an amphipathic structure. When a surfactant is dissolved in a solvent, the presence of the lyophobic group distorts the solvent liquid structure, resulting in an increase in the Gibbs free energy of the system. This acts to force the molecules out of the solution in an attempt to decrease the free energy of the system. However, the presence of the lyophilic group prevents the surfactant from being completely expelled from the solvent. The opposing forces give the surfactant molecule their surface active properties. In aqueous solution, the lyophilic group is known as the hydrophile and the lyophobic group is called the hydrophobe.

Surfactant molecules are typically depicted using a ball and stick representation as shown in Fig. 27.16. The stick portion represents the hydrophobic group or "tail" whereas the ball represents the hydrophilic "head group." Examples of common hydrophobes include alkylphenol, fatty alcohol, and alkylbenzene. Examples of common hydrophiles include ethylene oxide (EO), sulfates, sulfonates, and carboxylic acid groups. Depending on the charged nature of the head group, surfactants are classified as:

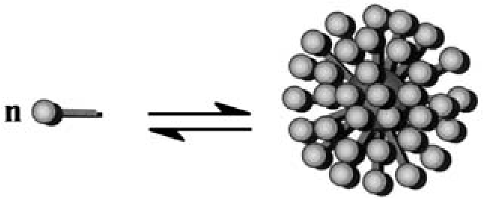
1. Anionic, the head group bears a negative charge.
2. Cationic, the head group bears a positive charge.
3. Zwitterionic, the head group has either a positive or negative charge depending on solution pH.
4. Nonionic, the head group bears no formal ionic charge.

Fig. 27.16 Representation of surfactant amphipatic structure.



Aside from their ability to adsorb at interfaces, the most important aspect of surfactants is their ability to form colloidal-sized aggregates in solution. In dilute solution, the surfactant molecules exist as individual hydrated species. Increasing the concentration promotes the formation of surfactant aggregates or "micelles," as shown in Fig. 27.17. The concentration at which micelles start to form is referred to as the critical micelle concentration (CMC). Micelle formation is an important phenomenon because interfacial effects such as detergency and solubilization depend on the existence of micelles in solution.

Fig. 27.17 Micelle formation in aqueous solution.



Surfactant molecules can be tailored for use in either aqueous or nonaqueous systems, depending on their solubility characteristics. The most common way to describe surfactant solubility is the hydrophilelipophile balance (HLB). In the HLB method, a number is assigned (040) related to the balance between the water-loving (hydrophile) and lipid-loving (lipophile) portions of the molecule.23,24 In some cases the HLB number is calculated from the structure of the molecule, in others it is based on experimental emulsification data. In general, there exists an optimum HLB for different applications as shown in Table 27.3. Matching the HLB of the surfactant to that of the application generally gives the best performance.

TABLE 27.3 Typical HLB Values for Different Applications

HLB Range	Typical Application
13	Reverse micelles
36	W/O emulsifier
710	Wetting agent
818	O/W emulsifier
1115	Detergent
>15	Solubilizing agent

For example, alkylbenzene is slightly surface active in nonaqueous media but is insoluble in water. Addition of a single SO3H group to the molecule imparts solubility and excellent surfactant characteristics in water but insolubility in petroleum solvents. If alkylbenzene is di- or trisulfonated, the compound becomes more and more water soluble and loses much of its surface activity. In both cases, the hydrophile balance is exceeded with insufficient lipophilic influence to cause micelle formation. However, neutralization of the SO3H group of a highly water-soluble alkylbenzene sulfonic acid with a long-chain amine produces an effective nonaqueous surfactant. A similar effect of decreasing water solubility and increasing nonaqueous detergency can be achieved by increasing the alkylbenzene carbon chain length from C12 to C18 or higher.

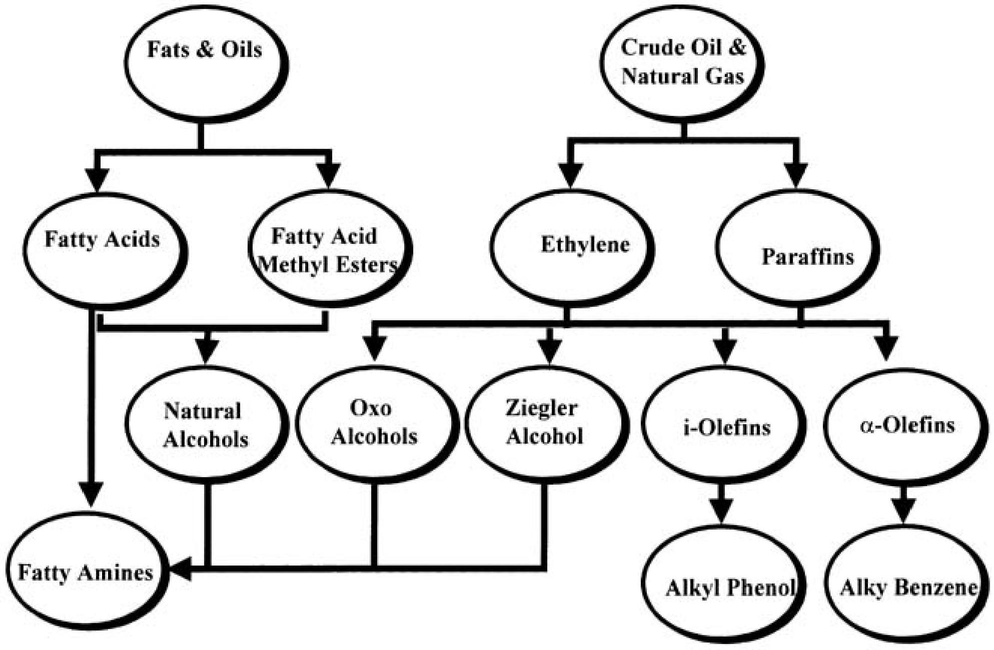
Most commercial nonionic surfactants are based on ethylene oxide (EO), which depend upon a multiplicity of hydrogen and oxygen bond linkages to effect the HLB. The carbon chain length and degree of branching can be varied to impact the hydrophobic characteristics of the molecule. The degree of ethoxylation is used to control the hydrophilic nature of the surfactant.

In surveying the literature, one might believe that commercial surfactants are pure compounds. In fact, commercially available compounds are complex mixtures of homologs. For instance, "lauryl sulfate" is a generic designation for a mixture of sulfates whose largest fraction is derived from the C12 alcohol but also contains higher and lower alkyl chain lengths. The homolog distribution depends upon the alcohol source and degree of fractionation. Alcohol ethoxylates are also complex mixtures of homologs. Not only is there a distribution of different alkyl chain lengths, but EO adds to the alcohol according to a normal distribution of adducts. Although it might seem that a competitive advantage would be gained by supplying a highly purified product, mixtures are frequently more effective than pure compounds. Long experience with soaps and synthetic detergents shows that except for highly specialized purposes, pure compounds are not competitive with commercial mixtures.

Synthetic Detergent Intermediates

Synthetic detergents are produced by various chemical routes, using either natural and/or synthetic sources as shown by the schematic in Fig. 27.18. Natural sources include fats and oils derived from vegetable and animal origins whereas synthetic sources include crude oil and natural gas. In many cases, the line between what is natural and synthetic is ill defined.

Fig. 27.18 Sources of raw materials for surfactant production.



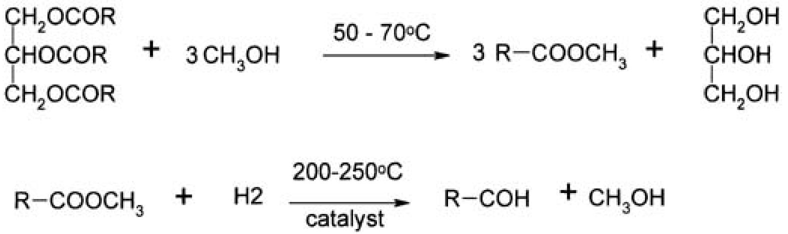
The current world production of fats and oils is around 65 MM metric tones of which 70 percent is of vegetable origin and 30 percent from animal sources. Oils with high content of C12 and C14 fatty acids are most commonly used by the detergent industry and are readily obtained from coconut and palm kernel. The most important animal fats are lard and tallow, which contain primarily C16 and C18 fatty acids. Fats and oils are composed of fatty acid esters of glycerin called triglycerides. Natural-based fatty acids possess an even number of carbon atoms and can be either saturated or unsaturated. After purification to remove phospholipids and sterols, the triglyceride is treated with alkali to produce soap and glycerin. Further treatment of the soap with inorganic acids gives the fatty acid. The fatty acid can be reacted with methanol to give the methyl ester. Alternatively, the triglyceride itself is treated with methanol using a base catalyst to give corresponding methyl ester and glycerin.

Crude petroleum oil consists of a complex mixture of long-chain hydrocarbons. Natural gas is a mixture of short-chain hydrocarbons rich in methane, ethane, propane, and butane. The exact composition of both depends on its source. Crude oil and natural gas are processed in different ways to give a variety of feedstocks suitable for surfactant manufacture. The most important feedstocks include ethylene, propylene, paraffin, olefin, phenol, and benzene.

One of the major factors leading to the development of modern synthetic detergents was the industrial manufacture of fatty alcohols. This was first achieved in Germany in the 1930s using the sodium reduction process. While this process was commonly used until the 1950s, it has been largely replaced by less hazardous processes. Today, there are three important technical processes for the manufacture of detergent grade fatty alcohols: high-pressure hydrogenation, the Ziegler process, and the oxo-synthesis.25 Numerous other processes have been developed but are of lesser technical and economic importance.

The most important process for the manufacture of natural fatty alcohols is the high-pressure hydrogenation of fatty acids or their corresponding methylesters. Common raw materials for the production of detergent range alcohols include coconut, palm kernel, palm stearin, soya, tallow, and lard. In the methylester process, the crude oil is reacted with an excess of methanol at 5070°C to give 3 moles of methylester and 1 mole of glycerin, as shown in Fig. 27.19. The reaction can be accelerated using sodium methoxide as a catalyst. The glycerol separates at the bottom of the reactor after an appropriate reaction time. The transesterification is complete when the methylester no longer contains glycerin.

Fig. 27.19 Preparation of natural fatty alcohols.



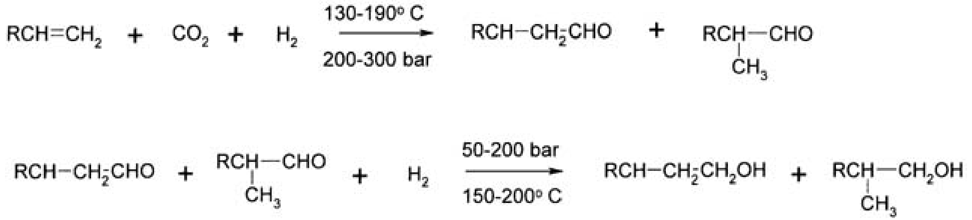
$$\begin{array}{l} CH_2OCOR \\ | \\ CHOCOR \\ | \\ CH_2OCOR \end{array} + 3\,CH_3OH \xrightarrow{50\text{ - }70^{o}C} 3\ R{-}COOCH_3 + \begin{array}{l} CH_2OH \\ | \\ CHOH \\ | \\ CH_2OH \end{array}$$

$$R{-}COOCH_3 + H2 \xrightarrow[\text{catalyst}]{200\text{-}250^{o}C} R{-}COH + CH_3OH$$

The hydrogenation reaction occurs at 200350°C and hydrogen pressures of 250300 bar. A variety of different metal oxide catalysts have been used, however, copper chromite is the most common. Under these reaction conditions, the ester group and any unsaturation are hydrogenated to give the saturated fatty alcohol. It is possible to produce unsaturated fatty alcohols using various zinc catalysts. A number of different process designs have been developed including the fixed bed, the trickle bed, and the suspension process. The various processes differ primarily in the way in which the catalyst contacts the feedstock and the specific catalyst used.

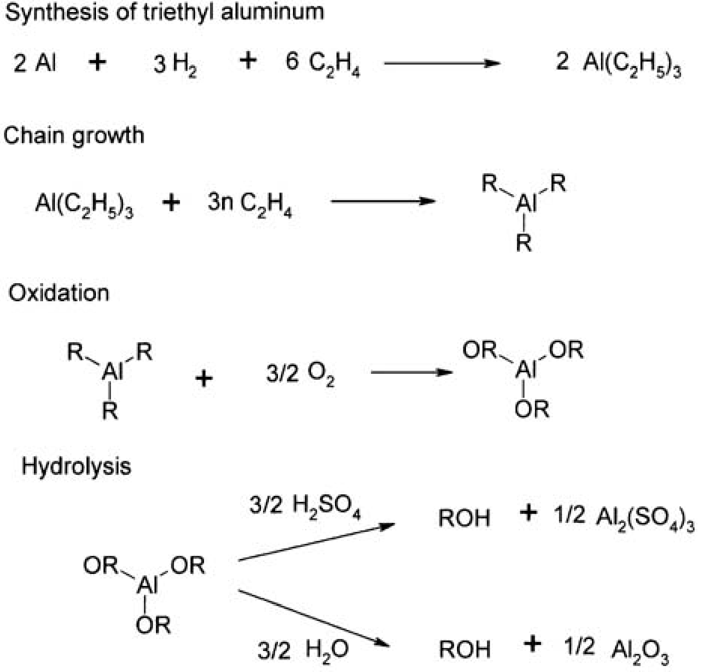
Both the oxo-synthesis and the Ziegler process are based on petrochemical feedstocks. The oxo or hydroformylation reaction was first discovered in Germany in the late 1930s, but was not used commercially until 1948, by what is now Exxon Chemical. The oxo process uses α-olefins obtained from oligomerization of ethylene, which are reacted with carbon dioxide and hydrogen at 200300 bar and 130190°C, to give a mixture of linear and branched aldehydes as shown in Fig. 27.20. The resulting aldehydes are hydrogenated at 50200 bar pressure and 150200°C using copper, zinc, or nickel catalyst to give the primary alcohol. Oxo alcohols contain both even and odd carbon numbers and are more highly branched than natural or Ziegler alcohols. Using a conventional cobalt catalyst, the amount of branching exceeds 50 percent.

Fig. 27.20 Oxo alcohol reaction chemistry.



The Ziegler process based on chemistry discovered in the 1950s produces linear primary alcohols with an even number of carbon atoms. The main reaction steps are shown in Fig. 27.21. Triethyl aluminum can be prepared using a one-step or two-step process. In practice, the one-step process yields more by-products and is seldom used. The reaction of triethyl aluminum with ethylene is highly exothermic. The distribution of chain lengths follows a Poisson distribution with only slight deviations due to competing side reactions. The aluminum alkyls are oxidized to the corresponding alkoxides using dry air and hydrolyzed using water or sulfuric acid to yield fatty alcohol. The Ziegler process produces alcohol similar to that obtained from natural sources.

Fig. 27.21 Ziegler alcohol reaction chemistry.



Another key surfactant intermediate is olefin, produced by thermal or catalytic cracking of paraffin wax or alternatively from oligomerization of ethylene. Olefins suitable for surfactant production are primarily linear with either terminal (α-olefins) or internal (i-olefins) double bonds. Olefins produced by the cracking process include α-olefins, di-olefins, cyclic olefins, aromatics, and paraffin wax. The major problem with the cracking process is separation of the desired olefins from the other by-products. Since the mid-1960s, n-paraffins have been used as starting materials for the manufacture of n-olefins. Today, n-paraffins are obtained from the distillation of paraffin-rich oils with the use of molecular sieves. Molecular sieves are aluminum silicates of the zeolite type. The pore diameter of the zeolite used for paraffin production is 5 Å, which allows the n-paraffins to enter the pore while the iso and cyclic paraffins are removed by distillation.

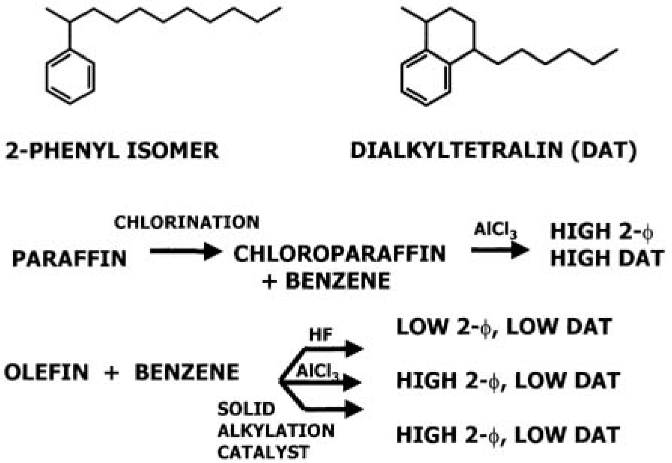
Oligomerization of ethylene can be performed in a number of different ways. Reaction of ethylene with triethyl aluminum using the Ziegler process consists of three steps. The first step is the reaction of ethylene with triethyl aluminum to give long-chain trialkyl aluminum. The second step is the transalkylation of the trialkyl aluminum with ethylene (Alfol Process) or a long-chain olefin (Ethyl Process) to give free olefin and regenerate the trialkyl aluminum. The third step is the separation of the olefin from the undesired by-products.

A slightly different oligomerization method is based on the Shell Higher Olefins Process (SHOP). The oligomerization reaction is carried out on a nickel phosphine complex that is separated and recycled by means of a solvent. The reaction product consists of α-olefins and i-olefins of different carbon chain lengths. The i-olefins are fed to the purification, isomerization, and disproportionation stages to obtain a wide range of α-olefins.

While olefins can be used directly to produce synthetic detergents, they are more commonly used as intermediate raw materials. Alkylbenzene consists of a mixture of phenyl substituted n-alkanes of 914 carbon atoms. Prior to 1965, alkylbenzene was synthesized from propylene tetramer, obtained by oligomerization of propylene, and reacted with an aluminum chloride or hydrogen fluoride catalyst and benzene. The resulting "hard" alkylate was a highly branched chain compound. However, the slow biodegradability of propylene tetramer-based materials became apparent and by 1965, most of the detergent industry had switched over to linear alkylbenzene (LAB). Extensive research has demonstrated biodegradation effectiveness in sewage treatment plants in excess of 95 percent.2629

Today, there are two main routes for producing linear alkylbenzene. (Fig. 27.22) The most prevalent route to alkylbenzene uses partial dehydrogenation of paraffins, followed by alkylation of benzene with a mixed olefin/paraffin feedstock using liquid hydrogen fluoride (HF), aluminum chloride (AlCl3), or solid alkylation catalyst. The Pacol process (paraffin conversion to olefin) uses n-olefins produced from n-paraffins by dehydrogenation over a heterogeneous platinum catalyst. The Pacol process produces a small amount of di-olefin, which lowers the LAB yield by forming alkylate bottoms. The Define process was developed to selectively hydrogenate di-olefins to mono-olefins. A second older process uses partial chlorination of paraffins, followed by alkylation of the chloroparaffin feedstock in the presence of an aluminum chloride catalyst. Distillation is used to separate alkylation reaction products into the targeted detergent alkylate homolog distribution. A coproduct stream of high molecular weight linear alkylbenzene and dialkylbenzene is also produced for use in lubricant detergents and industrial applications.

Fig. 27.22 Commercial alkyl benzene production.



Products with low 2-phenyl isomer content (1322%) are produced by HF alkylation of internal olefins from paraffin dehydrogenation. High 2-phenyl products (2535%) are produced from AlCl3 alkylation of chloroparaffins or olefins from paraffin dehydrogenation. Super high 2-phenyl products have been reported in the literature using shape selective catalysts (>80%) but are not commercially produced at this time. Generally, high 2-phenyl products are preferred for use in liquid detergents due to increased solubility. Low 2-phenyl products are typically used in powder detergents but can be used in liquid formulations with the appropriate choice of hydrotrope.

Alkyl phenols are produced by reacting an olefin with phenol using a strong acid catalyst. The most common alkyl phenols are 4-ter-octylphenol, 4-nonylphenol, and 4-dodecylphenol. Mono nonylphenol (MNP) is by far the most common hydrophobe. It is produced by the alkylation of phenol with nonene under acid conditions. All commercially produced MNP is made with nonene based on propylene trimer. Because of the skeletal rearrangements that occur during propylene oligomerization, MNP is a complex mixture of branched isomers.

Anionic Surfactants

Alkylaryl Sulfonates

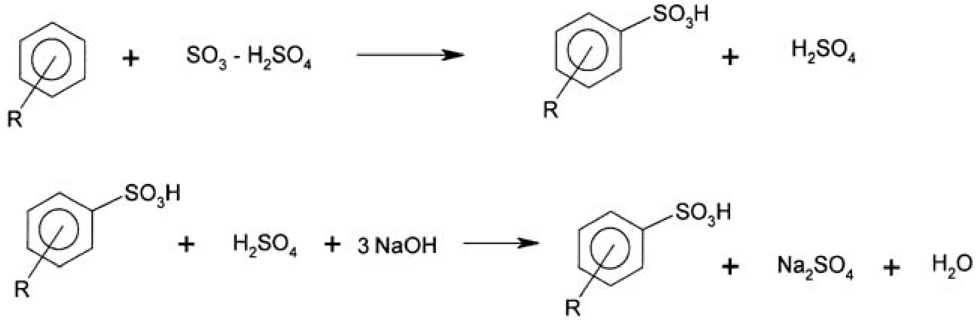
Alkylbenzene sulfonates have been the workhorse of the detergent industry for more than 40 years. They account for approximately 50 percent of the total synthetic anionic surfactant volume. Three basic grades of alkylate (alkylbenzene) are manufactured with molecular weight ranges of approximately 235, 240, and 260. In general, the lower molecular weight range alkylates are used for light duty liquid (LDL) systems whereas the higher molecular weight grades are used in heavy duty liquids (HDL) and heavy duty powder (HDP) detergents.

Sulfonation

For detergent applications, the alkylate must be sulfonated and then neutralized with a base such as sodium hydroxide to give a concentrated slurry, typically called "paste." The major uses of the paste are in the production of spray-dried powders and in both light and heavy duty liquid detergents. The paste can also be drum-dried to a powder or flake, or spray dried to light-density granules. Dried forms of LAS are used primarily for industrial and institutional cleaning applications.

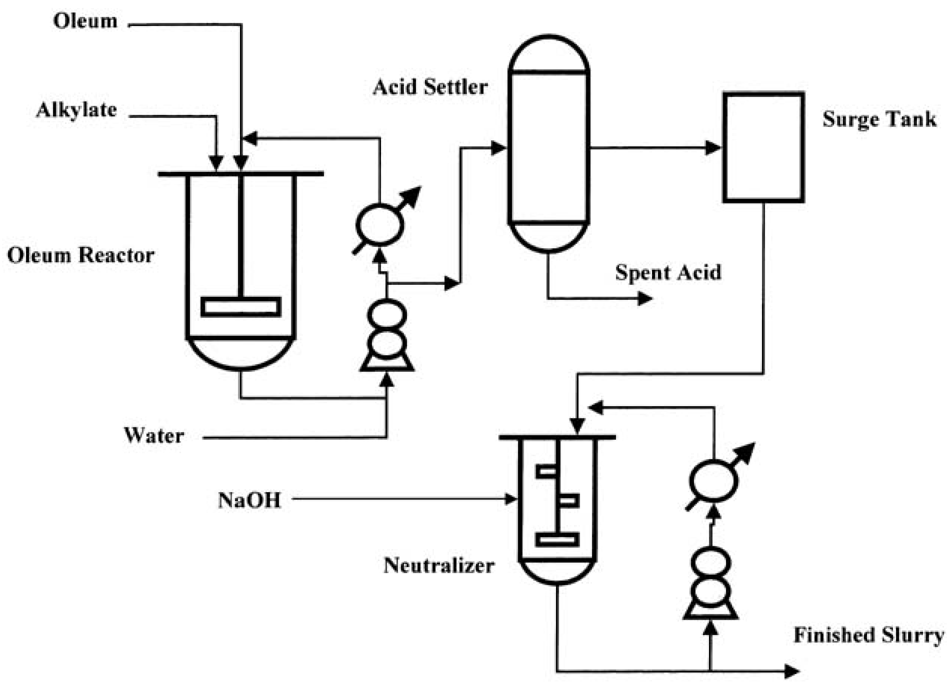
Oleum and sulfur trioxide (SO3) are commonly used for sulfonation, with the latter growing in popularity in recent years. Oleum sufonation requires relatively inexpensive equipment and can be run as batch or continuous processes. However, the oleum process has major disadvantages in the high SO3 cost/lb (as H2SO4), the need for spent acid stream disposal, and the potential corrosion problems due to sulfuric acid generation. The chemical reactions involved in oleum sulfonation are shown in Fig. 27.23. Typically, the oleum process yields 8891 percent purity sulfonic acid with the remainder consisting of approximately 610 percent sulfuric acid and 0.51.5 percent water and 0.51 percent unsulfonated oils.

Fig. 27.23 Oleum sulfonation process.



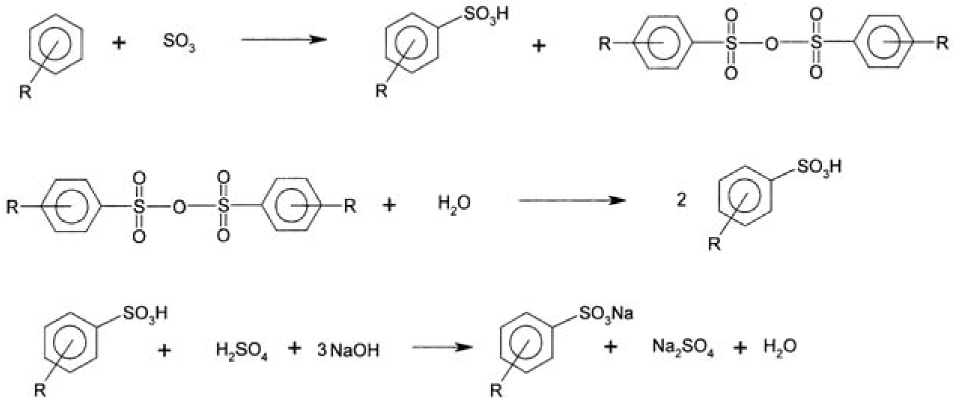
A typical batch oleum sufonation process is shown in Fig. 27.24. Four basic unit operations are utilized prior to neutralization: sulfonation, digestion, dilution, and phase separation. Mixing of the alkylate and oleum is done in the sulfonation stage. Heat removal from the highly exothermic reaction is achieved by reactor jacketing and/or recirculating heat exchanger loops. Key variables in controlling alkylbenzene conversion and acid color are the temperature, acid strength, reaction time, and the ratio of oleum to alkylate. The product leaving the sulfonation zone is aged or digested 1530 min to allow reaction completion. At this point in the process, the reaction product consists of a mixture of sulfonic and sulfuric acid. Following digestion, the mixture is diluted with water to quench the reaction. The reaction product is then fed to a separator to allow for gravity settling of the spent sulfuric acid layer from the lighter sulfonic acid. The separation time is dependent upon the separator vessel configuration, sulfonic acid viscosity, temperature, and the degree of aeration in the feed streams.

Fig. 27.24 Batch oleum sulfonation process.



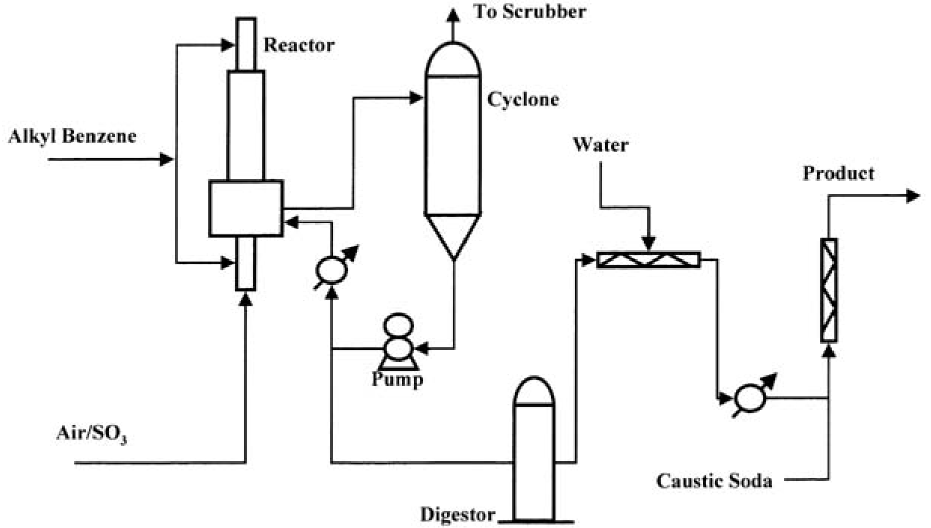
The gaseous air/SO_3 sulfonation process normally generates a product composed of 9598 percent sulfonic acid, 12 percent sulfuric acid, 0.51 percent water, and 12.5 percent unsulfonated oils. Although the initial capital costs are much higher than those for an oleum process, relative sulfonic acid yields and spent acid disposal costs are substantially lower. The chemical reactions involved in air/SO_3 sufonation are shown in Fig. 27.25. The SO_3 sulfonation reaction differs from oleum in that the reaction product consists of a mixture of sulfonic acid and the anhydride. The anhydride is reacted with water to produce 2 moles of sulfonic acid.

Fig. 27.25 Sulfur trioxide sulfonation process.



A typical continuous air/SO_3 sulfonation process is shown schematically in Fig. 27.26. Sulfur trioxide is produced from vaporization of liquid SO_3 or generated on-site by burning sulfur and then catalytically converting SO_2 to SO_3 over a vanadium catalyst. Increased shipping regulations and sensitivity to storage of the highly hazardous liquid SO_3 has increased the trend toward on-site sulfur trioxide generation. The vaporized SO_3 is diluted to 48 wt percent in dry air prior to sulfonation. The air serves as a diluent and heat sink for the highly exothermic sulfonation process. In the sulfonation reactor, the molar ratio of SO_3/alkylate is controlled between 1.03/1.0 and 1.0/1.0 to achieve optimum conversion to sulfonic acid without increased side reaction or color degradation. The Chemithon,30,31 Ballestra,32 and Mazzoni33 reactor designs each have unique falling film designs to provide intimate mixing of the alkylate and the air/SO_3 streams. The reactor temperature is controlled between 110 and 150°F, depending on the desired production rates and product quality targets. Digestion and hydrolysis temperatures are controlled between 110 and 125°F. During hydrolysis, approximately 2 parts of water per 100 parts of reaction product are added to convert acid anhydrides to sulfonic and sulfuric acid. Both the oleum and SO_3 sulfonation processes are quite exothermic and almost instantaneous. In order to prevent decomposition and maintain optimum product color, an efficient reactor heat removal system is necessary.

Fig. 27.26 Alkylbenzene sulfonation process.



Cold sulfonation is also practiced using SO_2 as a solvent and refrigerant.34,35 The sulfonic acid color is low because of low reaction temperature and rapid removal of the reaction exotherm via SO_2 vaporization. Equipment consists of fairly simple glass-lined reactors equipped with gaseous SO_2 recovery. Purging of SO_2 from the reaction product and control of SO_2 recovery are key to achieving a high-quality product. As in the case of air/SO_3 sulfonation, this process can be used to produce all detergent-grade surfactants including alkylbenzene, alcohol, and alcohol ethoxylates.

The Chemithon Corporation has patented a novel venture reactor.36,37 The organic to be sulfonated is injected into a stream of gas containing SO_3 at a venture. The reaction mixture is quenched and recycled downstream until sulfonation is complete. The reactor design can be used for highly viscous, high molecular weight specialty surfactants for industrial applications such as lubrication oil additives.

Neutralization of sulfonic acid is similar regardless of sulfonation process. The sulfonic acid can be neutralized with an aqueous solution of NaOH, KOH, NH_4OH or with alkanoamides. The sodium salts are used for formulating spray-dried powders for household laundry consumption. Ammonium and alkanolamine salts are commonly used in light duty liquid detergents. In general, neutralized salts from oleum sulfonation contain approximately 810 percent sodium sulfate whereas those from air/SO_3 sulfonation contain only 23 percent sodium sulfate. Sulfonic acid can also be processed to a dry powder form by in-situ neutralization with a base salt, such as sodium carbonate. The neutralized sodium salt can be dry blended with other detergent additives to produce a free flowing powder detergent.

Sulfation

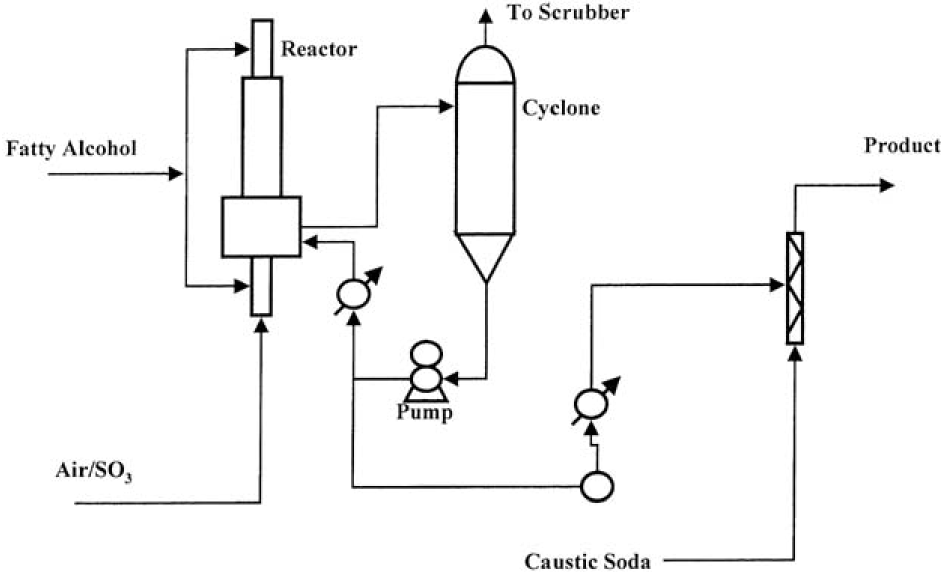
For the sulfation of alcohols or alcohol ethoxylates, falling film reactors are also used, however, cooling water and sulfation temperature are adjusted to lower values for these materials. The digestion and hydration steps are not required in sulfation and these systems are bypassed. The acids are neutralized immediately to minimize degradation and side reactions in storage. Typical SO_3 sulfation reactions for alcohols and alcohol ethoxylates are shown in Fig. 27.27. Typical processes for air/SO_3 sulfation of fatty alcohols is shown in Fig. 27.28.

Fig. 27.27 Sulfation of alcohols and alcohol ethoxylates.

$$R-OH \quad + \quad SO_3 \longrightarrow R-O-SO_3H$$

$$R\left[O-CH_2-CH_2\right]_n OH \quad + \quad SO_3 \longrightarrow R\left[O-CH_2-CH_2\right]_n O-SO_3H$$

Fig. 27.28 Fatty alcohol sulfonation process.



Nonionic Surfactants

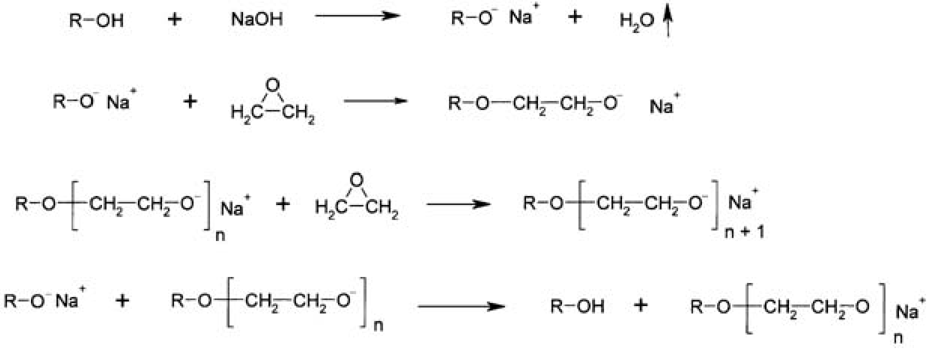
Nonionic surfactants are increasingly popular active ingredients. The majority of nonionic surfactants are ethylene or propylene oxide derivatives of alkylphenols or fatty alcohols, although fatty acids, alkyl amines, and alkanolamides are commonly used. In general, one ethylene oxide unit is required to solubilize each methylene unit in the hydrocarbon tails. By changing the ratio of ethylene oxide to fatty groups, the nature of the surfactant can be modified from oil-soluble to water-soluble products. This versatility makes applications possible in numerous areas, such as detergents, agricultural products, corrosion resistance for metals, emulsion polymerization, textile applications, deinking, and the like. Fatty alcohols are among the most useful intermediates for the production of nonionic surfactants, some of which are listed in Fig. 27.29. Discussion of each and every surfactant listed is beyond the scope of this work. Only the major surfactant classes will be covered. A more complete discussion of different surfactant classes and their properties is available.38

Fig. 27.29 Surfactants obtained from fatty alcohols.

Ethoxylation

Ethylene and propylene oxide are highly reactive species that will react with any molecule containing an active hydrogen (initiator). The alkoxylation reaction is best described as a two-step process. In the first step, the catalyst forms a charged complex with the initiator. In practice, the reaction is usually run using a base catalyst to avoid creating a large amount of polyethylene glycol (PEG) and other by-products. Alcohol and nonylphenol ethoxylates are two of the most common types of nonionic surfactant. The reaction chemistry for the alcohol ethoxylate is shown in Fig. 27.30. The degree of ethoxylation follows a normal Poisson distribution.

Fig. 27.30 Fatty alcohol ethoxylation chemistry.

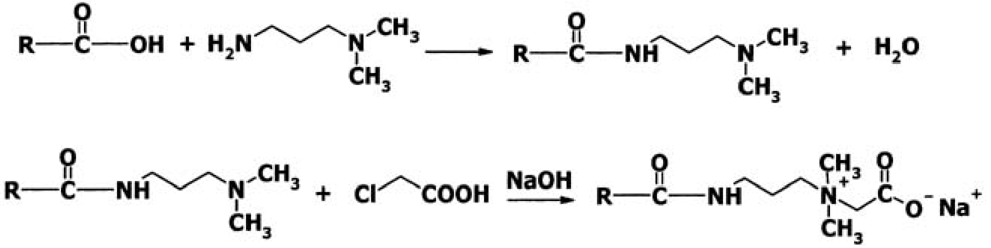


In processing, the initiator and catalyst are charged to a reactor and purged with an inert gas such as nitrogen. The partial pressure is adjusted so that the explosive limits of the gas mixture in the reactor during and after addition of the oxide are not exceeded. The reaction mixture is heated to 115200°C and the oxide is added until the desired pressure is reached. Heat removal rates control the rate of reaction. Excessive cooling can result in excessive absorption of oxide into the reaction mixture, which can result in a sudden reaction followed by overpressure. The reaction temperature and pressure are controlled until the desired quantity of oxide has been added. The batch is then cooled and the catalyst neutralized with a Lewis acid or removed by filtration.

Amphoteric Surfactants

These surfactants contain both cationic and/or anionic groups on the same molecule. Amphoteric surfactants are relatively expensive due to the raw material and processing costs. Typical applications include liquid dishwashing detergents, shampoos, body wash, and other personal care applications. In general, amphoteric surfactants detoxify and mitigate the irritation of anionic surfactants in the formulation. A typical reaction scheme for coco amide propyl betaine is shown in Fig. 27.31. Other examples of amphoteric surfactants include amine oxides, amphoacetates, and amino propionates. A detailed discussion of the various different types of amphoteric surfactants and their properties is available.39,40

Fig. 27.31 CAPB reaction chemistry.



Cationic Surfactants

Cationic surfactants tend to be specialty items that are tailored for diverse uses such as germicides, textile applications, corrosion inhibitors, and fabric softeners. Their primary growth has been in fabric softeners for home laundry detergents. Because cationics are antistatic agents, they perform well in removing static charges associated with synthetic fabrics.

Cationic surfactants have at least one hydrophobic group attached directly to a positively charged nitrogen atom. Cationic surfactants may be formed from nitriles, amines, amidoamines, or quaternary nitrogen bases. Significant literature reviews cover this subject in great detail.4143

Detergent Builders

Builders are typically added to a detergent formulation to extend or improve the cleaning performance of a formulation across a wide range of use conditions. The combination of builders and surfactants exhibit a synergistic effect to boost total detergency and cleaning efficacy, as compared with an equal amount of either compound alone. The major properties and characteristics that are desirable for a compound to be considered a detergent builder include:

1. the ability to control water hardness and other metal ions
2. a contribution to final product alkalinity
3. buffer capacity in the desired pH range
4. deflocculation capability
5. compatibility with other formulation ingredients and detergent additives
6. consumer safety
7. environmental acceptability
8. processibility
9. adequate cost/performance

In today's detergent products, a variety of builder materials are used singly or in combination with other builders to contribute unique properties for final product performance enhancement.

Phosphates

Tetrasodium pyrophosphate (Na4P2O7; TSPP) became commercially available in the mid-1930s and soon gained acceptance as the preferred builder for soap formulated products. The combination of TSPP and soda ash was more effective than either ingredient used separately. By the mid-1940s, sodium tripolyphosphate (Na5P3O10; STPP) was commercially available and had supplanted TSPP as the preferred builder because of superior detergent processing, solubility, and hard water sequestering properties. Sequestration is defined as the reaction of a cation with anion to form a soluble complex. The sequestration of Ca++ and Mg++ ions leads to softened water, which results in better surfactant solubility and improved cleaning performance. STPP was the major builder ingredient in heavy duty laundry detergents until the 1970s when phosphates were blamed for eutrophication of lakes and streams and thus has been almost completely removed from commercial laundry detergents. STPP is still widely used in automatic dishwashing detergents and in I&I cleaning applications.

Other possible builder salts include tetrapotassium pyrophosphate, sodium trimetaphosphate, trisodium and tripotassium phosphates, and sodium acid pyrophosphates. Significant literature reviews cover this subject in great detail.44,45

Silicates

Both sodium and potassium silicate solids or solutions have valued functionality including emulsification, buffering, deflocculation, and anti-redeposition ability. Silicates also provide corrosion protection to metal parts in washing machines, as well as to the surfaces of china patterns and metal utensils in automatic dishwashers. Silicates are manufactured in liquid, crystalline, or powdered forms and with different degrees of alkalinity. The alkalinity of the silicate provides buffering capacity in the presence of acidic soils and enhances the sequestration ability of the builder system in the formulation. The silicate/alkali ratios of the silicates are selected by the formulator to meet specific product requirements. Silicate ratios of 1/1 are commonly used in dry blending applications while silicate ratios of 2/1 and higher are commonly used in laundry and autodish applications.

Carbonates

In most areas of the United States, phosphates have been legally banned from use in home laundry detergents. Sodium carbonate, commonly called soda ash (Na2CO3), or a combination of soda ash and zeolite has replaced STPP as the builder in granular laundry products. Sodium carbonate softens water by precipitating hardness ions as insoluble carbonate salts. It is one of the most economical sources of alkalinity for detergents and provides some degree of soil dispersion and suspending action. Sodium carbonate is commonly used in powder laundry detergents, autodish powders, and hard surface cleaners.

Sodium Citrate

Sodium citrate (Na3C6H25O7) has builder characteristics due to its ability to sequester hard water ions and deflocculate soils. Sodium citrate's principal use is in liquid detergents and personal care products.

Zeolites

Zeolites are an important alternative builder for powdered laundry detergents and structured surfactant systems. Zeolites are crystalline hydrated aluminosilicates of Group I and Group II elements. They are a complex mixture of sodium, potassium, calcium, and magnesium salts, and appear in natural and synthesized forms. Zeolite 4A (Na2·Al2O3·2SiO2·4.5H2O, commonly called Zeolite A) is the most widely used form for laundry detergents.

Zeolites are not water soluble but they reduce calcium water hardness by ion exchange. The calcium ion passes through the zeolite pore openings and exchanges with the sodium ion. Magnesium ions are not significantly removed by zeolites because of the large atomic radium of the Mg++ ion. Zeolites are commonly used in combination with other builders such as sodium carbonate to control magnesium and other ions present in the wash water.

Sodium Nitrilotriacetate

One of the first compounds developed as a replacement for STPP was sodium nitrilotriacetate (N(CH2COONa)3·H2O; NTA), which has excellent sequestration and/or chelating properties. This product was quickly accepted by the detergent industry as an alternative builder. However, its use in detergents was suspended in the United States in 1970, when three major detergent manufacturers voluntarily agreed to discontinue its usage due to preliminary data suggesting that the compound could be teratogenic. Subsequent assessment by the EPA indicated that the risk to persons exposed to NTA from detergents is not significant. However, the EPA stated that NTA should not be used in products such as shampoos, hand-dishwashing detergents, foods, or other consumer products designed for direct dermal or oral exposure. Although NTA has not been used in U.S. household laundry products since the voluntary ban, it has been widely used as a detergent builder in Canada since 1970, and in the United States in I&I cleaning products. NTA is normally used in liquids as the primary builder or in combination with phosphates in spray-dried detergents.

Detergent Additives

In addition to detergent builders and surfactants, there are a number of different additives that perform specific functions in the detergent.

Sodium Sulfate

Although considered a weak builder by some manufacturers, sodium sulfate (Na2SO4) contributes little to detergent performance. It is commercially available from natural sources and as a by-product from rayon processing. In dry mix and agglomerated type products, sodium sulfate is used to improve finished-product flow characteristics. In spray-dried products, sodium sulfate acts as an inert filler and aids in density control.

Sodium Chloride

Sodium chloride (NaCl) is used by some detergent manufacturers, its main function being as an inert filler or diluent. In spray-drying operations, NaCl is used to control crutcher slurry viscosity and the density of the spray-dried bead or granules. In liquid formulations, NaCl is used to control product viscosity through the salt effect. Although salt is used to some extent in autodish formulations, it is not recommended because of potential machine and dishware corrosion.

Anti-Redeposition Polymers

The surfactants employed in synthetic detergents are effective in removing soil and stains from the surface of the washed fabric. Under heavy soil loading, there is a tendency for these soils to redeposit on the fabric. Various charged polymers have been found to prevent soil redeposition. Sodium carboxymethylcellulose (NaCMC) is effective in reducing soil redeposition at formulations level of 0.10.5 percent. The sodium salt of polyacrylic acid is also an effective anti-redeposition polymer. Molecular weights of 5000 gm/mol give a good combination of performance handling characteristics. Copolymer of styrene and maleic anhydride are also commonly used, although the optimum molecular weight is approximately 45,000. The polyacrylates are typically used at 0.51 percent in the formulation.

Optical Brighteners

Although the surfactant system in formulated detergents are effective at cleaning clothes, white fabrics tend to gray or lose the white appearance after multiple wash cycles. Optical brighteners are added to detergent formulations to visually improve the whiteness or brightness of the fabric.

All major liquid or powdered detergent products contain one or more fluorescent whitening agents.46 Substantive to most fabrics, these materials absorb ultraviolet radiation and convert it to visible bluewhite light that masks the natural yellowing or graying of the fabric. Most brightener systems are more effective on cotton than on synthetic fabrics. The usage range varies from 0.05 to 0.2 percent in the detergent formulation.

Hydrotropes

Hydrotropes are used primarily in liquid detergents to increase the solubility of sparingly soluble ingredients and improve storage stability. They are also used occasionally as viscosity modifiers in the slurry phase of manufacturing spray-dried detergent powders. Reduction of the detergent slurry viscosity improves its processing characteristics, reduces product bulk density, and insures crisp, free-flowing granules. The most common hydrotropes are the ammonium, potassium, or sodium salts of toluene, xylene, or cumene sulfonates.

Enzymes

A number of different enzymes have been developed for use as adjuvants in detergent systems to remove common soils and stains. Protease enzymes are used to cleave protein-based soils into smaller pieces more easily removed by the detergent system. Lipase enzymes are used to cleave lipid bonds resulting in fatty acid and glycerol. Lipase enzymes work better at elevated temperature and tend to become active in the drying operation. Noticeable removal of stains occurs after multiple wash cycles. Amylase enzymes tend to break carbohydrate bonds in sugar-based soils. Cellulase enzymes are designed to remove micro fibulas in cotton fabric to help retain color intensity and appearance. Enzymes are available in liquid and powder form. Care must be taken when handling solid enzymes to ensure that the powder does not become air borne where it can be inhaled by production workers. Enzymes are used in both liquid and powder laundry detergents.

PearlescentOpacifyingThickening Agents

These compounds are used in liquid detergent systems to produce a specific aesthetic effect. Pearlescent compounds, such as mica or ethylene glycol distearate (EGDS), are added to give a luxurious appearance to liquid formulations. Opacifying agents include water-soluble salts of styrene and maleic anhydride and are used to reduce translucence, modify the viscosity characteristics of liquids, and provide a cream-like texture to the product. Polymeric thickeners and clays are used to modify the rheological properties of liquids for various applications where cling to vertical surfaces is beneficial.

Anticaking Agents

The use of non-phosphate builders in laundry powders has changed the flow characteristics and caking tendencies. Products formulated with increased anionic surfactant levels and higher sodium carbonate content tend to exhibit compaction in storage. Anticaking agents have been added to the formulation to reduce this tendency, especially in warm, humid climates. A wide variety of products are available, including sodium benzoate, tricalcium phosphate, colloidal aluminium oxide, silicon dioxide, magnesium silicate, calcium stearate, and microctystaline cellulose.

Spray Drying

A major portion of powdered laundry detergents is spray dried. In this process, the detergent raw materials are mixed with water to form a thick paste or slurry, atomized into spherical droplets, and dried to produce a free-flowing granular product. The art of spray drying was first patented in 1883, but was not practiced to any extent until the 1930s. With the advent of synthetic detergents in the 1940s, spray drying became the dominant method of detergent processing. The advantages of spray drying are numerous: the product density can be varied from 0.20 to 0.80 gm/cm3, the detergent granule solubility is significantly improved, very high production rates are easily achieved, and multiple formulations can be prepared using the same equipment. The major disadvantages of spray drying are the initial capital investment required and the energy intensive nature of the process. Spray drying involves several processes: selection of formulations, slurry preparation, atomizing, drying conditioning of the product, post-additions, packaging, waste reclamation, and pollution control.

Spray-dried powders are formulated to give the best overall cleaning performance and flow characteristics at the lowest possible formulation costs. Typical spray-dried detergent formulations are shown in Table 27.4. Slurry preparation is considered to be the single most important factor in spray drying. Control of the raw material order of addition, slurry solids level, slurry temperature, viscosity, and aeration enables manufacturers to deliver the desired final product characteristics. The spray tower itself is simply a contained heat source that removes moisture at a uniform rate dependant upon the quality and the quantity of feed to the dryer. If feed variables are not controlled, variability in the density and moisture levels of the powder will vary. For example, if the slurry solids level is not constant (±2%), the evaporative load, production rate, particle sizing, and moisture levels of the spray-dried granules will vary. Compensating for nonuniform feed to the spray tower is difficult because of the interaction of control variables such as tower temperature, pump pressure, gas velocity, and so on. Although some minor tower adjustments are possible, the major control for routine production is consistent and uniform preparation of the detergent slurry.

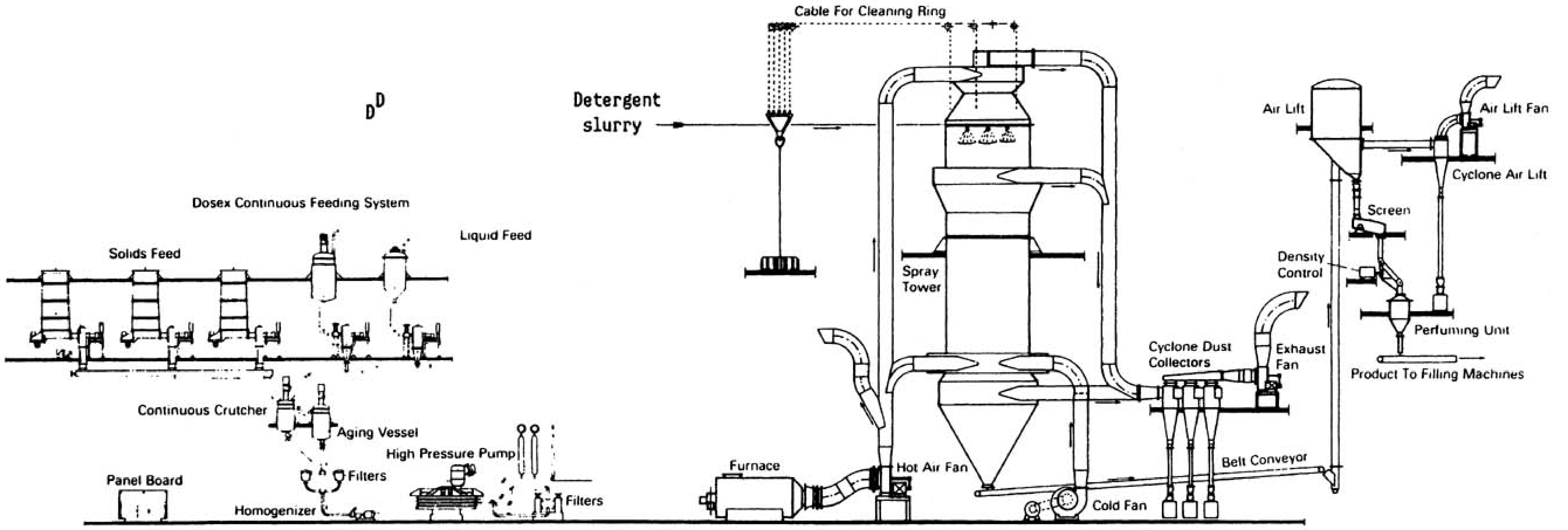
TABLE 27.4 Spray-Dried Detergent Formulation

	Phosphate Built Detergent Wt %	Non-Phosphate Built Detergent wt %
Sodium Tripolyphosphate	2035	0
Sodium Carbonate	510	510
Zeolite A	0	020
NaLAS	015	020
Alcohol Ethoxylate	015	020
NaAES	05	05
Sodium Silicate	36	36
Sodium Sulfate	1030	1030
NaCMC	0	0.11
Enzymes	0.11	0.11
Optical brightener	0.10.3	0.10.3
Bleach	05	05
Water	38	38

Slurries are prepared via batch, semi-continuous, and continuous crutching units. Solids levels are normally kept as high as possible (6472%) to reduce the heat requirements and maximize the tower throughput. After the detergent raw materials are mixed, the slurry is charged to a homogenizer and any large lumps or gritty particles are desized and screened to prevent plugging of the spray nozzle. The slurry is then charged to a booster pump, deaerated, and sent to a high-pressure pump (300120 psig) for atomization. Most detergent slurries are atomized into spherical droplets via high-pressure nozzles. The nozzles produce a hollow conical pattern with a spray angle of 4590°. The slurry emerges from the nozzle as a film, which is sheared into droplets by the centrifugal force developed by specially grooved cores inserted into the nozzle before the discharge point. A series of spray nozzles (225) are placed near the top of the spray tower and angled so that one edge of the desired spray pattern is vertical to the spray tower side walls. Some spray towers use two or three rings of nozzles, depending upon the formulation and drying capacity of the unit. Production rates of detergent spray towers range from 5000 to 80,000 lbs/hr. In most cases, the rates are controlled by the slurry feed facilities, solids levels, and the drying capacity. A schematic of a typical spray-drying facility is shown in Fig. 27.32.

Spray towers vary in size from approximately 10 to 35 ft in diameter and range from 40 to 200 ft in height. The majority of detergent spray towers use countercurrent airflow patterns although some use a concurrent design. In general, a countercurrent tower will produce a higher density particle than the concurrent type. In countercurrent systems, the heated gas stream (250270°C) is uniformly dispersed into the lower section of the tower through a series of plenum chambers or baffles equally spaced around the unit's periphery. The hot gases are cooled as they exhaust the top of the tower by the drying of atomized detergent droplets upon contact with the hot air stream. Because of the relatively low temperature (110140°C) at this point in the tower, the particles do not expand to their full capacity as water evaporates. This lower expansion rate results in a bead with increased bulk density and thicker shell. If a similar formulation were dried in a concurrent spray tower, the higher temperature (250270°F) and increased drying rate would increase the evaporative rate and result in a thin-walled bead of lighter density.

Fig. 27.32 Spray tower and slurry preparation.



In most spray towers, the airflow rate and the temperature are adjusted at the inlet ports to maintain balanced air/temperature patterns. When the atomized slurry is sprayed into the drying chamber, the airflow pattern is disturbed and becomes very turbulent. This effect can cause localized temperature gradients (hot or cold spots) and impact the particle drying characteristics. It is possible to have both over and under dried particles discharged at the same time from the tower because of poor adjustment of the airflow in the tower. Adjustment of the airflow patterns, as well as sending a balanced spray into the tower, can help counteract this situation. Increased impingement of the product on the tower walls is another negative effect of excessive turbulence and of excessive spray nozzle velocity. A ring of the product may form approximately 1020 ft below the spray nozzles and build in size until it cracks or scales off the walls. Ring removal methods, such as periodic cleaning using air or water lances, vibration or sonic shocks, or mechanical cleaning rings, are sometimes used to prevent a buildup that could result in blockage of the tower discharge.

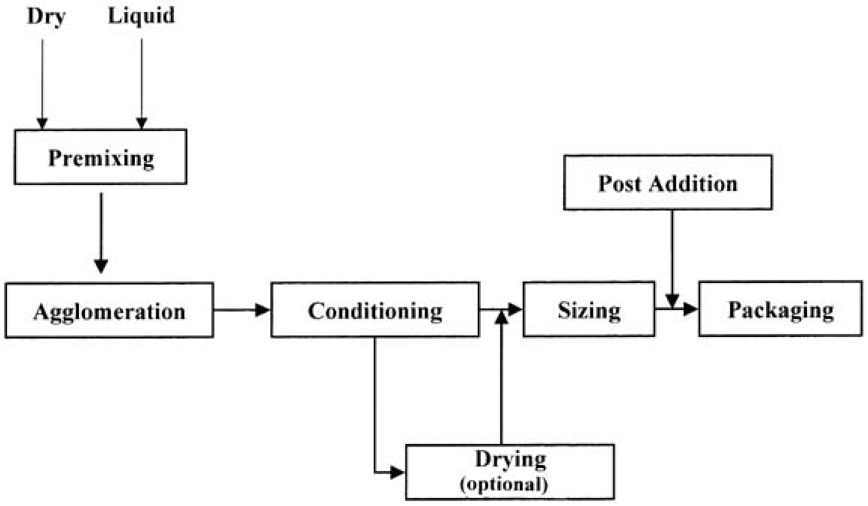
After drying in the spray tower, the detergent granules are conveyed to storage silos prior to packaging. In some processes, product storage is preceded by fluid-bed conditioning. The product for the tower is very warm (75120°C) and typically contains moisture levels of 13 percent. Conditioning is employed to prevent caking or product-flow problems due to either the high temperature or excessive free moisture content. Following conditioning, perfume and other heat-sensitive ingredients (enzymes, surfactants, bleaches) may be added to the product. Post-addition is accomplished by metering the additives into the spray-dried base in a rotary drum or other mixing device. The post-additions normally comprise 215 percent of the total detergent formulation. An excellent review of spray drying and key patents is available.47,48

Agglomeration Processing

Agglomeration is a technique of binding a mixture of granular and/or dry, powdered raw materials into an attractive granular form with enhanced aesthetic and flow characteristics. The majority of granular automatic-dishwashing compounds (ADWCs) and some laundry powders are manufactured using the agglomeration process. The advantages of agglomeration are low capital costs relative to spray towers and reduced energy consumption. Agglomeration also yields a higher-density product, which can result in reduced packaging costs. The major disadvantage of agglomeration is a limitation in surfactant loading due to raw material absorption characteristics, particularly in laundry applications. As energy costs have assumed increasing importance, and higher-density products have gained in popularity, agglomeration or combinations of agglomeration and spray-drying technologies have found wider use.

The typical process steps for formulating a laundry or auto-dish detergent are usually defined in the laboratory prior to continuous plant processing, however, they generally follow the scheme shown in Fig. 27.33. Frequently, the dry raw materials are premixed and then metered into the agglomeration equipment. In the agglomerator, the surface of the dry raw materials is exposed to atomized sprays of surfactant or silicates to begin building the agglomerated particle. Following addition of liquids to the dry raw materials, the product has a wet or pasty texture and requires conditioning to remove excess water prior to storage or packaging. Conditioned agglomerates are screened to remove oversize particles and then blended with other additives (bleaches, enzymes, dyes, and perfumes) that would not be stable under agglomeration conditions.

Fig. 27.33 Unit operations for batch agglomeration.



Numerous equipment types can be used for agglomeration, including horizontal ribbon, vertical mixers, rotary drums, zig-zag mixers, and pan agglomerators. Horizontal ribbon mixers (Fig. 27.34) are used primarily for dry mix formulating, where only limited agglomeration occurs. They are comprised of a U-shaped trough and a ribbon or paddle-type agitator. Liquids are sprayed onto the surface of the dry material and slowly dispersed into the mix. Mixer operations can be either batch or continuous. Vertical mixers are more commonly used due to low space requirements and high efficiency of operations. The vertical mixer (Fig. 27.35) utilizes high-speed agitation to intimately mix and uniformly blend liquids and dry raw materials prior to discharge. The contact time is reduced to less than 5 secs. The rotary drum units generally contain baffles that roll or lift the product for uniform dispersion or liquid feeds. Pan and zig-zag agglomerators are generally used for limited, specialty product manufacture.

Fig. 27.34 Batch or continuous ribbon mixer.

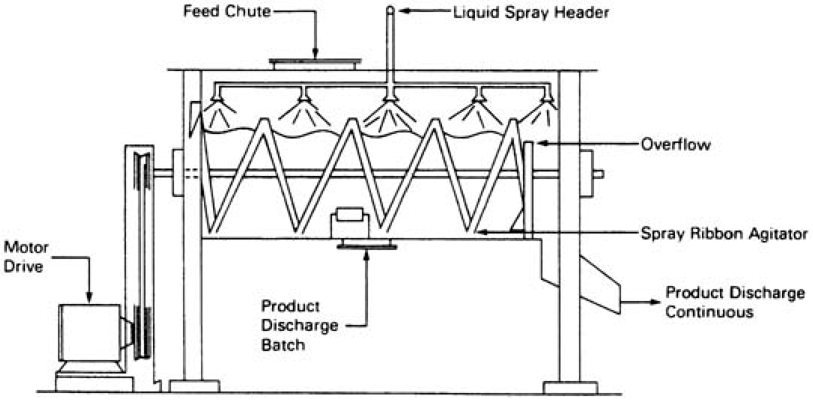


Fig. 27.35 Shugi Flex-o-Matic agglomerator.

The vertical and the rotary drum agglomeration equipment find most widespread use in the detergent industry because of improved agglomeration, increased liquid-feed dispersion levels, and optimized product uniformity. In addition to agglomeration of detergent and auto-dish detergents, these units can be used for hydration of STPP, dry neutralization of sulfonic acids, post-addition or raw materials and fines to spray-dried products, and particle encapsulation. An excellent review of detergent agglomeration is available.48,49

Liquid Detergent Processing

Liquid heavy-duty products account for approximately 60 percent of the 2001 U.S. laundry detergent market.50 The growth of heavy-duty liquids began in 1974, as the result of the phosphate ban in home laundry. Unbuilt liquids based on nonionic and anionic/nonionic surfactant combinations gradually increased in volume for various reasons including convenience, solubility, and improved performance versus non-phosphate containing powdered detergents. In the mid-1980s, the introduction of higher-performance liquids accelerated growth in this segment, with the market share essentially doubling from 1984 to 1990.51

Liquid products are approximately 5060 percent water, with the remainder being a combination of surfactants, builders, foam regulators, enzymes and enzyme stabilizers, hydrotropes, anti-redeposition polymers, optical brighteners, corrosion inhibitors, dye, and perfume. Two-in-one formulations also contain antistatic and fabric softening ingredients. Liquid systems require careful selection and blending of raw materials to achieve a stable product. Special attention is necessary for the following items:

1. Viscosity. The product must be pourable and must retain the same viscosity characteristics from batch to batch. This variable can be controlled by proper solvent or hydrotrope selection
2. Clearcloud point. The composition should have sufficient solubility to prevent hazing or separation of the product when subjected to storage in cool temperatures
3. Freezethaw stability. The formulation must be compounded to prevent phase separation or solidification at freezing temperatures

Most light-duty liquid (LDL) detergents are formulated using a combination of LAS and AES with various amphoteric foam stabilizers (CAPB, amine oxides, and so on). Heavy-duty liquid detergents often use a combination of actives such as LAS, AES, AE, and AOS. The manufacture of liquid products requires relatively basic equipment for mixing the various ingredients. The raw materials are added batchwise to a large mixing vessel or proportionated to in-line mixers before being fed to the filling lines. In 1986, liquid auto-dish detergents were introduced. These products were based on thixotropic slurries of STPP and soda ash. The rheology helped to solve issues with product dispensing rate and product physical stability. The slurries were generally viscous enough to prevent premature dispensing into the dishwasher prior to the wash cycle. The thixotropic nature of the product was achieved using natural clays of high molecular weight polymers. High-shear mixing equipment such as homogenizers is used to manufacture these products in batch or continuous stirred tank systems.

Detergent Trends

In the United States, liquid detergent sales continue to grow at the expense of powders. The trend is largely due to convenience as cost per load for liquids is significantly higher than for powders. In the United States, liquid products are clear and isotropic with vibrant colors. In Europe, liquid products have a lower market share and are typically based on structured surfactant systems with suspended STPP and/or zeolites. These systems tend to be opaque products and it remains to be seen if the U.S. market will accept these products. The performance of liquid laundry detergent performance has been increased through the use of multiple enzyme and ancillary ingredients. Recently, unit dose laundry tablets and liquid sachets have been introduced in the United States, which takes customer convenience to the next step. Presently, the market share for these products is small and it also remains to be seen how these products will fare the U.S. market.

The machines used to wash clothes are also changing. The DOE has mandated energy efficiency standards on new washing machines. In order to meet these efficiency standards, most manufacturers are offering horizontal axis washing machines similar to those that have been available in Europe for many years. The H-axis machines use less water and energy than conventional top loader machines, however, they are currently more expensive than conventional washing machines. The new machines generate more mechanical actions, which in turn generates more foam. Detergent manufacturers must now formulate lower-foam detergents for use in these machines.

The environmental movement, which began in Europe, is now beginning to appear in the United States. The use of nonylphenol ethoxylates in household products is declining. To a large extent, the NPEs have been replaced with alcohol ethoxylates. The use of surfactants that are environmentally acceptable continues to grow. Surfactants based on natural, re-growable raw materials are commercially available. Alkyl polyglycosides based on sugar and fatty alcohol and methyl ester sulfonates are examples of naturally derived surfactants. This trend can be expected to continue into the next decade.

27.3 References

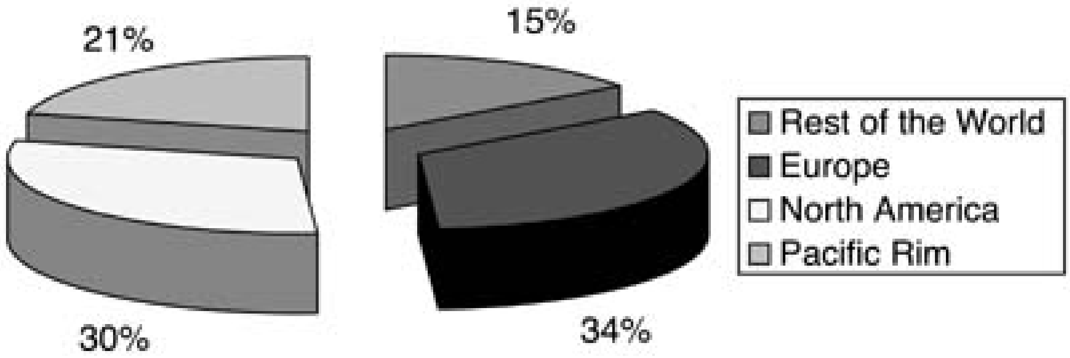
Soaps and Detergents, The Soap and Detergent Association, New York, 1981.
Spitz, L., Soaps and Detergents: A Theoretical and Practical Approach, AOCS Press, Champaign, IL, 1996
Gupta, S., "Chemistry, Chemical and Physical Properties and Raw Materials," in Soap Technology for the 1990's, L. Spitz (Ed.), pp. 4893, American Oil Chemists Society, AOCS Press, Champaign, IL, 1990.
Jungermann, E., "Soap" in Bailey's Industrial Oil and Fat Products, 4th ed., D. Swern (Ed.), pp. 511585, Vol. 1, John Wiley & Sons, New York, 1979.
Ghaim, J. B., and Volz, E. D., "Skin Cleansing Bars" in Handbook of Cosmetic Science and Technology, A. O. Barel, M. Payne, and H. I. Maibach (Eds.), pp. 485497, Marcel Dekker, Inc., New York, 2001.
By courtesy of G. Mazzoni S.p.A., Busto Arsizio, Italy. See also Spitz, L., J. Am. Oil Chem. Soc., 45, 423 (1967).
For general references, see (a) Dieckelmann, G., and Heinz, H. J., The Basics of Industrial Oleochemistry, Peter Pomp GmbH, Essen, W. Germany, 1988.
Woollatt, E., The Manufacture of Soap, Other Detergents and Glycerin, Ellis Horwood, Sussex, England, 1985.
By courtesy of Alfa-Laval, Tumba, Sweden.
Sonntag, N. O. V., J. Am. Oil Chem. Soc., 56, 729A, 861A (1979).
Markley, K. S., Fatty Acids, 2nd ed., Interscience, New York, 1967.
By courtesy of Lurgi GmbH, Frankfurt am Main, Germany.
Gosewinkel, L., and Knuth, M., Fett Wiss. Technol., 90, 155 (1988).
Stoiculescu, P., and Svet, V., Ind. Aliment. (Bucharest), 22, 320 (1971).
Leshchenko. P. S., Maslo-Zhir. Prom., 11, 21 (1973).
Linfield, W. M., Barauskas, R. A., Silvieri, L., Serota, S., and Stevenson, R. W., Sr., J. Am. Oil Chem. Soc., 61, 191 (1984).
Linfield, W. M., O'Brien, D. J., Serota, S., and Barauskas, R. A., J. Am. Oil Chem. Soc., 61, 1067 (1984).
Brady, C. L., Metcalfe, L., Slaboszewski, S., and Frank, D., J. Am. Oil Chem. Soc., 65, 917 (1988).
Kwon, D. Y., and Rhee, J. S., Korean J. Chern. Eng., I, 153 (1984).
Park, Y. K., Pastore, G. M., and de Almeida, M. M., J. Am. Oil Chem. Soc., 65, 252 (1988).
Holmberg, K., and Osterb, E., J. Am. Oil Chem. Soc., 65, 1544 (1988).
Mange, H. K., and Vaidya, S. D., J. Oil Technol. Assoc. India, 11, 73 (1979).
Vaidya, S. D., Subrahmanyan, V. V. R., and Kane, J. G., Indian J. Technol., 13, 528 (1975).
Berger, R., and McPherson, W., J. Am. Oil Chem. Soc., 56, 743A (1979).
Jach, K. W., and Stage, H., Fett Wiss Technol., 90, 501 (1988).
Stage, H., J. Am. Chem. Soc., 61, 204 (1984).
Jach, K. W., and Stage, H., Fett Wiss Technol., 90, 501 (1988).
Haradsson, G., J. Am. Oil Chem. Soc., 61, 219 (1984).
Zilch, K. T., J. Am. Oil Chem. Soc., 56, 739A (1979).
Luddy, F. E., J. Am. Oil Chem. Soc., 56, 759A (1979).
Haupt, D. E., Soap/Cosmetics/Chem. Spec., 60, 42 (1984).
Leikham, J. W., J. Am. Oil Chem. Soc., 89, (1988).
Griffin, W. C., J. Soc. Cosmet. Soc., 1, 311 (1949).
Griffin, W. C., J. Soc. Cosmet. Soc., 5, 249 (1954).
Henkel KGaA, Fatty Alcohols: Raw Materials, Process, Applications, Schneider & Hense GmbH, Dusseldorf, 1982.
Swisher, R. D., "Surfactant Biodegradation," Surfactant Science Series, Vol. 3, Marcel Dekker, New York, 1970.
Huddleston, R. L., and Nielsen, A. M., Household & Personal Products Industry, p. 72, 1979.
Gledhill, W. E., Huddleston, R. L., Kravetz, L., Nielson, A. M., Sedlak, R. L., and Vashon, R. D., Tenside, 26, 276 (1989).
Schoberl, P., Tenside, 26, 86 (1989).
Brooks, R. J. et al., U.S. Patent 3,257,715, June 21, 1966.
Brooks, R. J. et al., U.S. Patent 3,259,645, July 5, 1966.
Ballestra, M., U.S. Patent 3,180,699, April 27, 1965.
Lantini, A., U.S. Patent 3,931,272, January 6, 1976.
Hennig, H. et al., U.S. Patent 2,802,026, August 6, 1957.
Logan, R. S., U.S. Patent 3,188,183, June 8, 1965.
Brooks, R. J. et al., U.S. Patent 4,113,428, September 12, 1978.
Brooks, R. J., U.S. Patent 4,185,030, January 22, 1980.
Schick, M. J., "Nonionic Surfactants," Surfactant Science Series, Vol. 23, Marcel Dekker, New York, 1987.
Bluestein, B. R., and Hilton, C. L., Amphoteric Surfactants, Surfactant Science Series, Vol. 12, Marcel Dekker, New York, 1982.
Lomax, E. G., Amphoteric Surfactants, Surfactant Science Series, Vol. 59, Marcel Dekker, New York, 1996.
Swartz, A. M., and Perry, J. W., Surface Active Agents, Vol. 1, John Wiley & Sons, New York, 1949.
Jungermann, E., "Cationic Surfactants," Surfactant Science Series, Vol. 4, Marcel Dekker, New York, 1979.
Richmond, J. M., "Cationic Surfactants," Surfactant Science Series, Vol. 34, Marcel Dekker, New York, 1990.
Van Wazer, Phosphorus and Its Compounds, Vol. II, John Wiley & Sons, New York, 1961.
Sorgenfrei, M. E., "Increased Use of STPP," Soap/Cosmetics/Chem. Spec., 3(2), 60 (Feb. 1980).
Findley, W. R., JAOCS, 65 (4) (1988).
Sittig, M., "Detergent Mfg. Including Zeolite Builders," Chem. Tech. Review, No. 128, Noyes Data Corporation, New Jersey (1979).
De Groot, W. H. et al., The Manufacture of Modern Detergent Powders, Herman de Groot Academic Publisher, The Netherlands, 1995.
Dolan, M. J., "Detergent Agglomeration Technology," Soap and Chemical Specialties (Mar./Apr. 1988).
Marchie, M., "The Laundry Detergent Market," Happi, 76 (Jan. 2002).
Cannon, D. R., Chemical Week, p. 22 (Jan. 21, 1987).

28.1 Introduction

Change is constant in the coatings market. As mergers, acquisitions, and partnerships take shape, consolidation and globalization remain prominent. The 80/20 rule (20% of the firms accounting for 80% of business) takes effect as the need for regulatory and environmental compliance continues to plague the market. In 1975, the United States alone supported about 2,000 coatings companies. Today, there are less than half that many.

The paint and coatings market garnered $70.6 billion in sales (Fig. 28.1) in 2000. The major participants included: Akzo Nobel Coatings BV (Sassenheim, The Netherlands); PPG Industries, Inc. (Pittsburgh, PA); The Sherwin Williams Co. (Cleveland, OH); DuPont Performance Coatings (Wilmington, DE); ICI Paints (Berkshire, United Kingdom); BASF Coatings AG (Ludwigshafen, Germany); Sigma Coatings (SigmaKalon Group, BV); The Valspar Corporation (Minneapolis, MN); Nippon Paint Co., Ltd. (Kita-Ku, Osaka, Japan); and Kansai Paint Co. Ltd. (Chuo-Ku, Osaka, Japan). Smaller companies, being forced out of the market because they lack the capital required to meet compliance issues, are forced into consolidations. Globalization is also spurring more companies to form more partnerships. Akzo Nobel, together with joint-venture partner Chang Cheng Securities, will become the first manufacturer of powder coatings in Vietnam. This venture was made possible by the acquisition of Courtalds by Akzo Nobel in 1998, which helped strengthen the company's position in powder coatings.

Fig. 28.1 2000 global coatings market.1



The auto industry is also providing coatings companies with opportunities to form increasingly important relationships. PPG recently announced that it would be the first coatings manufacturer to be awarded an exclusive contract at the Nissan plant scheduled to open in Canton, Mississippi in 2003. PPG will provide everything from pretreatment to clear-coat.

In addition, PPG has formed a joint venture with a local minority-owned business, Systems Consultants Associates (Jackson, Mississippi) to coat chassis frames and other underbody components, as well as Nissan replacement parts. The joint venture, Systems Electro Coating, will be minority owned and will operate at a facility adjacent to the Nissan plant.

A Toyota plant in France, which opened in January 2001, will procure all of its coatings materials from DuPont Herberts Automotive Systems. This action was made possible through the acquisition of the Herberts Coatings business by DuPont in 1999.

The major coatings markets in the United States consist of architectural coatings (consumer and decorative), product finishes (maintenance/protective coatings for bridges, industrial machinery, etc.), and special purpose coatings such as automotive finishes (Original equipment manufacturer (OEM) and refinish) (Table 28.1).

TABLE 28.1 U.S. Sales of Paints and Coatings (Millions of Gallons)1

			Year		
Coating Type	1996	1997	1998	1999	2000
Architectural	640.3	655.6	631.6	660.2	645.6
OEM product	398.7	425.4	428.3	439.7	453.2
Special purpose	208.9	181.8	173.3	174.4	181.7

The coatings market is being plagued by environmental constraints imposed by local, state, and federal governments. The key concerns of government regulators and consumers include:1 treating the earth as one biosphere, having an infinite capacity for cleaning itself of toxins; cleaning air from pollutants that are killing lakes and causing a dramatic increase in respiratory illnesses; ensuring that smog-enshrouded cities do not become industrial nightmares; and making the agricultural and livestock products that people eat safe for consumption over the long term.

These concerns have led to legislation over the last several years that cause manufacturers to place a greater emphasis on research to meet and exceed government and consumer demands for clean air.

28.2 VOC Regulations

The environmental demands have lead to a nationwide program to limit volatile organic compounds (VOCs). The VOCs of paint are calculated as pounds of solvents per gallon of paint. Prior to 1970, the VOC content of most paints was well above 5 lb/gal. Current major industrial paints now are limited to approximately 3.5 lb/gal of VOC because of imposed environmental pressures. Stricter legislation in the future, however, will place requirements of less than 3.0 lb/gal of VOC on most markets. The issue of air quality compliance is complicated by different regulations in various states and regional areas. Although the federal government (through the U.S. EPA) has set minimum national standards, states and designated regional areas can adopt their own compliance rules.

All states and regions tend to restrict the solvent content in coatings. The following summaries indicate the wide variation in compliance requirements throughout the country.

West

California

This is where the regulations began. The regulatory activities of the South Coast Air Quality Management District (SCAQMD) have set a pattern that many other regulatory boards have chosen to follow. The district restricts application equipment and VOC content. The VOC requirements are continuously changing; so no attempt was made to list them here.

The SCAQMD in Southern California has mapped out a three-tier plan to bring the district into clean-air compliance. The plan will be revised periodically based on reports from over 40 monitoring stations throughout 13,350 square miles including Los Angeles, Orange, and Riverside counties and the urban portion of San Bernardino County. VOCs will not be monitored directly. Rather, ozone, which is formed when hydrocarbons (VOCs) react with nitrogen oxides in sunlight, will be measured for the presence of VOCs. The federal clean-air standard for ozone requires that ozone concentrations do not exceed 0.12 ppm more than one day a year.

Tier I of the three-tier plan involves 22 rules, which were adopted in 1993, to restrict the emission of VOCs from coatings. Control methods include the application of low-solvent coatings, use of high transfer efficiency (TE) techniques to apply coatings, and "to a certain extent" the use of add-on control devices (see Table 28.2).

TABLE 28.2 Application Equipment Requirements of SCAQMD Clean-Air Act2

	Rule									
Application Equipment	1104	1106	1107	1113	1115	1124	1125	1136	1145	1151
Dip		X			X	X	X			
Electrostatic application			X			X	X	X		X
Flow coat			X			X	X	X		
Hand application methods					X	X	X			
HVLP* spray		X			X	X	X		X	
Roll coater			X			X	X	X		
Other methods that are proven 65% transfer efficient			X			X	X	X		X
No specified application method	X	X		X	X				X	

*High-volume low-pressure application equipment.

Tier II (adopted in 2000) reduced the VOC emissions by 50 percent. Reductions were achieved by further tightening the VOC content of coatings (at times to levels lower than those stated in the Federal Control Technique Guidelines (CTG)) and increasing application transfer efficiency requirements (at times to levels more stringent than those stated in the EPA "Blue Book").

Tier III (to be implemented by December 31, 2007) calls for applying new technology to improve waterborne, UV-cure, and two-component coatings, and prohibiting "certain high-emitting uncontrollable coating processes."

Washington

No rule prohibits any method of spray finishing. VOC guidelines are determined by subtracting the amount of solids from the gross weight of the material used and determining whether the amount of solvent used exceeds the acceptable limits.

The maximum VOC limits (lb/gal) are: clear-coat, 4.3; color, 3.5; performance coatings, 3.5; all other coatings, 3.0. Metric limits (g/L) are defined for aerospace: primers, 650; topcoats, 600 (equivalent to 5.4 and 5.0 lb/gal, respectively).

Southwest

Texas

In Texas, regulators restrict emissions rather than equipment. Emissions are limited across the board for certain industries. Limits (lb/gal) are specified for particular coatings in other industries. For example: large appliances, 2.8; furniture, 3.0; coil coatings, 2.6; automotive OEMprimers, 1.2 if submersed and 2.8 if sprayed, topcoats, 2.8, and repair, 4.8; miscellaneous metalcolor, 2.5, top clear-coat, 4.3, and all others, 3.0.

South Carolina

VOC limits (lb/gal) include: metal furniture, 3.0; appliances, 2.8; miscellaneous metalcolor, 3.5, clear-coat, 4.3, and all others, 3.0; coil coating, 2.6.

North Carolina

The rules are applicable only in areas designated as non-attainment zones. If air pollution control equipment is not used, emission standards are based on the weight of solids used. High-solids coatings are automatically compliant regardless of spray equipment used. The limits (lb/gal) for automotive shops with air pollution control equipment include: primers, 1.4 if high solids, 1.2 if low solids; topcoats, 4.5 if high solids, 2.8 if low solids; repair, 13.8 if high solids, 4.8 if low solids. For automotive shops without air pollution control equipment, limits are: primers, 1.2; topcoats, 2.8; repair, 4.8. Air pollution control equipment is not precisely defined. Other limits include: metal furniture, 5.1; appliances, 2.8.

Georgia

VOC emissions are limited (lb/gal) by industry and coatings used: automotiveprimers, 1.2 if dipped and 2.8 if sprayed, topcoat, 2.8, and repair, 4.8; metal furniture, 3.0; appliance, 2.8; and miscellaneous metal-performance coatings, 3.5, high-performance coatings, 6.2, clear-coat, 4.3, and all others, 3.0; coil coating, 2.6.

Midwest

Indiana

Rules in Indiana are among the most complex in the nation. They require a baseline TE, specify coatings formulations, and limit VOC emissions (lb/gal) for automotive, metal furniture, and large appliances but then only specify acceptable equipment choices for wood. Limits are: automotive30 percent TE at 62 percent solids; metal furniture60 percent TE at 59.2 percent solids; large appliances60 percent TE at 62 percent solids. No TE test procedure is defined for the following VOC limits: automotiveprimer, 1.9, topcoat, 2.8, and repair, 4.8; metal furniture, 3.0; large appliances, 2.8; miscellaneous metalcolor, 3.5, specialty coats, 3.5, clear-coats, 4.3, coil coatings, 2.6, and all others, 3.0. In the wood industry, the state departs from these guidelines and identifies acceptable equipment choices: airless, heated airless, and air-assisted airless spray; manual and automatic electrostatic spray; and rollers brushes, or wipes. Significantly, no conventional air spray method, including HVLP, is recognized as compliant.

Illinois

A 30 percent minimum TE is required for automotive. Automotive VOC limits (lb/gal) are: primer/surface, 2.8; topcoat, 2.8; repair, 4.8. The rules allow higher VOC limits for automotive topcoats if the finishing method can be demonstrated to achieve a TE of over 65 percent. In such cases, topcoats may be used with a VOC of up to 3.6. For the wood industry, a minimum TE of 65 percent is required with the following VOC limits: washes, 6.5; stains, 4.7; top-stains, 6.6; sealers, 5.6; clear-coats, 5.6. No test procedure for TE has been defined.

Michigan

Although VOC guidelines are suggested, they are not stated clearly. In general, the finisher is expected to achieve minimal VOC levels when using metallic topcoats of waterborne coatings and performing touchup or repair work. Although conventional air spray is generally acceptable, some areas of Michigan require electrostatic systems with a TE of at least 40 percent and define VOC limitations.

East

New Jersey

A minimum of 65 percent TE is required for all industries except wood. Shops using waterborne coatings are automatically compliant regardless of TE. VOC limits (lb/gal) for all others are: automotiveprimers, 2.8, topcoats, 2.8, specialty coats, 5.0, and repair, 4.8; automotive refinishingbasecoats, 6.0, clear-coats, 4.4, and all others, 5.0; coil coating, 2.6; miscellaneous metalscolor, 4.3, performance coatings, 3.5, and all others, 3.0; leather, 5.8; urethane users, 3.8. Wood industry rules specify equipment by the shop size. Wood shops using less than 50 tons of coatings a year must achieve a minimum TE of 40 percent. Approved spray-finishing methods include airless, heated airless, and air-assisted airless. Wood shops that use 50 or more tons of coatings a year are required to achieve a TE of 65 percent or higher. Approved spray-finishing methods for these shops include airless, heated airless, air-assisted airless, electrostatic spray, or flat line finishing. VOC limits (lb/gal) are: wash-coat, 6.1; stain, 4.7; pigment, 5.0; sealer, 5.6; clear-coat, 5.6.

Pennsylvania

The rules specify only that wood cabinet and furniture finishers must use air-assisted airless or electrostatic spray.

New York

New York follows the same rules as Pennsylvania.

Ohio

Ohio has adopted VOC limitations and will require a TE of at least 30 percent.

28.3 Hazardous Waste Regulations

Included in regulatory compliance is the issue of hazardous wastes. On March 29, 1990, the U.S. EPA promulgated a final rule revising the procedure used to determine whether wastes are hazardous because of their potential to leach significant concentrations of toxic constituents to landfills. The extraction procedure (EP) toxicity characteristic procedure was replaced with the toxicity characteristic (TC), utilizing the toxicity characteristic leaching procedure (TCLP). The U.S. EPA also added to the list 25 organic chemicals commonly used in surface-finishing operations (i.e., methyl ethyl ketone, tetrachloroethylene, etc.).

Although the maximum allowable concentrations of the regulated eight metals (and pesticides, herbicides, and insecticides) do not change from those listed under the EP toxicity, the new TCLP analytical procedure may make it more difficult for listed wastes to stay under the allowable concentrations (Table 28.3).

TABLE 28.3 Maximum Concentration of Contaminants Allowable from TCLP[a,3]

EPA Hazardous Waste No.	Contaminant	Regulatory Level (mg/L)
D004	Arsenic	5.0
D005	Barium	100.0
D018	Benzene	0.5
D006	Cadmium	1.0
D019	Carbon tetrachloride	0.5
D021	Chlorobenzene	100.0
D022	Chloroform	6.0
D007	Chromium	5.0
D023	*o*-cresol	200.0
D024	*m*-cresol	200.0
D025	*p*-cresol	200.0
D026	Cresol	200.0
D027	1,4-dichlorobenzene	7.5
D028	1,2-dichloroethane	0.5
D029	1,1-dichloroethylene	0.7
D030	2,4-dinitrotoluene	0.13
D032	Hexachlorobenzene	0.13
D033	Hexachlorobutadiene	0.5
D034	Hexachloroethane	3.0
D008	Lead	5.0
D009	Mercury	0.2
D014	Methoxychlor	10.0
D035	Methyl ethyl ketone	200.0
D036	Nitrobenzene	2.0
D037	Pentachlorophenol	100.0
D010	Selenium	1.0
D011	Silver	5.0
D039	Tetrachloroethylene	0.7
D040	Trichloroethylene	0.5
D041	2,4,5-trichlorophenol	400.0
D042	2,4,6-trichlorophenol	2.0
D043	Vinyl chloride	0.2

[a] Excluding D012D017 and certain pesticides.

28.4 Technical Trends in Coatings

Through the years, improved paints have been developed because of competition from environmental and consumer activists. The objective of improving overall quality and reducing emissions of VOCs has led to a wider selection of paint types. The user, therefore, must select a coating type that meets the cost of required application, equipment, energy costs for using, pollution and waste collection expenses, and application and cured performance requirements of the coating.

Table 28.4 lists types of coatings markets served by leading paint and powder coating producers. By far, the leading type of coating manufactured for the OEM market produced by the suppliers surveyed is waterborne (87.4%), followed by low-solids solvent-borne (78.7%), high-solids solvent-borne (74.0%), and two-component coatings (69.3%). The percentage of powder coatings produced by the surveyed producers accounted for only 26.0 percent, while ultrahigh solids and UV/EB coatings added an additional 22.0 and 22.8 percent, respectively. The majority of those producers surveyed suggested that most of their R&D spending would be on waterborne coatings. While powder coating research continues to grow, E-coat and radiation cure research continues through suppliers in niche markets. Lower cosolvent, lower cost, and shortened cure times continue to be the hottest areas of research.

TABLE 28.4 Coatings Markets Served by Leading Paint and Powder Coating Producers4

Market Segment	Average Percent of Total Coatings Output
General industrial	34.4
Automotive	18.0
Metal furniture/business equipment	8.6
Metal containers/closures	9.0
Wood furniture/wood products	12.5
Appliance/HVAC	3.3
Lawn and garden	2.4
Aircraft	4.0
Other	7.9

A solvent-borne coating generally consists of seven major constituents:

1. resin (binder)
2. cross-linker
3. rheology control agent
4. catalyst
5. pigment
6. flow additive
7. solvent(s)
8. weathering stabilizers (if required)

The resin (often referred to as binder) is responsible for the main attributes of the coating when cured with the cross-linker (or cross-linking agent).

The curing, or network formation, often is accelerated through the use of a catalyst. Rheology control agents, often referred to as sag control agents, are added to impart shear-thinning behavior when under high shear (i.e., spray application) and rheology or anti-sagging behavior when under low shear (applied on substrate). Pigments, discussed later in this chapter, are introduced to impart color.

The amount of pigment utilized depends on the color and the hiding power required of the coating. The flow additive often is introduced to relieve surface tension variations between the coating and substrate, to eliminate pinholes or crater formation. Solvents are added as necessary to achieve flow under application conditions.

In order to achieve a VOC-compliant coating, one must consider the molecular weight of the main binder utilized in the coating formation, as well as the effect of additives on the final viscosity. After all, application viscosity, which in turn is affected by the rheology of the coating constituents, is most often controlled by freely adding additional solvent to achieve necessary flow and leveling upon application. This in turn affects the VOC emission of the coating. If one considers the molecular weight dependency of resins utilized in the various coating types, it is readily evident why VOC compliance is more easily achieved in radiation-curable and powder coatings. The molecular weight of nonaqueous dispersion resins, for example, exceeding 105 g/mol, results in very viscous materials, whereas the typical molecular weight of a radiation cure oligomer ranges from 102 to 103 g/mol, for a very low viscosity material, which requires little if any additional solvent to attain application specifications.

If we consider the evolution of coating technology over the last 70 years (see Table 28.5) we can see that by the mid-1950s, waterborne and higher solids coatings already were under development. Powder coatings, 100 percent reactive liquid coatings (electron/beam and UV-cure), electro-coatings, aqueous or waterborne (water-base) coatings, and high solids coatings can all be considered as novel "non-polluting" VOC-compliant coating systems.

TABLE 28.5 Evolution of Coatings Development5

1923	Low viscosity nitrocellulose
1923	Alkyds
1924	Titanium dioxide
1928	Phenolic resins (oil-soluble)
1930	Urea formaldehydealkyd mixtures
1933	Vinyl chloride copolymer
1934	Oil-based emulsion paints
1936	Thermoset acrylics
1937	Polyurethanes
1939	Melamine formaldehydealkyl mixtures
1944	Silicon-based paints
1947	Epoxy resins
1950	Polyvinylacetate and acrylic emulsions
1955	Powder coatings
1958	Latex house paint
1960	Water-based coatings
1962	Anodic electrodeposition
1963	UV- and electron-beam cure
1971	Cathodic electrodeposition
1974	Clear over color topcoats
1989	Supercritical CO2

28.5 Powder Coatings

Today, powder coatings are one of the fastest growing finishing technologies in North America, with a growth rate (in pounds sold) of 79 percent per year over the next 5 years. The major impetus for growth can be attributed to the virtual "pollution-free" finishing system as well as the dramatic leap forward in materials and equipment technology.

Powder coating resin systems have been developed to meet specific needs of the metal finishing operations. Epoxies were utilized almost exclusively during the early years of thermosetting powder coatings and continue to be utilized in a range of applications. However, because of the poor outdoor durability of epoxy resins, acrylics and polyester-urethane hybrids are finding increasing use.

Powder coating manufacturers continue to work on perfecting flow and leveling of the coatings, which often are prone to "orange peel." The other major disadvantage of thermosetting powder coatings is the high temperature (300350°F) required to cure these systems. Research seeks to lower the temperature of cure.

Electron Beam (EB) and Ultraviolet (UV) Curable Coatings

Radiation-curable coatings were developed by W. J. Burlant (of the Ford Motor Company) in 1962. Acceptance of finishes designed for radiation curing (high-solids materials formulated to be cured by UV and EB systems) is growing at roughly 8 percent per year with a market size of approximately 85 million lb.

The major advantages of radiation-curable coatings are speed of cure, reduced emissions, 100 percent solids systems, and elimination of ovens or other sources of thermal energy. The most rapidly growing areas in UV are UV-curable powder coatings and new applications for UV-curable coatings for automotive parts. The UV coating usually is comprised of unsaturated monomers that polymerize with UV light (in the presence of a photo-initiator, which absorbs radiation and initiates polymerization). Unsaturated polyesters co-polymerized with vinyl monomers or multifunctional acrylates often are utilized. Most recently, however, vinyl ether monomers for cationic curing with epoxies have shown promise due to reduced shrinkage of cured films and improved adhesion to substrates.

Major uses for UV curing (room temperature) include use of flat stock-fillers in particle boards and hardwood flooring, and for high-gloss (no-wax) protective coatings for floor tiles, wood finishes, and other specialty uses.

Current Automotive Coating Trends

Until 1965 automotive topcoats consisted of nothing more than acrylic lacquer technology. As technology improved to include solvent-borne nonaqueous dispersions in the mid-1960s and medium solids thermoset coatings in the 1970s, two competing factors were becoming more important: control of emissions (VOC) and the quality of the appearance of the topcoat. Today, the quality of the topcoat includes chip resistance, environmental etch resistance, and metallic glamour; so virtually all automotive procedures (except in truck plants) now utilize basecoat/clear-coat technology. The basecoat is the pigmented film that is responsible for providing color. It is covered, in a wet-on-wet spray process, with a clear-coat that provides chemical resistance, UV resistance, chip resistance, and distinctness of image (DOI) (often referred to as the "wet look").

The total automotive coating composite, however, consists of more than just the basecoat/clear-coat. The metal sheet stock that is formed into the automotive car body is treated with a corrosion-resistant primer applied by cathodic electrophoresis. This coating, often referred to as "ECOAT" or "uni-prime" is a hydroxy functional epoxy/blocked isocyanate dispersion that deposits onto the cathode of an electrolytic cell at a voltage of between 250 and 425 V. Film thickness, which is dependent upon the residence time, temperature, and coating voltage of the electro-coat bath (electrolytic cell), varies between 0.8 and 1.2 mils (1 mil = 25 µm). The curing temperature required to deblock the isocyanate and yield a cross-linked network varies from 275 to 375°F for 20 min. The function of this coating is to provide a corrosion-resistant film to the metal sheet stock.

A primer surfacer, composed most often of a polyester to be cross-linked with a blocked isocyanate or melamine cross-linker, is applied next at a dry film thickness of approximately 1.0 mil by conventional spray application. The function of this coating is to provide filling and leveling properties as well as some stone-chip resistance. Cure conditions range from 250 to 325°F for 20 min.

The solvent-borne basecoat, consisting of polyester or acrylic binder to be cross-lined with melamine or isocyanate, is applied at a dry film thickness of between 0.6 and 1.5 mils, depending upon the color. Poorer-hiding colors, such as red and white straight shades, require the higher film builds, whereas metallics and darker colors are applied at the lower film builds. The solvent-borne basecoat is allowed to flash-off solvent for approximately 35 min at ambient conditions prior to application or the solvent-borne clear-coat. This process is referred to as a wet-on-wet application because the basecoat is not in its cured or cross-linked state.

Optionally, a waterborne basecoat can be utilized to achieve better aluminum orientation in metallic films. Several papers6,7 have described metallic orientation and the advantages of waterborne basecoats, so this concept will not be thoroughly discussed here.

The waterborne basecoat consists of a urethane or an acrylic dispersion with passivated aluminum for metallic glamour. The aluminum must be passivated to prevent reaction with water and the evolution of hydrogen gas. Passivating agents can include polymers, hexavalent chromium (now mandated for removal), or phosphate esters. The basecoat is pre-baked (with infrared or thermal ovens) to release the majority of the water prior to clear-coat application.

The solvent-borne clear-coat is spray-applied over the basecoat at a dry film thickness of 1.82.0 mils. The clear-coat, responsible for UV protection, chemical resistance, chip resistance, and so on, is most often composed of a hydroxy functional acrylic binder together with a melamine or an isocyanate cross-linker. UV light absorbers and hindered amine light stabilizers also are added to achieve better free-radical oxidative degradation resistance when the coating is exposed to sunlight.

The clear-coat is referred to as a two-component (2K) clear-coat when isocyanate is used as the cross-linker. The isocyanate must be added to the hydroxy functional acrylic just prior to spray application to prevent premature cross-linking (or gelation). For this reason, the two-part mixing is referred to as 2K. Cure conditions for isocyanate-cured clear-coat range from room temperature (car refinish) to 250°F for 30 min.

One component, or 1K clear-coats are equivalent to 2K clear-coats with the exception that they are cross-linked with melamines. Cure conditions for 1K clear-coats are 250°F for 30 min. Blocked isocyanate versions of 1K clear-coats also are available, but cure requirements necessitate 275°F or higher.

Significant research is being conducted on waterborne clear-coats in attempts to achieve better VOC emission levels. Waterborne clear-coat research is limited to water-dispersible acrylics because high gloss and good weatherability can be attained. Removal of water from the film, however, is a problem, as coalescence often begins prior to through-cure, often resulting in "solvent-popping." Much work remains to be performed on optical formulation parameters (i.e., solvent blend, anti-cratering surfactants, rheology control, etc.) in order to attain defect-free coatings.

Coatings for Plastics

The use of plastics in automotive applications has expanded considerably over the past 20 years. In 2001, approximately 250 lb of plastics were used per automobile compared to less than 100 lb five years earlier. With the increased use of plastics, coating research has concentrated on the painting of these often difficult substrates. Among the numerous prerequisites for coatings on plastic are: wetting of the substrate; uniform film development; a cure response below the heat distortion temperature of the plastic; adhesion to the substrate; inhibition of substrate imperfections; and often a conductive nature to increase the TE of subsequent topcoat application.

Plastic types are variable, ranging from flexible to rigid and from thermoplastic to thermoset. Because of these variations, the type of coating applied to them also varies, depending on cure response, solvent sensitivity, and modulus. The type of coating also depends upon the end use of the coated plastic. For example, topcoats can be applied directly to the plastic without the use of a conductive primer if two-toning or blackout areas are not part of the styling latitude. Adhesion promoters also are often utilized if a particularly difficultto-adhere-to plastic is encountered (i.e., polypropylenes or thermoplastic polyolefins (TPOs), a blend of elastomer and olefin).

Conductivity in plastic coatings is achieved through the use of a conductive pigment, most often conductive carbon black. Conductivities of the order of 1 MΩ/cm2 are required for effective transfer efficiencies of subsequent topcoat applications. Without the use of a conductive primer, plastics in themselves would not be conductive, and hence would be very unreceptive to topcoat transfer efficiencies of greater than 20 percent.

The adhesion of a coating to plastic is related to the substrate wettability and also to physicalchemical forces. Both covalent (bond formation) and dispersive (London or van der Waals) forces are responsible for the adhesion of the coating due to physicalchemical forces.

Substrate wetting often is measured by contact angle measurements: the smaller the contact angle, the better the wetting of a coating to the plastic. The wettability of a coating on a plastic substrate is related to the surface tension of the substrate: the lower the surface tension of the plastic, the more difficult it is to wet.

Polypropylene, in particular, has a very low surface tension (less than 32 dynes/cm2). Because of the low surface tension, it is very difficult to wet out; and because of the apolar nature of the polypropylene backbone, it is a very difficult material to adhere to. Therefore, copolymers of ethylene propylene diene monomers (EPDM) with polypropylene have been developed. The higher the level of the EPDM in these TPOs, the more paintable is the substrate.

Conventional coating approaches to TPO, which rely on the use of substrate pretreatments such as flaming, corona discharge, plasma or chemical pre-treatment, have led to a substantial oxidation of the surface, which enhances adhesion. These pretreatments, however, have a limited service life before the surface reverts to its unoxidized, apolar condition.

Chlorinated polyolefins (CPOs) also are a mainstay of adhesion promotion to TPO. The CPO generally is applied as a dilute solution (535 wt. %) in aromatic solvents to the TPO to act as a "tie-coat," to which the topcoat is applied. These materials are often humidity sensitive and only adhere to TPOs with EPDM levels of greater than 5 percent.

New Cross-Linking Technologies

The traditional cross-linking technologies utilized in the coatings market range from hydroxyl functional binders (acrylics or polyesters) cross-linked with melamines (1K) or isocyanates (2K) to drying oils that cure through oxidation of unsaturation in the backbone, to epoxy resins cross-linked with amino resins (amido amines, amines, etc.).

New cross-linking technologies, however, are being developed with hopes of achieving more environmentally acceptable coatings. Fluoropolymer coatings, epoxy-anhydride new curing technology (NCT), and siliconized-acrylics new enamel technology (NET) are a few of the new coating chemistries being introduced. Increasing demands by automotive producers have necessitated these advances. The need for properties such as chip resistance and resistance to "acid rain" may result in further technological advances.

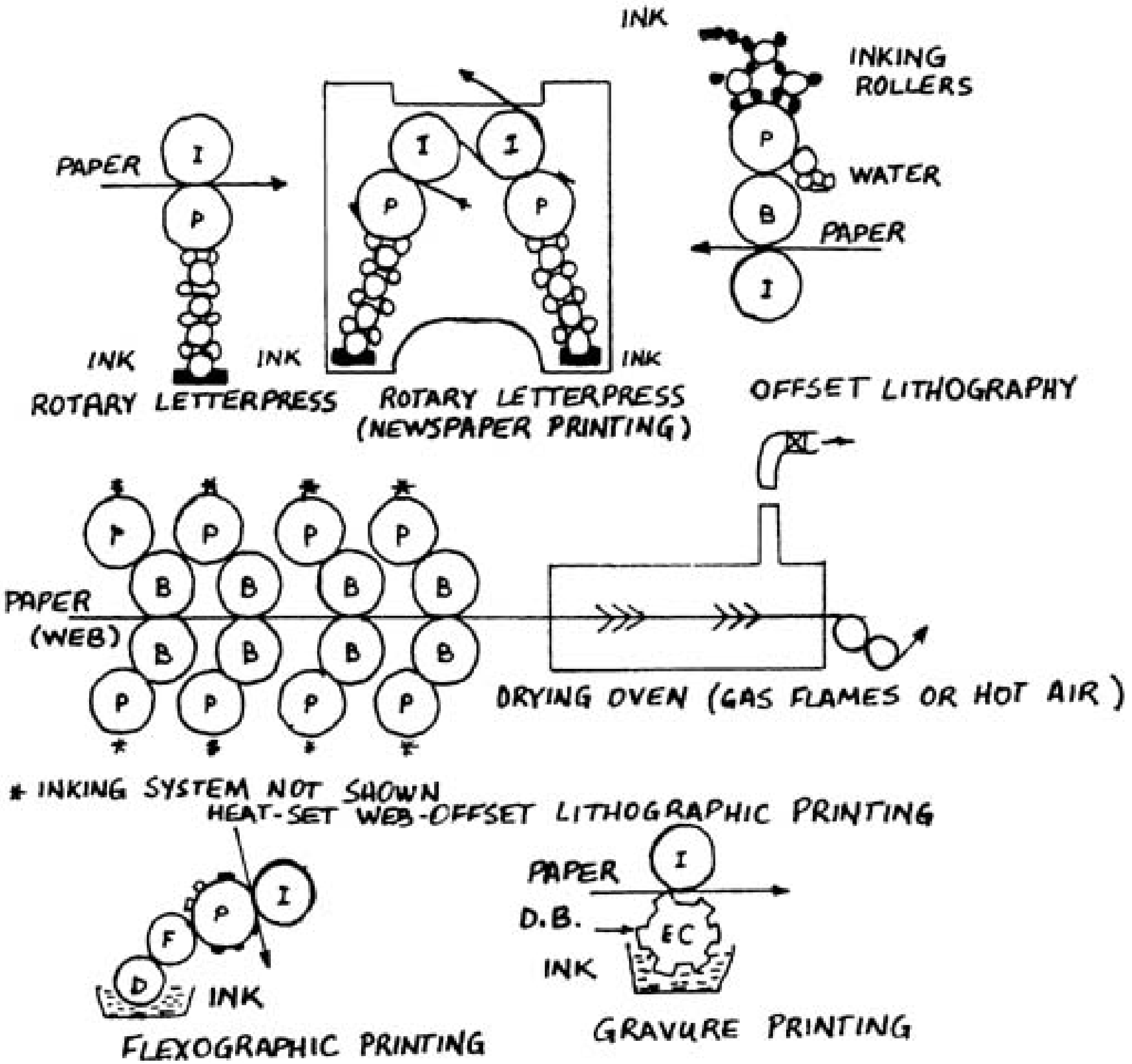
28.6 Printing Inks

Printing inks, utilized in printing and packaging applications, are colored, liquid materials consisting of essentially two components: the colorant, an insoluble solid or a dye, and the vehicle, a liquid in which the colorant is suspended or dissolved. Recent years have seen significant progress in water-based inks, but compared with the phenomenal growth in water-based paints, the advances in water-based inks have had a much smaller impact on the industry. The upsurge of interest in water-based inks, which began in the mid-1970s, was stimulated by the need to conserve expensive petrochemical-derived organic solvents, attention to environmental pollution problems, and increasing pressures from health and safety legislation.

Inks can be applied to a wide variety of substrates, including plastics, paper, textiles, metals, and glass. In contrast to paint films, inks are applied very thin, varying from 2 to 30 µm depending on the print process. The print process consists of application of the ink to the substrate by a printing press of varying design. As such, inks are classified by the type of process through which they are applied. There are five classifications: flexo-graphic, gravure, offset lithographic, letterpress, and screen.

The mechanisms of letterpress, offset lithographic/flexo-graphic, and gravure printing processes are outlined in Fig. 28.2.8 In letterpress and offset lithographic processes, a thin film of ink is carried to the printing plate by means of a train of rotating rubber rollers. Both of these processes preclude the use of volatile organic solvents in the inks because of the swelling effect or attack on the rubber rollers and blankets and because of their high evaporation rates, which could lead to premature drying or tack development of the inks on rollers, blanket cylinders, and printing plates. Water, too, evaporates more rapidly than the high-boiling aliphatic-type solvents normally used in letterpress and offset lithography, and the evaporation is accelerated by the heat that can be developed (temperatures of 30°C or more) in the rollers after a period of running.

Fig. 28.2 Mechanisms of letterpress, offset lithographic/flexo-graphic, and gravure printing processes.9



Several different types of drying mechanisms may be operative in printing of letterpress and offset lithographic inks (see Table 28.6).9 As shown in Table 28.6, the ink distribution systems of flexo-graphic and gravure presses are quite different from those of letterpress or offset lithography. There are no long roller trains, and because the drying mechanisms depend mainly upon evaporation of volatile solvent components, it is not surprising that most of the development work has been on inks for these processes.

TABLE 28.6 Letterpress and Lithographic InksDry Mechanisms9

		Principal Ink	
Type Ink or Process Remarks	Vehicle	Typical Substrates	Drying Mechanism
Letterpress W	Mineral oil	Newsprint A	Non-volatile mineral oil
Letterpress S	Oleoresinous and solvent	Papers, boards A, O	High-boiling range solvent
Letterpress S	Resin and glycol	Papers, boards P	High-boiling glycol
Letterpress S	Polyfunctional (acrylated) compounds	Papers, boards, R foils	No solvent ultraviolet curing
Offset litho W	Mineral oil and resin and solvent	Newsprint A	High-boiling solvent
Offset litho S	Oleoresinous and solvent	Papers, boards A, O	High-boiling solvent
Offset litho S	Oleoresinous and solvent	Tinplate O, E (stoving)	High-boiling solvent
Offset litho S	Polyfunctional (acrylated) compounds	Papers, boards, R foils, tinplate	No solvent ultraviolet curing
Offset litho W	Resin/solvent	Papers A, E (heat-set)	Gas flame hot air, high-boiling solvent

Notes: A: Absorption; O: Autoxidative/polymerization; P: Precipitation/gelation; R: Photochemical/polymerization, ultraviolet cure; S: Sheet-fed; W: Web-fed.

Inks that are printed by the flexo-graphic and gravure printing processes are characterized by their extremely fluid nature, and are generally termed "liquid inks." When compared to the fluid nature of flexo-graphic and gravure inks, lithographic and letterpress inks are much more viscous and paste-like.

In screen printing, the ink is forced through the open areas of a stencil supported on a mesh of synthetic fabric, stretched across a frame. The ink is forced through the mesh onto the substrate underneath by drawing a squeegee across the stencil. The inks have to remain fluid and coherent on the mesh before printing, transfer through the mesh without its sticking to the substrate, flow to make the ink film continuous and so avoid the pattern of the mesh but maintain the image, and then dry rapidly once applied to the substrate. Inks that are able to print by this method are of household paint consistency; they are intermediate in viscosity between the fluid flexo-graphic and gravure inks and the paste-like lithographic inks.

An ink must be transformed to a solid from its liquid supply form once on the substrate. This change of state is referred to as an ink drying, and can be physical (absorption or evaporation) or chemical (oxidative or photochemical radiation induced). There are several forms of radiation used to dry ink, including UV, infrared, EB, microwave, and radio frequency.

28.7 Pigments

Pigments are generally incorporated into paints to affect properties associated with appearance such as color, opacity, gloss, metallic look, and depth. In addition, pigments commonly are used to protect the substrate against corrosion, attack by microbes or barnacles, and to retard flammability. Other desirable properties can be obtained by the addition of pigments. For example, pigments can be used in paints to control flow and leveling as well as either spray viscosity or brush-ability, depending upon the type of system. In addition, pigments are added to paint to control the cost of raw materials.

One of the most important properties of pigments, however, is the ability to cover or "hide" the surface being painted. This property is commonly referred to as "hiding power" or opacity. The fewer coats required to hide the surface, the higher the hiding power or the pigment's opacity. Pigments of different colors have different degrees of opacity. In general, darker colors have a higher hiding power than lighter colors, in part because of their relative opacities. In practice, this trend is also dependent upon the refractive index of both the pigment and the binder; that is, the difference in refractive index between the pigment and the binder determines the hiding power of the paint. Hence, the larger the refractive index difference, the higher the hiding power. In addition, the particle size of the pigment critically affects the hiding power of the pigment. There is an optimum particle size range, 0.20.4 µm (which is approximately one half the wavelength of light in air), which affords maximum light scattering and hiding.

Pigments ideally should be non-soluble solids with low oil absorption to facilitate dispersion and allow maximum pigment loading, which would also reduce the VOC content of the paint. In addition, pigments should be chemically inert, thermally stable, pure in their composition, and nontoxic.

Pigments can be classified into two main categories: inorganic pigments and organic pigments. Among the more common inorganic pigments are white hiding pigments, extender pigments or fillers, blacks, colors, and metallic pigments. Organic pigments consist of colors and the carbon blacks.

Inorganic Pigments

Inorganic pigments generally hide better than organic pigments when dispersed in paint. Inorganic pigments impart flowability to paint because they do not tend to swell in the presence of common paint solvents. In addition, inorganic pigments do not bleed inorganic solvents, and they afford excellent color permanence, especially in exterior paints. Inorganic pigments are thermally stable and generally impart heat resistance to the coating.

White pigments are important not only as the principal pigment in white paint but also as hiding pigments in many shades and the tinting colors. The most commonly used of the white pigments are the non-reactive titanium dioxides and the reactive zinc oxide. These pigments are designed reactive or non-reactive, depending on whether or not they interact chemically with the binder or vehicle portion of the paint.

Titanium dioxide is available in two crystalline formsanatase and rutile. The anatase form has about 75 percent of the hiding power of the rutile form and is less resistant to chalking than the rutile form. Because of its high refractive index (2.76), titanium dioxide imparts excellent opacity to the paint. Because the hiding power of the titanium dioxide is so high, good opacity can be achieved by the addition of small amounts of the pigment to the paint, thus not adding significantly to the cost of the paint. The rutile form is available in grades designed specifically for enamels, flats, and solvent-based and water-based coatings.10

Although zinc oxide has about 15 percent of the hiding power of titanium dioxide and has a fairly high cost, it contributes a number of important properties to paint. Zinc oxide can form zinc soaps in some formulations, which can aid in pigment wetting and grinding. It also can be utilized to improve film hardness and color retention of the film on aging. The most common use of zinc oxide is in exterior finishes to reduce chalking and the growth of mildew in house paints.11

Extender pigments, because of their relatively low refractive index, exhibit poor hiding power and must be used in conjunction with pigments of high hiding power. The principal function of extender pigments in paint is to reduce the raw material cost. They also function to reduce gloss, as well as to control texture and viscosity. Calcium carbonate, more commonly known as whiting pigment, is available in a variety of particle sizes and surface treatments and is used to control sheen, degree of flow, degree of flatting, and tint retention. Magnesium silicate or talc generally is used because of its low cost and excellent durability. Talc imparts good anti-settling properties and low sheen. Aluminum silicate (China clay) is used mainly in water-based systems. China clay is easily dispersed with a high-speed dispenser and may even add to the hiding power of water-reducible paints. Other extenders such as diatomaceous silica, regular silica, barites, and mica also are frequently used to impart specific properties to coatings, which are determined mainly by the particle size and shape of the pigment.

Most of the black pigment used in coatings is organic carbon black. There are, however, a few inorganic black pigments such as black iron oxide, which is used as a tinting black having brownish tones and also is used in primers. In addition, some inorganic black pigments such as mineral or thermal blacks are used as low-cost black extender pigments.

Iron oxide pigments are low-cost, high-opacity pigments that impart excellent properties to coatings. These pigments are easy to disperse in both water and oil, are alkali- and chemical-resistant, and hence can be utilized in both water-based and solvent-based systems. Although both the red and yellow iron oxides afford rather dull colors, they possess high tinting strength and are popular shading colors. In addition, iron oxide pigments are extremely durable, colorfast, and non-bleeding.

Lead chromate pigments are mainly used as corrosion inhibitive pigments. These pigments are of low opacity and can be combined with higher opacity pigments for use in primers. Zinc chromate, also an anti-corrosion pigment, is used in metal pretreatments such as "wash primer," which also promotes adhesion to aluminum and steel.

Iron blue, one of the most widely used pigments, combines low cost, good hiding power, and high tinting strength. This blue pigment exhibits good exterior durability and good heat resistance. Iron blue should not be used in water-based paints because it has very poor alkali resistance.

Ultramarine blue has poor opacity but high tinting strength, high heat resistance, and good alkali resistance. It is primarily used as a tinting color, and when added to whites makes them look whiter by giving them a bluish cast.

Chrome greens range in color from light yellow greens to dark blue greens. They are bright, durable, economical, and have high hiding power. They are used less frequently of late because of their lead content. Chrome greens are not suggested for use in aqueous systems because of their poor alkali resistance.

Chrome oxide green is a dull pigment having excellent exterior durability and very good chemical as well as alkali resistance. Thus, this pigment is suitable for use in either water-based or solvent-based systems. Chrome oxide green has moderate hiding power and is easy to emulsify, and its infrared reflectance properties make it valuable for use in camouflage paints.

Organic Pigments

Organic pigments are insoluble solids, and thus are differentiated from dyes, which are colored organic compounds that are soluble in the media in which they are used. Pigments require a binder in order to be used as colorants. Organic pigments that are insoluble in their pure form are referred to as "toners." Organic pigments that require precipitation by an inorganic base such as a metal to be insoluble are referred to as "lakes."

Organic pigments have greater decorative value than inorganic pigments; that is, organic pigments have a greater esthetic appeal due to higher chroma. The higher tinting strength of organic pigments partially offsets their higher cost. Organic pigments have a smaller particle size than inorganic pigments and a higher oil absorption. In addition, organic pigments are generally less toxic than inorganic pigments.

Hansa yellows are insoluble azo pigments having a wide range of shades. They are generally excellent in masstone lightfastness although not good enough in tint lightfastness to warrant outdoor use. Compared with the chrome yellows, they have less hiding power and greater tint strength. The Hansa yellows are most commonly used in water-based coatings and air-dried coatings. Benzidine yellows are no longer used because they have been shown to be carcinogenic. Nickel azo yellows are transparent pigments used in greenish yellow enamels and tinting enamels. These pigments have better permanency than the benzidine yellows although they tend toward darker transparent masstones. Nickel azo yellows impart excellent alkali resistance and fairly good bleed resistance. The tinting strength of these pigments increases upon aging in the formulation. Vat yellows are comprised of flavanthrones and anthrapyrimidines. Of these, the flavanthrones are the more red in hue. The vat yellows have very good lightfastness in weak tints, and they impart very good heat resistance and bleed resistance to the coatings. Tetrachloroiso-indolinone-based pigments, although relatively expensive, are of very high quality, imparting good lightfastness to the coatings.

Benzidine orange pigments are bright, strong pigments that exhibit good alkali, heat, and bleed resistance. There are both opaque and transparent versions of these pigments, and they are fairly lightfast although they are not recommended for exterior use. Benzidine orange is no longer used because of its potential carcinogenic nature. Dianisidine orange pigments are more bleed-resistant than the benzidine oranges but less colorfast. Dinitroaniline orange pigments have better permanency than the two orange pigments described above. Although the dinitroaniline orange pigments bleed in most solvents, they are very stable in aqueous systems and exhibit superior lightfastness. Tetrachloroiso-indolinone, brominated anthanthrone, and perinone oranges are expensive, but very high-quality orange pigments, characterized by their excellent masstone and exterior lightfastness.

Copper phthalocyanine greens afford very clean, semi-transparent masstone, non-bleeding tints ranging from bluish green to yellowish green. These pigments have excellent heat and chemical resistance, are very lightfast in both masstone and tint, and have high tint strength. Pigment green B is an iron salt of nitro-sobetanaphthol that affords a semi-transparent masstone having a dull yellowish tint. Although this material is not lightfast enough for exterior applications, it does have excellent alkali resistance and is suitable for use in aqueous pigment dispersions.

Phthalocyanine blues are very clean, semi-transparent, non-bleeding pigments yielding high tint strength. As with the phthalocyanine greens, the blues have excellent heat and chemical resistance and are very lightfast in both masstone and tint. Also as with the greens, these blue pigments at high concentrations may bronze upon exterior exposure although this property is vehicle dependent. These characteristics apply to the crystallization and flocculation grades.

These pigments are highly recommended for use in water-based systems. Indanthrone blues are expensive vat-type, non-bleeding, semi-transparent masstone pigments having excellent masstone and tint lightfastness. Although they exhibit very good heat and chemical resistance, these pigments find limited use in coatings.

Carbazole dioxazine violets are semi-transparent masstone pigments having high tint strength, excellent heat and bleed resistance, and good lightfastness. Isoviolanthrone violets afford clean tints exhibiting excellent chemical resistance and very good lightfastness, but they bleed badly in paint solvents.

Toluidine reds are insoluble azo pigments ranging from light to deep reds having excellent chroma and masstone lightfastness. Although these pigments hide very well and exhibit very good alkali resistance, they are not suitable for exterior use and bleed in paint solvents. Toluidine reds have been replaced to a large extent by a combination of molybdate orange and quinacridone or rubine toners. This combination affords better gloss and gloss retention, better bleed resistance, and less haze. Compared with the toluidine reds, this combination does sacrifice some cleanliness and brilliance of tint as well as some masstone lightfastness. Para reds are insoluble azo pigments that are deeper, less bright, and not as masstone-lightfast as the toluidine reds. They have good opacity and good alkali resistance, but they bleed very badly in paint solvents. These pigments can be utilized in inexpensive air-dried water-based systems.

Quinacridones possess hues from gold to deep maroons and violets. These reds afford excellent bleed, heat, and alkali resistance, as well as very good tint and masstone lightfastness.

Carbon black is probably the most widely used pigment in the coatings industry. It is popular both as a straight shade and as a tinting color. It comes in a wide variety of particle sizes and surface treatments, the choice depending upon required jetness, tone, ease of dispersion, and of course, cost. Carbon black is characterized by excellent hiding power of high opacity, high tint strength, excellent durability, and very good lightfastness. Large-particle-size carbon blacks are chosen for their ease of dispersion and low resin demand, as well as for being less susceptible to floating and flooding than the finer particle sizes. They should, however, be used in applications that do not demand high levels of jetness. The finer particle sizes, higher surface area carbon blacks afford extremely jet, very durable, high gloss coatings. These carbon blacks would be very suitable in automotive topcoat applications. The fine-particle-size carbon blacks also are used frequently in conductive coatings. Although all carbon blacks are good conductors, differences exist based on surface area, structure, and volatile content. In general, however, conductivity of the carbon black increases as the surface area and structure increase and as the volatile content of the carbon black decreases. The structure of the carbon black is defined as the agglomeration of the fine particle into treelike or "dendritic" shapes.

Pearlescent Pigments

Pearlescent pigments are comprised mainly of mica platelets coated with a thin layer of titanium dioxide. Co-precipitating titanium dioxide and other metal oxides onto the mica results in pigments exhibiting a colored luster effect. These pigments have a very high refractive index, indicating high hiding power or high opacity. In addition, because of their light-reflective properties, they impart some exterior durability to the coatings; and because of their small, thin, flat structure and high refractive index, they give the coated object a three-dimensional effect, that is, the appearance of depth. This is quite different from the apparent surface gloss due to the use of metallic pigments. Pearlescent pigments are available in a range of particle sizes and characteristics, and provide a variety of effects. These include a stain-luster appearance resulting from fine-particle-size pigments. These pigments are alkali- and acid-resistant and extremely durable when coated with rutile titanium dioxide. Pearlescent pigments are readily dispersible in water and afford stable metal-like water-based coatings.

Aluminum Pigments

Aluminum pigments generally are provided in paste form to eliminate the hazards of dust explosion or ignition during handling and storage, as occurs with the aluminum powders. In addition, an aluminum paste is very economical because of its high hiding power, excellent durability, and low specific density.

Aluminum paint films provide excellent protection of substrates because the aluminum flakes align in many parallel layers. The layers of aluminum flake are staggered, thus yielding a tortuous path for the transmission of gases, chemicals, light, and moisture. Therefore, it takes longer for these materials to penetrate the film to get to the substrate. Aluminum paints made up of the leafing-type of aluminum paste reflect 7580 percent of light and radiant heat, and are commonly used as exterior maintenance-type coatings or decorative coatings. In fact, the minute flakes in aluminum paste impart a smooth, brilliant appearance to the film. The non-leafing type of aluminum flake can be tinted with other various colored pigments to effect brilliant finishes such as those used in automotive metallic topcoats.

28.8 Lacquers

Some time before 200 B.C., the Chinese used an exudate from the conifer *Rhus verniciflua* (later known as the varnish tree) as a coating. The process was further developed in Japan, where by A.D. 400, the product was known as Japanese lacquer. These lacquers are a type of oleoresin that dry by oxidation in a damp atmosphere.11

Lacquers are defined as non-convertible coatings; that is, lacquer film formation does not depend upon any chemical reaction. Film formation is dependent only upon the evaporation of solvents, and the formed film remains soluble in the parent solvent blend. Lacquers generally are spray-applied, taking advantage of the solvent's ability to dissolve the previously applied coat. Chlorinated rubbers and nitrocellulose esters are typical examples of lacquers.

Nitrocellulose lacquers can form glossy coatings having excellent adhesion when other suitable resins are added that are compatible in all proportions. Two types of cellulose nitrates are available, a low-viscosity type that affords high-solids, high-film-build coatings and a high-viscosity, low-solids type. The primary use of pigmented nitrocellulose lacquers has been in the car refinish and the wooden domestic furniture markets. The chlorinated rubber lacquers are used principally as protective coatings for steel and other vulnerable materials in particularly harsh environments.

Vinyl lacquers are used mainly where a high degree of chemical resistance is required; these lacquers are based on vinyl chlorides and vinyl acetates. Acrylic lacquers are based on methyl methacrylate and methyl acrylate polymers and copolymers. Other esters of acrylic and methacrylic acid also may be used to make non-convertible film formers. Judicious selection of these acrylic acid or methacrylic acid esters allows one to produce film formers with specifically designed properties such as hardness, flexibility, gloss, durability, heat, and chemical resistance. Acrylic lacquers, however, are not noted for their water resistance. The principal uses of acrylic-type lacquers are fluorescent and metallic paints, car refinish applications, clear lacquers and sealers for metals, and protective coatings for aircraft components and for vacuum-deposited metals, as well as uses in pigmented coatings for cabinets and appliances.

Vinyl emulsion paints that contain a non-convertible resin dispersed in water along with pigment can be considered as waterborne lacquers if film formation results from evaporation of the water followed by coalescence of the binder in which the pigment is dispersed. As environmental regulations become more strict, the role of traditional lacquers will essentially go the way of the dinosaur. The only allowable lacquers or non-convertible coatings eventually will be the water-based type. Indeed, as environmental regulations become more restrictive, the challenges facing the coatings chemist and the formulator become ever more demanding.

28.9 References

Sawyer, R., Paint and Coatings Industry, 25 (1990, September).
Industrial Finishing, 31 (1991, March).
Schulte, S., Products Finishing, 28 (1990, June).
Greissel, M., Industrial Paint and Powder, 18 (2001, February).
J. Coating Technology, 62(780), 83 (1990).
Besold, R., Reisser, W., and Roth, E., Waterborne and Higher Solids Symposium Proceedings, New Orleans, LA, 1991, February 68.
Akkerman, J. M., Prog. Org. Coatings, 17, 5368 (1989).
Hutchinson, G. H., JOCCA, 68 (12), 307 (1985).
Martens, C. R., in Riegel's Handbook of Industrial Chemistry, p. 787, J. Kent (Ed.), Van Nostrand Reinhold, New York, 1983.
Singer, E., Paint Handbook, pp. 3-233-32, McGraw-Hill, New York, 1981.
Skolnik, H., in Organic Coatings: Their Origin and Development, p. 16, B. Seymour and H. F. Mark (Eds.), Elsevier, New York, 1990.

29.1 Nitrogen

Characteristics

Nitrogen is colorless, odorless and slightly lighter than air with a density of 0.967 (air = 1.0). Some people consider nitrogen to be one of the most dangerous gases. This is because a person in an atmosphere of nitrogen can lose consciousness without any warning symptoms in as little as 20 sec. Death can follow in 34 min. One deep breath of 100 percent nitrogen can be fatal because nitrogen will displace carbon dioxide in the body completely. In the absence of a carbon dioxide signal to the brain, breathing stops.

No one should work in or enter atmospheres containing less than 19.5 percent oxygen, unless equipped with a self-contained breathing apparatus or a breathing air mask. This is also true of rescue personnel who can be overcome by the same oxygen-deficient atmosphere as the initial victim.[39]

The properties of nitrogen are listed in Table 29.1.

TABLE 29.1 Physical Properties of Nitrogen

Property	Value
Molecular weight	14.0067
Boiling point, °C (°F)	-195.8 (-320.5)
Freezing point, °C (°F)	-209.86 (-346.0)
Critical temperature, °C (°F)	-146.95 (-232.4)
Critical pressure, kPa (psi)	3393 (492.3)
Critical volume, cc/g-mol	89.5
Specific heat 20°C and 147 psia	
c_p (BTU/lb °F)	0.247
c_v (BTU/lb °F)	0.176
21°C (J/(kg-K)	1.046
$k = c_p/c_v$	1.41
Heat of formation of gas, ·H, (kJ/mol)	0.3604
Heat of fusion at melting point J/g	25.6
Heat of vaporization at boiling point J/g	199
BTU/lb	85.5
Solubility in water, g N_2 per 100 g H_2O at 760 mm	
0°C	0.00239
20°C	0.00189
40°C	0.00139
60°C	0.00105
Specific gravity	
Relative to air	0.967
-195.8°C (Liquid)	0.808
-252°C (Solid)	1.026
Specific volume (70°F and 1 atm), ft^3/lb	13.80
Density, kg/m^3	
20°C	1.16
Liquid at boiling point	808.6
Solid at melting/freezing point	1028
Thermal conductivity, W/m-K	
Gas at 21°C	0.026
Liquid at -200°C	0.14456

Although the ability of nitrogen to easily unite with other elements is quite limited, it does form some interesting compounds. When nitrogen combines with certain elements, it produces dyes that rival the rainbow in the brilliance of their colors. With other elements, nitrogen forms drugs such as morphine, quinine, and acetanilid that ease pain and combat diseases. With other compounds, nitrogen forms some of our most powerful explosives such as nitroglycerin and guncotton. Nitrogen can also be used to make some very deadly compounds such as ptomaines. Some nitrogen compounds provide pleasant-tasting flavors and sweet perfumes, while others are so vile in taste and odor that they are beyond description. In nitric acid, we find nitrogen in a strong acid, while in ammonium hydroxide we have it in a well-known base.[7]

Nitrogen Production Processes

Economical fixation of nitrogen from the atmosphere has been a never-ending quest. It has been a difficult task because elemental nitrogen is comparatively unreactive. It only combines with most elements under high pressure and/or high temperature. As a result industry meets its demand for high purity nitrogen by obtaining most of it from liquid air. See Comyns[1] and Chapter 13 (Industrial Gases).

Nitrogen production via pressure swing absorption (PSA) is based on the principle that nitrogen and oxygen have different absorption rates on carbon molecular sieves (CMS). Some of the nitrogen production processes that use this technology are described in Comyns,[1] Suzuki,[3] Suzuki,[4] *Eur. Chem. News*,[5] and *Chem. Mark. Rep.*[6]

29.2 Nitrogen Fixation

Nature uses nitrogen fixation to combine free nitrogen. Nodules on the roots of many plants contain bacteria that extract nitrogen from the air and convert it into soluble nitrates that fertilize the soil. Most nitrogen products are based on nitrogen oxides and ammonia; so the history of fixation of nitrogen as nitrogen oxides and ammonia is of interest.

Nitrogen Oxides

The five oxides of nitrogen are:7 Nitrous Oxide (N2O), Nitrogen Dioxide (NO2), Nitric Oxide (NO), Nitrogen Pentoxide (N2O5), and Nitrogen Trioxide (N2O3). Nitrogen trioxide and nitrogen pentoxide are the anhydrides of nitrous acid and of nitric acid.

Nitrous oxide (laughing gas) is a colorless gas. Since the 1840s the major use of nitrous oxide has been as an anestheticspecially by dentists. It is also used as a propellant in some aerosol cans, in atomic absorption spectrophotometry, in cryosurgery, and in racecar engines to provide extra power and acceleration.

Nitric oxide is a colorless gas that is insoluble in water. The most noticeable chemical characteristic of nitric oxide is the ease with which it combines with oxygen to form nitrogen dioxide. Nitric oxide is produced in considerable quantities in the industrial preparation of nitric acid.

In the late 1980s nitric oxide was discovered as a product of enzymatic synthesis in mammals. And in 1998, the Nobel Prize in Physiology was awarded to the scientist who discovered the role of NO as a biological messenger. Nitric oxide uses range from its role as a critical endogenous regulator of blood flow and thrombosis to a principal neurotransmitter mediating erectile function to a major pathophysiological mediator of inflammation and host defense. These major discoveries have stimulated research into a vast array of fields. The Nitric Oxide Society has been formed to coordinate meetings, and they also publish a research journal. More information is available at www.apnet.com/no.

Nitrogen dioxide is a red-brown gas that is very soluble in water. It has an unpleasant odor and is quite poisonous. When nitrogen dioxide is cooled, its brown color fades to a pale yellow.

$$\begin{array}{ccc} 2NO_2 & \leftrightarrows & N_2O_4 \\ \text{Nitrogen dioxide} & & \text{Nitrogen tetroxide} \\ \text{(Brown)} & & \text{(Yellow)} \end{array}$$

The importance of nitrogen dioxide lies in the fact that it combines readily with water to form nitric acid.

$$3NO_2 + H_2O \rightarrow 2HNO_3 + NO$$

This is the final step in the synthetic production of nitric acid. Nitrogen dioxide also acts as an oxidizing agent in the manufacture of sulfuric acid.

Nitrous oxide production technology can be licensed from the Sanghi Organization in Worli, India (www.sanghioverseas.com). Their technology is based on heating ammonium nitrate to 250°C and then purifying the resulting gas stream.

Other processes have been proposed for nitrogen fixation, but none have been commercialized. For example, it was shown that low energy ultraviolet photons can form nitric oxide at room temperature under oxidizing conditions on the surface of a titanium dioxide-containing material.8

Research continues on the formation of nitrogen oxides. The high temperature fixation of nitrogen as oxides remains an environmental problem rather than a commercially attractive process, as large quantities of nitrogen oxides are produced by fixation in high temperature combustion processes such as power plants, automobile engines, and home furnaces. The U.S. EPA promulgated regulations in 2000 (NO*x* SIP Call) that require several states in the Eastern United States to reduce their NO*x* emissions by a combined total of 35 to 40 percent by 2005 (see www.epa.gov/ttn/otag).

Ammonia

Ammonia was prepared by the alchemists by distilling leather scraps, hoof, horns, and other animal refuse. Since they found that the best ammonia was obtained by distilling deer antlers, the alchemists called ammonia, "Spirits of Hartshorn."7 Nature forms ammonia by decomposition of proteins. The odor of ammonia can often be detected around manure piles and other decaying organic matter.

A process for synthesizing ammonia from nitrogen and hydrogen, using high temperatures and pressures and an iron-containing catalyst, was invented by Fritz Haber at BASF in 1908.1 In 1909, C. Bosch of BASF built a pilot plant using an osmium-based catalyst, and in 1913 a larger plant was built in Germany.9,10,11,12,13,14 The chemistry for this process is:

$$N_2 + 3N_2 \leftrightarrows 2NH_3$$

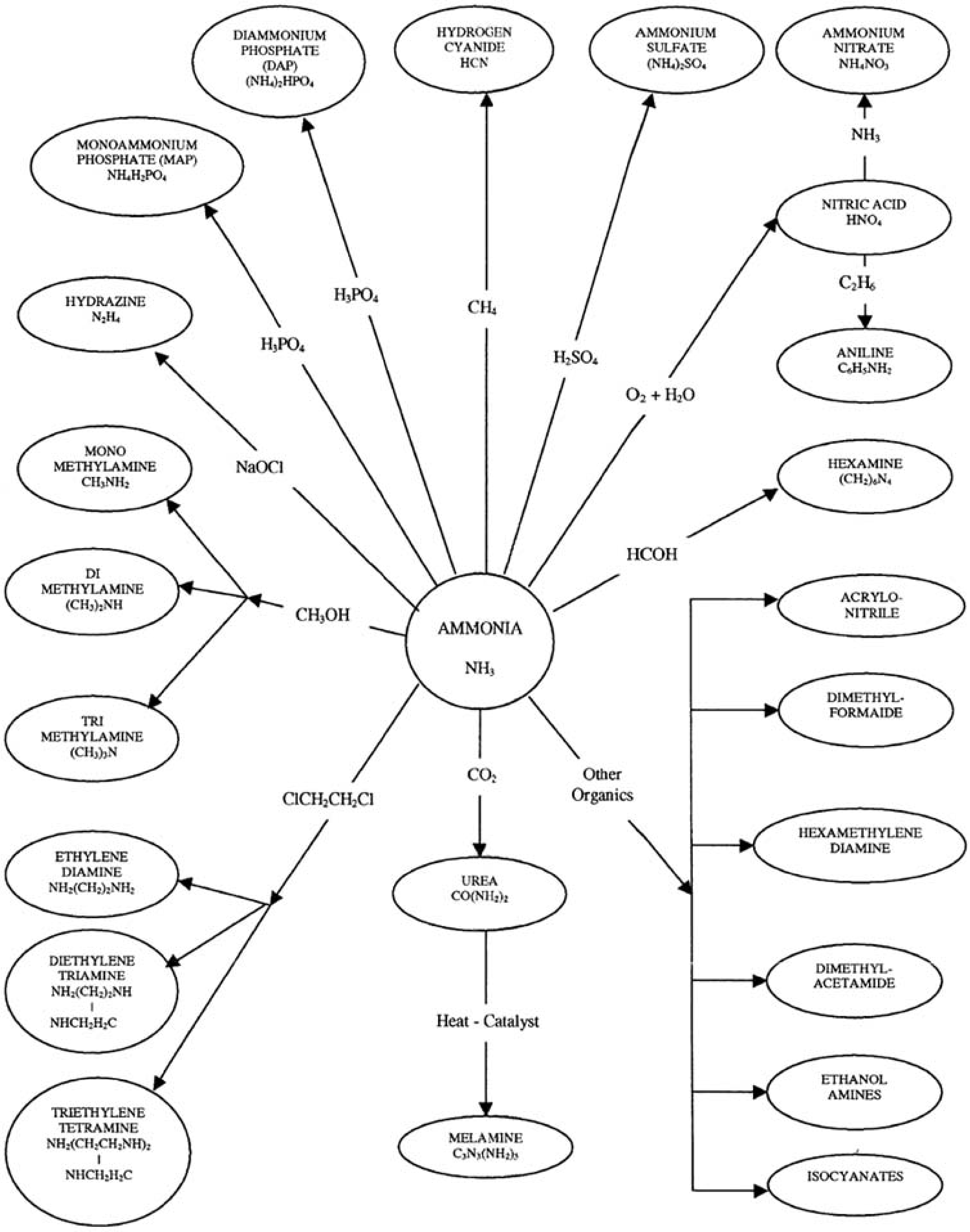
In 1883, the Mond gas process produced ammonia by gasifying coal at a relatively low temperature by using a mixture of air and steam.1 It was invented primarily to produce ammonia that was needed for the Solvay process that makes sodium carbonate.29,30

Some other processes that have been used to make ammonia are:

1. the Casale ammonia process in Switzerland in 19211,12,19
2. the high-pressure ClaudeCasale ammonia synthesis process in the 1920s1,12,20,21
3. the Fauser process in Italy in 192412
4. the Mont Cenis process that was similar to the HaberBosch process except that it used coke-oven gas.1 It was started up in the Netherlands in 192931,32
5. the Braun process is a variation on the classic ammonia synthesis process in which the synthesis gas is purified cryogenically.1 It has been widely used since the mid-1960s18
6. the AMV process for making ammonia was invented by ICI and announced in 1982.1 It is described in Livingston and Pinto,15 *Chem. Eng.*,16 and Hydrocarbon Processing.17
7. the Leading Concept Ammonia (LCA) is essentially a simplified form of the standard ammonia synthesis process that is more suitable for smaller plants. It is described in Comyns,1 *Chem. Eng.*,26,27 *Hydroncarbon Processing*,28 and Hill.29
8. the Bypass Ammonia Synthesis (BYAS) process can be used for economical expansion of existing ammonia synthesis plants. It is described in Comyns.1
9. the Kellogg Advanced Ammonia Process (KAAP) is the first high-pressure ammonia synthesis process that makes ammonia from nitrogen and hydrogen without the aid of an iron-containing catalyst.1 It is described in *Chem. Week*,22 *Eur. Chem. News*,23,24 and *Oil & Gas*.25

Ammonia and some of its salts and derivatives are used as fertilizers. Ammonia also is a very important building block for many other chemical commodities as shown in Fig. 29.1.

Fig. 29.1 Synthetic nitrogen products.



Other Processes

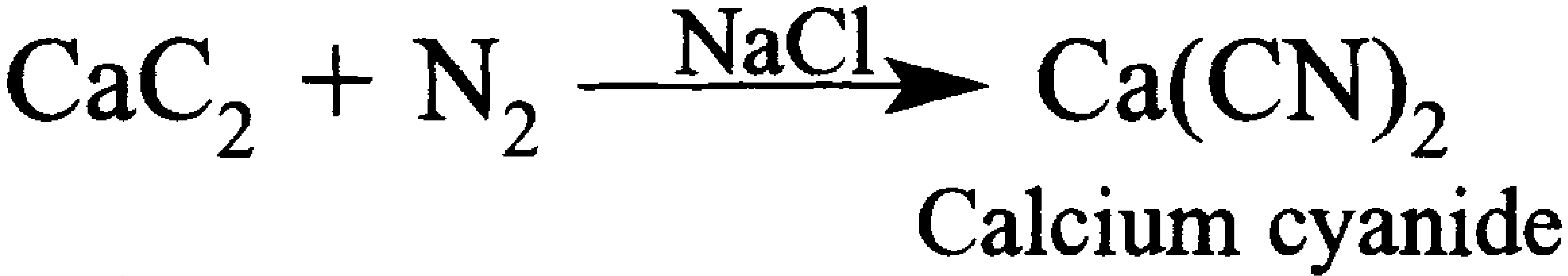
Atmospheric nitrogen can be fixed by the FrankCaro or Cyanamide process:

$$CaC_2 + N_2 \xrightarrow{1000°C} \underset{\text{Calcium cyanamide}}{CaNCN} + C$$

The calcium cyanamide can be hydrolyzed with steam to yield ammonia and calcium carbonate. A. Frank and N. Caro developed this process in Germany in 1895. It was used in Germany, Norway, and Italy until it was replaced by the HaberBosch process after World War I.1,12 In 2001, the FrankCaro process was still used in many countries including Germany, South Africa, Japan, and China.

The commercially formulated product contains approximately 65 percent calcium cyanamide. For most of the 20th century, it has been used as a fertilizer, and also as a cotton defoliant, herbicide, and soil insecticide. A 1999 study in France showed that it greatly improves the quality and yield of cauliflower crops by acting as both a herbicide and an insecticide.34

Calcium cyanide can be produced by the addition of sodium chloride as a "catalyst:"



$$CaC_2 + N_2 \xrightarrow{NaCl} Ca(CN)_2$$

The principal use for calcium cyanide is in gold mining.

29.3 Nitrogen Consumption

The demand for nitrogen in a chemically fixed form (as opposed to elemental nitrogen gas) drives a huge international industry that encompasses the production of seven key chemical nitrogen products. Nitrogen products had a total annual commercial value of the order of $50 billion in 1996. The cornerstone of this industry is ammonia. Virtually all ammonia is produced in anhydrous form via the Haber process. Anhydrous ammonia is the basic raw material in the manufacture of fertilizers, livestock feeds, commercial and military explosives, polymer intermediates, miscellaneous chemicals, and various process applications.35

On a worldwide basis, nitrogen fertilizer consumption accounts for more than 85 percent of the total nitrogen industry. In the United States, with its significant industrial markets, fertilizer use accounts for only approximately 80 percent of total nitrogen consumption.35 About 97 percent of nitrogen fertilizers are derived from synthetically produced ammonia. The remainder are produced as by-product ammonium sulphate from caprolactam processes, acrylonitrile processes, metal refining operations, and small quantities of natural nitrates, especially from Chile.35 Ammonia is the starting material for six primary nitrogen fertilizer products: urea, ammonium nitrate, ammonium phosphates, ammonium sulfate, nitric acid, and nitrogen solutions.36

Up through the 1960s, the development of the nitrogen industry took place in the developed countries of Western Europe, North America, and Japan. However, in the 1970s and early 1980s, the construction of new plants shifted to the gas-rich countries of the Caribbean and Middle East. Additional plants were also built in some large consuming countries such as China, India, Indonesia, and Pakistan. At the same time, many plant closures occurred in Western Europe and Japan. Western European share fell from 20 percent in 1980/81 to 11 percent in 1997/98. In 1980/81, the developing countries accounted for 31 percent of nitrogen fertilizer production. By 1997/98 their share had increased to 55 percent36 (see Fig. 29.2). The main regions that produced nitrogen fertilizers in 1997/98 are summarized in Table 29.2. Over one third of the world's production is in just 2 countries, China and India, but urea is produced in at least 60 countries.36

Fig. 29.2 Nitrogen fertilizer production1980/81 & 1999/2000.36 (*Reproduced by permission of International Fertilizer Industry Association*)

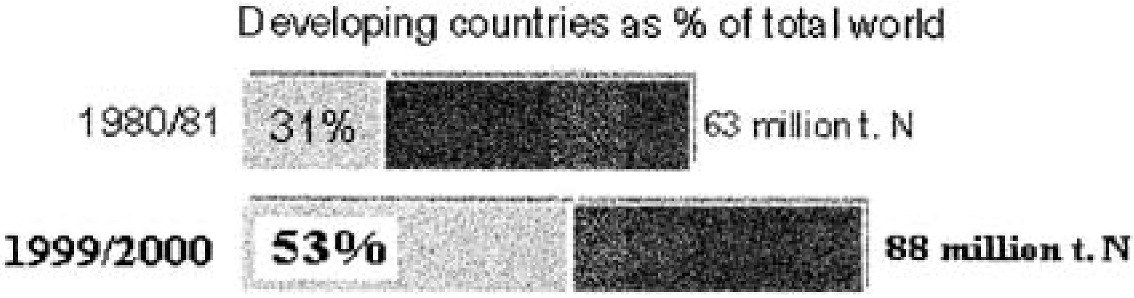


TABLE 29.2 Main Producing Regions for Nitrogen Fertilizers

Region	% of World Production
China	23
North America	18
South Asia	15
Western Europe	11
Former Soviet Union (FSU)	10
Other Countries	7
Central Europe	5
Middle East	5
Indonesia and Japan	4
Mexico and the Caribbean	2

World ammonia capacity increased by nearly 14 percent from 1984 to 1996 while capacity for urea, the primary downstream nitrogen product, increased by 45 percent. The increases were due primarily to: (1) a desire by some major importing countries to become more self-sufficient and (2) the construction of export-oriented capacity in the Middle East and in the Former Soviet Union (FSU). Ammonium nitrate capacity declined by 2 percent from 1984 to 1996 while ammonium sulfate capacity declined by 8 percent.

In the future developing nations are expected to continue to account for most of the increases in ammonia and urea capacity. Ammonia capacity is expected to increase by about 20 million tonnes and urea capacity by about 12 million tonnes of nitrogen between 1996 and 2002. The availability of relatively low-cost feedstock (usually natural gas) will be a major determinant as to where this new capacity is installed. According to SRI Chemical Economics Handbook, in 1996, ammonia accounted for almost 60 percent of the world production of nitrogen products.35

The apparent consumption of ammonia increased by about 15 percent between 1984 and 1996, while the apparent consumption of urea increased by about 54 percent. The developing countries are largely responsible for the increased consumption.

In 1996, urea accounted for almost 59 percent of apparent world consumption of the four major downstream nitrogen products. By 2002, the share is expected to increase marginally to about 61 percent. World urea consumption is forecast to increase at an average annual rate of 3.0 percent between 1996 and 2002. Only ammonium phosphate consumption, at 4.3 percent per year, is expected to grow more rapidly.35

Although world consumption grew substantially between 1984 and 1996, it actually peaked at 79.6 million tonnes of nitrogen in 1989 and declined through 1994 because of a sharp drop in fertilizer consumption in Eastern Europe, the FSU and Western Europe. Total world nitrogen fertilizer consumption resumed its long-term upward growth trend in 1995, and substantial growth is expected by 2002.35 It should be noted that Japan is one of the few countries where the use of slow-release fertilizers is developing on field crops, especially in crops grown in paddies.36

Estimates of world industrial nitrogen consumption by region show that the four largest markets in 1984, 1996, and 2002 have been and continue to be the United States, Socialist Asia, Western Europe and the FSU.35

Trade is an important component of the world nitrogen industry, and trade in each of the nitrogen products analyzed (ammonia, urea, ammonium nitrate, ammonium sulfate, and ammonium phosphates) increased substantially between 1984 and 1996. The percentage of production that is traded internationally varies from 10 percent for ammonia up to 40 percent for ammonium phosphates.35

29.4 Environmental Issues

Since 1980, the nitrogen industry has made major improvements in environmental controls to reduce the release of atmospheric emissions and liquid effluents.

The primary pollution problem in nitric acid manufacture is the abatement of nitrogen oxides (NO_x) in tail gases. In the United States, gaseous emissions from newly constructed nitric acid plants must be limited to 1.5 kg of NO_x per tonne of nitric acid (100% basis) produced, with a maximum stack opacity of 10 percent. Modern acid towers, with extended sections, can reduce NO_x emissions to less than 200 ppm.35

The most important environmental concern about the use of nitrogen fertilizers is the loss of nitrogen to groundwater. This is having a significant impact on the use of nitrogen fertilizers, particularly in Western Europe. There is a lesser but still significant concern about nitrate levels in groundwater in the United States. Other than this runoff aspect, which most seriously affects ammonium nitrate, no serious environmental concerns exist with the use of the major nitrogen products as fertilizer materials.35

Environmental issues do play an important role in the ammonium sulfate industry. This is because a significant portion of the world's ammonium sulfate production is the direct result of the necessity to remove sulfur dioxide from stack gases at various metal smelting and refining operations in order to conform to government regulations on sulfur dioxide emissions. A large potential source of additional by-product ammonium sulfate production is sulfur dioxide recovery from coal-fired electrical generating stations. It is estimated that a 500 MW plant that is burning 3.5 percent sulfur coal could generate 250 thousand tonnes of ammonium sulfate per year.35 This is equivalent to •7 percent of the total worldwide ammonium sulfate production in 2000.

29.5 Ammonia Production

Since no economical nitrogen fixation process that starts with nitrogen oxides has been discovered, ammonia has developed into the most important building block for synthetic nitrogen products. Anhydrous ammonia is produced in about 80 countries.36

In the United States only 80 percent of the ammonia is used to make fertilizers. Chemical intermediates (such as acrylonitrile and caprolactam) account for 19 percent of the ammonia use while the remaining 1 percent is used in pulp and paper, metals, and refrigeration applications.37

Worldwide, ammonia capacity grew from 119 million tonnes in 1980 to a peak of 141 million tonnes in 1989.36 Ammonia demand grew at a rate of 3.2 percent per year from 1993 to 1998, but the growth rate is expected to slow to 0.5 percent per year from 1998 through 2002.37

In 2001, several major North American ammonia producers unexpectedly curtailed some of their ammonia and urea production. Although periodic downtimes of plants are not uncommon, the combination of import competition that depressed prices and very high feedstock costs in many industrialized countries during 2000 has resulted in shutdowns of some high-cost plants which resulted in a decline in world capacity in 2000.38,57

In 1998, most of the ammonia production capacity was located in the regions shown in Table 29.3.36,57

TABLE 29.3 Ammonia Capacity by Region *(Thousands of Metric Tons N)*

	1987		1998	1999		2005	
Region	Quantity	Share (%)	Share	Quantity	Share (%)	Capacity	Share (%)
China	18,675	16.9	22	30,450	23.6	33,460	24.6
Former Soviet Union	21,725	19.7	17	19,340	15.0	18,455	13.6
North America	16,390	14.8	15	18,955	14.7	18,410	13.6
South Asia	8,935	8.1	12	15,750	12.2	16,705	12.3
Western Europe	15,635	14.1	9	11,870	9.2	11,255	8.3
Middle East	4,100	3.7	7	5,950	4.6	7,795	5.7
Central Europe	9,830	8.9	6	7,560	5.9	6,820	5.0
Indonesia and Japan	5,800	5.3	4	7,725	6.0	8,340	6.1
Mexico and the Caribbean (including Venezuela)	5,705	5.2	3	6,415	5.0	8,265	6.1
Other Countries	3,700	3.4	5	4,790	3.7	6,330	4.7
Total	110,495			128,805		135,835	

About 90 percent of world ammonia production is processed or used in the countries where it is produced. The remaining 10 percent of production enters international trade. The major net exporters in 1997 were gas-rich regions such as Russia and the Ukraine (43% of world exports), Trinidad (22%), and the Middle East (13%). The major net importers were the United States (43% of world imports) and Western Europe (26%).36

In the mid-1990s, the ammonia industry accounted for about 5 percent of the worldwide natural gas consumption. For economic and environmental reasons, natural gas is the feedstock of choice. However, processes for ammonia production can use a wide range of energy sources. For example, 60 percent of China's nitrogen fertilizer production is currently based on coal. At present natural gas is the most economic feedstock for the production of ammonia as shown in Table 29.4.53

TABLE 29.4 Feedstock Economics for Ammonia Production: Relative Consumptions and Economics for Different Feedstocks

	Natural Gas	Heavy Oil	Coal
Energy consumption	1.0	1.3	1.7
Investment cost	1.0	1.4	2.4
Production cost	1.0	1.2	1.7

Source: Reproduced by permission of European Fertilizer Manufacturers Association.

Prior to World War II, ammonia production capacity remained relatively stable. During the war, the need for explosives caused an increase in the production of ammonia for nitric acid manufacture. Then, after the war, available ammonia capacity was directed to the manufacture of fertilizers. Accordingly, there was a rapid increase in fertilizer consumption.

From 1940 to 1950, the number of ammonia plants doubled; then from 1950 to 1960, the number more than doubled again. Since 1963, the advent of large single-train plants has resulted in a large increase in production capacity, the shutdown of a number of smaller plants, and a reduction in manufacturing costs. Capacity tripled in the period from about 1958 to 1968. U.S. capacity declined in the 1970s, and it continued to decline during the 1990s by 510 percent to about 13 million tonnes per year. During the 1990s, U.S. ammonia capacity was between 13 million and 14 million tonnes per year with operating rates over 95 percent. The U.S. capacity is forecast to reach 15 million tonnes per year by 2002, and worldwide ammonia capacity is forecast to grow at over 2 percent per year and exceed 140 million tonnes per year by 2002.35

All commercial ammonia manufacturing processes are based on the synthesis of ammonia from nitrogen and hydrogen. Thus, the many processes are differentiated by the method of producing the hydrogen and nitrogen for the synthesis. Preparation of nitrogen can be by separation from air in a standard "air" plant. Or air can be fed to the ammonia unit where the oxygen is used to burn a portion of the hydrocarbon feed, leaving the nitrogen for ammonia synthesis. The latter is the dominant type. Most of the hydrogen for ammonia manufacture is prepared from hydrocarbon feeds.

Some physical properties of ammonia and densities of aqueous solutions of ammonia are given in Tables 29.5 and 29.6, respectively.

TABLE 29.5 Physical Properties of Ammonia

Property	Value
Molecular weight	17.03
Boiling point, °C	-33.35 at 760 mmHg
Freezing point, °C	-77.7
Color	Colorless
Critical temperature, °C	133.0
Critical pressure, kPa (psi)	1425 (1657)
Specific heat, J/kgK	
0°C	2097.2
100°C	2226.2
200°C	2105.6
Heat of Formation of gas, ΔH, kJ/kgmol (BTU/lb-mol)	
0°K	-39,222 (-16,862)
298°K	-46,222 (-19,872)
Solubility in water, wt. %	
0°C	42.8
20°C	33.1
25°C	31.8
40°C	23.4
60°C	14.1
Specific gravitya	
-40°C	0.690
0°C	0.639
40°C	0.580
Vapor pressure, mmHg	7500 at 25°C
Vapor density	0.6 (air = 1.0) at 0°C

Flammable limits in air, % by volume	
Lower explosion limit (LEL)	15
Upper explosion limit (UEL)	28
Autoignition temperature	651°C (1204°F)

[a]Anhydrous ammonia.

TABLE 29.6 Densities of Aqueous Ammonia at 15°C

Ammonia, wt. %	Density, g/L
8	0.970
16	0.947
32	0.889
50	0.832
75	0.733
100	0.618

29.6 Hydrogen Production

Ammonia production is the single largest consumer of hydrogen for use in chemical manufacture. Hydrogen production is typically integrated into the overall plant design, so most ammonia producers regard themselves as consumers of natural gas (or another feedstock) rather than hydrogen. Worldwide, a few ammonia plants have been located near by-product sources of hydrogen. In the United States, Coastal Refining & Marketing, Inc. brought a new ammonia plant on stream in late 1997 in Freeport, Texas using purchased pipeline hydrogen. Air Liquide America purifies 45 million standard cubic feet per day of by-product hydrogen from The Dow Chemical Company for pipeline delivery to Coastal's Freeport facility. This is the only U.S. ammonia plant using purchased hydrogen.40 Hydrogen requirements in the main world regions are shown in SRI International40 and Hydrogen41 and more details are given in Chapter 13 (Industrial Gases).

To help meet increased hydrogen requirements, a methodology has been developed for assessment of hydrogen sources that is based on an analogy with the problem of process heat recovery. A hydrogen surplus diagram allows the engineer to find the "hydrogen pinch" and to set targets for hydrogen recovery, hydrogen plant production, and import requirements. This method also gives insights into the effective use of hydrogen purification units. It has been shown that a purification unit should not be placed below the hydrogen pinch. Purifying gas above the pinch may have some benefits, but placing the purifier across the pinch is the best option.43

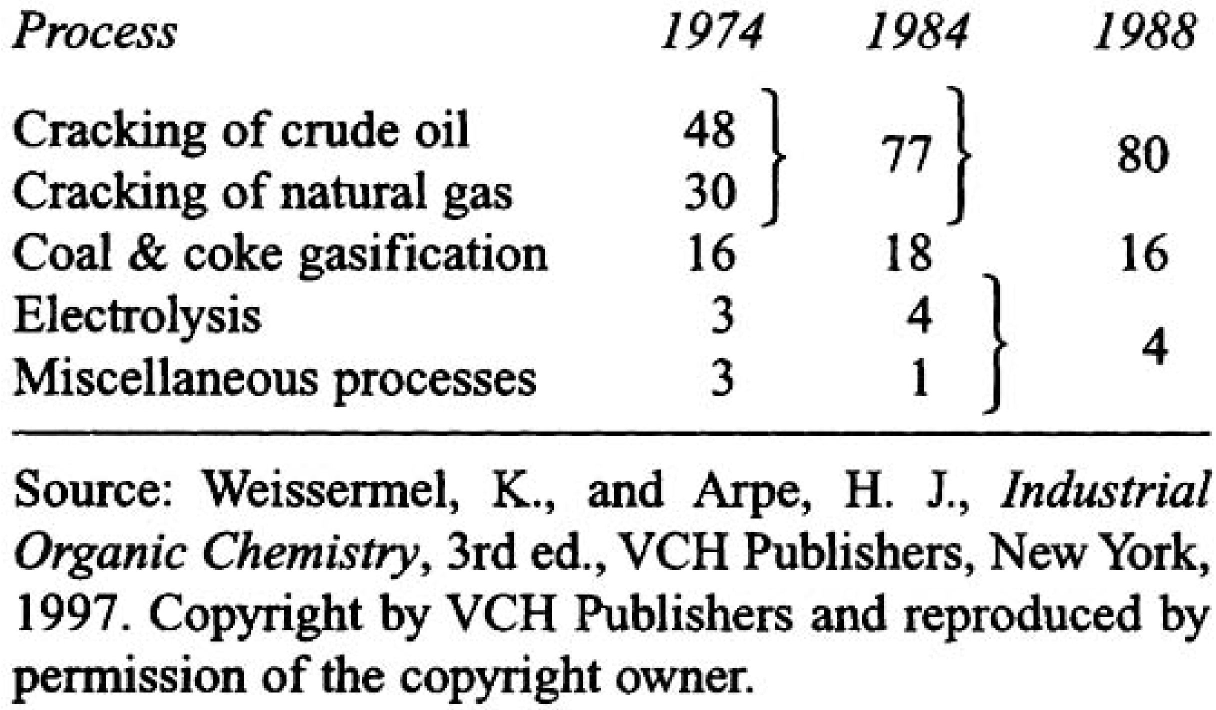
An improved "hydrogen pinch" approach has been developed that can account for pressure and makes best use of the existing compressors. This method is mathematically based and can account for costs and tradeoffsincluding hydrogen production, compression power, fuel value, and piping costs.44

Hydrogen Pinch Technology has been applied commercially by several companies including AspenTech, BP-Amoco, Engineers India, Exxon, ICI, Linnhoff-March, M.W. Kellogg, and UOP.45

Hydrogen is present in fossil fuels and water in sufficient amounts that it can be produced from these sources on a large scale by three different methods: (1) petrochemical processes, (2) coal-based chemical processes, and (3) electrochemical processes (electrolysis).

In Table 29.7, the percentage of hydrogen production is broken down by type of manufacturing process.46

TABLE 29.7 World Hydrogen Production and Type of Manufacturing Process (in %)



Process	1974	1984	1988
Cracking of crude oil	48	77	80
Cracking of natural gas	30		
Coal & coke gasification	16	18	16
Electrolysis	3	4	4
Miscellaneous processes	3	1	

Source: Weissermel, K., and Arpe, H. J., *Industrial Organic Chemistry*, 3rd ed., VCH Publishers, New York, 1997. Copyright by VCH Publishers and reproduced by permission of the copyright owner.

In the Petrochemical Processes, both natural gas and crude oil fractions can be converted into synthesis gas using two basically different methods:46

1. With the allothermal steam reforming method, catalytic cracking takes place in the presence of water vapor. The necessary heat is supplied from external sources.
2. With the autothermal cracking process, heat for the thermal cracking is supplied by partial combustion of the feed. Water is used and carbon dioxide (CO_2) may be recycled to attain a desired CO/H_2 ratio.

The processes for the manufacture of synthesis gas were originally based on the gasification of coke from hard coal and low temperature coke from brown coal by means of air and steam. After World War II, the easy-to-handle liquid and gaseous fossil fuelsoil and natural gaswere also employed as feedstocks. Their value lies in their high hydrogen content that produces higher molar ratios of hydrogen (H_2) to carbon monoxide (CO) in the synthesis gas (see Table 29.7). By using excess steam in the reforming of methane, the H_2-to-CO molar ratio can be as high as 5.0 (see Table 29.8).

TABLE 29.8 H_2/CO Molar Ratio in Synthesis Gas

Method of Manufacture	H_2/CO Molar Ratio
OxygenCokeSteam	0.6
AirCokeSteam	0.9
OxygenCoalSteam	1.0
OxygenFuel OilSteam	1.0
PropaneSteam	1.33
MethaneOxygen	1.7
OilSteam	2.1
Petroleum etherSteam	2.4
MethaneSteam	3.05.0

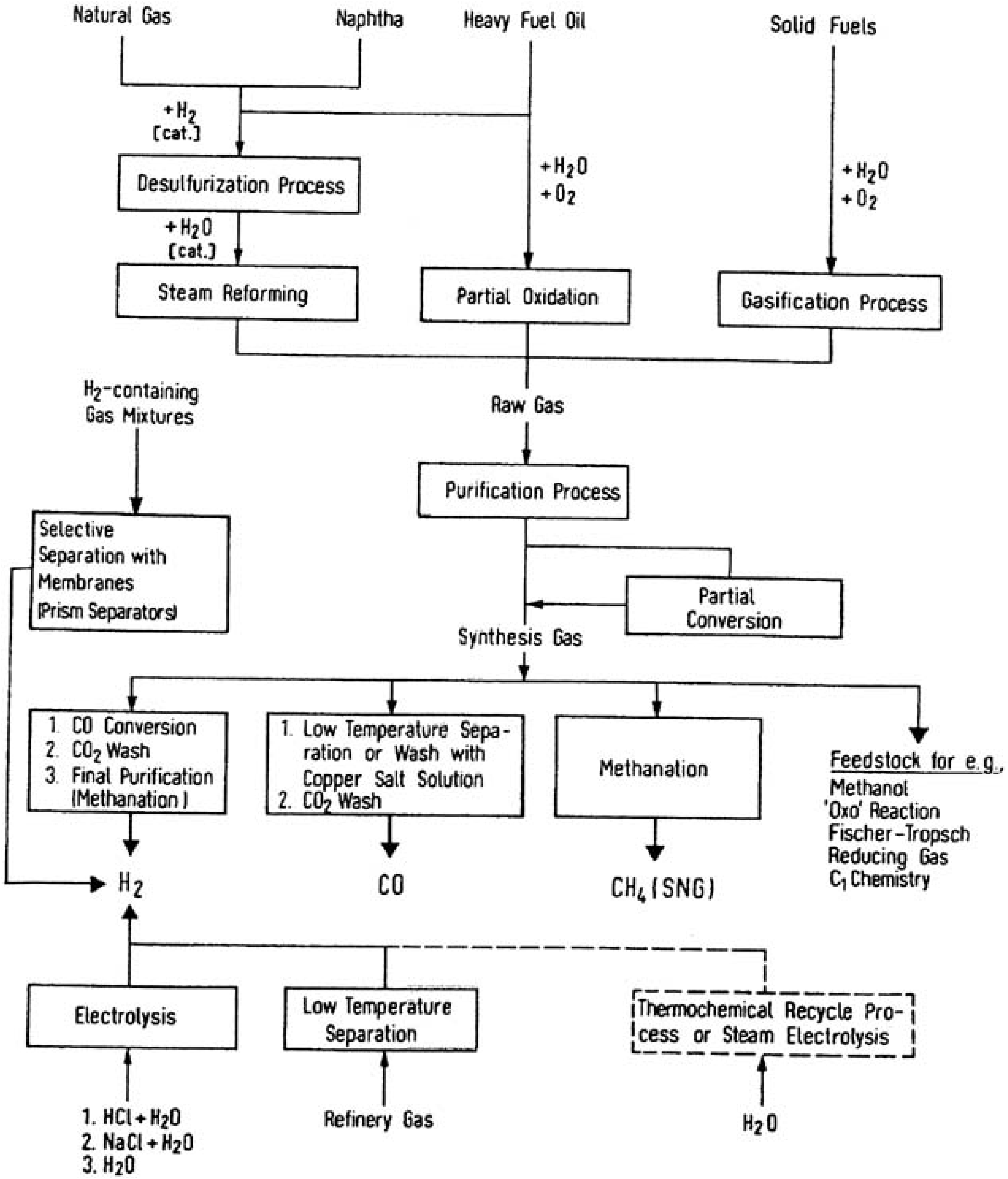
Selection of a process for hydrogen manufacture from hydrocarbons and coal therefore depends on the raw material and its cost, the scale of operation, the purity of the synthesis gas to be produced, the pressure level of the natural gas feed, and the number and type of downstream processes that will consume the carbon monoxide and hydrogen.

Hydrogen is also manufactured industrially by direct electrolysis of H_2O, HF, and 2225 percent hydrochloric acid (HCl). However, the hydrogen produced by electrolysis only accounts for a small percentage of the total hydrogen production (see Table 29.7). In contrast to the steam reforming of hydrocarbons, the hydrogen from electrolysis is very pure (>99 vol. %) which eliminates the costly purification steps.

Manufacturing Processes

Hydrogen is manufactured by four principal processes (see Table 29.7 and Fig. 29.3): (1) steam reforming of natural gas, (2) partial combustion of natural gas or oil with pure oxygen, (3) gasification of coal or coke with air (or oxygen) and steam, and (4) recovery of by-product hydrogen from petroleum refinery gases or other cracking operations. Small amounts of hydrogen also are manufactured by electrolysis. These processes are discussed in more detail in the following sections.

Fig. 29.3 Hydrogen manufacturing process steps.46 (Weissermel, K., and Arpe, H.-J., *Industrial Organic Chemistry*, 3rd ed., VCH Publishers, New York, 1997. Copyright by VCH Publishers and reproduced by permission of the copyright owner.)



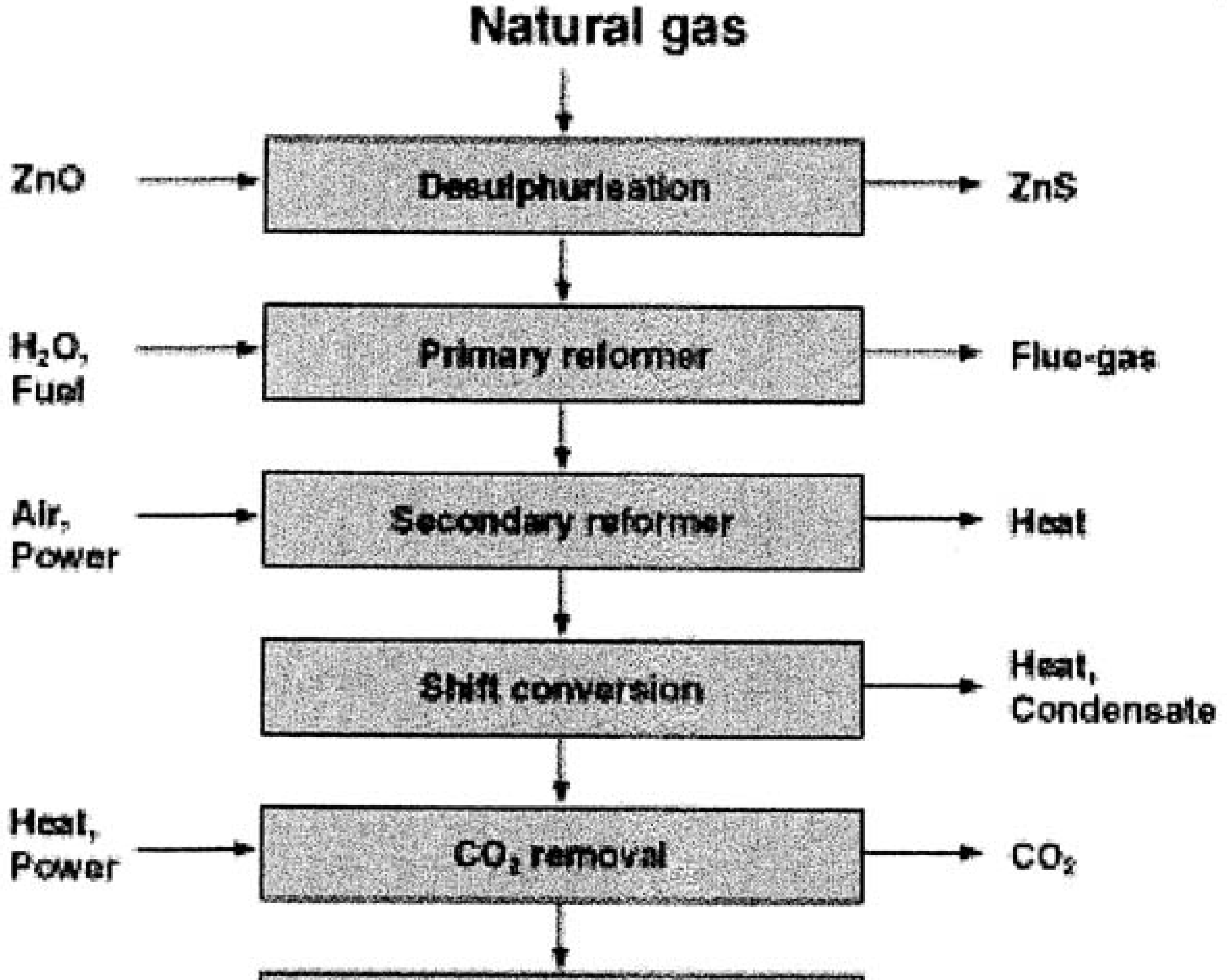
Reforming

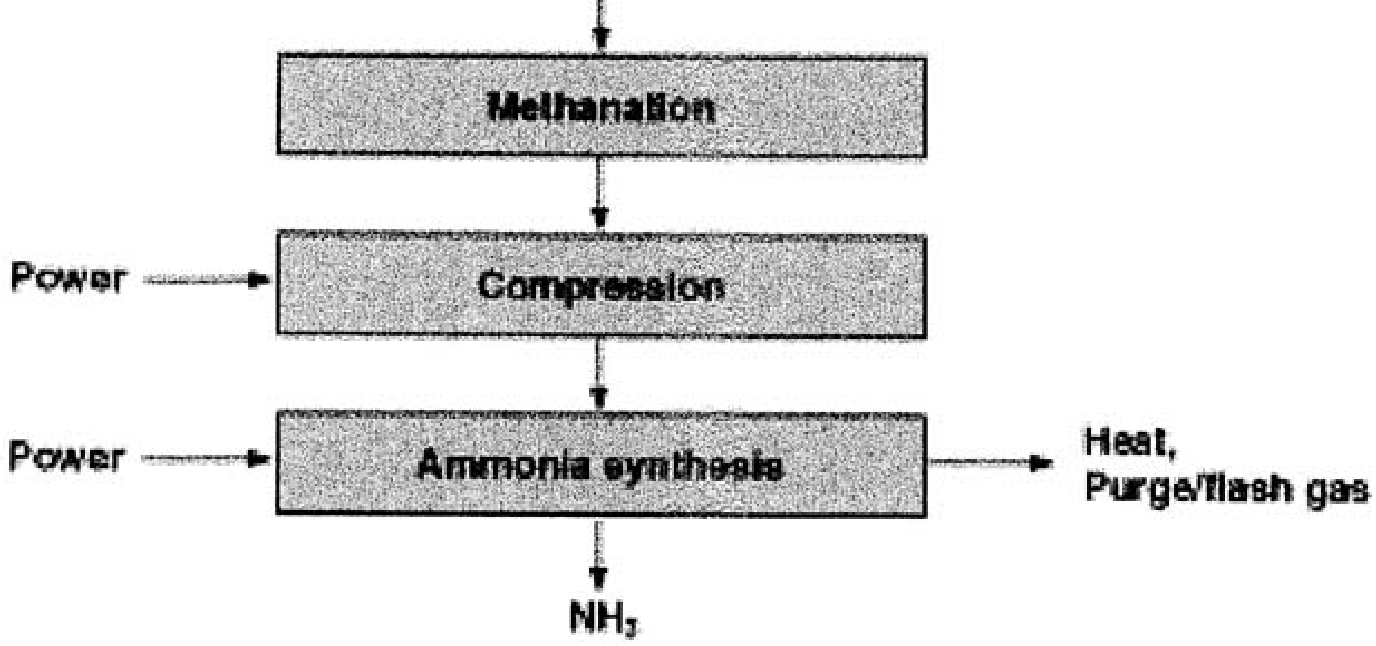
Reforming is a general name for the reaction of a hydrocarbon, such as methane, with water and/or carbon dioxide, to produce a mixture of carbon monoxide and hydrogen. The different reforming processes can be divided into the following types:53

1. conventional steam reforming with a fired primary reformer and stoichiometric air secondary reforming (stoichiometric H/N ratio).
2. steam reforming with mild conditions in a fired primary reformer and excess air secondary reforming (under-stoichiometric H/N ratio).
3. heat exchange autothermal reforming, with a process gas heated steam reformer (heat exchange reformer) and a separate secondary reformer, or in a combined autothermal reformer, using excess or enriched air (under-stoichiometric or stoichiometric H/N ratio).

The typical steps in the reforming process and how they tie into ammonia production are shown in Fig. 29.4.

Fig. 29.4 Block diagram of steam/air reforming process.53 (*Used by permission of European Fertilizer Manufactures Association.*)





If water is used, the process is called steam reforming or steam cracking. The reforming reaction (29-1) is endothermic and requires a catalyst.1

$$\mathrm{CH_4 + H_2O \rightleftarrows 3H_2 + CO}$$

$$\Delta H = +49\,\mathrm{kcal/mol\ or\ 205\,kJ/mol} \qquad (29\text{-}1)$$

Other reactions that proceed at the same time as the reforming reaction are:46

$$\mathrm{CO + H_2O \rightleftarrows H_2 + CO_2}$$

$$\Delta H = -10\ \mathrm{kcal/mol\ or\ 42\,kJ/mol} \qquad (29\text{-}2)$$

(The homogenous water gas reaction or water gas shift.)

$$\mathrm{CH_4 \rightleftarrows 2H_2 + C}$$

$$\Delta H = +21\,\mathrm{kcal/mol\ or\ 87\,kJ/mol} \qquad (29\text{-}3)$$

$$\mathrm{2CO \rightleftarrows CO_2 + C}$$

$$\text{Boudouard reaction}$$

$$\Delta H = -38\,\mathrm{kcal/mol\ or\ 162\,kJ/mol} \qquad (29\text{-}4)$$

The equilibrium composition of the synthesis gas depends on the steam-to-gas ratio entering the reactor, the reaction temperature, the reaction pressure, and the quantity of inerts in the reaction mixture. To avoid carbon formation as indicated by reactions (29-3) and (29-4), the steam-to-gas ratio must be kept high enough to favor the reforming reaction (29-1) and the water gas shift reaction (29-2) over the reactions that form carbon.

In 2001, the most well-known, large-scale steam reforming process is licensed by Synetix. Hydrocarbon feeds with boiling points up to 200°C can be treated, and the process consists of three steps: feed pretreatment, catalytic reforming, and reforming of residual methane. The advantage of the Synetix process is that no soot formseven with liquid crude oil fractions as feed. This makes catalyst regeneration unnecessary.46 Because of these advantages, the Synetix technology was use in over 400 reformers in over 30 countries. When North Sea gas and other natural gas reserves around the world were developed, the use of naphtha as a feed stock declined and the need for this technology was reduced.

Feed Pretreatment

In the Feed Pretreatment step, the naphtha feed must be purified because the reforming catalyst is very sensitive to impurities. The naphtha is treated with hydrogen at 350450°C until the sulfur content is less than 1 ppm and in some cases less than 0.1 ppm. Simultaneously, any olefins are hydrogenated.46 The same catalyst similarly converts any organo-chloride species to give HCl and also acts as an absorbent for most problematic metal species.70 Recommended materials of construction are discussed in Nair.88

Pre-Reformer

A pre-reformer partially completes the steam reforming reactions upstream of the main steam reformer at a much lower temperature by using a more active catalyst. One advantage in new plant designs is that the conventional steam reformer furnace can be smaller.70 All the ethane, the heavier hydrocarbons, and some of the methane are decomposed endothermically and at temperatures below those in the main reformer. The product gas from the pre-reformer is an equilibrium mixture of hydrogen, methane, steam, and carbon oxides. The pre-reformer can reduce fuel consumption by up to 10 percent and increase throughput by up to 15 percent.71

Another advantage is that a pre-reformer may be preferred in flowsheets for processing heavier feedstocks. This is especially true if: (1) the end product is CO or town gas, (2) a range of hydrocarbon feeds are to be processed, or (3) minimum steam export is a requirement. The pre-reformer also converts heavier hydrocarbons to methane before they are fed to the steam reformer. This allows the steam reformer to be designed for methane service.70

Reformers

Catalytic reforming takes place across a nickel catalyst that is packed in tubes in a fired furnace. The tube furnace is called the "primary reformer," and the lined chamber reactor is called the "secondary reformer." Typical reaction conditions in the primary reformer are 700830°C and 1540 bar. Additional details about the primary reformer can be found in Weissermel and Arpe.46 A portion of the gas is combusted in the secondary reformer with added air or oxygen whereby the gas mixture reaches a temperature of over 1,200°C. Methane reacts with steam at this temperature until only an insignificant amount remains. The methane concentration depends upon the pressure, the temperature, and the quantity of nitrogen and steam present.46

Catalysts

The catalyst that is used in the secondary reformer does not need to be as active as that in the primary reformer. Hence, the usual nickel concentration is about 15 percent, compared with 25 percent in the primary reformer catalyst.

The steam reforming catalyst selection depends on the feedstock types and the operating conditions. The main categories of catalyst are for (1) light duty reforming, (2) intermediate duty reforming, and (3) heavy duty reforming. The light duty catalysts are for natural gas, refinery off-gas and pre-formed feeds.70

Intermediate duty catalysts are for feeds with a significant content of components from ethanes up to liquid petroleum gas (LPG). Heavy duty catalysts are primarily for naphtha feeds that have even more tendency for carbon deposition.70

Silica and silica-bearing materials cannot be used with the shift catalyst. This is because the silica volatilizes and migrates from the hotter zone to lower temperature zones downstream. Usually it deposits on the waste heat boiler tubes after the secondary reformer.

Normally, the nickel oxide is reduced to nickel and water by the hydrogen that is produced in the operation. In some cases the reduced nickel can be reoxidized to nickel oxide when large amounts of steam and small amounts of hydrogen are present:47

$$\mathrm{Ni + H_2O \rightleftarrows NiO + H_2} \qquad (29\text{-}5)$$

The addition of hydrogen to the natural gas feed keeps the nickel in the reduced state, thus making it more active. Also, the hydrogen will retard the formation of nickel sulfide, thus avoiding poisoning of the catalyst:

$$\mathrm{NiS + H_2 \rightleftarrows Ni + H_2S} \qquad (29\text{-}6)$$

Carbon can also reduce the effectiveness of the catalyst. When conditions are favorable for the following reactions, carbon will deposit on the catalyst:

$$\mathrm{CH_4 \rightleftarrows 2H_2 + C}$$

$$\Delta H = +21\,\mathrm{kcal/mol\ or\ 87\,kJ/mol} \qquad (29\text{-}3)$$

$$\mathrm{2CO \rightleftarrows CO_2 + C}$$

$$\Delta H = -38\,\mathrm{kcal/mol\ or\ 162\,kJ/mol} \qquad (29\text{-}4)$$

Thus, if insufficient steam is present, carbon will be deposited on the catalyst and reduce its effectiveness. Carbon will gasify with steam at 800°C or higher in the reformer:

$$\mathrm{C + H_2O \rightleftarrows H_2 + CO}$$

$$\Delta H = +28\,\mathrm{kcal/mol\ or\ 119\,kJ/mol} \qquad (29\text{-}1)$$

Reformer operation and changes in operating rate should always be carried out in a way that ensures sufficient steam is over the catalyst. Thus, if the feed rate is to be increased, the steam rate should be increased first. Conversely, if the feed rate is to be decreased, the natural gas rate should be decreased first.

Reformer Materials of Construction

The primary reformer is a steam-hydrocarbon reforming tubular furnace that is typically externally fired at 2535 bar and 780820°C on the process side. Previously SS 304, SS 310, SS 347, HK 40 alloy, and HP 25/35 modified alloys were used as tube materials. However, these materials developed various operating problems as rates increased and longer service lives were needed for economical operation.88

HP micro alloys were developed during the 1990s. The micro alloys enhanced carburization resistance and improved high-temperature creep-rupture resistance.88

For reformer outlet manifolds, the normal metallurgy choice is a wrought type of Alloy 800 H. Hot reformed-gas transfer lines are usually refractory-lined with an interior of Alloy 800 sheathing.88

In the secondary reformer, air is added to the process stream at operating conditions of 2830 bar and 9551025°C. The refractory-lined vessel has an outer shell of a low-alloy steel containing 0.5 Mo. Metal dusting occurs in the secondary reformer outlet sections. With hot gases containing a high CO content, carbon will diffuse into the FeCrNi alloy. This phenomenon can lead to local mechanical fracturing of surface layers and failures by pitting.88

Materials such as SS 304 and Alloy 800 are very susceptible to metal dusting in the range of 500800°C. Besides temperature, carbon activity (the CO/CO2 ratio in the gas) and gas partial pressure also affect metal dusting. Severe attacks occur when the carbon activity is in the range of 310. Recirculating CO2 into the primary reformer along with feedstock can maintain a low CO/CO2 ratio and avoid the severity of this attack. By maintaining a high steam-to-hydrogen ratio in the gas, the metal dusting can also be minimized.88

Hydrogen embrittlement is another important corrosion problem that is encountered in reformed-gas pipelines. The Nelson curves list the operating limits that should be followed to avoid decarburization and fissuring of steel in hydrogen service.88

Waste Heat Recovery

The waste heat recovery system is associated with flue gas from the reformer furnace and process gas from the secondary reformer. It generates high-pressure steam in specially designed boilers. Proper material selections and stringent water quality are two proactive loss-prevention methods.88

Hydrogen Production Costs

The theoretical energy requirement per mole of hydrogen produced for the overall steam reforming process is 40.75 kJ/mole of hydrogen. The process does hot have any by-product credits, and it does not look environmentally benign due to large CO2 emissions. The capture and disposal of CO2 (8085% of CO2 is captured from the concentrated streams) adds about 2530 percent to the cost of hydrogen produced by steam reforming.60

The cost for making hydrogen by steam reforming of natural gas depends primarily on the cost of natural gas and can be found in the 1998 Hydrogen Product Review that is published by SRI International.40,52

Other Reforming Processes

The RKN process uses steam reforming to make hydrogen from hydrocarbon gases. This process was developed by Haldor Tops⊘e in the 1960s. By 1974, 24 plants based on this technology were operating.2

The Steam Methane Advanced Reformer Technology (SMART) makes hydrogen by the steam reforming of methane, and it houses the catalyst in a proprietary heat exchanger. The process was developed by Mannesmann KTI in 1996, and the first installation was planned in Maryland in 1997.2

The Autothermal Reforming (ATR) process makes CO-enriched syngas. It combines partial oxidation with adiabatic steam-reforming. It was developed in the late 1950s for ammonia and methanol synthesis, and then further developed in the 1990s by Haldor Tops⊘e.2

The Combined Autothermal Reforming (CAR) process is used to make syngas from light hydrocarbons, and the heat is provided by partial oxidation in a section of the reactor. It was developed by Uhde and commercialized in Slovakia in 1991.2

The Kellogg Reforming Exchanger System (KRES) is a reforming process that provides syngas to the KAAP process. The Kellogg Advanced Ammonia Process (KAAP) process is a high-pressure process that makes ammonia from its elements and does not use an iron-containing catalyst. MW Kellogg developed the process in 1990, and Ocelot Ammonia installed the first plant in British Columbia, in 1991. Another plant was installed at Ampro Fertilizers in Louisiana in 1996.2

The Haldor Tops⊘e Convection Reformer (HTCR) is a relatively small piece of equipment that combines the radiant and waste heat sections of the conventional reformer. It is best for small- and medium-sized hydrogen plants (50010,000 Nm3/hr).

Some reforming processes use less primary reforming by moving some of the heat duty to the secondary reformer. Some of the revisions to conventional reforming are: decreased firing in the primary reformer, increased process air flow to the secondary reformer, and lower inert levels in the make-up syngas.53

Decreased Firing in the Primary Reformer

Decreased heat supply in the primary reformer means that the process outlet temperature is lowered to about 700°C; the firing efficiency increases; and the size and cost of the primary reformer are reduced. The milder operating conditions prolong catalyst life, catalyst tube life, and outlet header service life.53

Increased Process Air Supply to the Secondary Reformer

In conjunction with "Decreased Firing in the Primary Reformer," the process air requirement is about 50 percent higher than in the conventional process. This means increased compression capacity and energy use. The process air compressor is usually driven by a gas turbine with the exhaust gas from the turbine being used as combustion air in the primary reformer. Some excess steam is available for export when using a gas turbine.53

Decreased heat supply in the primary reformer means that increased internal firing is necessary to achieve approximately the same degree of total reforming. A somewhat higher methane slip, and thus a lower secondary reformer outlet temperature, is acceptable and preferable in this type of process.53

Heat Exchange Autothermal Reforming

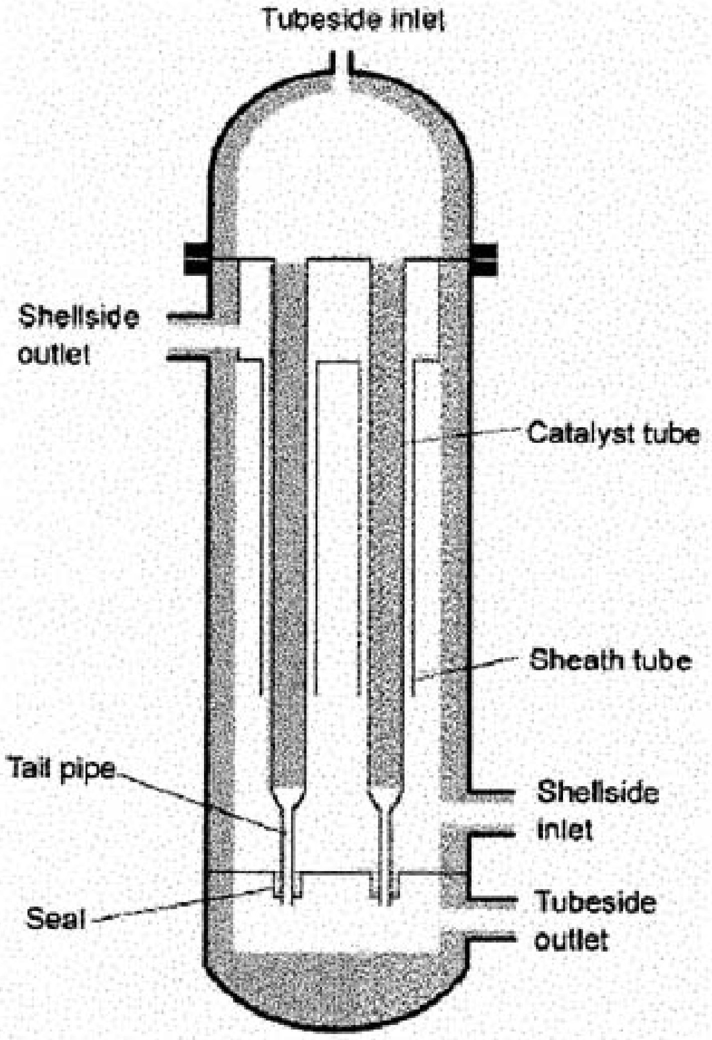
Heat exchange autothermal reforming was developed in 1997. In this process, the heat content of the secondary reformer gas is used in a newly developed primary reformer. This newly developed reformer is a gas-heated, heat exchange reformer rather than the conventional fired furnace design. Surplus air or oxygen-enriched air is required in the secondary reformer to meet the heat balance in this autothermal concept.53

NO*x* emissions may be reduced by 50 percent or more compared to conventional steam reforming by eliminating the flue-gas from the primary reformer. The level of NO*x* reductions depends on the extent of auxiliary combustion in the plant.53

Two processes of this kind are in operation, and some others are at the pilot stage. So far, single train capacities have been limited to 1000 tonnes per day or less.53 As of 1999, four plants had been installed that use Synetix gas heated reformer (GHR) technology.72

Synetix has developed a modified design of the GHR known as the Advanced GHR or AGHR. BHP Petroleum of Australia became the first operator of an AGHR in 1998. The key difference in the AGHR is that it uses a single-pass tube arrangement while the GHR uses a bayonet tube reformer. Based on the experience at BHP Petroleum, the AGHR design results in a reformer that is lower in cost, easier to operate, easier to fabricate, and allows scale-up to capacities in excess of current world-scale throughputs (see Fig. 29.5).72

Fig. 29.5 Synetix's advanced gas heated reformer (AGHR). (*Reproduced by permission of Imperial Chemical Industries PLC.*)



A hydrogen plant based on a new design was installed in Venezuela at the Compagnia Hidrogeno de Paraguana (CHP) plant to supply hydrogen to a refinery. The plant is a joint venture between BOC and Foster Wheeler and was started up in September 1997. This plant has a reduction in total fuel firing that is estimated to be 4 percent, and an installed cost that is estimated to be 23 percent below that of previous plants of the same size. The plant uses steam reforming with purification by PSA to produce 50 million SCFD of hydrogen at 99.5 percent purity and 400 psig.87

Small-Scale Reforming

Small-scale reforming systems are being pursued by a number of companies developing fuel cell electrical generation systems for home and automotive fuel cell applications. Long before these systems become affordable and simple enough to be used as home electrical system components, they may be cost-effective for industrial hydrogen supply systems.48

Small-scale reforming systems are relatively complex because they need fuel and air-feed systems, the reformer, a hydrogen purification system, and various cooling and water processing ancillary systems to make it all work. Additionally, the systems have to employ a specific hydrocarbon that is available at a reasonable cost at a customer's location. These systems probably work best for customers with hydrogen consumption rates in the 150010,000 SCF/hr range. They are most cost-effective when employed to produce hydrogen of a 99.9 percent or lower assay. As of 2001, small-scale reforming systems are not commercial. But they are in the advanced development stage.48

Technology Suppliers

Companies that provide technology to produce hydrogen via steam reforming are in the United States, Europe, and Japan. Many of these companies are listed in SRI International.40

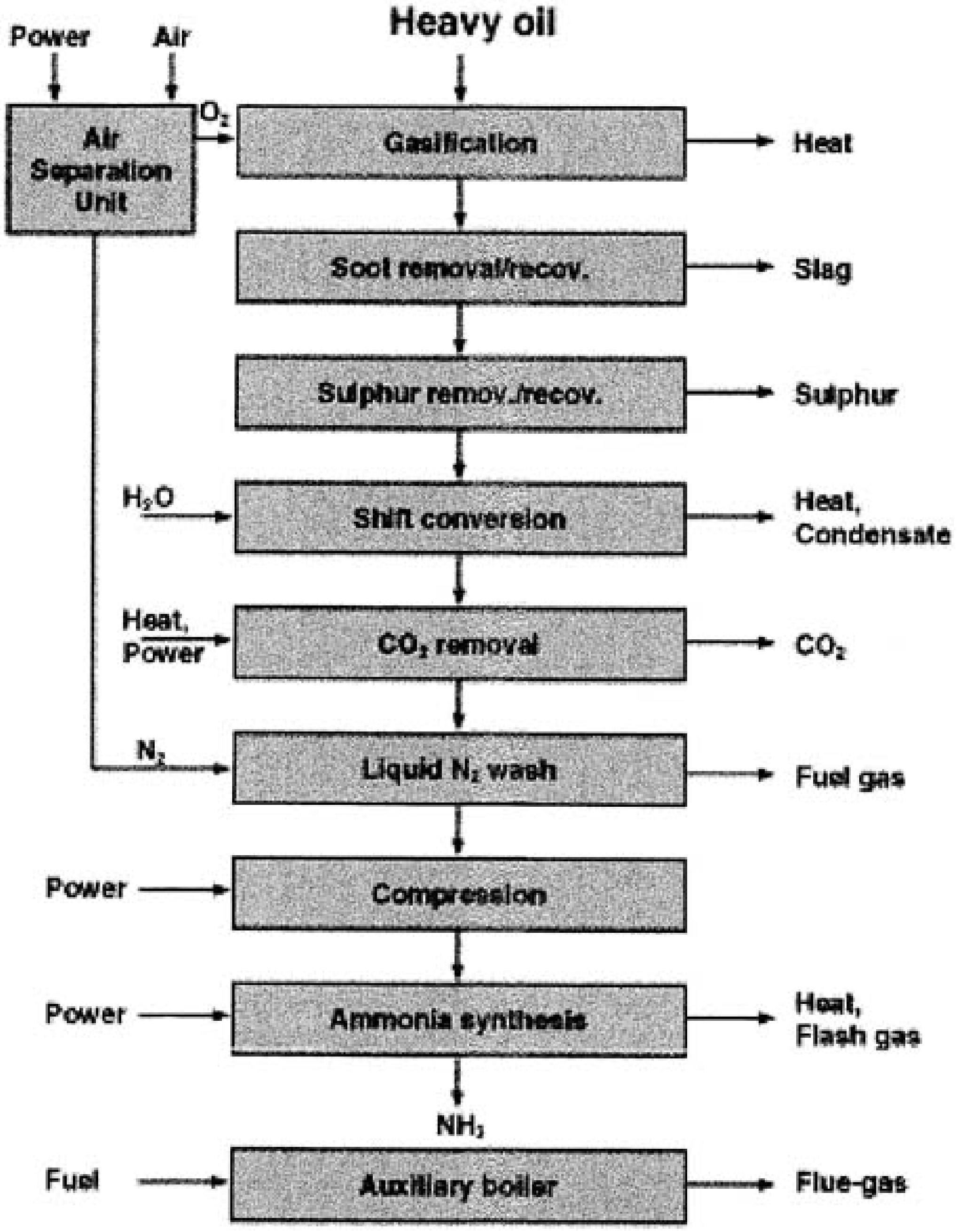
Partial Oxidation

Noncatalytic partial oxidation processes react hydrocarbons with gaseous oxygen at very high temperatures in a refractory-lined, pressurized reactor to produce synthesis gas. While the process can handle a wide range of feedstocks, it was developed mainly to utilize the low-value "bottom of the barrel" and waste feedstocks that typically have high sulfur and high metal contents.40

Synthesis gas manufacture by partial oxidation or autothermal cracking of crude oil fractions was developed by BASF/Lurgi, Texaco and Hydrocarbon Research. Heat for the thermal cracking is supplied by partial combustion of the feed in the presence of water. Recycled CO2 may also be added to the combustion to attain a desired CO/H2 ratio. Shell developed a modified version (which does not use a catalyst) that is referred to as the gasification process.46

A typical partial oxidation process is shown in Fig. 29.6.

Fig. 29.6 Block diagram of partial oxidation process.53 (*Used by permission of European Fertilizer Manufacturers Association.*)



As illustrated in Table 29.8, the H2 to CO molar ratio in the synthesis gas product stream depends on the raw materials and the operating conditions. By burning natural gas with a limited quantity of oxygen, a synthesis gas that has a H2 to CO molar ratio of approximately 2 to 1 can be produced:

$$CH_4 + \tfrac{1}{2}O_2 \rightarrow 2H_2 + CO \qquad (29\text{-}7)$$

If steam is introduced into the reaction, the H2 to CO molar ratio can be increased to over 2.0. This reaction is exothermic and can reach temperatures of 12001500°C.49 Typical operating costs are shown in SRI International.40

If fuel oil or crude oil is used as the raw material in partial combustion, care must be taken to provide for the removal of sulfur compounds and ash-containing materials in the crude oil (see Fig. 29.6). With crude-oil or fuel-oil partial combustion, the carbon content is much greater than in the case of natural gas, and special design considerations are necessary to produce a satisfactory gas.

Typical gas compositions from the partial combustion of various fuels are given in Table 29.9. The nitrogen that is needed to produce the desired H2/N2 ratio for ammonia production is usually introduced later in the processing sequence.

TABLE 29.9 Product Gas Compositions from Various Feeds by Partial Combustion (Percent by VolumeDry)

Feedstocks & Gas Components	Naural Gas	Light Naphtha	Heavy Fuel Oil	Vacuum Resid.	Propane Asphalt	Coal
Hydrogen	61.80	52.09	46.74	44.80	43.69	34.39
Carbon monoxide	33.75	42.59	48.14	49.52	50.09	44.22
Carbon dioxide	3.10	4.88	3.80	4.16	4.45	18.64
Methane	1.00	0.30	0.30	0.30	0.30	0.38
N2 + A	0.27	0.13	0.23	0.18	0.30	0.68
H2S		0.01	0.76	1.00	1.12	1.46
COS			0.03	0.04	0.05	0.10
NH3					0.13	
H2/CO, mol/mol	1.83	1.22	0.97	0.90	0.87	0.78

Partial Oxidation Process

The Electrochemical Partial Oxidation (Electropox) process is an electrochemical process for oxidizing methane to syngas. It combines partial oxidation and steam reforming of methane with oxygen separation in a single stage. It was invented in 1988 at BP Chemicals. An industrial-academic consortium to develop the process was formed in 1997.1

The Koppers Hasche process is a cyclic process for converting methane to syngas by partial oxidation over an alumina catalyst. The process operates in a pair of horizontal catalyst chambers that alternate their functions at one-minute intervals.1

In the 1990s, the University of Orleans (France) developed a non-catalyzed, plasma-assisted partial oxidation (or POX) process for making syngas.1,50

The Texaco Hydrogen Generation Process (THGP) makes pure, high-pressure hydrogen from various gaseous and light hydrocarbons. This process can be licensed from the Texaco Development Corporation.1,51

Technology Suppliers

The companies that license partial oxidation technology can be found in the United States and Europe. Some of these companies are listed in SRI International.40

Coal and Coke Gasification

In coal gasification, the exothermic partial combustion of carbon and the endothermic water gas formation represent the actual gasification reactions:46

$$C + O_2 \rightleftarrows 2CO \quad \Delta H = -60\,\text{kcal/mol or } 246\,\text{kJ/mol} \qquad (29\text{-}8)$$

Partial combustion

$$C + H_2O \rightleftarrows H_2 + CO \quad \Delta H = +28\,\text{kcal/mol or } 119\,\text{kJ/mol} \qquad (29\text{-}1)$$

Heterogeneous water gas reaction

Some other important reactions are:

$$C + CO_2 \rightleftarrows 2CO \quad \Delta H = +38\,\text{kcal/mol or } 162\,\text{kJ/mol} \qquad (29\text{-}9)$$

Boudouard reaction

$$CO + H_2O \rightleftarrows H_2 + CO_2 \quad \Delta H = -10\,\text{kcal/mol or } 42\,\text{kJ/mol}$$

Homogenous water gas reaction (water gas shift)

$$(29\text{-}2)$$

$$C + 2H_2 \rightleftarrows CH_4 \quad \Delta H = -21\,\text{kcal/mol or } 87\,\text{kJ/mol} \qquad (29\text{-}10)$$

Hydrogenative gasification

$$CO + 3H_2 \rightleftarrows H_2O + CH_4 \quad \Delta H = -49\,\text{kcal/mol or } 206\,\text{kJ/mol} \qquad (29\text{-}11)$$

Methanation

Gasification processes can be characterized on the one hand by the type of coal used and by the coal's physical and chemical properties. On the other hand, the processes differ in the technology involved.46 The Shell gasification process uses a dry coal feed while other processes feed the coal as a wet slurry.76 The coal gasification processes that have been developed are described in more detail under Coal Technology (Chapter 17).

Suid-Afrikaans Sintetiese Olie (Sasol) in South Africa has one of the largest coal gasification production operations in the world with the three plants (Sasol One, Sasol Two, and Sasol Three).1 The Sasol One plant produces only chemicals while the other plants produce both liquid fuels and chemical feedstocks from coal. Sasol One recovers ammonia from the raw gas that leaves the gasifiers, and ammonia is also produced from the gas that leaves the low-temperature slurry phase distillate reactor. In 1994, a new 240,000 ton per year ammonia plant was commissioned to bring Sasol's total capacity to over 500,000 tons per year.54

The only commercial size coal gasification plant in the United States is the Great Plains Synfuels plant that is operated by Dakota Gasification in Beulah, ND. This plant consumes over 5.5 million tonnes of coal per year and produces over 54 billion standard cubic feet (1.4 billion Nm3) of natural gas, 365,000 tonnes per year of ammonia, 24 million gallons of liquid nitrogen per year (68 million kg or 150 million lb) plus several other chemicals.89 (See www.dakota.com for more details.)

By-Product Hydrogen

Large volumes of by-product hydrogen are generated from a variety of production processes. Some of these processes are listed in SRI International.40 Some of the processes that can be used to obtain purified hydrogen are summarized below.

The Hydrogen Texaco (HyTex) process makes pure hydrogen from waste gases in oil refineries in a three-step process.1 The process was developed by Texaco and announced in 1991. The first commercial unit started up in Anacortes, WA in 1993.

Platforming operations make a gas that contains as much as 9095 percent hydrogen. This gas is usually purified by low temperature fractionation or washing with liquid nitrogen.

Kvaerner Oil & GasNorway has introduced a new process for manufacturing carbon black that generates substantial quantities of by-product hydrogen. In 1998, the process was installed in Montreal, Québec (Canada) to produce 20,000 tonnes of carbon black and 50 million Nm3 per year (5.2 million SCF per day) of hydrogen. The process appears to be economic only when the carbon black can also be sold as an end product.40

Air Products and Chemicals, Inc. has been selected to supply a hydrocarbon and nitrogen recovery system for a new polyethylene manufacturing plant in Baytown, TX. The plant will be owned by Chevron Phillips Chemical Company and Solvay Polymers, Inc. The recovery system uses partial condensation in conjunction with Air Products' pressure swing adsorption technology to recover hydrocarbons in the polyolefin plants, and recycle nitrogen with a purity of greater than 99 percent.55

A few ammonia plants have been located where a hydrogen off-gas stream is available from a nearby methanol or ethylene operation (e.g., Canadian plants at Kitimat, BC and Joffre, Alberta). The capital cost of such a plant is about 50 percent of the cost of a conventional plant of similar capacity because only the synthesis portion of the plant is required. However, by-product carbon dioxide is not produced and downstream urea production is therefore not possible.56

Electrolysis

Electrolytic hydrogen production yields the highest purity hydrogen (up to 99.999%), benefits from widespread raw materials availability (electricity and water), boasts simple system architecture, and can be scaled economically to serve applications ranging from the smallest hydrogen uses to the larger volume uses. The main disadvantage of water electrolysis is that electricity is an expensive "fuel." As a result, the technology generally is practical only for systems with hydrogen requirements of approximately 2000 SCF per hour and less.48

By 2001, several manufacturers had introduced advanced water electrolysis systems that are standardized, compact in size, need minimal operator intervention, and require little maintenance. New cell designs, materials of construction, standardized designs, and manufacturing techniques have enabled manufacturers to decrease the fixed costs of electrolysis technology. In addition, these new systems operate automatically and require very little maintenance.48,58

Developing regions are the largest market for electrolysis systems. Electrolysis currently accounts for a very small portion of the hydrogen generated in developed countries that have a commercial hydrogen infrastructure. However, electrolysis can be economic for small-scale generation in areas with inexpensive electricity, and a few plants still produce hydrogen for small-volume ammonia production.57 In 1997, Messer-MG Industries announced two hydrogen gas plants based on water electrolysis in the United States. In Canada, a portion of the capacity installed at HydrogenAl's merchant hydrogen plant in Québec is based on electrolytic cells.40 Companies with technology to build electrolytic plants can be found in North America, Europe, and Japan. Some of these companies are listed in SRI International.40

Other Processes.1

The Hypro process makes hydrogen by catalytically decomposing hydrocarbons to carbon and hydrogen. The carbon is burned to provide the heat for the reaction. This process was developed by UOP.

The Lane process makes hydrogen by passing steam over sponge iron at approximately 650°C. The iron is converted to magnetite.

The Methanol Reformer Hydrogen (MRH) process for generating hydrogen from methanol and separating it by PSA. This process was developed by the Marutani CPE Company.

The Proximol process makes hydrogen by reforming methanol. This technology is offered by Lurgi.

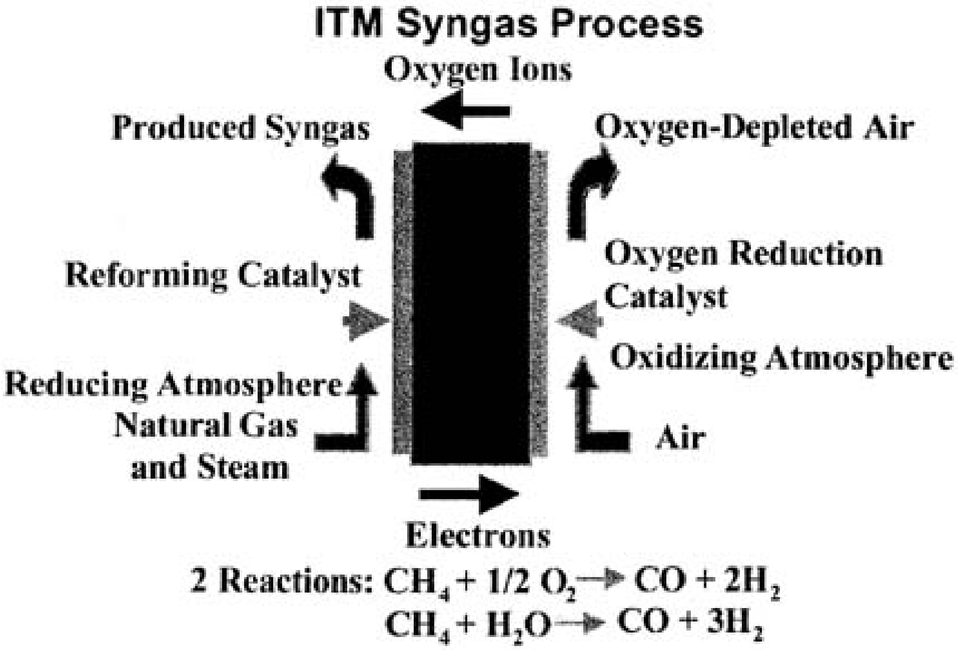
The Société Belge de l'Azote-Haldor Tops∅e (SBA-HT) process is a combination of both steam reforming and partial oxidation. The process converts LPG to syngas that is rich in hydrogen. This process was operated in France and Belgium in the 1960s.

New Developments

Several research organizations are developing low-cost methods of hydrogen production. Much of this work is sponsored by the U.S. Department of Energy. Some of the research programs can be found in Gronich.59

The ITM Syngas Process is the subject of one of these research programs. It involves the direct conversion of methane to synthesis gas (see Fig. 29.7). The process utilizes a mixed, conducting ceramic membrane and partial oxidation to produce the synthesis gas.40 The goal of this advanced reformer technology is to reduce the cost of hydrogen production by over 25 percent.59,79

Fig. 29.7 ITM syngas process.



Two projects were announced in the late 1990s to develop the ITM Process and another related technology. Air Products will lead an eight-year, $US 85 million research project on the ITM Process. Members of the team include Babcock & Wilcox, Ceramatec, Eltron Research, ARCO, Argonne National Lab, Pacific Northwest National Lab, Penn State University, and the University of Pennsylvania. A separate alliance includes Amoco, British Petroleum, Praxair, Statoil, and Sasol. This alliance is working to develop ceramic membrane technology for conversion of natural gas to synthesis gas.40

Another process is the CO2-free production of hydrogen via thermocatalytic decomposition of hydrocarbon fuels. The process involves a single-step decomposition of hydrocarbons over carbon catalysts in an air- and water-free environment. Preliminary assessments of the process indicated that hydrogen could be produced at a cost of $US5.00 per million BTU (if carbon sold at $US100/ton). This production cost is less than that for a steam reforming process coupled with CO2 sequestration.60

Some additional new developments in hydrogen production are summarized below.

NGK Insulators of Japan were issued U.S. Patent 5,741,474 in 1998 for a "Process for Production of High Purity Hydrogen." This process is a combination of reforming, partial oxidation, and membrane separation.

Northwest Power Systems obtained U.S. Patent 5,997,594 in 1999 for a "Steam Reformer with Internal Hydrogen Purification."

BP and Kvaerner Process are finalizing the demonstration of their "Compact Reformer Technology" in 2001. This technology involves the integration of combustion, heat transfer, and catalytic reaction within a simple tubular module.90

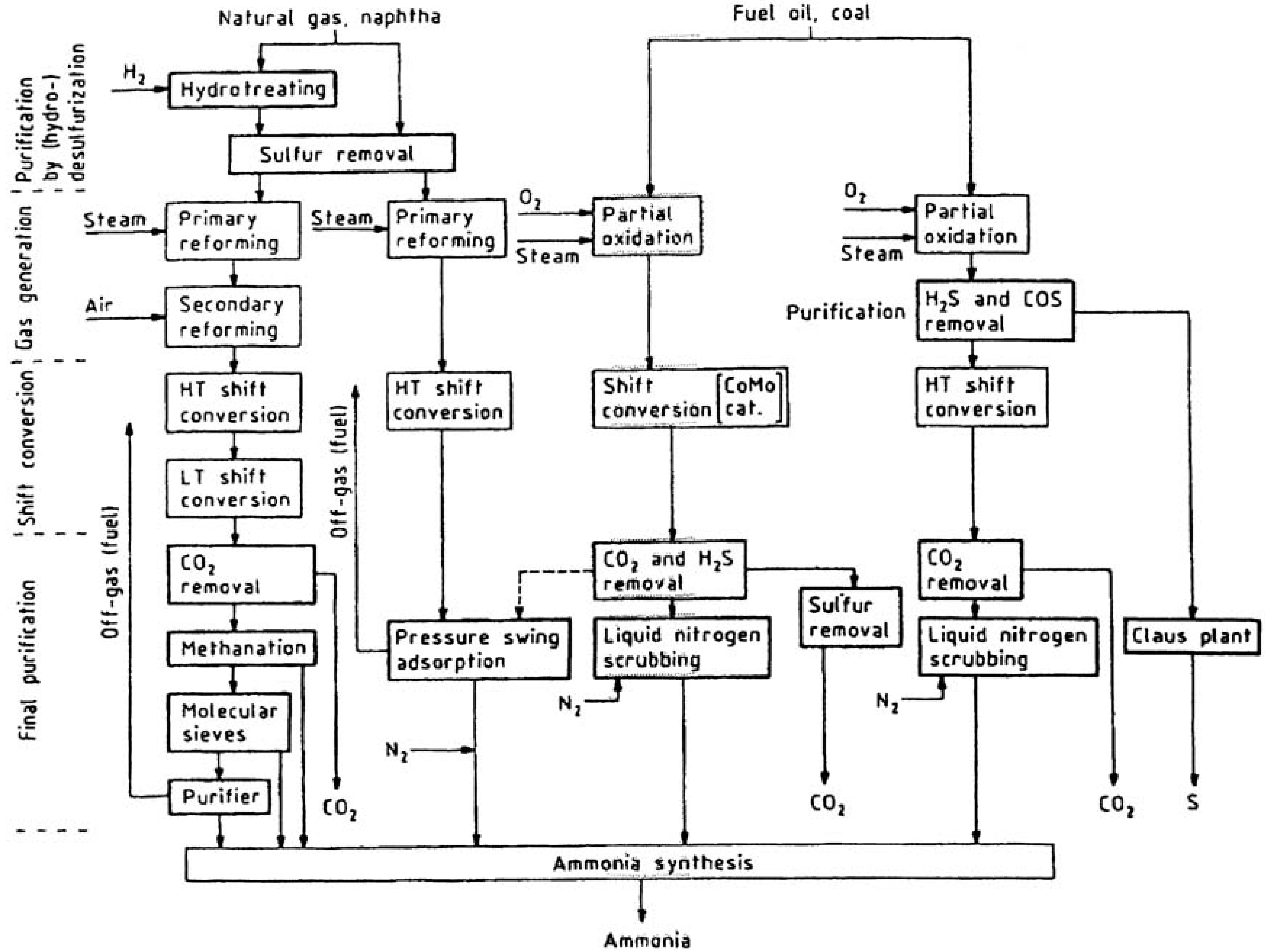
Battelle Pacific Northwest National Laboratories are developing microreactors that produce synthesis gas. These reactors can be mass-produced to yield efficient, compact, and cost-effective systems.61

A related effort is Fast Oxidation Reaction in Si-technology-based Microreactors (FORSiM), which is funded by the Dutch Technology Foundation and is a cooperative venture between the University of Twente and the Technical University of Eindhoven. The objective of this work is to build and operate the first microreactor for catalytic partial oxidation for small-scale and on-demand hydrogen production.61

Initial Purification of Synthesis Gas

Synthesis gas from the gasification of fossil fuels is contaminated by several gaseous compounds that would affect its further use in different ways. The purification steps depend on the type of syngas process and the feedstock. Table 29.9 illustrates how the synthesis gas stream can vary with different feedstocks. Figure 29.8 illustrates the different process steps that may occur before ammonia synthesis takes place.74

Fig. 29.8 Alternative process steps for generation and purification of synthesis gas. (Bakemeier, H., Huberich, T., et al., *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. A2, 5th ed., Reinhold Publishing, New York, 1985. Copyright by Reinhold Publishing and reproduced by permission of copyright owner.)



If sulfur is present as H2S or COS, it is a poison for many catalysts and will partly or completely inhibit the catalyst activity.46 Carbon monoxide (CO) and carbon dioxide (CO2) can poison the ammonia synthesis catalyst so both of these compounds must be removed.53

After the sulfur is removed, most traditional ammonia processes have employed the purification steps that are shown in Figs. 29.4 and 29.6. However, ammonia plants have been built that use hydrogen purification via PSA, membrane separation with polymeric membranes, and cryogenic separation. PSA achieves the greatest product purities (about 99.999% pure hydrogen), but it is capital-intensive. Polymeric membranes are the least capital-intensive, but they achieve the lowest purity (about 9698%). In 1998, PSA was considered the standard process for high-purity hydrogen recovery from raw synthesis gas. It has replaced the traditional steps of the shift conversion of carbon monoxide followed by carbon dioxide removal by absorption and then final purification by methanation. Characteristics of these three purification technologies are outlined in Table 29.10.40

TABLE 29.10 Hydrogen Recovery Technology Characteristics40,64

Characteristic	Membrane	Absorption	Cryogenics
Hydrogen purity (%)	<95	99.9+	9599
Hydrogen recovery (%)	<90	7590	9098
Hydrogen product pressure	< Feed pressure	Feed pressure	Variable
By-products available	No	No	Yes
Feed pressure (psig)	2501800		250500

Purification with PSA and Polymeric Membranes

The first purification plant that used PSA was developed by Union Carbide Corporation (UCC) and was built at the Yokkaichi Plant of Mitsubishi Petrochemical Industries in 1971.4 The process is now licensed by UOP, and more than 400 units were operating worldwide in 1992.1

The HYSEC Process was developed by Mitsubishi Kakoki K. and Kansai Coke & Chemicals. It has basically the same PSA unit as the UCC Process. After the main PSA beds, trace amounts of remaining oxygen are removed by a deoxo catalytic convener followed by a zeolitic dehumidifier.4

The Last OutFirst In (LOFIN) includes a unique gas-retaining vessel that preserves the concentration gradient in one stream before using it to repressurize another bed. This process was jointly developed by Toyo Engineering and Essex Corporation.4

The Sumitomo-BF PSA process uses CMS as the selective adsorbent.4

Monsanto and Ube (Japan) developed membrane processes for purification of hydrogen gas mixtures. This process is based on the selective diffusion of hydrogen through semi-permeable membranes in the form of hollow fibers. The Monsanto process (referred to as PRISM® separators) uses a polysulfone fiber while Ube uses an aromatic polyimide fiber.46 The PRISM® technology is now owned and marketed by Air Products.

Conventional polymeric hydrogen separation membranes yield hydrogen at low pressure. Air Products has demonstrated a carbon membrane on an alumina support that removes hydrocarbons from hydrogen/hydrocarbon mixtures and leaves the hydrogen at high pressure.40

$$CO + H_2O \rightleftarrows H_2 + CO_2 \qquad \Delta H = -10\,\mathrm{kcal/mol\ or\ 42\,kJ/mol}$$

Homogenous water gas reaction (water gas shift)

(29-2)

For ultrahigh-purity hydrogen (up to 99.9999% pure), palladium membranes are used. Johnson Matthey is believed to be the leading supplier of these membranes. In 1996, Generex (Tryon, NC) licensed newly developed palladium membrane technology from Los Alamos National Laboratories. In 1998, Wah Chang, an Allegheny Teledyne Company, purchased all patents and rights to the hydrogen metal membrane technology that Bend Research developed.40

U.S. Patent 6,183,542 was issued in 2001 for a palladium membrane process. This process provides an apparatus that can handle high flow rates of gas while using a minimal amount of hydrogen-permeable material.

Proton Energy Systems (Rocky Hill, CT) received U.S. Patent 6,168,705 in 2001 for an electrochemical gas purifier system that can purify and simultaneously compress hydrogen from a dirty gas stream without relying on moving parts. Final hydrogen gas pressures can exceed 2000 psig.

The companies that supply PSA and membrane separation systems can be found in North America, Western Europe, and Japan. Some of these companies are listed in SRI International.40

Carbon Monoxide Shift

The water gas shift conversion or the carbon monoxide shift reaction is one of the traditional purification steps that is found in many ammonia plants. The CO must be removed because it acts as a poison to the catalyst that is used in ammonia synthesis.

The carbon monoxide shift removes most of the CO from the synthesis gas and also produces more hydrogen.

The "shift" from CO to CO2 occurs in two steps. In the high temperature shift (HTS) conversion, the synthesis gas is passed through a bed of iron oxide/chromium oxide catalyst at around 400°C. The CO content of the gas is reduced to about 3 percent (on a dry gas basis). Additional details about the HTS step are given in references.53,70

The gas from the HTS is cooled to increase the conversion, and then it is passed through the low temperature shift (LTS) converter. The LTS converter is filled with copper oxide/zinc oxide-based catalyst and operates at about 200220°C. The residual CO content is about 0.20.4 percent (on a dry gas basis).53 Some LTS reactors operate with an inlet temperature of 190210°C and reduce the CO level to 0.10.2 mole percent (dry).70 Additional details about the water-gas shift catalysts are given in Lombard,62 Lundberg,63 Carstensen,73 and Thompson.83,84

Removal of Sulfur Compounds and Carbon Dioxide

If sulfur is present as H2S or COS or if CO2 is present, any of these compounds will be a poison for many catalysts and will partly or completely inhibit catalyst activity. The point at which sulfur removal is employed depends on the synthesis gas process that is used. Table 29.11 lists many of the processes that are available.46

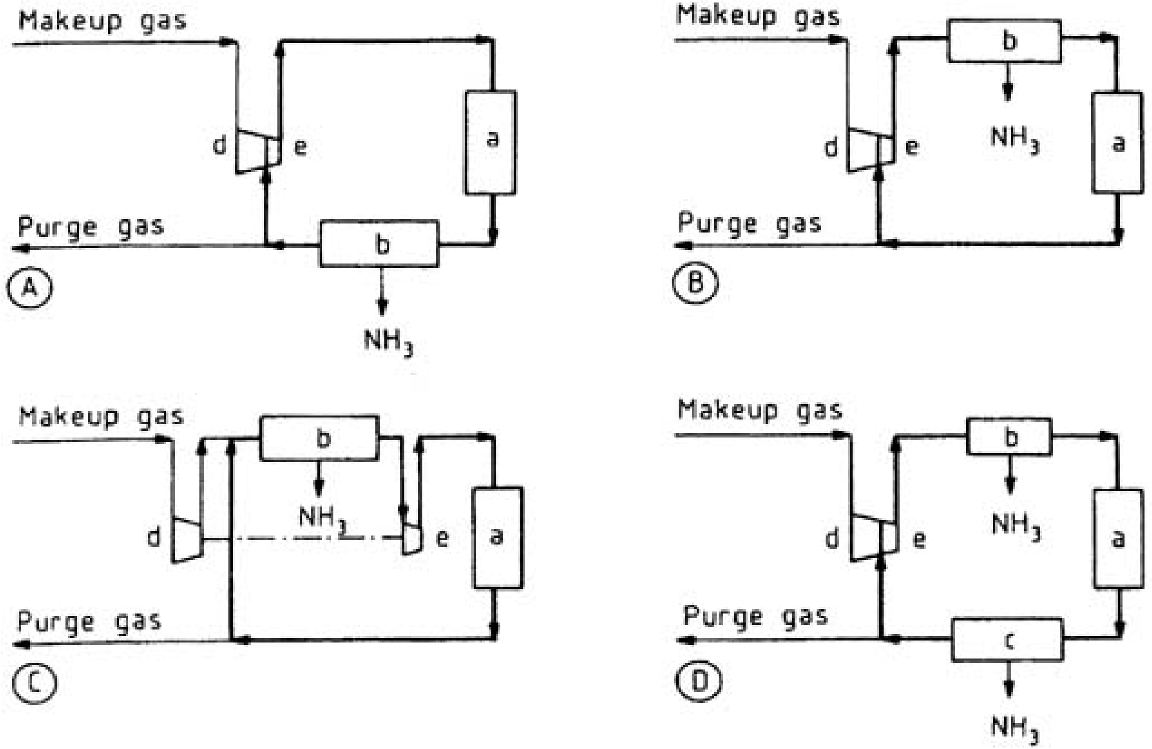
TABLE 29.11 Processes for Removal of Carbon Dioxide and Sulfur Compounds from Synthesis Gas

Process Name	Key Chemical(s)
The Alkazid Process	Alkali salts of amino acids (N-methylaminopropionic acid)
The AMDEA Process	Activated methyl diethanolamine
The Benfield Process	Hot potassium carbonate
The Carsol Process	Potassium carbonate
The Catacarb Process	Hot potassium carbonate
The Fluor Solvent Process	Propylene carbonate
The GiammarcoVetrocoke Process	Potassium carbonate with arsenite
The Hi Pure Process	Two-stage hot potassium carbonate washing
The Purisol Process from Lurgi	N-methyl-2-pyrrolidone (NMP)
The Rectisol Process	Low temperature methanol
The Selexol Process	Dimethyl ethers of polyethylene glycol
The Shell Sulfinol Process	Di-isopropanolamine dissolved in sulfolane and water
Pressurized Washing	Monoethanolamine (MEA) or diglycolamine

The Amine Guard is a corrosion inhibitor that was developed by UCC. It is added to the MEA solvent and has the further advantage that it allows MEA concentration to be increased. A higher MEA concentration leads to a lower circulation flow rate and a lower energy demand during regeneration.74

As shown in Figs 29.5, 29.7, and 29.9, the CO2 removal step is normally after the shift conversion step. The process gas from the LTS converter contains mainly hydrogen, nitrogen, CO2, and excess process steam. The gas is cooled and most of the excess steam is condensed before it enters the CO2 removal system. This condensate normally contains 15002000 ppm of ammonia and 8001200 ppm of methanol. Therefore, it should be stripped or recycled.

Fig. 29.9 Schematic flow diagrams of typical ammonia synthesis loops.74 (A) Synthesis loop for pure and dry makeup gas; (B) Product recovery after recycle compression; (C) Product recovery before recycle compression; (four-nozzle compressor design); (D) Two stages of product condensation; (a) Ammonia converter with heat exchangers; (b) Ammonia recovery by chilling and condensation; (c) Ammonia recovery by condensation at ambient temperature; (d) Synthesis gas compressor; (e) Recycle compressor. (Bakemeier, H., Huberich, T., et al., *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. A2, 5th ed., Reinhold Publishing, New York, 1985. Copyright by Reinhold Publishing and reproduced by permission of copyright owner.)



The heat that is released during the cooling/condensation can be used: to regenerate CO2 scrubbing solution, to drive an absorption refrigeration unit, or to preheat boiler feedwater. The amount of heat released depends on the process steam to carbon ratio (see Tables 29.12 and 29.16). It all this low-level heat is used for CO2 removal or absorption refrigeration, then high-level heat has to be used for the feedwater system. An energy-efficient process should therefore have a CO2 removal system with a low heat demand.53

TABLE 29.12 Health Effects of Different Ammonia Concentrations

Ammonia Concentration (ppm)	Health Effecta
5	Threshold detection limit
50	Easily perceived
5072	No significant impairment to respiration
100	Irritation to nose and throat Burning sensation in eyes
200	Headache and nausea
250500	Rapid heart beating
700	Immediate onset of burning sensations in the eyes
1000	Immediate coughing

aThe TLV has been set at 25 ppm (17 mg/m3) as an 8 hr time-weighted average (TWA) with a short-term exposure limit (STEL) of 35 ppm (24 mg/m3).

The CO2 is removed in either a chemical or a physical absorption process. Residual CO2 contents are usually in the range of 501000 ppmvdepending on the type and design of the removal unit. The physical absorption processes may be designed for zero heat consumption, but for comparison with the chemical processes, the mechanical energy requirements have to be considered.53

Alkazid Process

The Alkazid process removes sulfur compounds from gas streams. All the sulfur compounds are first catalytically hydrogenated to hydrogen sulfide (H2S) using a cobalt/molybdena catalyst. The H2S is then absorbed in an aqueous solution of a potassium salt of either methylamino propionic acid ("Alkazid M"), or dimethylamino acetic acid ("Alkazid DIK"). This solution is heated to regenerate the hydrogen sulfide as a concentrate. This concentrate is then treated by the Claus process to recover the sulfur.1

AMDEA Process

The Activated Methyl Diethanolamine (AMDEA) process removes CO2, H2S, and trace sulfur compounds from natural gas and syngas via a pressurized wash with activated diethanolamine. This process was developed by BASF and in 1997 it was used in more than 90 plants.1

Benfield Process

The Benson and Field (Benfield) process removes carbon dioxide, H2S, and other acid gases from industrial gas streams by scrubbing with hot aqueous potassium carbonate that contains activators. The chemical reactions are:1,46

$$\mathrm{K_2CO_3 + CO_2 + H_2O \rightleftarrows KHCO_3} \qquad (29\text{-}12)$$

$$\mathrm{K_2CO_3 + H_2S \rightleftarrows KHS + KHCO_3} \qquad (29\text{-}13)$$

Process details are given in *Chem. Eng. Prog.*65 and Nair.88 The process was invented in 1952 is now licensed by UOP. More than 650 plants were operating in 1996.1

Catacarb Process

The Catalyzed Removal of Carbon Dioxide (Catacarb) process removes carbon dioxide and hydrogen sulfide from gas streams by adsorption in a hot potassium carbonate solution that contains a proprietary catalyst. The process was developed and licensed by Eickmeyer and Associates based on work at the U.S. Bureau of Mines in the 1950s. More than 100 plants were operating in 1997.1

Fluor Solvent Process

The Fluor Solvent Process removes CO2 from natural gas and various industrial gas streams by dissolving the CO2 in a propylene carbonate solvent.1 Additional details are given in *Chem. Eng.*67 and *Hydrocarbon Processing.*68 The process was invented in 1958 by the Fluor Corporation and can be licensed from Fluor Daniel. By 1985, 13 plants were operating.1

GiammarcoVetrocoke Process

Two processes are known by the GiammarcoVetrocoke name. Both processes use an aqueous solution of sodium or potassium carbonate and arsenite to absorb acid gases. Some variations of the process use glycine instead of arsenite to activate the potassium carbonate solution. In one process, the solution is used to extract carbon dioxide from natural gas or synthesis gas. In the other, hydrogen sulfide is extracted from coke-oven or synthesis gas and yields elemental sulfur from a complex sequence of reactions. In 1992, more than 200 plants were operating.1,66

Hi Pure Process

The Hi Pure Process is a variation of the Benfield process. It uses two stages of scrubbing by hot potassium carbonate solution to reduce the CO2 content of gases to very low levels.1

Purisol Process

The Purisol Process removes H2S from gases by selective absorption in N-methyl-2-pyrrolidone (NMP). It was developed and licensed by Lurgi, particularly for desulfurizing waste gases from Integrated Gasification Combined Cycle (IGCC) coal gasification plants. However, the technology might be applied to some synthesis gas processes. In 1996, seven Purisol units were either in operation or under construction.1

Rectisol Process

The Rectisol Process was originally developed to remove sulfur compounds from gas mixtures that are produced from the partial oxidation of hydrocarbons. It is based on pressurized washing with low-temperature methanol, which results in the physical absorption of the sulfur compounds in the methanol. The process was originally developed in 1951 by Lurgi for the

SASOL coal gasification plant in South Africa. This process was further developed by Linde and is now offered for licensing by both companies for the removal of CO2, H2S, HCN, C6H6 (benzene), and gum-forming hydrocarbons from syngas and fuel gas. In 1990, over 70 units were in operation or under construction.1

Selexol Process

The Selexol Process removes acid gases from hydrocarbon gas streams by selective absorption in polyethylene glycol dimethyl ether (DMPEG). It absorbs H2S, CO2, COS, and mercaptans. The process has been used to remove carbon dioxide from syngas, natural gas, and coal gas. The process was developed by Allied Chemical in the 1960s and is now offered for license by UCC. Over 50 units were operating in 1996.1

Shell Sulfinol Process

The Shell Sulfinol Process removes H2S, CO2, COS, and organic sulfur compounds from natural gas by scrubbing with di-isopropanolamine dissolved in a mixture of sulfolane (C4H8SO2) and water. It was developed in the 1960s by Shell. In 1996, over 180 commercial units were operating or under construction.1

Pressure Washing with Monoethanolamine (MEA)

In this process, 1530 percent solution of MEA in water is used to absorb the CO2 under pressure. The solution is then regenerated by heating it in a stripper to release the CO2. This process is characterized by good CO2 absorption properties at low pressure. However, high regeneration energy consumption limits its use.

If the Amine Guard corrosion inhibitor is used, the MEA concentration in the circulating solution can be increased to 30 percent from a normal 20 percent. Hence, the circulating rate can be decreased by 33 percent and the heat requirements are decreased by 43 percent.69 Some pressurized wash processes also use diglycolamine to remove sulfur compounds and CO2.46

Final Purification of Synthesis Gas

Before the synthesis gas enters the ammonia synthesis loop, essentially all of the oxygen compounds must be completely removed to: (1) avoid poisoning the ammonia synthesis catalyst and (2) keep CO2 from forming carbamates and ammonium carbonate in the synthesis loop. It is also advantageous to remove the inert gases (methane, argon, etc.) to achieve a higher synthesis conversion per pass.74

Methanation

In a steam reforming process that includes CO shift conversion and CO2 removal, the synthesis gas still contains 0.10.2 mole percent CO and 1001000 ppmv of CO2. The following reactions are the simplest method for eliminating these small concentrations of oxygen compounds.

$$\mathrm{CO} + 3\mathrm{H}_2 \rightleftarrows \mathrm{H}_2\mathrm{O} + \mathrm{CH}_4 \quad \Delta H = -49.27\,\mathrm{kcal/mol\ or\ }206\,\mathrm{kJ/mol} \quad (29\text{-}11)$$

$$\mathrm{CO}_2 + 4\mathrm{H}_2 \rightleftarrows 2\mathrm{H}_2\mathrm{O} + \mathrm{CH}_4 \quad \Delta H = -39.44\,\mathrm{kcal/mol\ or\ }165\,\mathrm{kJ/mol} \quad (29\text{-}12)$$

Methanation

The normal methanation operating temperature is 250300°C, and a large excess of hydrogen is present. The equilibrium lies far to the right side of the above reactions so the CO and CO2 impurities can be reduced to about 5 ppm.70 Additional catalyst and process details are given in Bakemeier et al.,74 Synthesis Gas Technologies,75 Bajus• and Goodwin,77 and Hoost.78

Cryogenic Purification

In the cryogenic purifier, all the methane and the excess nitrogen are removed from the synthesis gas as well as a part of the argon. The cooling is produced by depressurization and no external refrigeration is needed. The purified syngas is then practically free of all impurities, except for a small amount of argon. The cryogenic unit also receives the purge from the ammonia synthesis section and delivers an off-gas for fuel. The combination of higher conversion per pass and reduced purge flow result in a more efficient process.53 The KBR Purifier is an example of this process.81 Fifteen KBR Purifier plants have been built since 1966. These plants range in capacity from 680 to 1750 tonnes per day, and as of 2001, all 15 plants are still operating.81

Dehydration

If the makeup gas to the ammonia synthesis loop is absolutely free of catalyst poisons, such as H2O and CO2, it can flow directly to the ammonia synthesis converter. This leads to the most favorable arrangement from a minimum energy point of view. This can be accomplished by allowing the gas that leaves the methanation step to pass through beds of molecular sieves to remove water.74

Liquid Nitrogen Wash

In many partial oxidation syngas processes, liquid nitrogen scrubbing is used to remove the carbon monoxide that remains after the shift conversion step. The CO content may be as high as 35 percent in plants that have only a HTS conversion. The liquid nitrogen wash: (1) delivers a gas to the ammonia synthesis loop that is free of all impurities (including inert gases) and (2) adds all or part of the nitrogen that is required for ammonia synthesis. The nitrogen is obtained from the air-separation plant that provides the oxygen for the partial oxidation process (see Fig. 29.6).74

Adjust Hydrogen to Nitrogen Ratio

The optimum ammonia synthesis reaction rate depends on several factors including pressure, temperature, H2-to-N2 molar ratio, and catalyst activity. Therefore the H2-to-N2 molar ratio is adjusted to suit the requirements in ammonia synthesis. This adjustment occurs before the compression step.

Compression

Ammonia synthesis is normally carried out at a pressure that is higher than that for synthesis gas preparation. Therefore, the purified synthesis gas to the ammonia synthesis loop must be compressed to a higher pressure.74

Due to several major developments in ammonia process technology, ammonia plants with 10001500 tonne per day capacities have became the industry standard for new plant construction. In 2001, plants as large as 2000 tonnes per day have become common. These plants have much lower production costs than the earlier generation of smaller plants mainly because centrifugal compressors are used rather than reciprocating compressors.57,74

29.7 Ammonia Synthesis

The ammonia synthesis reaction is:74

$$N_2 + 3H_2 \rightleftarrows 2NH_3$$

$$\Delta H_{298} = -45.72\,\mathrm{kJ/mol}\ \text{or}\ 10.94\,\mathrm{kcal/mol} \qquad (29\text{-}14)$$

The reaction normally takes place on an iron catalyst. The reaction pressure is in the range of 100250 bar, and temperatures are in the range of 350550°C. At the usual commercial converter operating conditions, the conversion achieved per pass is 2030 percent.53 In most commercial ammonia plants, the Haber recycle loop process is still used to give substantially complete conversion of the synthesis gas. In this process, the ammonia is separated from the recycle gas by cooling and condensation. Next the unconverted synthesis gas is supplemented with fresh makeup gas, and returned as feed to the ammonia synthesis converter.74

Synthesis loop arrangements differ with respect to: (1) the points in the loop at which the make-up gas is delivered, (2) where the ammonia is taken out, and (3) where the purge gas is taken out53 (see Fig. 29.9). The best arrangement is shown in Fig. 29.9(A). After the gas leaves the ammonia synthesis converter, ammonia is condensed/removed by cooling and the recycle gas is returned to the recycle compressor. This represents the most favorable arrangement from a minimum energy point of view. It results in the lowest ammonia content at the entrance to the converter and the highest ammonia concentration for condensation.74 The advantages and disadvantages of the other arrangements in Fig. 29.9 are described in Synthesis Gas Technologies.74

Conventional reforming with methanation as the final purification step produces a synthesis gas that contains inerts (CH_4 and argon) in quantities that do not dissolve in the condensed ammonia. Most of the inerts are removed by taking a purge stream out of the synthesis loop. The size of this purge stream controls the level of inerts in the loop at about 1015 percent. The purge gas is scrubbed with water to remove ammonia and then it can be used as fuel or sent to hydrogen recovery. The best point at which to take the loop purge is discussed in EFMA.53

Reaction Rate

Knowledge of the macrokinetics is important for solving the industrial problem of designing ammonia synthesis reactors, for determining the optimal operating conditions and for computer control of ammonia plants. Some of the considerations are: high pressure promotes a high rate of ammonia formation; high ammonia concentration in the synthesis gas (recycle gas) restricts ammonia formation (see Fig. 29.10); the rate of formation initially increases with rising temperature but then goes through a maximum as the system approaches thermodynamic equilibrium (see Fig. 29.11); and with lower temperatures, the maximum rate shifts to a lower hydrogennitrogen ratio (see Fig. 29.12).74

Fig. 29.10 Reaction rate for ammonia synthesisdependence on ammonia concentration at various pressures. (Bakemeier, H., Huberich, T. et al., *Ullmann's Enclyclopedia of Industrial Chemistry*, Vol. A2, 5th ed., Reinhold Publishing, New York, 1985. Copyright by Reinhold Publishing and reproduced by permission of copyright owner.)

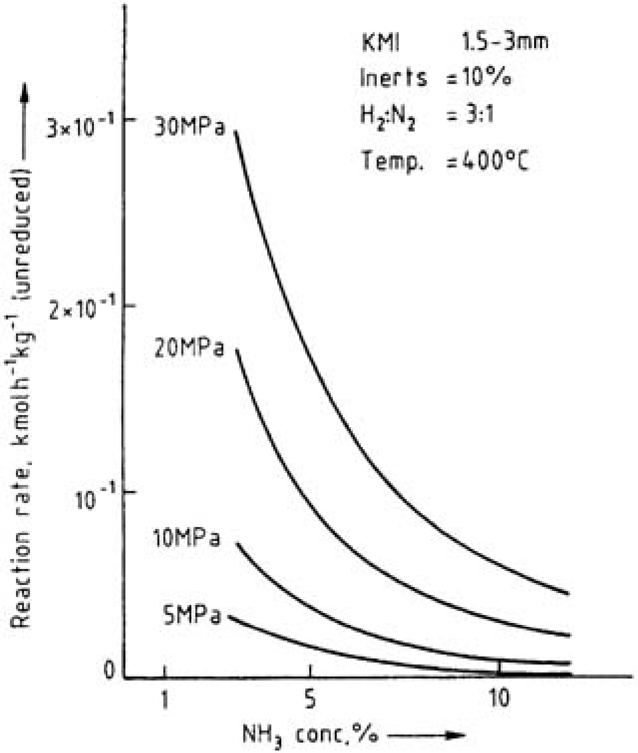


Fig. 29.11 Reaction rate for ammonia synthesisdependence on temperature at various pressures. (Bakemeier, H., Huberich, T. et al., *Ullmann's Enclyclopedia of Industrial Chemistry*, Vol. A2, 5th ed., Reinhold Publishing, New York, 1985. Copyright by Reinhold Publishing and reproduced by permission of copyright owner.)

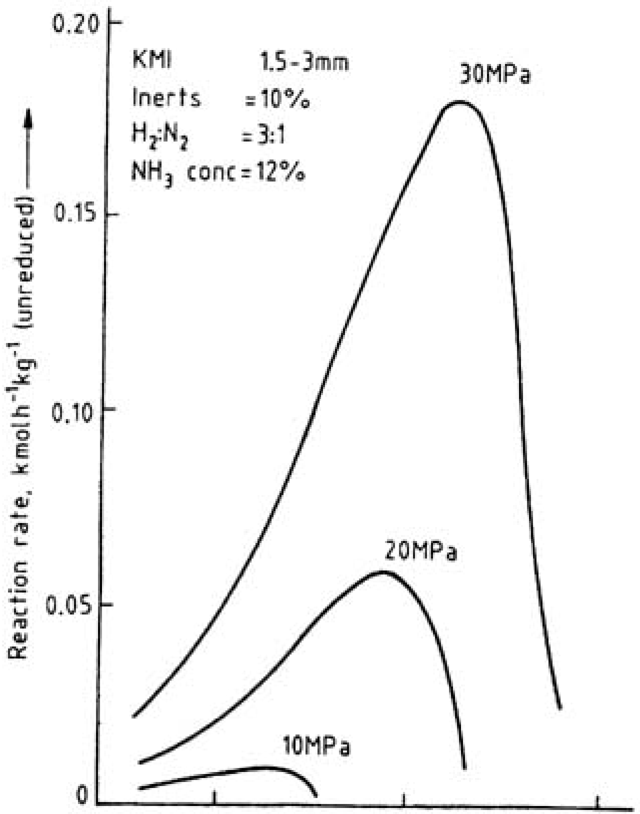
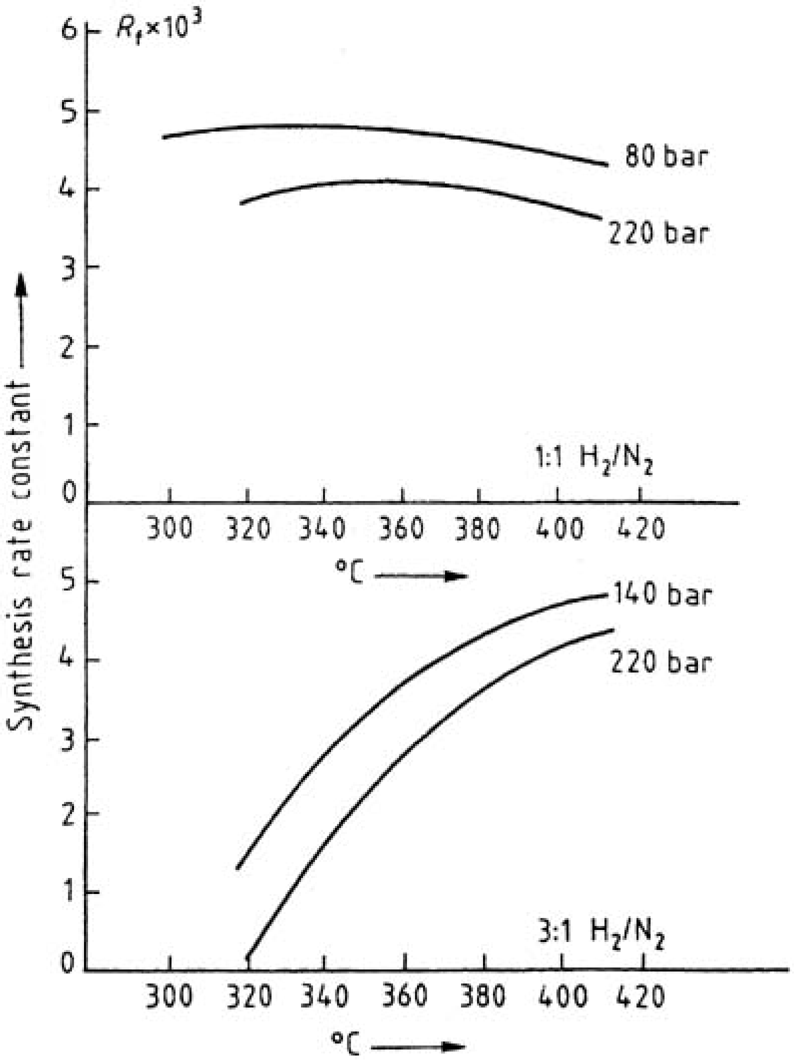


Fig. 29.12 Ammonia synthesis rate constant dependence on hydrogennitrogen ratio. (Bakemeier, H., Huberich, T., et al., *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. A2, 5th ed., Reinhold Publishing, New York, 1985. Copyright by Reinhold Publishing and reproduced by permission of copyright owner.)

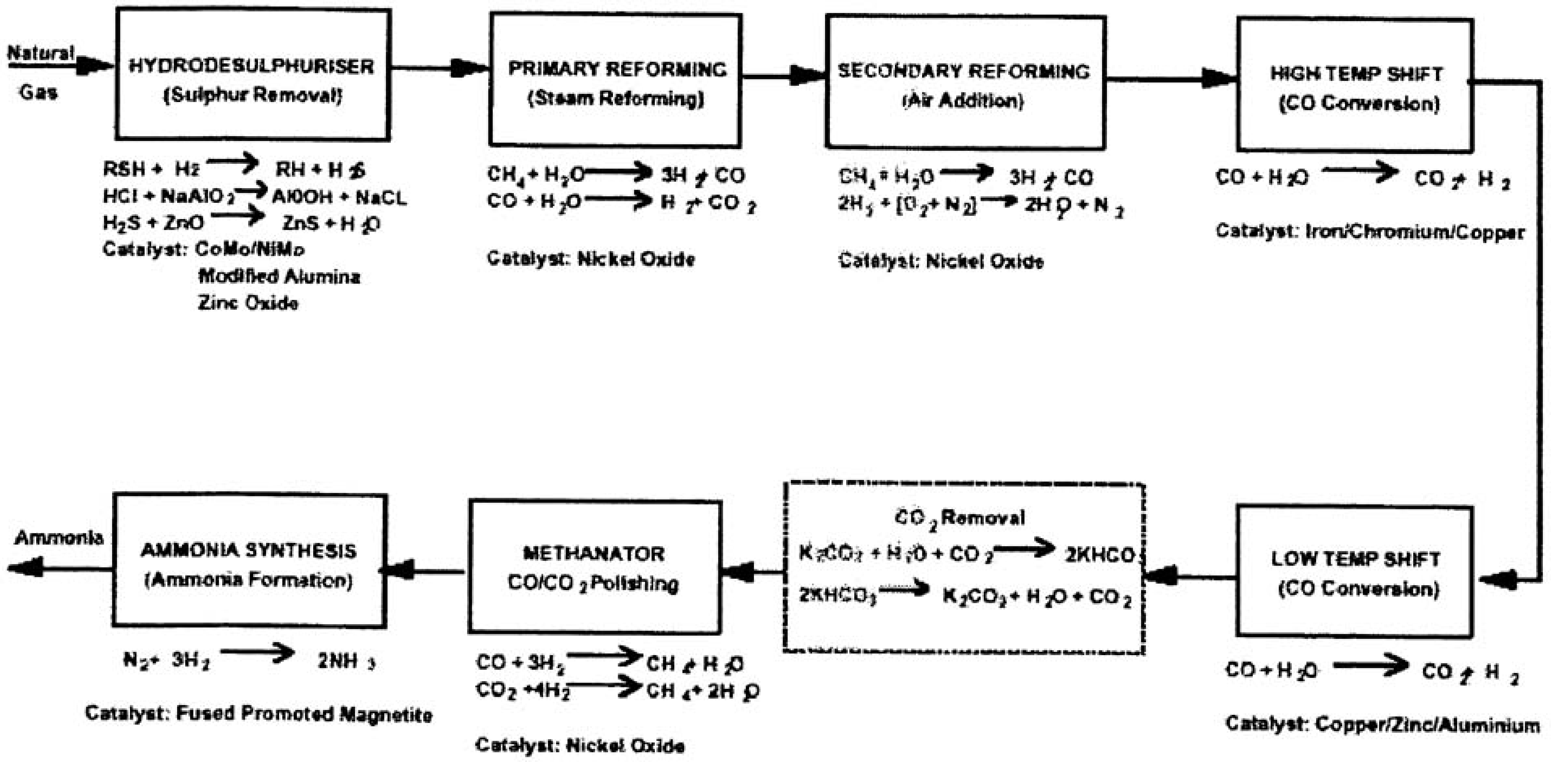


In 2001, Hyprotech and Synetix announced an ammonia plant simulation that can be used for modeling, on-line monitoring, and optimization of the plant. The simulation includes Synetix reactor models, customized thermodynamic data and information to simulate the performance of a range of catalysts. The reactor models in the simulation include primary and secondary reformers, high temperature shift converter, low temperature shift converter, methanator, and ammonia synthesis converter.80

Catalysts

When ammonia is made from natural gas using the steam reforming process, several reaction stages are needed and catalysts are key to the economic operation of each stage. The chemistry of the process and the basic materials that make up the catalysts are shown in Fig. 29.13.

Fig. 29.13 Chemistry of ammonia process and basic catalyst materials.82 (*Reproduced by permission of Imperial Chemical Industries PLC*)



For a given operating pressure and a desired production rate, the catalyst determines: (1) the operating temperature range, (2) recycle gas flow, and (3) refrigeration requirements. It also indirectly influences the makeup gas purity requirements.74

Industrial catalysts for ammonia synthesis must satisfy the following requirements: (1) high catalyst activity at the lowest possible reaction temperatures, (2) the highest possible insensitivity to oxygen- and chlorine-containing catalyst poisons, (3) long life, and (4) mechanical strength.

The choice of particle size and shape of commercial ammonia catalysts is determined mainly by: (1) the catalyst performance and (2) the pressure drop. From the standpoint of spacetime yield, it is desirable to use the finest possible particle, which is about 12 mm.74

For processes operating at pressures of 2545 MPa (250450 bar) and space velocities of 800020,000 m3/m3 h (STP), a grain size of 610 mm is preferred. Space velocity is defined as the number of reactor volumes of feedat specified conditionsthat can be treated in unit time. Larger granulations (815 mm or 1420 mm), are used only in plants where the lowest possible pressure drop is essential because of very high gas velocities. In catalyst zones in which the ammonia formation rate is so high that the allowable temperature limits are exceeded, it may be advantageous as well to use coarse particles for suppressing the reaction.74

Some of the poisons of ammonia catalysts are discussed in Bakemeier et al.74

The activity of damaged catalysts may be almost completely restored by reduction with clean synthesis gas at a relatively low temperature. It has also been found that the degree of poisoning rises as the partial pressure ratio, *P*H2O/*P*H2 increases and the degree of poisoning falls with increasing temperature.74

Energy Efficiency

Due to increased feedstock costs, some new ammonia plant designs use fuel more efficiently but their capital cost may be higher. The recovery of hydrogen and ammonia from the synthesis purge gas by a cryogenic unit or a membrane system results in an ammonia capacity increase of about 5 percent.57

Substantial improvements have been made in the energy efficiency of CO2 removal systems. The first large-scale ammonia plants in the 1960s typically used monoethanolamine (MEA) as a solvent. Energy input was over 50,000 kcal/kg-mol of CO2 removed. In 2001, plants use improved solvents and designs that can reduce the energy input to about 10,000 kcal/kg-mol of CO2 removed.57

A typical world-scale plant that was built in the 1970s consumed about 42 billion BTU of natural gas per tonne of ammonia produced. Retrofitting such a plant to improve fuel efficiency can reduce gas consumption to about 36 million BTU per tonne. Ammonia plants that were built in the late 1990s use only about 30 million BTU per tonne of ammonia, are easier to operate, and have slightly lower conversion costs. Some new plants also recover more than one million BTU per tonne by generating electricity from waste heat.57

Ammonia Plant Design

A simplified flowsheet for an ammonia plant that processes natural gas via steam reforming is shown in Fig. 29.14. A block diagram of this same plant is shown in Fig. 29.15. This diagram lists typical stream compositions, typical operating conditions, catalyst types (recommended by Synetix), and catalyst volumes.

Fig. 29.14 Simplified ammonia plant flowsheet.82 (*Reproduced by permission of Imperial Chemical Industries PLC*)

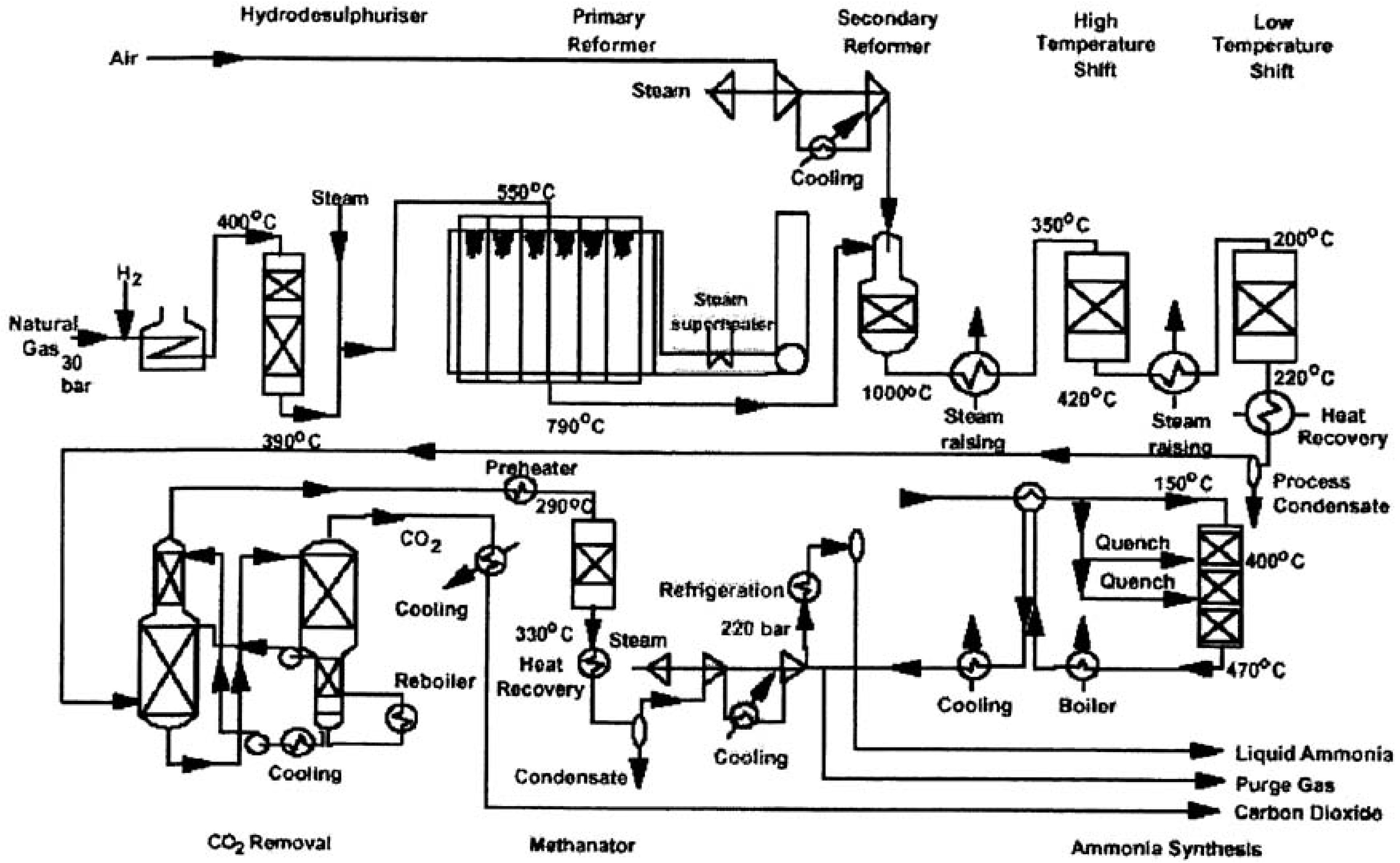
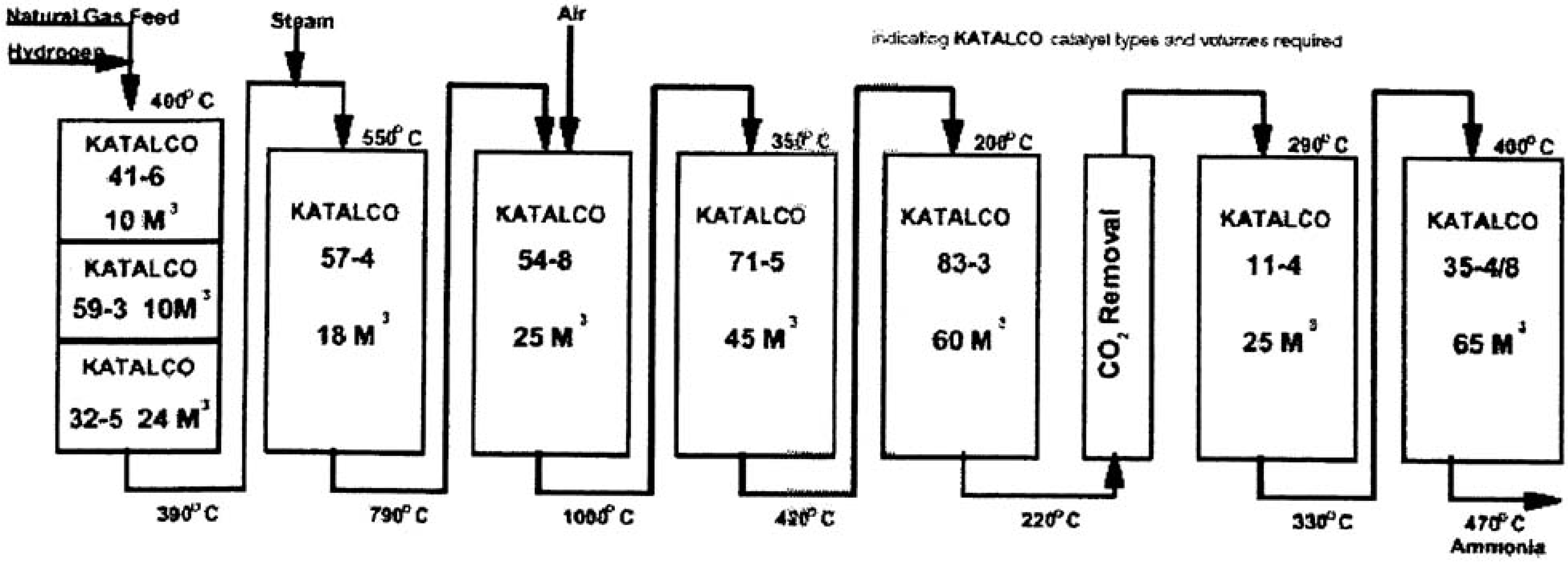


Fig. 29.15 Block diagram of 1000 tonne/day ammonia plant.82 (*Reproduced by permission of Imperial Chemical Industries PLC*)



Exit Gas Composition	Purification	Primary Reforming Volume %	Secondary Reforming Volume %	High Temperature CO Shift Volume %	Low Temperature CO Shift Volume %	CO_2 Removal Volume %	Methanation MUG Volume %	Ammonia Synthesis Loop Volume %
N_2		69.8	56.5	60.3	61.4	74.6	74.2	63.4
CO		9.3	12.9	3.0	0.3	0.4	-	-
CO_2		10.5	7.5	15.6	17.8	0.1	-	-
CH_4		10.4	0.3	0.3	0.2	0.3	0.8	9.5
N_2 + A		-	22.8	20.8	20.3	24.6	25.0	24.9
CO + CO_2		-	-	-	-	-	5ppm	<5ppm NH3 - 2.5

The Kellogg Brown & Root Advanced Ammonia ProcessPlus (KAAP*plus*™) is an ammonia process design made up of commercially proven technologiesKBR Reforming Exchanger System (KRES), KBR Purifier, and KBR Advanced Ammonia Process (KAAP) for synthesis. This process uses a ruthenium-based catalyst that is: (1) much more active than the iron-based catalysts and (2) improves synthesis efficiency because synthesis pressure is lowered from a typical pressure of 150 bar (2175 psig) to 90 bar (1305 psig).57,81 Capital cost savings of 38 percent are achieved with KAAP*plus*™.81

Haldor Tops⊘e's ammonia synthesis technology is based on the S-200 ammonia converter. This is a two-bed radial flow converter with indirect cooling between the beds. This converter concept has been used extensively to upgrade existing converters (Tops⊘e or other designs) in modification projects to achieve higher capacity (up to 20%) and/or better energy efficiency.85

Ammonia Production Costs

Ammonia production costs are very dependent on feedstock cost; which is normally natural gas. Natural gas prices can vary significantly from region to region and by the cycles in the world energy market. Typical ammonia production costs are listed in SRI International.57 The natural gas cost makes up 70 percent to over 85 percent of the cash production costs as the natural gas cost varies from $2.00 per million BTU up to $6.00 per million BTU. In general, the ammonia cash production costs in dollars per tonne are a factor of 4050 times the natural gas price in US dollars per million BTU.

Reliability and Revamps

Reformer

Where an increase of 1050 percent of existing steam reforming capacity is required, revamping is often more cost-effective than a new plant or the purchase of hydrogen over-the-fence. In most revamps, the reaction area is only slightly affected by operating rate, and existing reactors are often able to operate at increased rate. Existing heat exchangers are often adequate as capacity is increased. However, air coolers in a CO2 removal unit often need to be modified by adding surface area or increasing fan horsepower.86

Separation equipment tends to have fixed limits and can be costly to change. However, more capacity can often be obtained by changing column internals or the solvent composition.86

Convection Section

The effects on the reformer of increased gas flow and temperature are multiplied in the convection section because additional load is placed on convection coils, fans, and the steam system. These areas commonly limit the operating rate of the reformer. The tube supports also need to be considered because they are exposed to hot flue-gas without the cooling effect of process fluids.86

One of the most effective reformer modifications is to use heat from the convection section to preheat radiant section feed. This will reduce radiant section heat load, reduce radiant section firing rate, and potentially unload other areas such as steam generation. This option has been used to increase capacity by 10 percent without increasing the arch temperature in the radiant section.86

Combustion Air Preheat

Combustion air preheat reduces the flue-gas flow through the furnace which unloads the fans and the entire convection section. The best known option is to use heat exchange with flue-gas. For new units this can have the advantage of very high efficiency because the heat sink is ambient air. Another option is the use of steam to preheat air. This option was used to obtain a 12 percent increase in capacity at the same fuel firing rate and combustion air flow. A third option is to use circulating boiler feedwater to preheat the air.86

Radiant Section

Changes in the radiant section can be difficult to justify based on capacity alone. However, if the radiant tubes are near the end of their useful life, an upgrade to this section may provide more capacity. Changes in metallurgy since the mid-1970s have allowed changes in radiant section operating conditions:

Time Period	Tube Material	Operating Pressure (psig)	Operating Temperature (°F/°C)

1970s	HK-40	300	1450/790
1980s	HP	325	1550/845
1990s	Microalloy	450	1575/855

The microalloy tubes allow increased flux rates and higher reformer outlet temperatures. This in turn can make it possible to reduce the steam-to-carbon ratio while the hydrogen purity remains the same.86

CO2 Removal System

In a CO2 removal system that uses wet scrubbing, the existing towers are the major limit to more capacity because they are expensive to replace. In an amine system, absorption increases as amine concentration increases. But a higher amine concentration requires: (1) more filtration to clean the solution and (2) the addition of corrosion inhibitors. Another option is to change from MEA to methyl diethanolamine (MDEA).86

In a potassium carbonate system, different additives can be used to increase CO2 absorption rate. In any wet scrubbing system, a change from random to structured packing can lead to higher solvent circulation rates and improved mass transfer.86

Ammonia Synthesis

If the synthesis gas contains traces of carbon oxides, ammonium carbamate will form upon mixing with the ammonia in the recirculating gas from the synthesis loop. The carbamate will clog and/or corrode downstream equipment. To avoid this condition, the carbon oxide levels in fresh makeup gas should be less than 5 ppm.88

Many ammonia synthesis converters are subject to nitriding and hydrogen embrittlement. The internal catalyst baskets are made of SS 321 material.88 The nitriding effect is more pronounced in low-alloy steels above 450°C. Austinic steels with a high-nickel content offer considerably more resistance. Alloys of the CrNiMo type are usually used for the gas side. Atomic hydrogen is absorbed in the metal by diffusion. The subsequent reaction with carbon in the steel results in blistering and cracking from decarburization.88

29.8 Uses of Ammonia

Anhydrous ammonia is produced in about 80 countries. About 85 percent is used for nitrogen fertilizer production, including about 6 percent that is directly applied to the fields.36 In the United States, the distribution of ammonia use differs slightly from the worldwide uses. In the United States, only 80 percent of the ammonia is used to make fertilizers. Chemical intermediates account for 19 percent of the ammonia use while the remaining 1 percent is used in pulp and paper, metals, and refrigeration applications.37

Chemical Production and Other Uses

The uses of ammonia are in:

1. *Acrylonitrile*. Acrylonitrile is used in the production of acrylic fibers and various resins. Acrylonitrile is produced by the catalytic oxidation of propylene and ammonia.57
2. *Caprolactam*. Caprolactam is used principally in the production of nylon 6 fibers. In the production of caprolactam, ammonia is used to maintain a reaction pH of 7 in one step, and in another step it is used to neutralize sulfuric acid. Both of these steps result in the production of a low-grade ammonium sulfate fertilizer. In the classical process, 5 kg of (NH4)2SO4 is produced per kilogram of caprolactam. In 1992, about 33 percent of the worldwide caprolactam capacity was based on DSM's caprolactam process that only produces 1.8 kg of (NH4)2SO4 per kilogram of caprolactam.46
3. *Miscellaneous Amines and Nitrites*. A small but significant amount of ammonia is consumed in the production of various amines and nitriles.57
4. *Metallurgy*. In this application, ammonia is cracked to produce a "dissociated ammonia" that consists of 75 percent hydrogen and 25 percent nitrogen. Dissociated ammonia is used in a number of metal treatment processes.57
5. *Refrigeration*. Anhydrous ammonia is used as a refrigerant in industrial, closed-circuit refrigeration systems. Ammonia's high latent heat, low vapor density, chemical stability, and low iron corrosivity promote its use. Ammonia is used in large industrial systems (rather than halocarbon refrigerants) because ammonia absorbs about six times as much heat per unit weight and energy costs of operation are substantially lower.57
6. *Pulp and Paper*. In the production of ammonia-based sulfite pulp, the amount of ammonia used per ton of pulp varies from mill to mill. Ammonia is also used for stack gas scrubbing and wastewater treatment.57
7. *Phosphate Ore Flotation*. A small amount of ammonia is used as a modifying reagent in the froth flotation of phosphate ores.
8. *Household Cleaners*. Household ammonia, liquid window cleaners, liquid all-purpose cleaners, and various other household cleaning products constitute this small market for ammonia. Ammonia is also used to make surfactants that go into liquid dishwashing detergents.57
9. *Air Pollution Control*. A number of processes use ammonia for scrubbing sulfur oxides (SO_x) and/or nitrogen oxides (NO_x) from industrial and electrical power plant stack gases.57
10. *Uranium Concentrate Production*. Ammonia is used to precipitate the final uranium concentrate salt (yellowcake) prior to drying or calcination. It is generally economically preferred over sodium hydroxide or hydrogen peroxide.57
11. *Water Purification*. Ammonia is combined with chlorine to purify some municipal and industrial water supplies.57
12. *Waste Treatment*. Small amounts of ammonia are added when needed as the nitrogen source for the bacteria in industrial and municipal biological waste treatment systems. Other industrial plants use minor amounts of ammonia to neutralize acid in plant wastes.57
13. *Duplicating Processes*. Ammonia is used as a process chemical in photographic copy machines.57
14. *Fabric Treatment*. Permanent-press, 100 percent cotton fabrics are produced via a licensed process involving mercerization in a bath of liquid anhydrous ammonia. This Sanfor-Set process was commercialized in 1975.57

29.9 Distribution and Storage

Ammonia Toxicity

Ammonia is a strong local irritant, and the primary target organ is the pulmonary system. Ammonia or ammonium hydroxide can penetrate the cornea rapidly and lead to damage of the iris, cataract, and glaucoma. Oral ingestion of aqueous ammonia can corrode the mucous membranes of oral cavity and constrictions of the esophagus may result.74

When ammonia enters the body, it is converted to urea and excreted by the kidneys. The capacity for detoxification via urea is sufficient to eliminate the ammonium ion when ammonia is inhaled in non-irritating concentrations. Ammonia is not considered to be carcinogenic nor is it mutagenic. The effects of different ammonia concentrations are summarized in Table 29.12.74

Ammonia Quality

Most industrial uses of ammonia require a higher purity level of product than agricultural uses. Typical ammonia specifications are shown in Table 29.13.74

TABLE 29.13 Ammonia Quality Specifications

	Commercial Grade		Refrigeration	
	USA	Germany	USA	Germany
Purity, wt. %, minimum	99.5	99.5	99.98	99.98[a]
Watery, wt. %, maximum	0.5	0.2	0.015	0.02
Inerts[b], mL/g, maximum	Not specified		0.1	0.08
Oil, ppm by weight	5.0	5.0	3.0	Not specified
Free of H_2S, pyridine and naphthenes				

[a] Allowable boiling point change on vaporization of 597% of the test sample, 0.9°C.

[b] The non-condensable gases dissolved in ammonia are H_2, N_2, CH_4, and Ar. Their amounts depend on the methods of synthesis and storage. The inerts amount to about 50 mL/kg for atmospheric storage.

Commercial grade is produced by ammonia synthesis while refrigeration grade is normally made from industrial raw ammonia by distillation. For ammonia shipped or pipelined in the United States, water content must be at least 0.2 wt. percent to inhibit stress corrosion cracking of the carbon steel.57

Various concentrations and purities of aqueous ammonia are on the market. The typical concentration is 2530 percent ammonia and the iron content is less than 10 ppm. If the ammonia content is above 25 percent, it must be shipped in a pressure vessel because of its elevated vapor pressure.74

Distribution and Storage

Ammonia production requires storage facilities to smooth over fluctuations in production, usage, and shipments. Ammonia is stored and distributed to point of use almost exclusively as a liquid. Gaseous ammonia sometimes is used within plants.74

Three methods have been used for storing liquid ammonia:74

1. pressure storage at ambient temperature in spherical or cylindrical pressure vessels having capacities up to about 1500 tonnes
2. atmospheric storage at -33°C in insulated cylindrical tanks for up to about 50,000 tonnes per vessel
3. reduced pressure storage at about 0°C. Usually in insulated, spherical pressure vessels for quantities up to about 2500 tonnes per sphere

Ammonia is delivered in small containers, tank trucks, tank cars, barges, and via pipeline. The most common small containers are: (1) cylindrical steel bottles and pressurized flasks that contain about 20200 kg and (2) polyethylene canisters and metals casks.74 Trucks have ammonia capacities up to 100 m3 while jumbo rail cars hold up to 150 m3. Additional guidelines about ammonia shipments are given in SRI International57 and Bakemeier et al.74

Several ammonia pipeline systems appeared in the United States in the 1960s. Their construction resulted from the construction of large plants near the sources of natural gas and from developing a large market for liquid anhydrous ammonia in direct fertilization. Pipelines are the lowest cost method to move ammonia over the long distances between the producers along the Gulf Coast and the consumers in the upper Midwest. Some pipelines are up to 3000 km long and up to 10 in. in diameter.74

An integrated pipeline system has not been developed in Europe. Only a few pipelines with lengths of less than 50 km have been installed. However, in the FSU, a 2424 km long pipeline is in use.74

Ammonia Price

Direct sales by basic producers are generally at wholesale prices. Basic producers also sell to distributors and jobbers for resale to end users buying smaller quantities. Distributors of ammonia to industrial end users have their own storage facilities, packaging plants, and distribution networks. They are able to supply end users that buy in less-than-truckload quantities (a minimum of 10002000 lb or 450910 kg) or in cylinders or bottles.57

Another important aspect of the ammonia business is seasonality. The major end use for anhydrous ammonia is as a direct application fertilizer. The application season, particularly in Canada and the northern United States, is limited and significant storage capability is needed. As a result, it is normal to expect a plant to produce a downstream nitrogen product (such as urea) that is: (1) easier to store and handle and (2) does not have such a limited application season. Industrial markets are normally nonseasonal.57

Long-term fixed-price contracts are not considered workable due to the volatility in the market and the uncertainty regarding future natural gas prices. The normal long-term contract simply guarantees continuity of supply and calls for prices to be negotiated every three to six months.57 Historical prices are listed in SRI International.57

29.10 Nitric Acid

Nitric acid is a strongly acidic, corrosive liquid that is produced by the oxidation of ammonia and subsequent reaction of the oxidation products with water. Pure nitric acid is colorless. The characteristic yellow-brown color generally associated with concentrated solutions is due to dissolved nitrogen dioxide.91

Nitric acid is an active compound and its salts are found in all fertile soils. The alchemists obtained nitric acid by heating alum and copper sulfate with nitrate in a retort. Owing to its powerful corrosive action, they named it aqua fortis or "strong water."7

Being a powerful oxidizing agent, nitric acid reacts violently with many organic materials and the reactions may be explosive. As a general rule, oxidizing reactions occur primarily with the concentrated acid and favor the formation of nitrogen dioxide (NO2). The acidic properties tend to dominate the dilute acid and this results in the preferential formation of nitrogen oxide (NO).53

Nitric acid reacts with all metals except the precious metal series and certain alloys. Although chromium, iron, and aluminum readily dissolve in dilute nitric acid, the concentrated acid forms a metal oxide layer that protects the metal from further oxidation.53

Physical Properties

Pure anhydrous nitric acid (100%) is a colorless liquid that solidifies at -41.6°C to form white crystals. It boils at 84.1°C. When it boils in light, a partial decomposition occurs with the formation of NO2 via the following reaction:

$$HNO_3 \rightleftarrows H_2O + 2NO_2 + \tfrac{1}{2}O_2 \qquad (29\text{-}15)$$

Anhydrous nitric acid should be stored below 0°C to avoid decomposition. The nitrogen dioxide remains dissolved in the nitric acid and creates a yellow color at room temperature and a red color at higher temperatures. While the pure acid tends to give off white fumes when exposed to air, acid with dissolved nitrogen dioxide gives off reddish-brown vapors which leads to the common name "red fuming acid."53

Nitric acid is miscible with water and distillation results in an azeotrope with a concentration of 68.4 percent HNO3 and a boiling temperature of 121.9°C at atmospheric temperature. Two solid hydrates are knownthe monohydrate (HNO·H2O) and the trihydrate (HNO·3H2O).53

Nitrogen oxides are soluble in nitric acid, and this property influences all the physical characteristics that depend on the concentration of the oxides. This mainly includes the vapor pressure above the liquid and the boiling temperature as well as the color. "Red Fuming Nitric Acid" is a term that, according to "Sax's Dangerous Properties of Industrial Materials," applies to nitric acid that contains more than 17 percent NO2 and has a density of more than 1.48 g/cm3.53 White fuming nitric acid usually contains 9099 percent by weight HNO3, from 0 to 2 percent by weight dissolved NO2 and up to 10 percent by weight water.

The TLV for nitric acid has been set at 2 ppm as an 8 hr time-weighted average (TWA) with a short-term exposure limit (STEL) of 4 ppm.

Some properties of nitric acid are given in Tables 29.14 and 29.15.

TABLE 29.14 Physical Properties of Nitric Acid Solutions92,93,94

					Partial press. (at 20°C) Pab			
wt. % HNO3	Density (at 20°C) g/cm3	Freezing point °C	Boiling point. °C	Specific heat (at 20°C) J/(g·K)a	HNO3	H2O	(=cP)	W/m·K)c
0.0	0.99823	0	100.0	4.19		2333	1.0	0.61
10.0	1.0543	-7	101.2	3.73		2266	1.1	0.57
20.0	1.1150	-17	103.4	3.39		2026	1.2	0.54
30.0	1.1800	-36	107.0	3.18		1760	1.4	0.50
40.0	1.2463	-30	112.0	3.01		1440	1.6	0.47
50.0	1.3100	-20	116.4	2.85	27	1053	1.9	0.43
60.0	1.3667	-22	120.4	2.64	120	653	2.0	0.40
70.0	1.4134	-41	121.6	2.43	387	347	2.0	0.36
80.0	1.4521	-39	116.6	2.22	1400	120	1.9	0.35
90.0	1.4826	-60	102.0	1.97	3600	27	1.4	0.31
100.0	1.5129	-42	86.0	1.76	6000	0	0.9	0.28

aTo convert J/(g·K) to cal/(g·°C), divide by 4.184.
bTo convert Pa to atm, divide by 1.013 × 105.
cTo convert W (m·K) to BTU (h·ft·°F), divide by 1.7307.

TABLE 29.15 Thermodynamic Properties of Nitric Acid and Its Hydrates95

	HNO3	HNO3·H2O	HNO3·3H2O
Nitric acid, wt. %	100.0	77.77	53.83
Freezing point, °C	-41.59	-37.62	-18.47
Heat of formation (at 25°C). kJ/mola	-173.35	-472.07	-888.45
Free energy of formation (at 25°C). kJ/mola	-79.97	-329.29	-810.99
Entropy (at 25°C), kJ/(mol·K)a	155.71	217.00	347.17
Heat of fusion, kJ/mola	10.48	17.52	29.12
Heat of vaporization (at 20°C). kJ/mola	39.48		

aTo convert J to cal, divide by 4.184.

Processes

The BirdelandEyde or Arc process made nitric acid by: (1) passing air through an electric arc, (2) forming nitric oxide, (3) oxidizing NO with air, and (4) absorbing the resulting oxides of nitrogen in water. The reactions are:

$$N_2 + O_2 \rightarrow 2NO \qquad (29\text{-}16)$$

$$2NO + O_2 \rightarrow 2NO_2 \qquad (29\text{-}17)$$

$$2NO_2 + H_2O \rightarrow HNO_2 + HNO_3 \qquad (29\text{-}18)$$

$$3HNO_2 \rightarrow HNO_3 + 2NO + H_2O \qquad (29\text{-}19)$$

The process was operated in Norway from 1905 to 1930, and it was first used in the United States in 1917.1 The process used hydroelectric power, but it was made obsolete by the ammonia oxidation process.

The Ostwald process is the basis for the modern processes that make nitric acid by the catalytic oxidation of ammonia. It was discovered in 1900 by Wilhelm Ostwald, a German physical chemist. The process was used by Germany during World War I to make explosives after the Allied blockade cut off the supply of nitrites from Chile and other places.96

Some of the processes that have been based on the Ostwald process are:

1. The Concentrated Nitric Acid (CNA) process is a general name for processes that make nitric acid more concentrated than the 70 percent made in conventional processes.
2. The CONIA process makes nitric acid simultaneously at two different concentrations.
3. The Direct Strong Nitric (DSN) process is a general name for processes that concentrate nitric acid from the 5070 percent range up to the 98 percent range.
4. The HOKO process also makes nitric acid simultaneously at two different concentrations.
5. The Hycon process makes nitric acid at a range of concentrations. It was developed in 1968 and patented (U.S. Patent 3,542,510) by the Chemical Construction Corporation.
6. The Strong Acid By Azeotropic Rectification (SABAR) process makes nitric acid by the atmospheric oxidation of ammonia. Davy McKee developed the process and built plants based on this technology from 1974 to 1986.

Chemistry

In the Ostwald process, ammonia is catalytically oxidized. The key steps in the process are: (1) oxidation of ammonia to nitric oxide (NO), (2) oxidation of NO to nitrogen dioxide (NO2), and (3) absorption of NO2 in water to produce nitric acid. Overall, the principal reactions may be summarized as follows:91,97

Ammonia reacts with air on platinum/rhodium alloy catalysts in the oxidation section of nitric acid plants.

$$4NH_3 + 5O_2 \rightarrow 4NO + 6H_2O$$

$$\Delta H_{298} = -54\,\text{kcal/mol or } -226\,\text{kJ/mol} \qquad (29\text{-}20)$$

Simultaneously N2O, N2, and H2O are formed in side reactions according to Equations 29-21 and 29-22.

$$4NH_3 + 3O_2 \rightarrow 2N_2 + 6H_2O$$

$$\Delta H_{298} = -75.8\,\text{kcal/mol or } -317\,\text{kJ/mol} \qquad (29\text{-}21)$$

$$4NH_3 + 4O_2 \rightarrow 2N_2O + 6H_2O \qquad (29\text{-}22)$$

Some of the process energy recovery options are discussed in EFMA.97 The yield of nitric oxide in reaction (29-20) depends on pressure and temperature as indicated in Table 29.16.97

TABLE 29.16 Effect of Reaction Pressure and Temperature on Nitric Oxide Yield

Pressure (bar)	Temperature (°C)	Nitric Oxide Yield %
Below 1.7	810850	97
1.76.5	850900	96
Above 6.5	900940	95

Source: Reproduced by permission of European Fertilizer Manufacturers Association.

The space velocity in the ammonia oxidation reactor should be kept high to maximize yield and avoid decomposition of ammonia in the feed gas. Another potential yield loss is shown in reaction (29-23).

$$4NH_3 + 3NO_2 \rightarrow \tfrac{7}{2}N_2 + 6H_2O \qquad (29\text{-}23)$$

The effect of reactor residence time on nitric oxide yield is shown in Table 29.17.98

TABLE 29.17 Effect of Reactor Residence Time on Nitric Oxide Yield

Reactor Residence Time, sec	Nitric Oxide Yield, Percent NH3 Oxidized
0.28	82.1
0.11	85.7
0.061	90.2
0.023	91.8

NO is oxidized to NO2 as the combustion gases are cooled.

$$2NO + O_2 \rightarrow 2NO_2$$

$$\Delta H_{298} = -13.6\,\mathrm{kcal/mol\ or\ } -57\,\mathrm{kJ/mol} \qquad (29\text{-}17)$$

$$4NO \rightarrow 2N_2O + O_2 \qquad (29\text{-}24)$$

$$2NO_2 \rightleftarrows N_2O_4$$

$$\Delta H_{298} = -6.8\,\mathrm{kcal/mol\ or\ } -28.6\,\mathrm{kJ/mol} \qquad (29\text{-}25)$$

For the reaction (29-17), secondary air is added to the gas mixture obtained from the ammonia oxidation to increase the oxygen content to such a level that the waste gas leaving the plant has an oxygen content between 2 and 4 percent by volume. The rate of reaction is proportional to pressure cubed (see reaction (29-26)), and the volume or size of equipment needed to oxidize the nitric oxide is inversely proportional to pressure cubed:

$$\mathrm{d}(P_{\mathrm{NO}})/\mathrm{d}t = -k(P_{\mathrm{NO}})^2(P_{\mathrm{O2}}) \qquad (29\text{-}26)$$

The absorption reaction (29-27) is exothermic and continuous cooling is required within the absorber.

$$3NO_2 + H_2O \rightarrow NO + 2HNO_3$$

$$\Delta H_{298} = -6.8\,\mathrm{kcal/mol\ or\ } -28.6\,\mathrm{kJ/mol} \qquad (29\text{-}27)$$

The nitric acid produced in the absorber contains dissolved nitrogen oxides and is then bleached by the secondary air.

The overall reaction is:

$$NH_3 + 2O_2 \rightarrow HNO_3 + H_2O$$

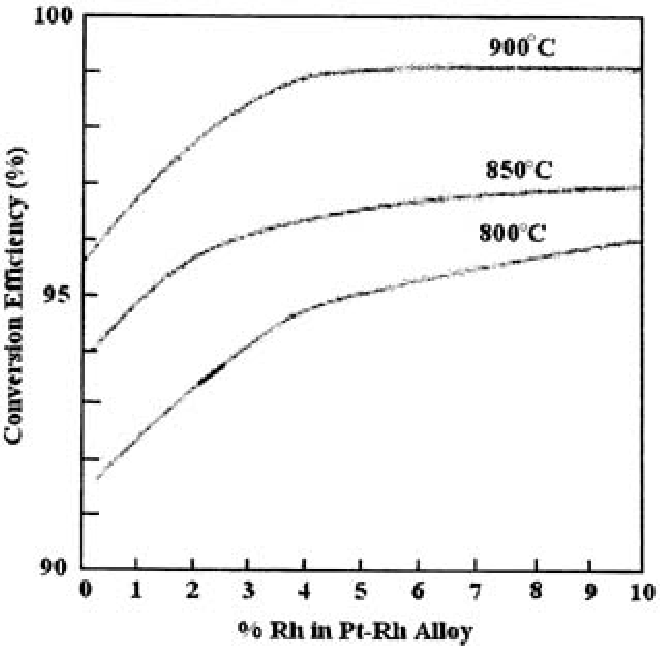
$$\Delta H = -98.7\,\mathrm{kcal/mol} \qquad (29\text{-}28)$$

Several process variations are described in SRI International.91

Catalysts

The catalyst typically consists of several woven or knitted wire gauze sheets with a composition of about 90 percent platinum, 5 percent rhodium, and 5 percent palladium. The primary functions of rhodium are to add mechanical strength to the gauze and to improve yield. The palladium is present to replace a portion of the more expensive rhodium.91 Figure 29.16 shows that conversion efficiency does not improve if more than 5 percent rhodium is present in the catalyst.143

Fig. 29.16 Effect of rhodium content on catalyst efficiency in ammonia oxidation.



Catalyst life depends on several variables. High-pressure oxidation operations (which also operate at higher temperatures, as per Table 29.16) require more frequent catalyst regeneration. The presence of trace amounts of iron, calcium, MoS_2 (lubricants), and phosphorus in the ammonia feed has been shown to have a deleterious effect on catalyst life. Deposits of iron oxides tend to catalyze the conversion of ammonia and oxygen to nitrogen and waterrather than to nitric oxide. These effects, as well as poor ammonia-air mixing and poor gas distribution across the catalyst, may reduce the yield by up to 10 percent.91,97

Some weight loss due to catalyst erosion and vaporization occurs and is more pronounced at the higher converter temperatures that are employed in the higher pressure processes. Lost catalyst is usually recovered by "getter" systems that use palladium alloy gauzes and recover 6080 percent of the lost catalyst. The "getter" gauze is placed as close to the platinum catalyst pack as possible so that the catalyst components can be recovered while they are in the vapor form. Although pure palladium is the most effective "getter," it becomes too brittle so a palladium alloy is used. Additional catalyst is recovered during periodic cleaning of the ammonia oxidation reactor and during cleaning of downstream heat exchangers.91,97

When the catalyst pack no longer provides the desired reaction rates or product composition, new sheets of catalyst may be added to the old sheets. Or the old sheets may be replaced by a set of new sheets, and the old sheets are sent back to the supplier to be refurbished. If any catalyst that is lost from the gauze is trapped in the downstream filter or recovered during equipment cleaning, it is also returned to be reprocessed.91

U.S. Patent 6,073,467 was issued in 2000 for a three-dimensional, knitted, noble metal gauze for nitric acid production. This catalyst design claims to provide elevated product yields, longer catalyst service life with the small noble metal losses.

Operating Pressure

All nitric acid plants are based on the same basic chemical operations: (1) oxidation of ammonia with air to give nitric oxide, (2) oxidation of the nitric oxide to nitrogen dioxide, and (3) absorption in water to give a solution of nitric acid. The efficiency of the first step is favored by low pressure whereas that of the second step is favored by high pressure. Therefore, both single pressure and dual pressure nitric plants are operating.97

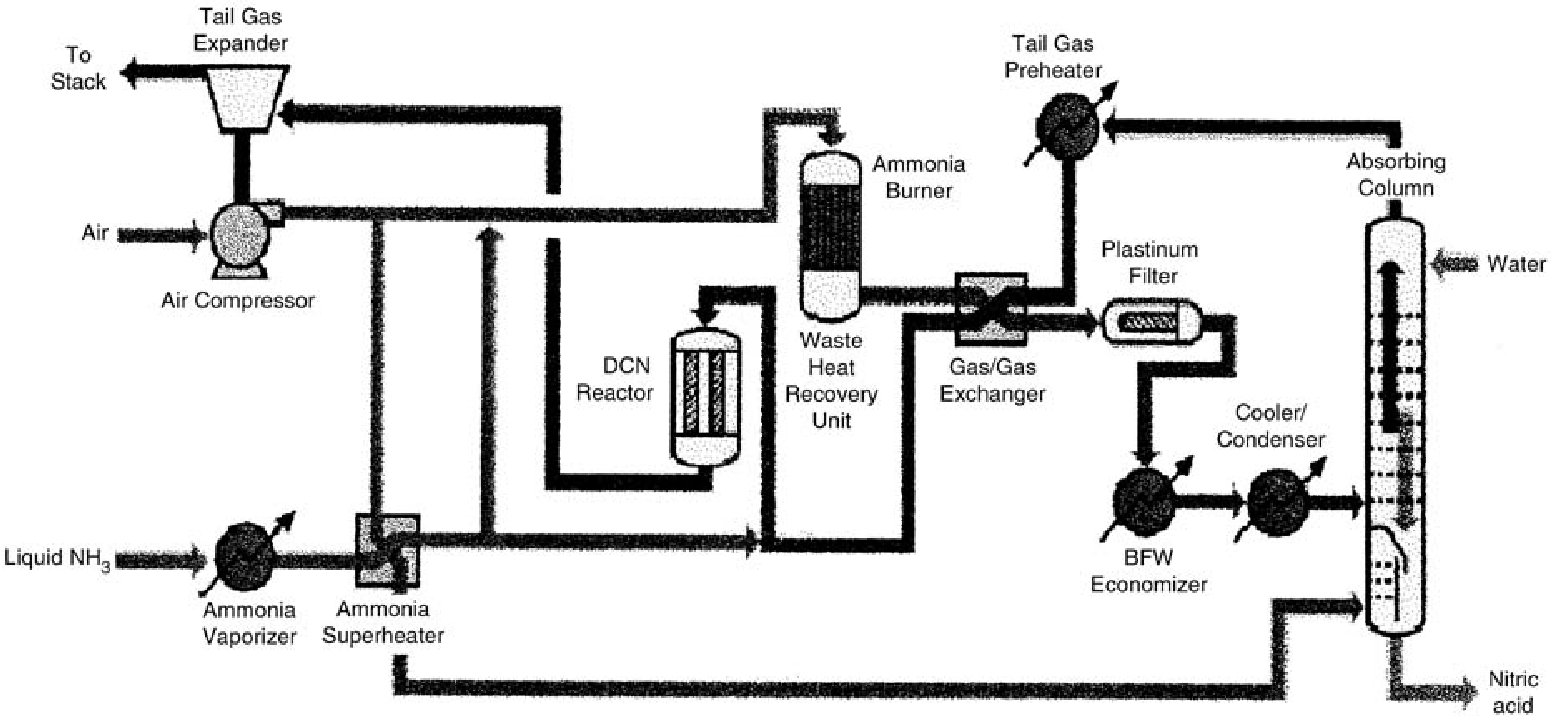
In the single pressure plant, the oxidation and absorption steps take place at essentially the same pressure. In dual pressure plants, absorption takes place at a higher pressure than the oxidation stage. The oxidation and absorption steps can be classified as:97 low pressure (below 1.7 bar), medium pressure (between 1.7 and 6.5 bar), and high pressure (between 6.5 and 13 bar).

The main unit operations in nitric acid plants are:97 (1) ammonia evaporation, (2) ammonia filtration, (3) air filtration, (4) air compression, (5) air/ammonia mixing, (6) catalytic reaction of ammonia and air, (7) energy recovery by steam generation, and/or gas reheating, (8) gas cooling, (9) dual pressure only → NO_x compression, (10) absorption with nitric acid production, and (11) tail gas energy recovery.

Most new plants built in the 1990s have a dual-pressure design. Ammonia oxidation is operated at a medium pressure while absorption is operated at a high pressure. High-pressure systems offer the advantage of lower capital costs (as much as 30% lower than some lower pressure processes) and are preferred in the United States for this reason. Relative disadvantages are higher catalyst loss and lower ammonia conversion. Lower pressure oxidation designs have the best efficiencies in ammonia conversion and catalyst conservation. These plants have been preferred in Europe and other areas where ammonia has been relatively expensive and where allowable payout times have been longer.91

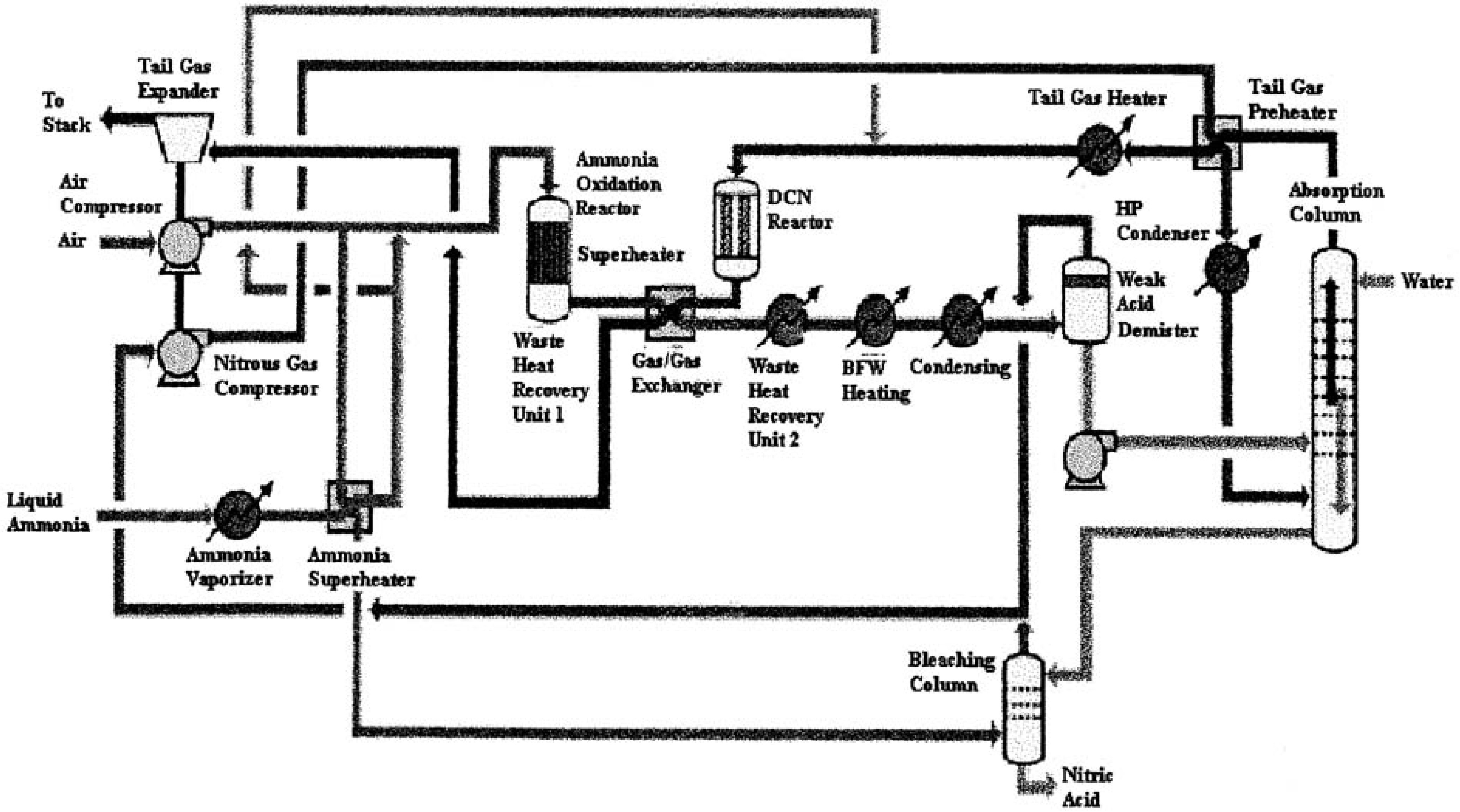
The single pressure design (see Fig. 29.17) is generally used in smaller plants or in larger plants, where minimization of capital is a critical design consideration. Ammonia oxidation and absorption of NO_2 occur at the same relative pressure. This reduces the complexity and capital cost compared to the dual pressure plant. However, efficiency may be lower in both the front and back ends of the plant. Single pressure plants can be designed to operate: (1) at low to medium pressure to favor the ammonia oxidation reaction, (2) at high pressure to favor the absorption reactions, or (3) at any operating pressure in between.99

Fig. 29.17 Single pressure nitric acid plant.99 (*Reproduced by permission of Monsanto Enviro-Chem.*)



The dual pressure design (see Fig. 29.18) is generally used in larger plants, or in mid-size plants, where higher utility/raw material costs dictate a minimization of operating expense. Ammonia oxidation occurs at low or medium pressure. The result is an increase in efficiency of the ammonia oxidation reaction and lower catalyst loss. Absorption of NO_2 occurs at high pressure to maximize the partial pressure of the gas reactants. A nitrous gas compressor boosts the pressure from the front end of the process (ammonia oxidation) to the back end (absorption).91

Fig. 29.18 Dual pressure nitric acid plant.99 (*Reproduced by permission of Monsanto Enviro-Chem.*)



In Table 29.18 the raw material and utility consumption figures are given for both single and dual pressure processes.99

TABLE 29.18 Raw Material and Utility Consumptions in Different Nitric Add Plant Designs

	H&H	M&M	M&H
Operating pressure, bar-a	11	9	4.5/12
Typical run time, days	70	120	210
Ammonia usage,a ton/ton	0.291	0.285	0.279
Platinum usage,b mg/ton	50	35	30
Power usage,c kW-hr/ton	1	2	2
HP steam export,d ton/ton	0.51	0.60	0.54
LP steam import,e ton/ton	0.27	0.21	-0.06
Cooling water,f m3/ton	130	120	110

Notes:

aIncludes catalytic NO*x* abatement to 200 ppmv and average ammonia conversion efficiency over the entire gauze run length.

bDescribed as net of platinum recovery systems.

cAdd 25 kW for HVAC, lighting and instrumentation.

dSteam conditions are 42 kg/cm2 G and 400°C.

eSteam conditions are saturated at 3.5 kg/cm2 G.

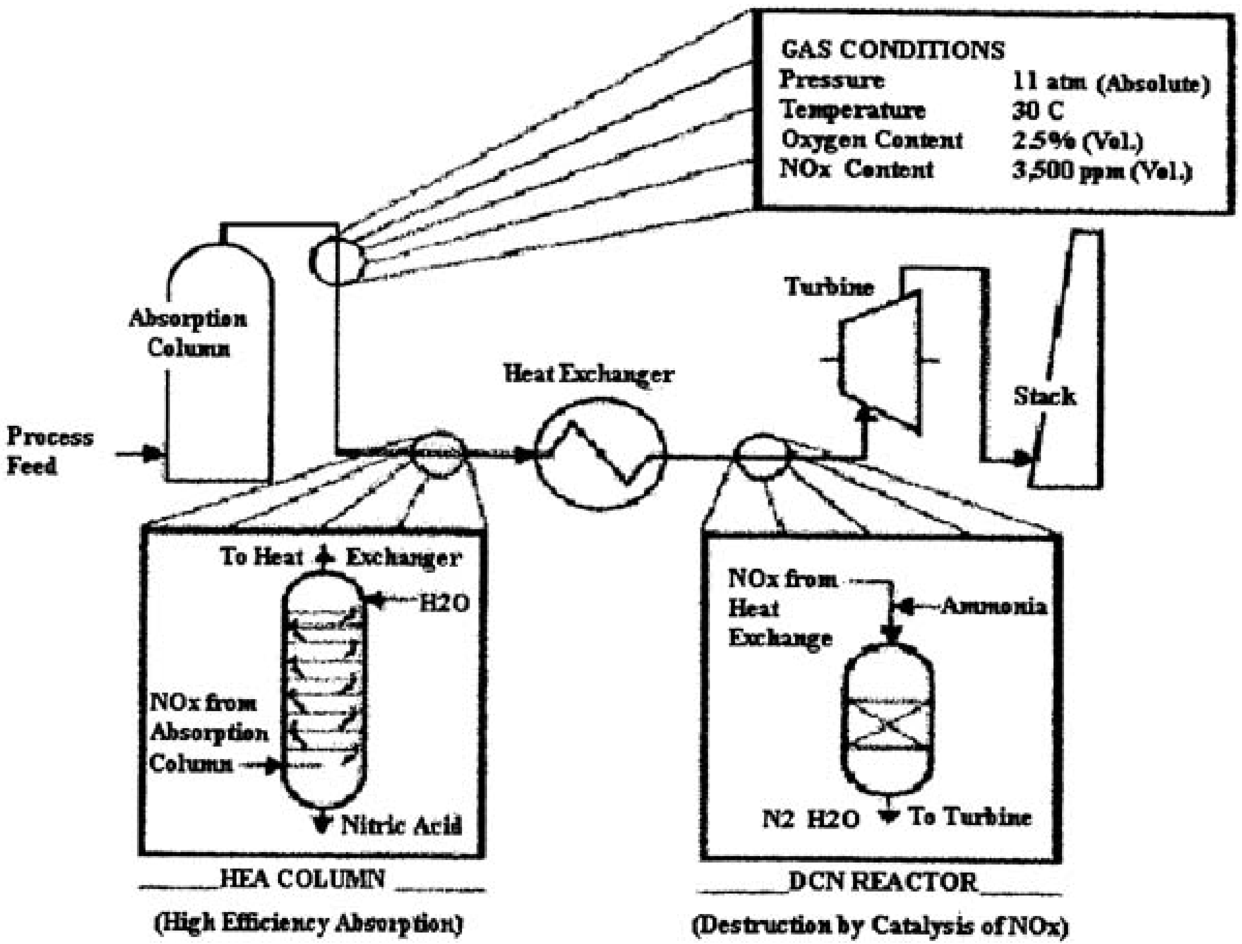
fAt 11°C temperature rise and 2.5 kg/cm2 pressure allowance.

Source: Reproduced by permission of Monsanto Enviro-Chem.

Single Pressure Plant

The single pressure nitric acid plant (see Fig. 29.17) is described in Monsanto Enviro-chem,99 and one type of NO*x* abatement system, the Destruction by Catalysis of NO*x* (DCN) reactor is shown in Fig. 29.19.99

Fig. 29.19 NO*x* abatement system with Destruction by Catalysis of NO*x* (DCN). (*Reproduced by permission of Monsanto Enviro-Chem.*)



Dual Pressure Plant

The dual pressure nitric acid plant is shown in Fig. 29.18, and its use of DCN is also described in Monsanto Enviro-chem.99

Concentration of Nitric Acid

Two processes that make both concentrated and dilute nitric acids have the following designs:

1. Mono-pressure process operating at high pressure
1. ammonia oxidation and NO2 absorption are performed at high pressure
2. Dual-pressure process (middle and high pressure)
1. ammonia oxidation at middle pressure and NO2 absorption is performed at high pressure

The steps in both processes are essentially the same and are described in yet2.com Inc.100

Essentially two basic methods exist to convert nitric oxide into nitric acid. In both, nitric oxide is first oxidized to nitrogen dioxide, which exists in equilibrium with dinitrogen tetroxide (N2O4). In traditional processes, the nitrogen dioxide is absorbed in water to yield nitric oxide (which is recycled) and "weak" acid (5060% HNO3). The "weak" acid can be used directly for the manufacture of ammonium nitrate and other products or further concentrated.91

A process exists that can directly produce concentrated nitric acid from dehydrated and concentrated nitrogen dioxide, weak acid, and oxygen. Generally, the process operates at high pressure and follows these chemical reactions:91

$$2NO_2 \rightleftarrows N_2O_4 \qquad (29\text{-}25)$$

$$N_2O_4 + H_2O + \tfrac{1}{2}O_2 \rightarrow 2HNO_3 \qquad (29\text{-}28)$$

Few U.S. plants employ this method for production of concentrated acid. Most production of concentrated HNO3 (i.e., concentrations above the azeotrope) is based on weak-acid dehydration via distillation with a drying agent. The most common dehydrators have been concentrated sulfuric acid and magnesium nitrite.91

An example of the dehydration process is the Kvaerner Chemetics process that uses concentrated sulphuric acid to eliminate the azeotrope and enable colorless, strong nitric acid up to 99 percent to be distilled from the mixed acids. The residual acid leaving the bottom of the processing tower is approximately 70 percent sulphuric acid.101

Alternative methods of integrated concentration have been developed that do not require separate dehydrating agents. These processes differ from the direct process in that weak acid reacts with concentrated nitrogen dioxide to produce an acid that is sufficiently superazeotropic that distillation into concentrated acid is economically feasible. The weaker azeotropic acid may be recycled for concentration or used as it is.91 The magnesium nitrate process is described in Kirk-Othmer.104

Uhde developed another process for the production of highly concentrated nitric acid (9899%) that uses oxygen but no dehydrating agents. The NO from ammonia oxidation is converted to NO2 by highly concentrated HNO3. The NO2 is cooled with brine, and after removal of the residual water, it is physically absorbed in chilled, highly concentrated nitric acid. Then the NO2 is separated from the concentrated nitric acid in a distillation column and liquefied by refrigeration. The liquid NO2 is dimerized in a N2O4/HNO3/H2O mixture that reacts with supplemental oxygen at a pressure of about 720 psi to form highly concentrated nitric acid. The acid that leaves the reactor contains about 20 percent dissolved N2O4, which is separated in a distillation column and returned to the reactor. The concentrated nitric acid product is bleached and sent to storage, and a portion is recycled to the absorber.

Stabilizers

Over a period of time, concentrated nitric acids tend to decompose according to reaction (29-29) and pressure will build up in storage vessels:

$$HNO_3 \rightarrow 4NO_2 + H_2O + O_2 \qquad (29\text{-}29)$$

Since nitric acid is also very corrosive, some stabilizers and/or corrosion inhibitors are used.

Corrosion of aluminum by red fuming nitric acid is reduced by adding 4 percent by weight of hydrogen fluoride. Decomposition of concentrated acid is reduced by such substances as quaternary ammonium compounds, organic sulfones, inorganic persulfates, and organic sulfonium compounds.102

Pollution Abatement

The primary pollution problem in nitric acid plants is the abatement of NO*x* in tail gases. The processes that have been developed to reduce emissions at existing and new plants can be classified into three general categories: absorption, adsorption, and catalytic reduction.91

Absorption abatement refers to modifications that involve the addition of increased absorption capacity or optimization of the existing absorption system.91

Adsorption abatement uses acid-resistant molecular sieves to absorb the NO*x* from the tail gas. The adsorbant is periodically regenerated and the NO*x* recovered, converted to NO2, and recycled for recovery as nitric acid.91

Catalysts have been developed to permit selective reduction of NO*x* to N2 at relatively low temperatures using ammonia as the reducing agent. Although additional ammonia is consumed in the process, it provides economical control NO*x* emission.91

The main environmental factor that affects nitric acid process selection is the concentration of NO*x* in the tail gas. In the United States, gaseous emissions from newly constructed nitric acid plants are limited to 1.5 kg NO*x* per tonne of nitric acid produced with a maximum opacity of 10 percent. In Western Europe, NO*x* emissions are limited by EU regulations to 200 ppm.91

The gas at the outlet of the absorber may vary within the following limits during stable operation: NO*x* = 1003500 ppmv, N2O =3003500 ppmv, O2 = 1.54 percent by volume, H2O = 0.32 percent by volume, flow = 31003400 Nm3/tonne of HNO3.

The minimum emission levels achieved in 2001 without added pollution abatement are:

1. for medium pressure absorption 10002000 ppmv
2. for high pressure absorption 100200 ppmv

Whereas a dual-pressure plant or a high-pressure, single-pressure plant (with high-pressure absorption) may give acceptable emission levels, the medium-pressure absorption plant must be followed by an abatement system.97

Selective catalytic reduction (SCR) is normally used in new nitric acid plants. In this process, ammonia reacts with nitric oxide and nitrogen dioxide and to a lesser extent with oxygen as shown below:

$$6NO + 4NH_3 \rightarrow 5N_2 + 6H_2O \qquad (29\text{-}30)$$

$$6NO_2 + 8NH_3 \rightarrow 7N_2 + 12H_2O \qquad (29\text{-}31)$$

$$3O_2 + 4NH_3 \rightarrow 2N_2 + 6H_2O \qquad (29\text{-}32)$$

Additional process details can be found in EFMA.97

Monsanto Enviro-Chem offers NO*x* abatement technology that is licensed from Rhodia of France. It includes a high efficiency absorption (HEA) section for extended absorption and a catalytic reduction section (SCR) for catalytic destruction of NO*x* (i.e., the DCN technology).99 Additional process details are given in Monsanto Enviro-Chem.99 The operating conditions for the steps in the Monsanto technology are compared in Table 29.19.

TABLE 29.19 Operating Conditions in NO*x* Abatement Systems

	DCNa Only	HEA Plus DCN	HEAb Only
NO*x* gas to HEA, ppmv	N/A	3500	3500
NO*x* gas to DCN, ppmv	3500	700	N/A
NO*x* gas to turbine, ppmv	200	200	200
Nitric acid recovered, kg/tonHNO3	N/A	27	31
Ammonia consumed, kg/tonHNO3	8.5	1.3	N/A
Gas temperature rise, °C	43	6.5	0
Pressure drop, kg/cm2	0.07	0.2	0.35
Capital cost (ratio)	1.0	2.5	3.5

aDCN: Destruction of catalysis of NO*x*.

bHEA: High efficiency absorption.

Source: Reproduced by permission of Monsanto Enviro-Chem.

Production

An estimated 65 percent of the worldwide nitric acid production is used to make ammonium nitrate (AN) for use in fertilizers. Other AN uses (mainly explosives) account for about 15 percent of the market. The balance is consumed in a variety of non-AN industrial applications. World nitric acid production in 1999 is estimated to be 53 million tonnes per year. Captive use consumes most of the production, and the merchant market probably involves only about 10 percent of the total.91

World demand for nitric acid will continue to be largely dependent upon demand for solid AN fertilizer and nitrogen fertilizer solutions that incorporate AN. A trend toward the use of solid urea rather than solid ammonium nitrate fertilizers has been developing. This has restricted growth in nitric acid production, but compensating growth in liquid fertilizers containing AN is anticipated. World nitric acid production declined by about 5 percent between 1987 and 1999, but it is projected to increase marginally by 2005. The major producing regions are Western Europe, the United States, the FSU, and Eastern Europe. However, production in the FSU and Eastern Europe has dropped by as much as 35 percent since 1987.91

Nitric acid is sold commercially in various concentrations that are expressed in degrees Baumé (Be), which is an arbitrary specific gravity scale. Various grades of concentrated or fuming nitric acid (95%, or more HNO3/NO2) are available.91

Storage and Distribution

Nitric acid is normally stored in flat-bottomed, roofed tanks that are made from low-carbon, austenitic stainless steel. Most concentrations of nitric acid are transported in tank cars and by truck. Stainless steel is necessary for concentrations up to 8085 percent. Stronger solutions are less corrosive and may be stored in aluminum. Approximately, 90 percent of nitric acid production is consumed on site to make AN fertilizers and some industrial explosives. Thus, the merchant market for nitric acid accounts for only 10 percent of the total. The production of AN fertilizers and most industrial explosives do not require acid concentrations in excess of the azeotropic composition of 68.8 percent.91

In the United States, the Department of Transportation (DOT) defines three categories of nitric acid: (1) nonfuming, more than 70 wt. % acid; (2) nonfuming, less than 70 wt. % acid; and (3) red fuming nitric acid. All must be labeled "corrosive."

Nitric acid is subject to self-decomposition, depending on the concentration and temperature. The general rule is: the higher the concentration or the temperature, the faster the decomposition rate. Nitric acid must be transported from the manufacturer to the user within the shortest possible time, particularly in the summer months.97

Uses of Nitric Acid

Some of the major uses for nitric acid are:

1. AN production for fertilizers and explosives
2. *Adipic acid*. Nitric acid oxidizes cyclohexanonecyclohexanol mixtures to adipic acid, nylon 66, other resins, and plasticizers
3. *Nitrobenzene*. Nitrobenzene is made by the direct nitration of benzene with nitric/sulfuric acid mixtures
4. *Toluene diisocyanate (TDI)*. Toluene diisocyanates (2,4- and 2,6-isomers) are produced from toluene diamine derived from dinitrotoluene, which is produced by the nitration of toluene with nitric/sulfuric acid mixtures. TDI is used mainly in polyurethane foams
5. Potassium nitrate production for tobacco fertilizer and industrial markets
6. Sodium nitrate production for fertilizers, explosives, and as a heat transfer medium

Some of the other uses for nitric acid are: other metal nitrates, in the steel industry, in the electronics industry, to make nitrochlorobenzene, cellulose nitrate, in several nitrate paraffins (e.g., nitromethane, nitroethane, and nitropropane), in non-military explosives, to digest crude uranium concentrates, in mixed fertilizers, and in concentrated nitric acid.

29.11 Ammonium Nitrate

AN was first prepared in the 16th century. Its early industrial development was primarily for use in explosives. However, after 1940 its use as a fertilizer developed rapidly. It is made by the reaction between gaseous ammonia and aqueous nitric acid:

$$\mathrm{HNO_3(l) + NH_3(g) \rightarrow NH_4NO_3(aq)}$$

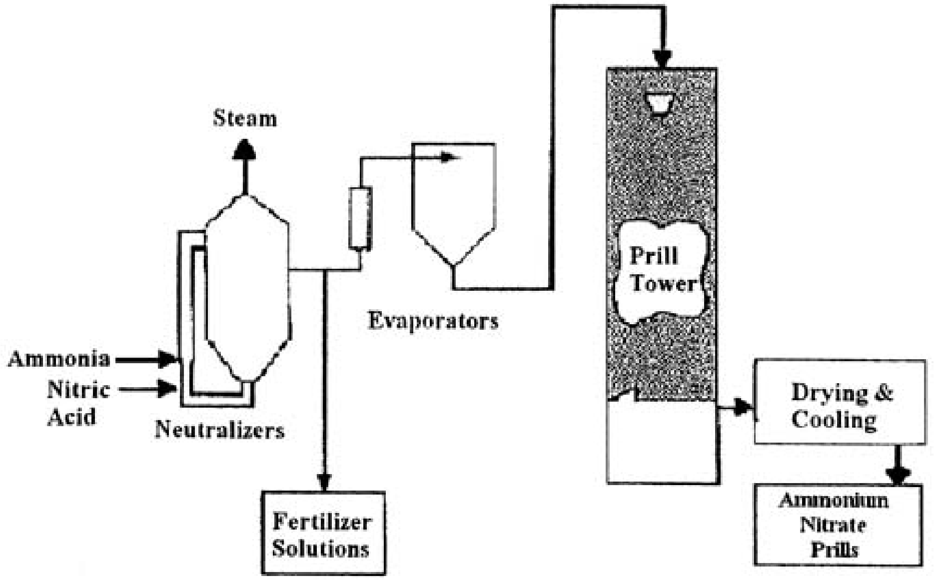
$$\Delta H = +26\,\mathrm{kcal/g\text{-}mol} \qquad (29\text{-}33)$$

The resulting AN solution may be handled in various ways.103 It can be stored as a solution, used in downstream plants or sold as a solution; it can be formed into solid AN by prilling or granulation, or it can be mixed with a solid filler. The most common filler is calcium carbonate in the form of ground limestone, dolomite, or by-product calcium carbonate. This product is know as "Calcium Ammonium Nitrate" (CAN) and can be prilled or granulated.

Processes

As shown in Fig. 29.20, the typical AN production process has three main unit operations: neutralization, evaporation, and solidification (prilling and granulation).103,105 Additional details about AN production are given in *Chemical Week*.107

Fig. 29.20 Ammonium nitrate process sketch.105 (*Reproduced by permission of IMC-Agrico Company*.)



Neutralization

The exothermic neutralization of nitric acid with ammonia gas produces AN solution and steam. The nitric acid is commonly preheated if the available concentration of nitric acid is in the lower end of the 5070 percent range.103

Neutralization can be performed in a single stage or in two stages. A two-stage neutralizer operates with a low pH in the first stage and a neutral pH in the second stage. In most neutralizers, the pressure, temperature, and concentration are linked by the boiling point characteristics of AN solutions with only two of these variables being independent.103

Neutralizers may be free-boiling vessels, circulating systems, or pipe reactors. The water in the nitric acid is evaporated by the heat of reaction (550620 BTU/lb of AN, depending on the acid strength). At least 10 different types and designs of neutralizers are used in Europe. The factors that influence the choice of neutralizer design are described in EFMA.103

Steam Purification

The steam that leaves the neutralizer can be purified, or it can be condensed and then purified. Some steam purification techniques are described in EFMA.103 The steam may be used: (1) in the evaporator; (2) to preheat and evaporate ammonia; or (3) to preheat the nitric acid.103

Evaporation

The evaporator is normally needed to remove water from the AN solution. It must produce a solution with the required concentration at a temperature that avoids crystallization. The acceptable water content is normally below 1 percent for a prilled product and up to 8 percent for the feed to some granulation processes. Evaporators in commercial use include circulatory systems, shell and tube heat exchangers, and falling film-type evaporators.103

Prilling

Prilling refers to the formation of granules by the solidification of droplets of AN. The AN solution is concentrated to 9699 percent at a pH above of 4.5 to form a "melt." The melt may be mixed with an additive or nucleating agent that stabilizes the prills against temperature cycling through the crystal transition phases. Then the melt is transported to the top of the prilling tower and sprayed through a system of fixed orifices or dropped through a rotating, perforated bucket into a rising air stream. The droplets crystallize into hard, spherical "prills" that are dried, cooled, and sized for shipment. If CAN is made, ground calcium carbonate (limestone or dolomite) is added to the melt prior to the formation of the droplets.103,105

Sasol in South Africa produces a porous, prilled ammonium nitrate (PPAN) that finds its widest application in a mix with fuel oil. This mixture is used as an explosive and is commonly known as Ammonium Nitrate Fuel Oil (ANFO). Additional details about PPAN are given in Sasol SMX.106

Granulation

Granulation refers to techniques using processes such as agglomeration, accretion, or crushing to make a granular fertilizer. In contrast to the prilling technique, granulation requires a more complicated plant and a variety of equipment are usedrotating pans and drums, fluidized beds, and other equipment. Granular products can be made in a wider choice of particle sizes than prills.103

The AN is added in the granulator as a spray of hot concentrated solution. No further drying of the granules will normally be required. The granules are screened and the fines and crushed oversize returned to the granulator.103

Typical CAN granulators include drums and pugmills. The calcium carbonate may be mixed with the AN solution before granulation or in the granulator itself. Granules from this process will normally require drying in a fluidized bed or rotary drier. It may not be necessary to add any additional heat when drying CAN as the granules can have sufficient heat to provide the necessary driving force. The granules are screened after the drier.103

Air Abatement Equipment

Emissions from the prilling and granulation sections of AN and CAN plants can be treated by a range of abatement equipment. Particulate material from some granulation plants is relatively coarse in particle size, whereas the prill tower emissions contain very fine particles. Candle filters are normally required for prill tower emissions. They can abate particulate emissions down to 15 mg/m3 of air. For coarser material, dry devices such as bag filters or dry cyclones are used. Particulate emissions can have higher concentrations, perhaps up to a range of 3050 mg/m3, but the recovered material is a solid that can more readily be recycled.

Production

Production of fertilizer grade AN is concentrated mainly in Europe and North America. International trade amounts to about 4 million tonnes (on a nitrogen equivalent basis) per year. The trade is divided equally between concentrated AN (33.534.5% nitrogen) and CAN (2628% nitrogen). The FSU is the largest net exporter with 1.3 million tonnes exported between 1996 and 1998 compared to zero prior to 1990.36 Additional information can be found in Chapter 11 (fertilizers).

Some of the physical and chemical properties of AN are given in Table 29.20.103

TABLE 29.20 Ammonium NitratePhysical and Chemical Properties

Property	
Appearance	White or off-white granules or prills
Odor	Odorless
pH (10 g/100 ml)	>4.5
Melting point	160170°C (depending on moisture content)
Decomposition point	>210°C
Molecular weight	80.05
Solubility in water	
0°C	118.3 g/100 g
20°C	190.0 g/100 g
80°C	576.0 g/100 g
Bulk density	8301100 kg/m3
Nitrogen content	35% (100% ammonium nitrate)

Storage and Distribution

Solid AN in packages must be stored in a general warehouse that has been approved for AN duty. Bulk AN and CAN must be protected from moisture as both products are hygroscopic. Large bulk warehouses may be air-conditioned due to the local climate. Some additives can also reduce the water uptake. The anti-caking agents may be used that are internal to the finished particle or applied as a coating.103

AN solution must be stored at a temperature above the crystallizing temperature of the solution. Gaseous ammonia is normally added in small quantities to maintain the solution at the correct pH because AN solutions lose ammonia during storage.103 Steps should also be taken to avoid contamination by decomposition catalysts such as chlorides and organic materials.

AN decomposes in two ways. Controlled decomposition with careful heating is the commercial process for producing nitrous oxide:

$$NH_4NO_3 \xrightarrow{200\text{–}260°C} N_2O + H_2O \qquad \Delta H = -6.7\ \text{kcal/g-mol} \qquad (29\text{-}34)$$

Explosive decomposition by heating or by explosive shock is the basis for its use as an explosive:

$$2NH_4NO_3 \rightarrow 2N_2 + 4H_2O + O_2 \qquad \Delta H = -28.2\ \text{kcal/g-mol} \qquad (29\text{-}35)$$

The oxygen that is released in reaction (29-35) can be used to more than triple the explosive effects by mixing an organic, such as fuel oil, with AN. This mixture, known as ANFO, is much cheaper and much safer to handle than dynamite:

$$3NH_4NO_3 + (CH_2)_n \rightarrow 3N_2 + 7H_2O + CO_2 \qquad \Delta H = -102.5\ \text{kcal/g-mol} \qquad (29\text{-}36)$$

29.12 Urea

Urea was discovered in 1773, and it was first synthesized from ammonia and cyanic acid in 1828.108 In the early 1890s, urea was produced on an industrial scale by the hydration of cyanamide, which was obtained from calcium cyanamide:109

$$CaCN_2 + H_2O + CO_2 \rightarrow CaCO_3 + CNNH_3 \qquad (29\text{-}37)$$

$$CNNH_3 + H_2O \rightarrow CO(NH_2)_2 \qquad (29\text{-}38)$$

After development of the ammonia process by Haber and Bosch in 1913, the production of urea from ammonia and carbon dioxide developed rapidly. In 2001, urea was prepared on an industrial scale exclusively by this method.109

$$\underset{\text{Ammonia}}{2NH_3} + \underset{\substack{\text{Carbon}\\\text{dioxide}}}{CO_2} \rightleftarrows \underset{\substack{\text{Ammonium}\\\text{carbamate}}}{NH_2COONH_4} \qquad (29\text{-}39)$$

$$\underset{\substack{\text{Ammonium}\\\text{carbamate}}}{NH_2COONH_4} \rightleftarrows \underset{\text{Urea}}{CO(NH_2)_2} + \underset{\text{Water}}{H_2O} \qquad (29\text{-}40)$$

Reaction (29-39) is fast and exothermic and essentially goes to completion under the high pressure reaction conditions. Reaction (29-40) is slower, is endothermic, and does not go to completion. The conversion (on a CO2) basis is usually 5080 percent. The conversion increases with increasing temperature and NH3/CO2 ratio. It decreases with increasing H2O/CO2 ratio.110

When urea melt is subjected to heat, some biuret is formed.

$$\underset{\text{Urea}}{2CO(NH_2)_2} \rightleftarrows \underset{\text{Biuret}}{NH_2CONHCONH_2} + \underset{\text{Ammonia}}{NH_3} \qquad (29\text{-}41)$$

Biuret is undesirable for some industrial and fertilizer applications. Prilled urea normally contains more biuret (0.81.5%) than granular urea because prilled urea goes through an extra evaporation step.

Urea is an odorless, white solid that contains 46.65 percent nitrogen. It is not flammable, but it will melt and decompose in a fire to give off ammonia. When it is dissolved in water, it hydrolyzes very slowly to ammonium carbonate and eventually decomposes to ammonia and carbon dioxide. Some of the other properties of urea are given in Table 29.21.108,110

TABLE 29.21 Physical Properties of Urea108

Chemical formula	NH2CONH2
Molecular weight	60.06
Freezing/melting point	132.6°C
Boiling point	Decomposes
Density, d420	1.3230 g/cm3
Heat of solution in water	+251 J/g (60 cal/g)
Bulk density	0.74 g/cm3
Specific heat, J/kg-K	
0°C	1.439
50°C	1.661
100°C	1.887
150°C	2.109

Properties of Saturated Aqueous Solutions of Urea

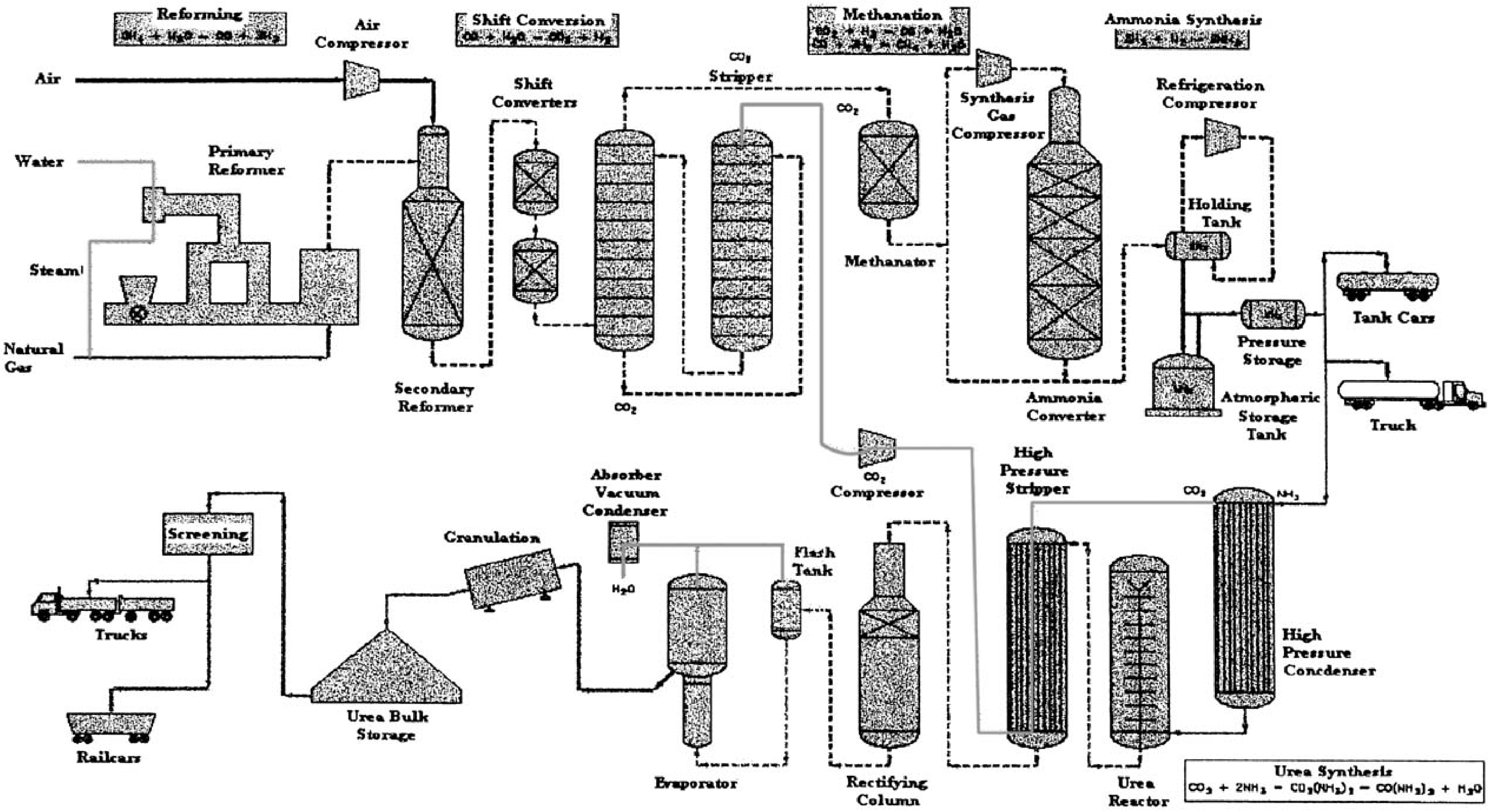
Temperature (°C)	Solubility in Water (g/100 solution)	Density (g/cm3)	Viscosity (mPa-sec = cP)	Water Vapor Pressure (kPa)
0	41.0	1.120	2.63	0.53
20	51.6	1.147	1.96	1.73
40	62.2	1.167	1.72	5.33
60	72.2	1.184	1.72	12.00
80	80.6	1.198	1.93	21.33
100	88.3	1.210	2.35	29.33
120	95.5	1.221	2.93	18.00
130	99.2	1.226	3.25	0.93

Processes

In 2001, most new urea plants are licensed by Snamprogetti (SNAM), Stamicarbon (STAC), or Toyo. SNAM utilizes thermal stripping while STAC and Toyo use CO2 stripping. At the end of 1996, about 70 SNAM plants, 125 STAC, and 7 Toyo plants had been built. STAC will design plants for over 3000 tonnes per day; SNAM designs plants for about 2800 tonnes per day; and Toyo designs plants for about 2300 tonnes per day as single train units.108

Since urea is made from ammonia and carbon dioxide, all urea plants are located adjacent to or in close proximity to an ammonia plant. Figure 29.21111 shows how ammonia and urea plants can be integrated.

Fig. 29.21 Agrium's Carseland nitrogen operations: ammonia and urea production. (*Reproduced by permission of Agrium.*)



The SNAM process, the synthesis section of the STAC process, and the synthesis section of the Toyo-Advanced Process for Cost and Energy Saving (Toyo-ACES) process are described in Kirk-Othmer108 and EFMA.110 An updated version of the STAC process, Urea 2000plus, is offered by DSM (Stamicarbon's parent company).108,110

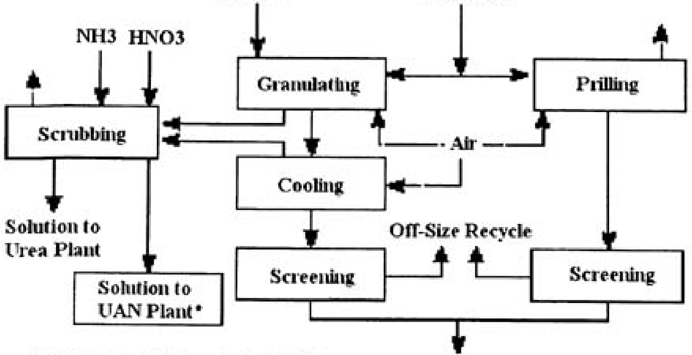
The Isobaric Double-Recycle (IDR) urea process was developed by Montedison. In this process, most of the unconverted material leaving the reactor is separated by heating and stripping at synthesis pressure using two strippers in series. The unconverted ammonia, carbon dioxide and carbamate in the urea solution are recycled to the synthesis loop.108,110

The Heat Recycle Urea Process (HRUP) was developed by Urea Technologies in the 1970s and is now offered by Monsanto Enviro-Chem. This process is described in EFMA110 and Monsanto Enviro-Chem.112

Prilling

A prilling plant is show in Fig. 29.22 and is described in Kirk-Othmer108 and EFMA.110

Fig. 29.22 Block diagram for urea granulation and prilling processes.110 (*Reproduced by permission of European Fertilizer Manufacturers Association.*)



Granulation

Almost all new plants make granules and the Hydro-Agri process, which can be built to make over 3,000 tonnes per day, is used in the majority of plants. The basic principle of the process involves the spraying of the melt onto recycled seed particles or prills circulating in the granulator. The process is described in more detail in Kirk-Othmer,108 Meessen and Petersen,109 and EFMA.110

Another process is the C&I Girdler drum system, but it cannot compete in today's market because of capacity restrictions. Toyo has developed a spouting-fluid bed technology and by 1996 three plants were in operation. STAC will also license a fluid-bed plant that is similar to the Hydro-Agri design.108

Production

The ammonia cost makes up 62 percent to over 70 percent of the cash production costs of urea as the natural gas cost varies from $2.00 per million BTU up to $6.00 per million BTU. In general, urea cash production costs in dollars per tonne are a factor of 1.41.6 times the ammonia price in US dollars per tonne.113

Storage and Distribution

Urea may be supplied as a solid or in a solution, and solid urea is classified as granular or prilled products. Prilled products have less desirable qualities than granules and are usually marginally cheaper than granulated product.109

The majority of urea is designated as fertilizer grade. Technical grade does not contain any additives. Low-biuret grade contains less than 0.3 percent biuret and is used on citrus crops. Feed grade is fed directly to cattle. It is free of additives and is supplied as microprills with a diameter of about 0.5 mm. Slow-release grades use coatings and additives to increase the amount of nitrogen absorbed by crops. Urea supergranules have diameters up to 15 mm and are used in wetland rice and forest fertilization.109

The shift from bagged to bulk transport and storage of prilled and granulated urea has called for warehouse designs in which large quantities of urea can be stored in bulk. Caking and subsequent product degradation at unloading are the result of water absorption. To avoid these problems, the warehouse should be airtight and thoroughly insulated. The caking can also be reduced by adding small amounts of formaldehyde (up to 0.6 wt. %) to the urea melt or by adding surfactants to the solid product.109

Uses

Urea is used in:113 liquid and solid fertilizers, ureaformaldehyde resins that make adhesives and binders mostly for wood products, livestock feeds, melamine for resins, NO*x* control from boilers and furnaces, and numerous chemical applications.

< previous page

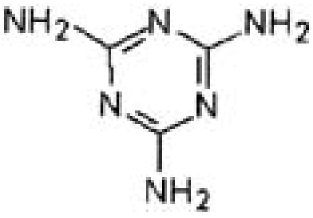
next page >

29.13 Melamine

Melamine is characterized by its three amino (-NH2) groups and by its 66.6 percent nitrogen content. Its chemical formula can be written as $C_3H_6N_6$ or $C_3N_3(NH_2)_3$. It was first prepared in 1834, but it was not until the 1930s that it received commercial attention.

Some of the properties of melamine are listed in Table 29.22.114

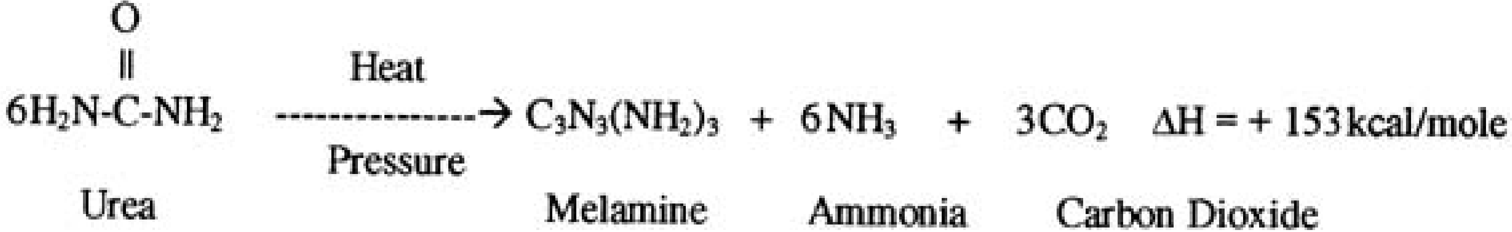
TABLE 29.22 Properties of Melamine

Appearance	Fine white crystalline powder
Molecular weight	126.13
Specific density, g/cc	1.573
Melting point	354°C
Boiling point	Decomposes above 300°C and releases ammonia vapors
Bulk density	750 kg/m3
Toxicity	Mild/low
Stability	Stable
Solubility	Insoluble in most inorganic compounds Very soluble in water
Melamine structure	

Processes

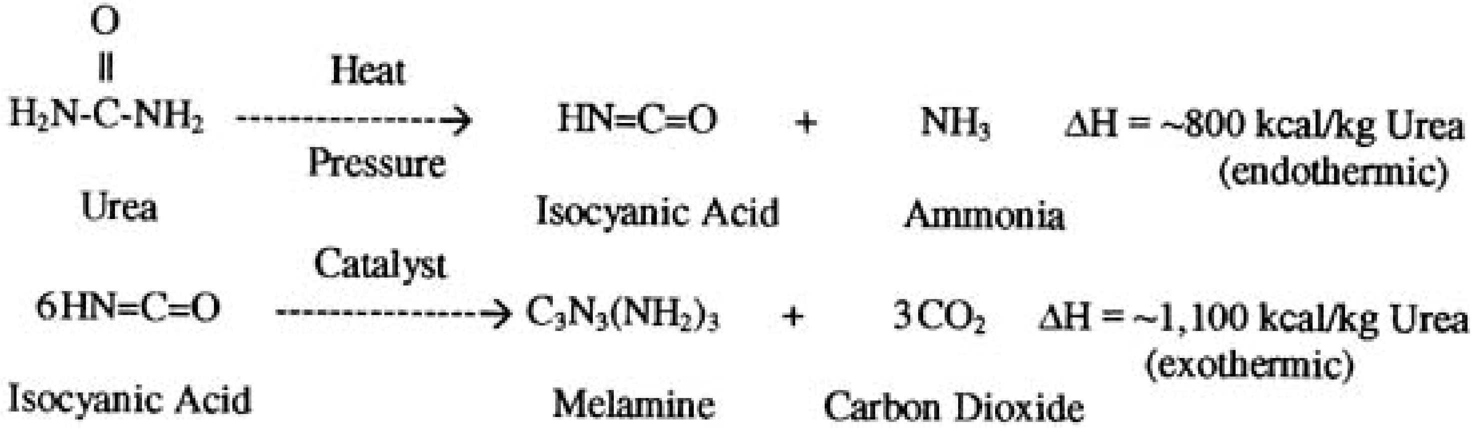
Melamine can be produced from urea, dicyandiamide, or hydrogen cyanide, but in 2001 all commercial-scale production is based on urea.114

Melamine is produced from urea by either a high- or a low-pressure process, and either process can consist of one or two stages. The net reaction is:114



$$\underset{\text{Urea}}{6H_2N\text{-}\overset{\overset{O}{\parallel}}{C}\text{-}NH_2} \xrightarrow[\text{Pressure}]{\text{Heat}} \underset{\text{Melamine}}{C_3N_3(NH_2)_3} + \underset{\text{Ammonia}}{6NH_3} + \underset{\text{Carbon Dioxide}}{3CO_2} \quad \Delta H = +153\,\text{kcal/mole}$$

The reactions in the two-stage process are shown in reactions (29-43) and (29-44). The products are quenched with water or an aqueous mother liquor. This stream is then processed by filtration, in centrifuges or by crystallization.114



$$\underset{\text{Urea}}{H_2N\text{-}\overset{\overset{O}{\parallel}}{C}\text{-}NH_2} \xrightarrow[\text{Pressure}]{\text{Heat}} \underset{\text{Isocyanic Acid}}{HN{=}C{=}O} + \underset{\text{Ammonia}}{NH_3} \quad \Delta H = \sim 800\,\text{kcal/kg Urea (endothermic)}$$

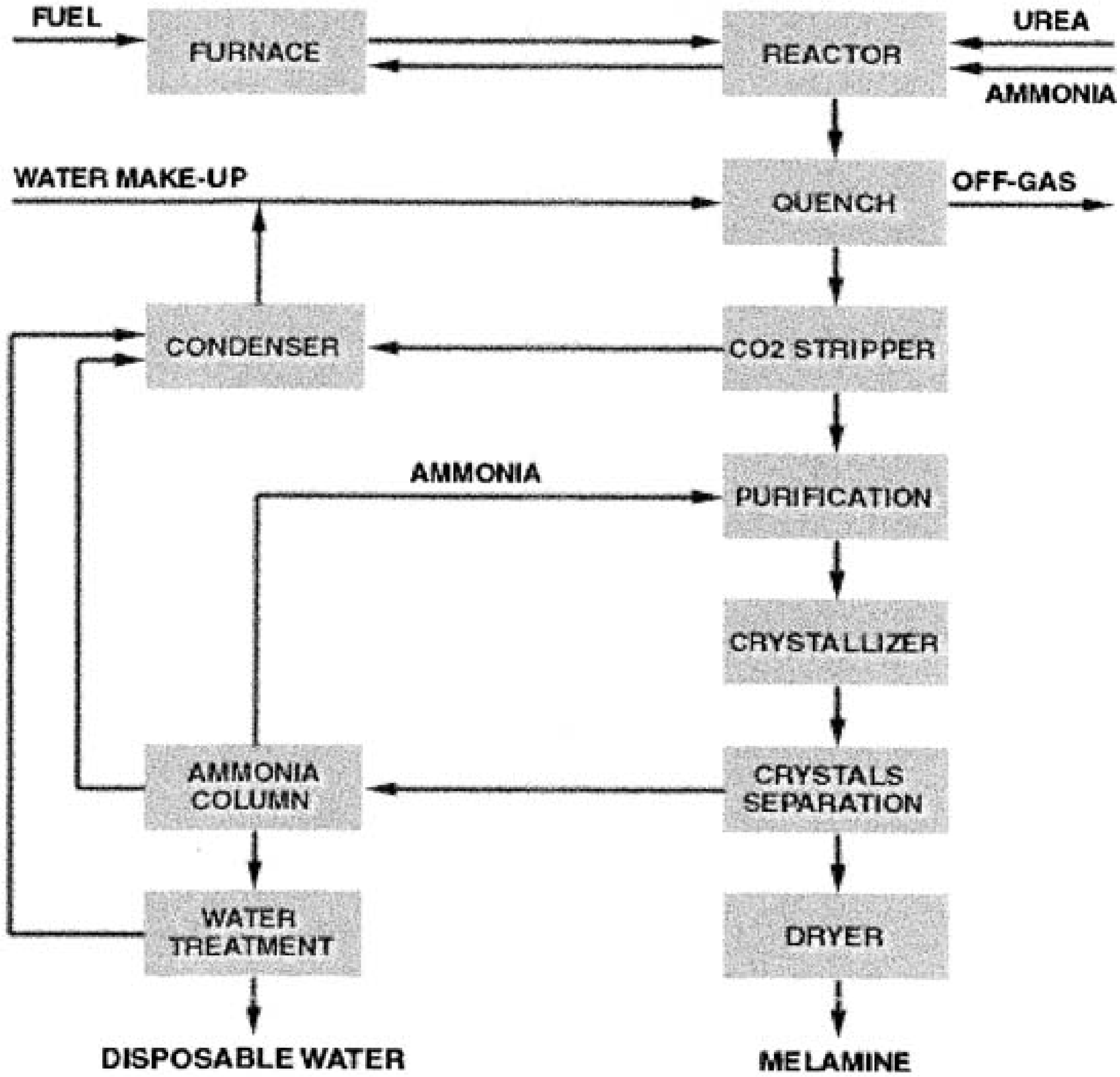
$$\underset{\text{Isocyanic Acid}}{6HN{=}C{=}O} \xrightarrow{\text{Catalyst}} \underset{\text{Melamine}}{C_3N_3(NH_2)_3} + \underset{\text{Carbon Dioxide}}{3CO_2} \quad \Delta H = \sim 1{,}100\,\text{kcal/kg Urea (exothermic)}$$

In a typical high-pressure process such as the one licensed by Eurotecnica (see Fig. 29.23 and www.eurotecnica.it/melamine.htm), the reaction is carried out in the liquid phase (without a catalyst) and at 90150 bar and 380450°C. Under these conditions, urea forms cyanuric acid which then reacts with ammonia to form melamine. The first step is the same as reaction (43) while the second step is shown in reactions (45) and (46).46,114

$$\underset{\text{Isocyanic acid}}{3HN{=}C{=}O} \rightarrow \underset{\text{Cyanuric acid}}{C_3N_3(OH)_3} \qquad (29\text{-}45)$$

$$\underset{\text{Cyanuric acid}}{C_3N_3(OH)_3} + \underset{\text{Ammonia}}{3NH_3} \rightarrow \underset{\text{Melamine}}{C_3N_3(NH_2)_3} + 3H_2O \qquad (29\text{-}46)$$

Fig. 29.23 Typical melamine process.



In a typical low-pressure process, the reaction is carried out in the vapor phase at 110 bar and 350400°C. This process uses a catalyst such as modified aluminum oxide or aluminosilicate. Under these conditions, urea forms isocyanic acid as shown in reaction (29-43). On the catalyst, the isocyanic acid is converted to cyanamide or carbodiimide, which is subsequently converted to melamine (see reactions (29-47) and (29-48).46,114

$$\underset{\text{Isocyanic acid}}{2HN{=}C{=}O} \rightarrow (\underset{\text{Carbodiimide}}{HN{=}C{=}NH} \text{ or } \underset{\text{Cyanamide}}{H_2N{=}C - N}) + \underset{\text{Carbon dioxide}}{CO_2} \tag{29-47}$$

$$(\underset{\text{Carbodiimide}}{HN{=}C{=}NH} \text{ or } \underset{\text{Cyanamide}}{H_2N{=}C - N}) \rightarrow \underset{\text{Melamine}}{C_3N_3(NH_2)_3} + \underset{\text{Carbon dioxide}}{CO_2} \tag{29-48}$$

Carbon dioxide and ammonia are obtained as by-products in both the low- and high-pressure processes and are usually recycled to the urea process to improve process economics. With recycling, about 1.5 lb of urea are used to produce one pound of melamine which corresponds to a yield of 95 percent (compared to a yield of only 49 percent without recycle).114

DSM Melamine planned to start up a new 30,000 tonne per year melamine plant toward the end of 2002. This plant will use the high-pressure shortened liquid phase (SLP) process developed by DSM. The SLP process is expected to increase efficiency by 25 percent and have the same costs as a 100,000 tonne/year production facility. The new process requires only 3 or 4 processing steps, in contrast to the 10 steps in conventional processes.115

The major melamine process technologies and the suppliers of these technologies are listed in SRI International.114

Production

The world supply/demand for melamine is summarized in SRI International.114 In 1998, Western Europe accounted for 35 percent of world capacity, followed by Asia (excluding Japan) with 24 percent, and both the United States and Japan with 16 percent each.114

Since the 1980s, melamine list price in the United States has ranged from the low $0.40s to $0.70/lb. In 1998, the price was about $0.50/lb.114

Uses

Melamine is supplied in bags, carlots, and truckloads. Some of the melamine uses are:114 surface coatings, laminates, paper treating, molding compounds, textile treating, wood adhesives, and other uses such as ceiling tiles, tire cord, flame retardants, and many other small-volume applications.

< previous page

next page >

29.14 Aliphatic Amines

Amines are derivatives of ammonia in which the hydrogen atoms in the ammonia have been displaced by monovalent hydrocarbon radicals. Depending on the number of hydrogen atoms displaced, amines are classified as primary (RNH2), secondary (R2NH), or tertiary (R3H). The amines can be categorized further according to the type of hydrocarbon radicals that substitute for the hydrogen atoms: aliphatic (saturated or unsaturated), aromatic, heterocyclic, alicyclic, or any combination of these.

One group of compounds that are similar to amines are imines. Imines contain an ammonia molecule in which two hydrogen atoms are displaced by bivalent hydrocarbon radicals (R=NH). Another group of compounds that are similar to amines are nitriles. In nitriles all the hydrogen atoms in ammonia are displaced by a trivalent hydrocarbon radical (RC≡H).

Methylamines

Methylamines are colorless liquids that are volatile at normal atmospheric conditions. They have threshold odor limits of less than 10 ppm, and at low concentrations they have a fishy smell. At high concentrations they smell like ammonia. The physical properties are given in Tables 29.23 and 29.24.

TABLE 29.23 Physical Properties of Anhydrous Methylamines

	MMA	DMA	TMA
Chemical formula	CH3NH2	(CH3)2NH	(CH3)3N
Molecular weight	31.06	45.08	59.11
Freezing point, °C	-93.5	-92.02	-117.3
Boiling point, °C	-6.3	6.9	2.9
Vapor density (1 atm, 25°C) g/L	1.3	1.9	2.5
Specific gravity, air = 1	1.1	1.6	2.0
Liquid density (25°C), g/ml	0.6562	0.6496	0.6270
lb/gal	5.48	5.42	5.23
Liquid viscosity (25°C), cP		0.190	0.175
Critical temperature, °C	156.9	164.5	160.1
Critical pressure, atm	73.6	52.4	40.2
Enthalpy of formation, -H°f (25°C), kcal/mol			
Gas	-5.49	-4.41	-5.81
Liquid	-11.3	-10.5	-11.0
Solution (1000 moles H2O)	-16.78	-17.3	-18.6
Vapor pressure, psia			
25°C	50	30	32
100°C	400	230	205
Heat capacity, Cp (25°C), cal,/°C·mol			
Gas	12.7	16.9	21.9
Liquid		32.9	32.31
Heat of fusion (m.p.), cal/g	47.20	31.50	26.46
Heat of vaporization (b.p.), cal/g	198.6	140.4	92.7
Autoignition temperature, °C	430	400	190
Flammable limits, vol. %			
Lower	4.9	2.8	2.0
Upper	20.7	14.4	11.6

Notes: MMA: Monomethylamine; DMA: Dimethylamine; TMA: Trimethylamine.
Source: DuPont Methylamines Bulletin.

TABLE 29.24 Physical Properties of Aqueous Methylamines

Properties	40% MMA	40% DMA	60% DMA	25% TMA
Boiling point, °C	48	54	36	43
Freezing point, °C	-38	-37	-74.5	6
Liquid density (25°C)				
g/ml	0.897	0.892	0.829	0.930
lb/gal	7.49	7.44	6.92	7.76
Vapor pressure (25°C)				
psia	5.8	4.2	9.7	6.6
mmHg	300	215	500	340
kPa	40	29	67	45
Flash point, closed cup, °C	-12	-18	-52	6

Conventional Processes

The reaction of an alcohol with ammonia provides the most common commercial routes to alkylamines. Capacities of many plants depend on the product mix of mono/di/tri products as well as the variety of amines (ethyl, propyl, and butyl).116

In alcohol amination, methanol and excess ammonia react at 350500°C and 1530 bar in the presence of aluminum oxide, silicate, or phosphate catalysts according to the following reactions:46

$$NH_3 + CH_3OH \xrightarrow{Cat} CH_3NH_2 + H_2O \quad \Delta H = -5.358\ \text{kcal/mol} \tag{29-49}$$

$$CH_3NH_2 + CH_3OH \xrightarrow{Cat} (CH_3)_2NH + H_2O \quad \Delta H = -9.598\ \text{kcal/mol} \tag{29-50}$$

$$(CH_3)_2NH + CH_3OH \xrightarrow{Cat} (CH_3)_3N + H_2O \quad \Delta H = -14.098\ \text{kcal/mol} \tag{29-51}$$

$$NH_3 + (CH_3)_3N \xrightarrow{Cat} CH_3NH_2 + (CH_3)_2NH \quad \Delta H = +8.740\ \text{kcal/mol} \tag{29-52}$$

$$NH_3 + (CH_3)_2NH \xrightarrow{Cat} 2CH_3NH_2 \quad \Delta H = +4.240\ \text{kcal/mol} \tag{29-53}$$

$$CH_3NH_2 + (CH_3)_3N \xrightarrow{Cat} 2(CH_3)_2NH \quad \Delta H = +4.500\ \text{kcal/mol} \tag{29-54}$$

Reactions (29-49)(29-51) are known as the alkylation reactions. They are exothermic and highly irreversible, except for reaction (29-51). Reactions (29-52)(29-54) are known as disproportionation reactions. They are reversible and are endothermic. The alkylation reactions dictate the rate of consumption of methanol and are somewhat faster than the disproportionation rates that govern the selectivity of the three amines. Guidelines for optimizing the reactions are discussed in Weissermel and Arpe46 and SRI International.116

Figure 29.24 shows a typical methylamines process diagram.117 The raw material and utility use are summarized in Table 29.25.17

Fig. 29.24 Typical methylamine process.

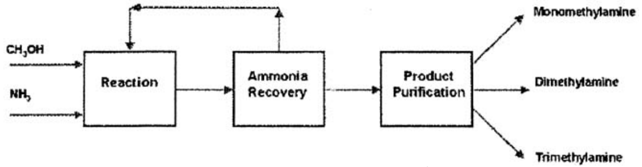


TABLE 29.25 Methylamine Production: Raw Material and Utility Use (Tonne per Tonne of Product Leaving the Reactor)

	MMA	DMA	TMA
Methanol	1.053	1.451	1.660
Ammonia	0.559	0.385	0.294
Steam	3.0		
Cooling water, m3	150		
Electricity, kWh	150		

Source: Hydrocarbon Processing, 1991, 70(3), 132136. Reproduced by permission of Hydrocarbon Processing.

Although companies can produce a range of C2 to C6 amines on a campaign basis in a higher alkylamine plant, methylamine plants are usually designed and dedicated solely for production of a given chain-length product. The methylamine plants also use a different catalyst from that in a higher amines plant.116

Mitsubishi Rayon has operated a process since 1984 in which the equilibrium of methylamine formation is shifted to make more dimethylamine by use of an acid zeolite catalyst. The product stream contains 7 mol. % MMA, 86 mol. % DMA, and 7 mol. % TMA, and the investment as well as the operating costs are lower than the conventional Leonard process that is used by most companies. Since DMA has the highest sales volume, the process might be appealing.

Aldehyde-Amine and Ketamine Hydrogenation Process

Imines can be produced by the addition of ammonia to a carbonyl compound, and the imines are hydrogenated to make amines as per reaction (29-55):116

$$\begin{aligned} RCHO + NH_3 &\rightarrow RCH{=}NH \qquad (29\text{-}55) \\ &\overset{H_2}{\rightarrow} RCH_2NH_2 \end{aligned}$$

Nitrile Reduction

The choice of catalyst and variations of the temperature, pressure, and concentration of ammonia can control the product mix of amines from hydrogenation of a nitrile. The reaction is shown in (29-56):

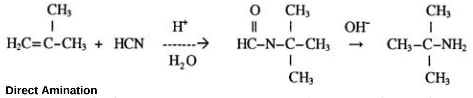
$$\begin{aligned} RC - N + H_2 &\rightarrow RCH_2NH_2 + (RCH_2)_2NH_2 \\ &+ (RCH_2)_3N \qquad (29\text{-}56) \end{aligned}$$

A large excess of ammonia suppresses formation of di- and tri-alkylamines. SRI International116 describes the process in more detail.

Other Alkyl Amines

Ritter Reaction

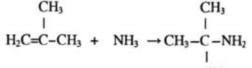
Mitsubishi Rayon and Sumitomo Chemical in Japan make *t*-butylamine reaction (29-57). Average yields are about 85 percent based on isobutylene. Rohm and Haas in the United States makes *t*-octylamine from hydrogen cyanide and diisobutylene by a similar route.116



$$H_2C{=}\overset{\overset{\displaystyle CH_3}{|}}{C}{-}CH_3 + HCN \xrightarrow[H_2O]{H^+} H\overset{\overset{\displaystyle O}{\|}}{C}{-}N{-}\underset{\underset{\displaystyle CH_3}{|}}{\overset{\overset{\displaystyle CH_3}{|}}{C}}{-}CH_3 \xrightarrow{OH^-} CH_3{-}\underset{\underset{\displaystyle CH_3}{|}}{\overset{\overset{\displaystyle CH_3}{|}}{C}}{-}NH_2$$

Direct Amination

Direct amination of olefins is a special case and has been successfully developed for only a few amines. BASF has developed a process for the direct amination of isobutylene to yield *t*-butylamine (see reaction (29-58)), and has operated a commercial plant in Western Europe since 1993.116



$$H_2C{=}\overset{\overset{\displaystyle CH_3}{|}}{C}{-}CH_3 + NH_3 \rightarrow CH_3{-}\underset{\underset{\displaystyle CH_3}{|}}{\overset{\overset{\displaystyle CH_3}{|}}{C}}{-}NH_2$$

Hydrogenation of Anilines

A manufacturing process that is specific for cyclohexylamines is the catalytic hydrogenation of anilines or phenols in the presence of ammonia. Process details are given in SRI International.116

Manufacturing Costs

Manufacturing costs for alkylamines depend on the specific amine and the process. SRI International116 gives the estimated production costs for two different processes.

Production

Methylamines are commonly sold as anhydrous gases or as 25, 40, 50, and 60 percent aqueous solutions. The other alkylamines are usually sold on a 100 percent basis although some are also marketed as aqueous solutions. In 1997, MMA and TMA prices increased to $0.73 per lb and DMA went up to $0.77 per lb. Since that time, market prices have declined somewhat as methanol prices have receded.116 From 1995 to 2001, the price of cyclohexylamine has varied from a low of $1.40 per lb to a high of $1.52 per lb.118

Methylamine growth should remain strong as applications in the largest market segments are growing two to three times quicker than GNP.119 However, worldwide growth in alkylamines should average a modest 23 percent annually from 2000 through 2004.116

The major world producers of methylamines include BASF, Air Products, DuPont, and UCB. The major producers of C2 to C6 alkylamines include Air Products, BASF, and Celanese.116

The largest consumer of alkylamines is the United Statesfollowed by Western Europe (at •70% of the U.S. consumption) and Japan (at •25% of the U.S. consumption).116

Uses

The three methylamines are important intermediates for the manufacture of solvents, insecticides, herbicides, pharmaceuticals, and detergents. DMA is the most important due to its use in the manufacture of N, N-dimethylformamide and N, N-dimethylacetamide, which find wide application as solvents.46,116

MMA ranks second in terms of demand. It is mainly used to make dimethyl urea and N-methylpyrrolidone, as well as for methyltaurine, which is employed in CO2 washes or as a raw material for detergents.46,116

TMA plays only a minor role. It is used in the manufacture of choline chloride, quaternary salts, and epichlorohydrin.46,116

Some of the other alkyl amines and their uses are:

1. ethylamines: pesticides and catalysts
2. propylamines: herbicides, detergents, and catalysts
3. butylamines: pesticides, plasticizers, insecticides, catalysts, and herbicides
4. cyclohexylamines: corrosion inhibitors, catalysts, antibiotics, herbicides, and cyclamate sweeteners in Asia

29.15 Ethanolamines and Secondary Products

The main use of ethanolamines is for the manufacture of detergents by reaction with fatty acids. The principal secondary products of ethanolamines are morpholine, ethylenimine, and ethylenediamine.

Ethanolamine Process

Ethylene oxide reacts exothermically with 2030 percent aqueous ammonia at 60150°C and 30150 bar in a tubular reactor to form the three possible ethanolamines (MEA, diethanolamine (DEA), and triethanolamine (TEA)) with high selectivity.

$$\underset{\text{Ethylene Oxide}}{H_2C{-}CH_2 \text{ (O)}} + \underset{\text{Ammonia}}{NH_3} \rightarrow \underset{\text{MEA}}{H_2NC_2H_4OH} + \underset{\text{DEA}}{HN(C_2H_4OH)_2} + \underset{\text{TEA}}{N(C_2H_4OH)_3}$$

The composition of the reaction product stream can be influenced by temperature and pressure. Higher ratios of excess ammonia lead to a higher MEA content as shown in Table 29.26. The high TEA content that results from an equimolar mixture of reactants indicates that the primary reaction with ammonia is slower than the secondary reactions.46

TABLE 29.26 Effect of Excess Ammonia on Ethanolamine Selectivity

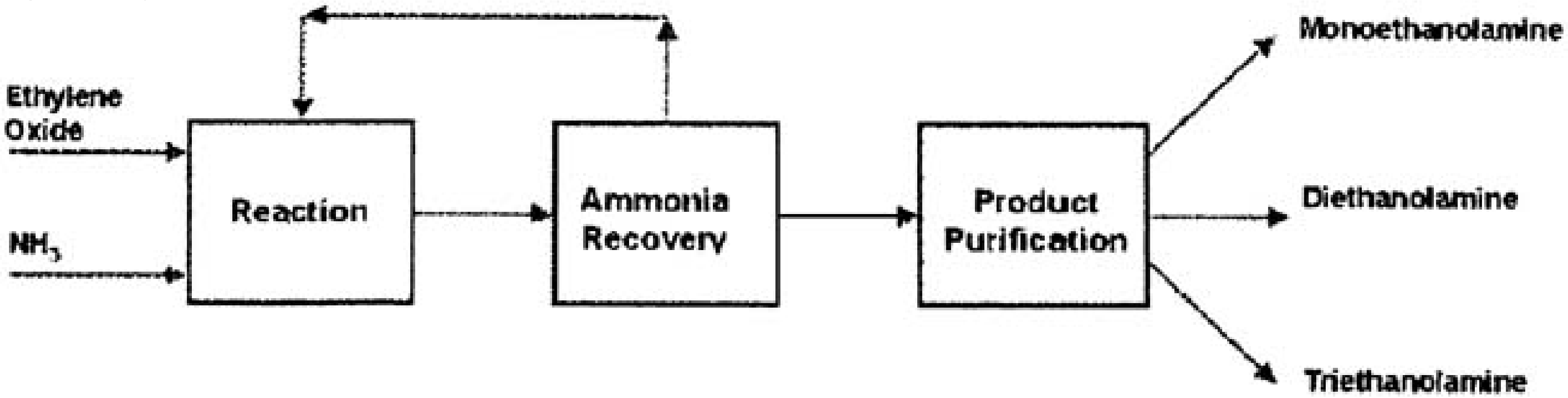
	Selectivity Ratios		
Molar Ratio (NH3 to EO)	MEA	MEA	TEA
10:1	75	21	4
1:1	12	23	65

Notes: MEA = monoethanolamine; DEA = diethanolamine; TEA = triethanolamine.

Source: Weissermel, K., and Arpe, H.-J., *Industrial Organic Chemistry*, 3rd ed., VCH Publishers, New York, 1997. Reproduced by permission of VCH Publishers.

A typical ethanolamine flowsheet is shown in Fig. 29.25.117 Nippon Shokubai has developed technology that uses a zeolite catalyst that suppresses the formation of TEA and produces more MEA and DEA.120 Another process flowsheet is described in Garcia et al.121 along with detailed process conditions.

Fig. 29.25 Typical ethanolamine process.117



Himtek Engineering in St. Petersburg, Russia offers ethanolamine (EOA) technology that operates at 4070°C and 1535 atm. The ammonia and part of the MEA are recycled. Three vacuum stills are used to separate the three products. This technology is described in Russian Patents No. 2,141,475 and No. 2,063,955.122

Production

All EOA plants make a mixture of products. The most important is MEA, which accounts for about half of total production. It is followed by DEA with 3035 percent of production. Dow is the largest global producer followed by BASF, Huntsman, Ineos. EOA demand is growing between 4 and 6 percent annually with MEA and DEA growth rates substantially exceeding that for TEA. In 2001, EOA supply exceeded demand and the oversupply situation continues through 2002 with startups of more new plants. About 50 percent of EOA capacity is in the Americas, more than 30 percent is in Europe, and the balance is in the Pacific Rim and the Middle East. Table 29.27 summarizes world consumption and growth rates for EOA.38,120

TABLE 29.27 World Consumption of Ethanolamines and Growth Rates120 Consumption (Thousands of Tonnes)

Region	1995	2000	2005
Americas	279	358	418
Western Europe	260	329	385
Asia	99	157	197
Other	27	60	90
Total	665	904	1,090
	Growth Rates (%)		
Americas		5.1	3.1
Western Europe		4.8	3.2
Asia		9.7	4.6
Other		17.3	8.4
Total		6.3	3.8

EOAs are corrosive and will attack some plastics and rubbers. They are a moderate fire hazard. MEA and TEA are clear, viscous liquids with a mild ammonia-like odor. DEA is crystalline or a viscous liquid. All are soluble in water and ethanol. Vapor is irritating to the eyes, skin, and respiratory tract and depression of the central nervous system can occur. They can also be absorbed by the skin in toxic amounts.120

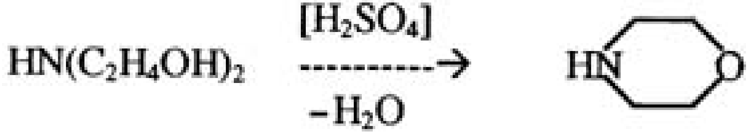
Uses

EOA production is consumed in the following areas: Detergents (32%), Ethyleneamines (20%), Gas Purification (17%), Metal Cleaning (10%), Herbicide Intermediates (10%), and Miscellaneous (6%). The specific uses of MEA, DEA and TEA are listed in *Chem. Mark. Rep*.123

Secondary Products of Ethanolamine

Morpholine

Morpholine is a solvent and intermediate for optical brighteners and rubber chemicals. In one process it is obtained from DEA by dehydration with 70 percent H2SO4 to close the ring:



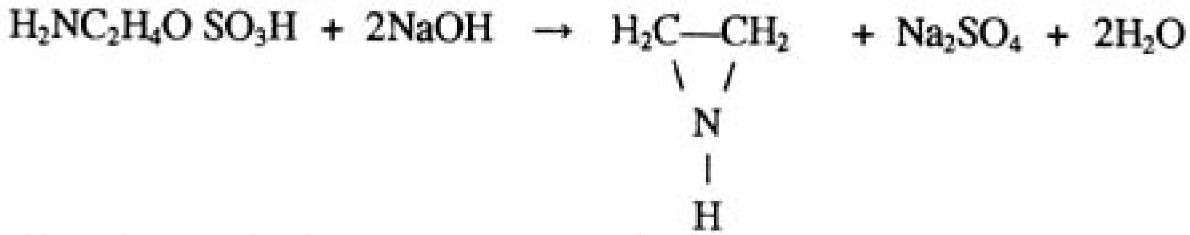
$$HN(C_2H_4OH)_2 \xrightarrow[-H_2O]{[H_2SO_4]}$$

In the United States, a newer production route is used in which diethylene glycol, ammonia, and hydrogen react at 150400°C and 30400 bar over a catalyst containing Ni, Cu, Cr, or Co to give morpholine.46

Ethylenimine

Ethylenimine can be made from MEA via a two-step process that BASF and Hoechst developed.

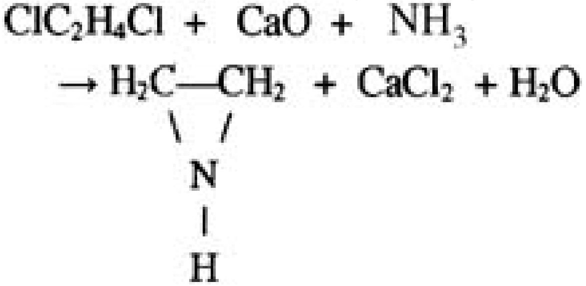
$$H_2NC_2H_4OH + H_2SO_4 \rightarrow H_2NC_2H_4OSO_3H + H_2O \qquad (29\text{-}61)$$



$$\mathrm{H_2NC_2H_4O\,SO_3H + 2NaOH \rightarrow \underset{\substack{\backslash\ /\\ N\\ |\\ H}}{H_2C{-}CH_2} + Na_2SO_4 + 2H_2O}$$

Additional process details are given in Weissermel and Arpe.46

Dow makes ethylenimine by reacting 1,2-dichloroethane with ammonia in the presence of CaO at about 100°C:



$$\mathrm{ClC_2H_4Cl + CaO + NH_3 \rightarrow \underset{\substack{\backslash\ /\\ N\\ |\\ H}}{H_2C{-}CH_2} + CaCl_2 + H_2O}$$

Most ethylenimine is converted into polyethylenimine, a processing aid in the paper industry. It also serves as a reaction intermediate when ureas are formed with isocyanates.46

Ethylenediamine

Commercial ethylenediamine (EDA) is produced by two different routes:

1. From 1,2-dichloroethane (EDC) and ammonia by Cl/NH2 substitution.
2. From MEA and ammonia by OH/NH2 substitution.

The reaction in the first process is shown in reaction (63), and the process details are given in Weissermel and Arpe.46

$$\begin{aligned}&\mathrm{ClC_2H_4Cl + 2NH_3}\\ &\quad\rightarrow \mathrm{NH_2C_2H_4NH_2} + \text{higher amines}\\ &\qquad + \mathrm{NH_4Cl} + \text{amine chlorides} \qquad (29\text{-}63)\end{aligned}$$

In 1969, BASF developed a process to make EDA by the reaction of MEA and ammonia (see reaction (64)):

$$\mathrm{NH_2C_2H_4OH + NH_3 \xrightarrow[H_2]{Ni} NH_2C_2H_4NH_2 + H_2O} \qquad (29\text{-}64)$$

Process details are given in Weissermel and Arpe.46

The properties of EDA and some of the higher amines are given in Table 29.28.

TABLE 29.28 Properties of Ethylenediamine and Higher Amines

	EDA	DETA	TETA	TEPA	PEHA
Molecular weight	60.11	103.17	146.24	189.31	232.36
Melting point, °C	1139	-35	-30	-26	
Boiling point, °C	117	207	277	decomp. at 340°C	
Density, lb/gal	7.67	7.96	8.19	8.33	8.39
Flash point (O.C.), °C	38.102	143	185	185	
Specific gravity	0.898				
pH	11.9 (25% solution in water)				
Vapor density	2.07				
Nitrogen content, wt. %.	46.6				
Viscosity, mPa-sec	2.0 (at 20°C)				
Autoignition temperature, °C	385				
Physical state	Colorless liquid with amine-like odor				

Notes: EDA: Ethylenediamine; DETA: Diethylenetriamine; TETA: Triethylenetetramine; TEPA: Tetraethylenepentamine; PEHA: Pentaethylenehexamine.

Ethylenediamine Production

In 1999, the EDA capacity in the United States was 390 million lb per year of ethyleneamines. Demand was expected to grow from 310 million pounds per year in 1998 to 370 million pounds per year in 2003. These estimates include exports that average 7080 million lb per year and imports that total about 1012 million lb per year. From 1989 to 1998, growth averaged 4 percent per year, but it is expected to slow to 3 percent per year through 2003.124 Prices are kept confidential, but the 2001 U.S. EDA price is estimated to be $0.800.85 per lb.118

Ethylenediamine Uses

EDA is used in the following products: lube oil and fuel additives (20%), chelating agents (16%), wet-strength resins (15%), epoxy curing agents and polyamides (13%), surfactants (10%), oil field chemicals (7%), and miscellaneous (19%).

29.16 Hexamethylenetetramine (Hexamine)

Hexamine was first prepared in 1859. It is a white, crystalline powder with a slight amine odor. It is soluble in water, alcohol, and chloroform, but it is insoluble in ether. However, the aqueous solutions exhibit inverse solubility, that is, less hexamine dissolves as the temperature increases. The hydrate, $(CH_2)_6N_4 \cdot 6H_2O$ can be crystallized from the aqueous solution at temperatures below 14°C. Some additional properties are listed in Table 29.29.

TABLE 29.29 Properties of Hexamine/Hexamethylenetetramine

Molecular weight	140.19
Sublimation temperature	285295°C
Flash point	250°C
Density of solid at 20°C	1.33 g/cm3
Bulk density	700800 g/L
Particle size	700 µmMaximum
Specific heat	36.5 cal/°C
Heat of formation at 25°C	28.8 kcal/mol
Heat of combustion at 25°C	1003 kcal/mol
Solubility in water	
20°C	874 g/L
25°C	867 g/L
60°C	844 g/L
pH of 10% aqueous solution	89
Vapor pressure at 20°C	0.0035

Hexamine Processes

Hexamine is made by the liquid phase reaction of ammonia and formaldehyde:

$$4NH_3 + 6HCHO \rightarrow (CH_2)_6N_4 + H_2O$$

$$\Delta H = -55\,\text{kcal/mol hexamine} \qquad (29\text{-}65)$$

Process details are given in Kothari Phytochemicals International.125

Production

Hexamine is produced as a granular and free-flowing powder as well as a 42.5 percent solution. The solution is shipped in tank trucks, railcars, and drums. Solids are packed in bags, fiber drums, and super sacks. Hexamine is sensitive to moisture. Therefore, it should be stored in an atmosphere with a relative humidity below 60 percent.126

Uses

Hexamine is used as a stabilizer and surface treatment in numerous applications and industries. Specific applications are described in Saudi Formaldehyde Chemical Company Ltd.127

29.17 Hydrazine

Hydrazine, NH2NH2, is the simplest diamine. Anhydrous hydrazine was first prepared in 1894. In 1953, Arch Hydrazine (formerly Olin Chemical) pioneered U.S. production of hydrazine to fuel the first Titan rocket.

Hydrazine is a clear, hygroscopic fuming liquid with the odor of ammonia. It is a mild base that is miscible in polar solvents such as water, ammonia, amines, and alcohols. Additional properties are shown in Table 29.30.135

TABLE 29.30 Physical Properties of Hydrazine

Molecular weight	
Anhydrous	32
Hydrate	50
Boiling point, °C	113.5
Melting point, °C	1.4
Specific gravity, g/ml	
0°C	1.025
15°C	1.014
25°C	1.004
50°C	0.982
Critical temperature, °C	380
Critical pressure, atm	145
Vapor pressure, mmHg	
25°C	14
31°C	20
36°C	100
Viscosity, cp	
5°C	1.2
25°C	0.9
Heat of vaporization, kcal/mol	9.6
Heat of solution, kcal/mol at 25°C	-3.9
Heat capacity, J/molK at 25°C	98.87
Heat of combustion, kcal/mol	-146.6
Heat of formation, kcal/mol	
Liquid	12
Gas	23
Flash point, °C	52
Explosive limits in air by vol. %	4.7100
Azeotrope	
Boiling point (68% hydrazine), °C	120.5

	Hydrazine Concentration (wt. %)132					
	100	64	51.2	35.2	22.4	15.4
Melting point, °C	2.0	-51.7	-59.8	-64.6	-26	-14
Boiling point, °C	113.5	120.5	117.2	108	107	103
Density, 25°C, g/ml	1.0045	1.0320	1.0281	1.0209	1.0132	1.0083
Viscosity, 20°C, µPa-sec	0.974	1.5	1.44	1.10	1.08	1.04
pH		12.75	12.10			10.5

Processes

Five processes are available for the commercial production of hydrazine: (1) The Raschig process, (2) the Raschig/Olin process, (3) The Hoffmann (urea) process, (4) Bayer ketazine process, and (5) the peroxide process from Produits Chimiques Ugine Kuhlmann (of France).

Raschig Process

The Raschig process was discovered in 1907 and then modified into the Olin process. The chemical reactions take place in the liquid phase and involve three steps:

$$NaOH + Cl_2 \rightarrow NaOCl + HCl \qquad (29\text{-}66)$$

$$NH_3 + NaOCl \rightarrow NH_2Cl + NaOH \qquad (29\text{-}67)$$

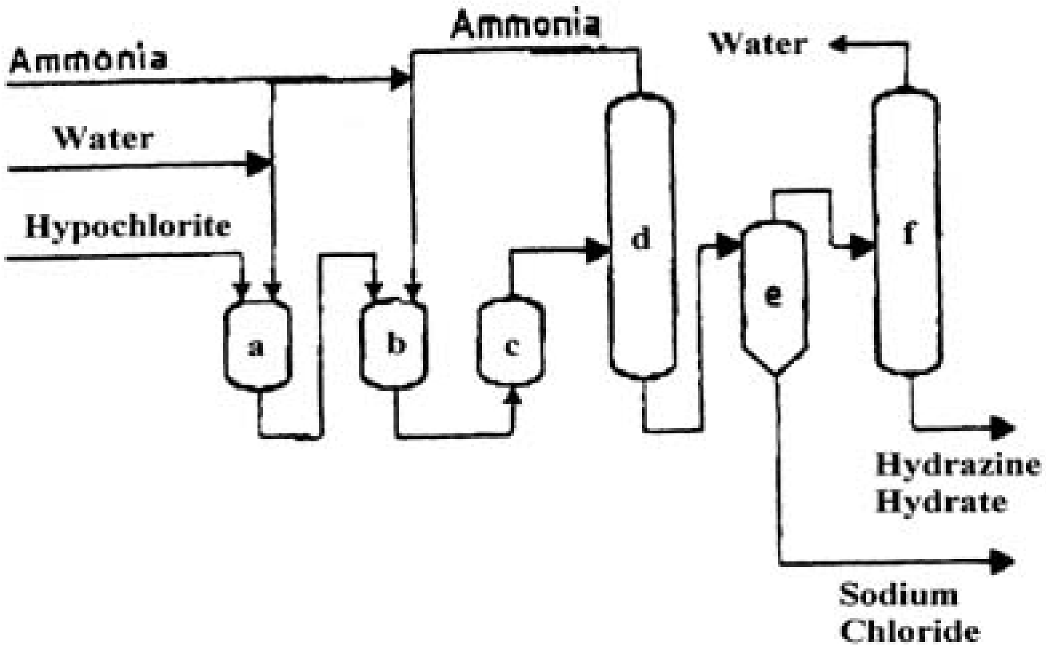
$$NH_2Cl + NH_3 + NaOH$$

$$\rightarrow NH_2{-}NH_2 \cdot H_2O + NaCl \qquad (29\text{-}68)$$

Hydrazine is produced in the hydrated form with one mole of water added. Although a significant fraction of hydrazine is used as the hydrate, numerous applications (such as rocket propulsion) require anhydrous hydrazine. Because of the azeotrope at 68 percent hydrazine, reactive distillation or extractive distillation must be used to produce pure hydrazine.

Additional process details are given in Schirmann.132 The Raschig process is shown in Fig. 29.26.

Fig. 29.26 Raschig process for hydrazine production. (a) Chloramine reactor; (b) and (c) hydrazine reactors; (d) ammonia evaporator; (e) hydrazinesodium chloride separator; (f) hydrazine hydrate concentrator. (Shirmann, J. P., *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. A13, pp. 177190, Reinhold Publishing, New York, 1996. Copyright by Reinhold Publishing and reproduced by permission of copyright owner.)



The Raschig process can also be used to react amines with chloramine to make monosubstituted or unsymmetrical disubstituted hydrazines.

$$NH_2Cl + RNH_2 \rightarrow RNHNH_2 \qquad (29\text{-}69)$$

$$NH_2Cl + R_2NH \rightarrow R_2NNH_2 \qquad (29\text{-}70)$$

Rasching/Olin Process

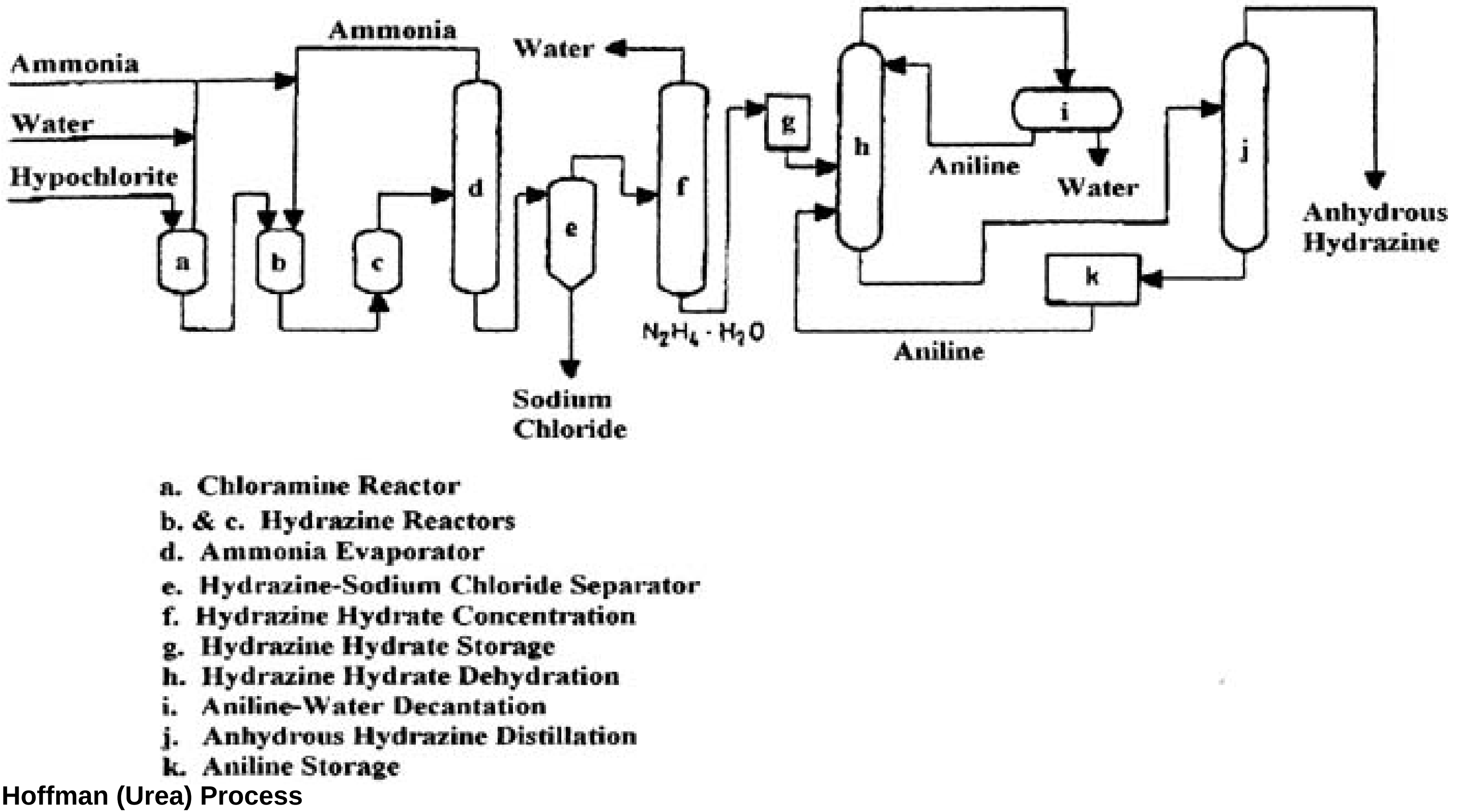
The Raschig/Olin process is used to make anhydrous hydrazine. In this process, reactions (66)(68) also occur. However, the refining area has a different design from the Raschig process. Additional details are given in Schirmann.132

The refining area has five steps: (1) ammonia removal, (2) sodium chloride concentration and removal, (3) hydrazinewater distillation to reach azeotropic concentration, (4) extractive distillation with aniline to break the azeotrope, and (5) distillation to separate hydrazine from aniline. Sometimes 50 percent caustic replaces aniline in the extractive distillation. If ultra pure hydrazine is needed, freeze crystallization is used to remove the supernatant fluid. Assays between 99.5 and 99.99 percent have been achieved.

The overall yield based on chlorine is 65 percent. The combined yield for reactions (66) and (67) is about 95 percent, whereas the yield for reaction (68) is 70 percent. Japanese patent SHO-62-83308 claims to increase hydrazine yield by almost 10 percent by high frequency heating immediately after the reactants are mixed in reaction (67).

A sketch of the Raschig/Olin process is shown in Fig. 29.27.132

Fig. 29.27 Raschig/Olin process for hydrazine production.132 (a) Chloramine reactor; (b) and (c) hydrazine reactors; (d) ammonia evaporator; (e) hydrazinesodium chloride separator; (f) hydrazine hydrate concentration; (g) hydrazine hydrate storage; (h) hydrazine hydrate dehydration; (i) anilinewater decantation; (j) anhydrous hydrazine distillation; (k) aniline storage. (Shirmann, J. P., *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. A13, pp. 177190, Reinhold Publishing, New York, 1996. Copyright by Reinhold Publishing and reproduced by permission of copyright owner.)



Hoffman (Urea) Process

The overall reaction for the Hoffman process is:

$$\underset{\text{Urea}}{CO(NH_2)_2} + \underset{\text{Hypochlorite}}{NaOCl} + 2NaOH \rightarrow \underset{\text{Hydrazine}}{NH_2{-}NH_2} + NaCl + Na_2CO_3 + H_2O \qquad (29\text{-}71)$$

Process details are given in Schirmann.132

This process is not being operated in 2001 although it has operated commercially in the past. Compared with the standard Raschig process, it was the most economical method for low production levels. However, rapid growth in plant size made it obsolete.132

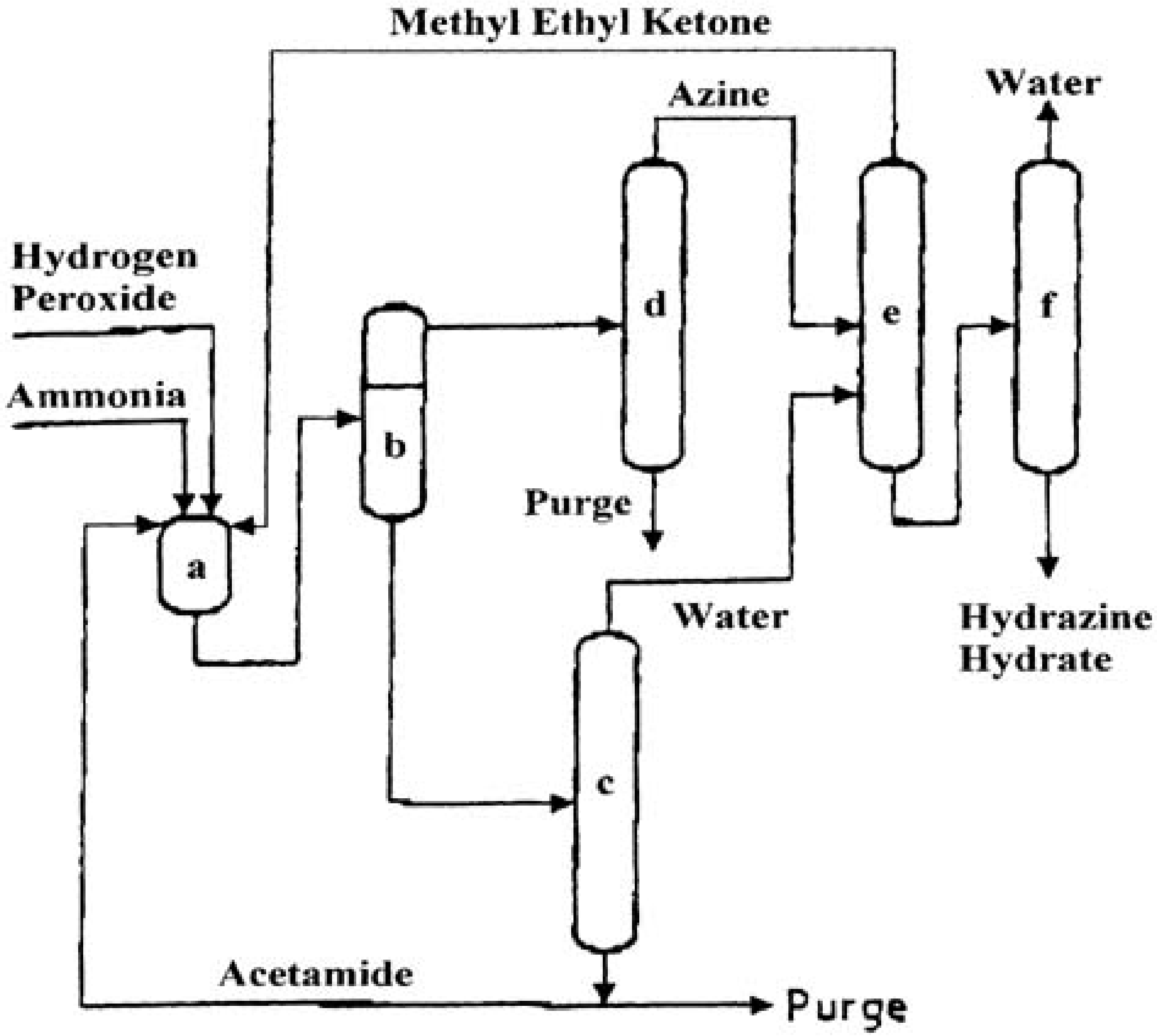
Bayer Ketazine Process

The Bayer ketazine process is based on the reaction of chloramine with ammonia in the presence of acetone at pH 1214. NaOCl, acetone, and a 20 percent aqueous solution of ammonia (at a mole ratio of 1:2:20, respectively) are fed to a reactor at 35°C and 200 kPa to make the aqueous dimethyl ketazine solution. Excess ammonia and acetone are removed in a series of columns and recycled to the reactor. The ketazine solution is distilled to make a hydrazine hydrate containing 64 percent hydrazine.132

Peroxide Process

This process (see Fig. 29.28) was invented by PCUK and is operated by AtoFina in France. The reaction is carried out in the presence of methyl ethyl ketone (MEK) at atmospheric pressure and 50°C. The molar ratio in the feed of hydrogen peroxide (H2O2), MEK, and NH3 is 1:2:4. Additional process details are given in Schirmann.132

Fig. 29.28 Peroxide process for hydrazine production.132 (a) Azine reactor; (b) phase separator; (c) aqueous phase concentration; (d) azine purification; (e) azine hydrolysis; (f) hydrazine hydrate concentration. (Shirmann, J. P., *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. A13, pp. 177190, Reinhold Publishing, New York, 1996. Copyright by Reinhold Publishing and reproduced by permission of copyright owner.)



Production

Hydrazine is marketed as anhydrous hydrazine, as monohydrate (64% hydrazine), and as a 54.5 percent aqueous solution (85% hydrazine monohydrate). The largest U.S. producers of hydrazine based on a 1988 report134 are: Arch Chemical (21 million lb per year), Mobay (14 million lb per year), and Fairmont Chemical (1 million lb per year).

About 29 million lb are sold in the United States while the manufacturers retain the remainder for internal use. The space industry uses only 5 percent of all hydrazine produced in the United States.134 Total capacity in the western world in 1988 was estimated to be about 130 million lb per year.132

The Permissible Exposure Limit (PEL) for hydrazine hydrate on an 8 hr, TWA is 0.1 ppm. The ammonia-like odor of hydrazine hydrate normally cannot be detected until the concentration is above 35 ppm.135

Uses

Hydrazine is used directly as an oxygen scavenger and as an energy source. Hydrazine is consumed in the following applications: blowing agents (33%), pesticides (32%), water treatment (18%), and miscellaneous (17%). Some of the specific uses are listed in Arch Hydrazine.128

29.18 Hydrogen Cyanide

Hydrogen cyanide, HCN, is also known as hydrocyanic acid, prussic acid, and formonitrile. Although HCN is a colorless liquid, some iron cyanides are blue (Prussian blue) and their presence as a slight contaminant usually gives HCN a bluish tint that led to the German name Blausäureblue acid. Selected physical properties of hydrogen cyanide are shown in Table 29.31.

TABLE 29.31 Physical Properties of Hydrogen Cyanide

Molecular formula	HC≡N
Molecular weight	27.03
Melting point	-13.24°C
Triple point	-13.32°C
Boiling point, 1 atm	25.70°C
Density, g/ml	
0°C	0.7150
10°C	0.7017
20°C	0.6884
Specific gravity of aqueous solutions (at 18°C compared to water at 18°C)	
10.04%	0.9838
20.29%	0.9578
60.23%	0.8290
Vapor specific gravity, at 31°C, air = 1	0.947
Vapor pressure, kPa	
-29.5°C	6.697
0.0°C	35.24
27.2°C	107.6
Liquid viscosity at 20.2°C	0.2014
Specific heat, J/mol	
-33.1°C, liquid	58.36
16.0°C, liquid	70.88
27.0°C, gas	36.03
Heat of formation, at 25°C, 1 atm, Δ*H*f, kJ/mol	
Gas at 25°C	-130.5
Liquid at 25°C	-105.4
Heat of fusion at -14°C, kJ/mol	7,100
Latent heat, kJ/mol	23.8
Heat of combustion, net, kJ/mol	642
Conductivity, S/cm	3.3 × 10-6
Heat of vaporization, kJ/mol	25.2
Heat of polymerization, kJ/mol	42.7
Flash cup, closed cup, °C	-17.8
Explosive limits in air at 100 kPa and 20°C, vol. %	641
Autoignition temperature, °C	538

Safety

Some people can detect hydrogen cyanide by odor or taste sensation at a concentration of 1 ppm in air while most people can detect 5 ppm. OSHA has set 4.7 ppm as the maximum, average safe exposure limit for a 15 min period. Exposure to 20 ppm of HCN in air causes slight warning symptoms after several hours; 50 ppm causes disturbances within an hour; 100 ppm is dangerous for exposures of 3060 min; and 300 ppm can be rapidly fatal unless prompt, effective first aid is administered. A small concentration of cyanide always exists in a person's body, and the body has a mechanism to continuously remove small amounts of cyanide.129

HCN may also present some processing hazards. HCN undergoes an exothermic polymerization at a pH of 511 to form solid black compounds. This polymerization can become explosively violent, especially if it is confined. To avoid polymerization, stored HCN should contain less than 1.0 wt. % water; it should be kept cool; and it should be inhibited with sulfuric, phosphoric, or acetic acid.129

Explosively violent hydrolysis can occur if an excess of a strong acid (H2SO4, HNO3, or HCl) is added to HCN. The reaction is fastest at or near stoichiometric ratios, for example, 12 moles H2SO4 per mole HCN and can cause severe equipment damage if confined.129

An important environmental issue is the fate of cyanide. HCN, if spilled, evaporates quite readily. That which does not evaporate is soon decomposed or rendered nonhazardous by complexing with iron in the soil, by biological oxidation, or by degradation from photolysis.129 Several methods for treatment and disposal of cyanide compounds are described in Kirk-Othmer129 and Botz.133

Manufacture

Six processes are presently of economic significance: (1) the Andrussow process is currently the principal HCN manufacturing process in the world, (2) the BMA process is practiced by two companies and provides high yield and selectivity by using a complex reaction system, (3) the Fluohmic process is of interest in locations where electricity is inexpensive, (4) the Formamide process is useful for sites with inexpensive carbon monoxide, (5) the Sohio acrylonitrile process produces HCN as a by-product, and (6) the Methanol process.

Andrussow Process

The Andrussow process was patented in Germany by Dr. L. Andrussow in 1933. Its main advantages are low converter investment, low maintenance cost, and high natural gas yields. The Andrussow process produces HCN by the reaction of ammonia, air, and natural gas at 10001200°C in the presence of a platinum/rhodium catalyst. The reaction is:

$$NH_3 + CH_4 + 1.5O_2 \xrightarrow[Pt/Rh]{1200°C} HCN + 3H_2O \quad \Delta H = -115.2 \text{ kcal/mol HCN} \qquad (29\text{-}72)$$

More reaction details are given in Kirk-Othmer.129 The heat of reaction is recovered in the waste-heat boiler to generate about 5 lb of steam per pound of HCN produced.

The precious metal catalyst is usually 90 percent platinum and 10 percent rhodium in gauze form. The catalyst life can range from 4000 hr to as much as 10,000 hr. A catalyst pack is usually replaced because it has been contaminated by minor impurities in the feed gases. Very little catalyst metal is consumed or lost during the life of a catalyst.

The reaction takes place under fuelrich conditions to maintain a nonflammable feed mixture. Typical feed composition is 1315 percent ammonia, 1113 percent methane, and 7276 percent air on a volumetric basis. Control of feed composition is essential to guard against deflagrations as well as to maximize the yield. The yield from methane is approximately 60 percent of theoretical.

The reactions in the Andrussow process are more complex than that shown in reaction (29-72).130 Most of the heat required for HCN formation is supplied by combustion of methane. This results in an overall reaction that is exothermic.129 The converter off-gas typically has the following molar composi-tion: HCN = 612 percent, NH3 = 1.53 percent, H2 = 711 percent, N2 = 5055 percent, H2O = 2025 percent, CO = 34 percent, CO2 = 0.3 percent, and CH4 = 0.1 percent.

Two processes can be used to remove NH3 from the converter off-gases. Figure 29.29 shows a process that recovers unreacted ammonia for recycle to the converter while the process in Fig. 29.30 uses sulfuric acid to produce ammonium sulfate as a by-product. Ammonium sulfate can create a disposal problem, but the recycle system is capital and energy intensive. Ammonia must be removed from the off-gas before HCN can be recovered because the ammonia promotes polymerization of the HCN.129 Additional process details are given in Kirk-Othmer.129

Fig. 29.29 Andrussow HCN process with ammonia recycle.129 (Kirk-Othmer, *Encyclopedia of Chemical Technology*, 4th ed., Vol. 7, pp. 753782, John Wiley & Sons, Inc., New York, 1993. Copyright by John Wiley & Sons, Inc. and reproduced by permission of copyright owner.)

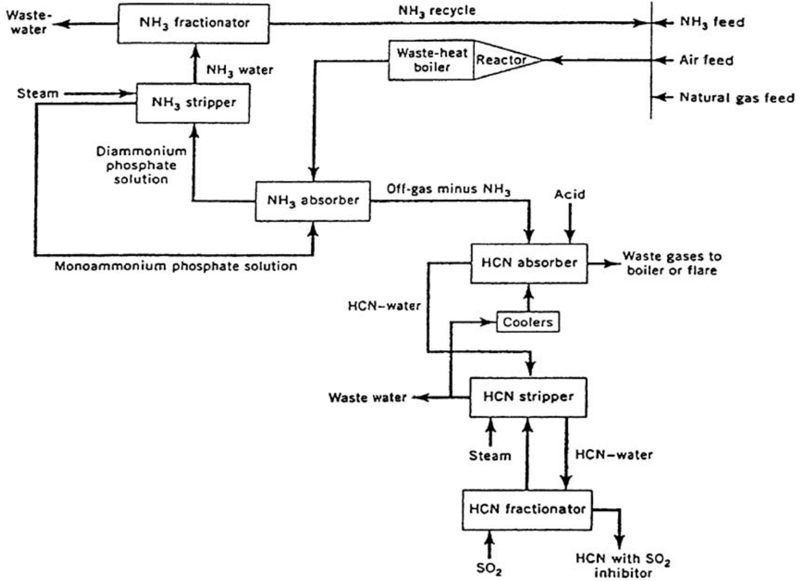
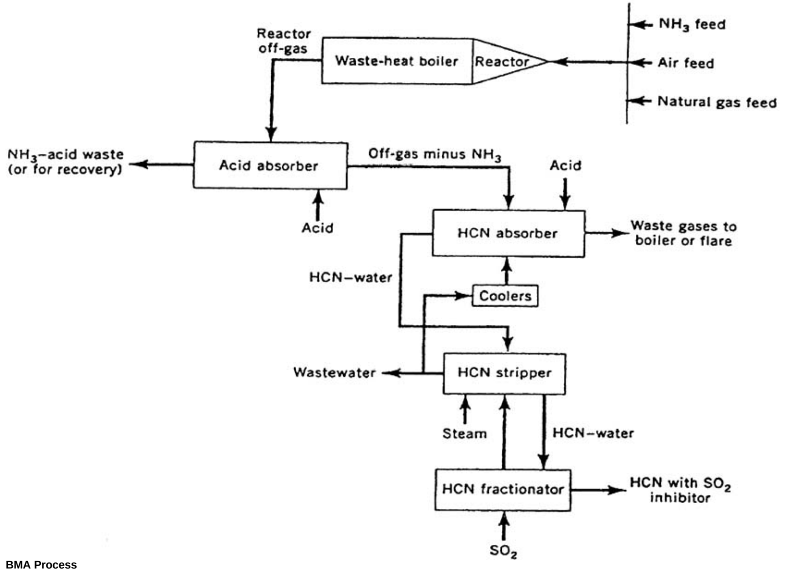


Fig. 29.30 Andrussow HCN process with ammonia removal.129 (Kirk-Othmer, *Encyclopedia of Chemical Technology*, 4th ed., Vol. 7, pp. 753782, John Wiley & Sons, Inc., New York, 1993. Copyright by John Wiley & Sons, Inc. and reproduced by permission of copyright owner.)



BMA Process

The Blausäure-Methan-Ammoniak (BMA), or HCNmethaneammonia process was developed by Degussa around 1949. In the process, ammonia reacts with methane in the absence of oxygen in a gas-fired tubular reactor. U.S. Patent 5,785,942 describes the BMA process and improvements that were made during the mid-1990s. The reaction is:46

$$NH_3 + CH_4 \xrightarrow[Pt/Ru]{1300°C} HCN + 3H_2 \quad \Delta H = +60 \text{ kcal/mol HCN} \qquad (29\text{-}73)$$

In a reactor that is similar to a reformer, the reaction occurs in tubes that are heated externally to supply the endothermic heat of reaction.129 Sintered corundum (α-Al2O3) tubes with an internal layer of platinum/ruthenium catalyst are normally used. To achieve adequate heat transfer, the tubes may be only 3/4 in. in diameter and 61/2 feet long. Selectivities of 9091 percent for methane and 8384 percent for ammonia are reached at 12001300°C reaction temperatures.

A typical converter is made up of multiple furnaces, each of which contains 810 reactors. Each reactor contains 1030 tubes with a catalyst lining. Catalyst life is approximately 10,000 hr.129

The converter off-gas typically has the following molar composition: HCN = 2023 percent, NH3 = 23 percent, H2 = 7072 percent, N2 = 0.51 percent, and CH4 = 12.4 percent.

The unreacted ammonia is removed from the reactor off-gas by scrubbing with sulfuric acid to make ammonium sulfate. Just like the Andrussow process, ammonia must be removed from the off-gas before HCN can be recovered because the ammonia promotes polymerization of the HCN. After ammonia is removed from the converter off-gases, the remaining gas stream is processed in a way similar to the Andrussow process.129

In the BMA process, the gas that leaves the HCN absorber has the following molar composition: H2 = 96 percent, N2 and O2 = 1.5 percent, and CH4 = 2.5 percent. The vent gas is used either as a chemical feedstock or as fuel gas in the HCN furnaces.

The advantages of the BMA process are high ammonia yields, high natural gas yields and the useful hydrogen in the vent gas. But the high investment and high maintenance costs for the converter are a disadvantage. The complexity of the reaction system also makes the process less attractive for large-scale production.129 The BMA process is used at four plants.131

Fluohmic Process

The Fluohmic process was developed around 1960 by Shawinigan Chemicals of Canada. This process involves the reaction of ammonia with hydrocarbons (usually propane or butane) in an electrically heated, fluidized bed of coke. The reaction is:

$$3NH_3 + C_3H_8 \xrightarrow{1370°C} 3HCN + 7H_2$$
$$\Delta H = +151 \text{ kcal/mol HCN} \qquad (29\text{-}74)$$

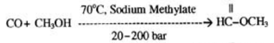
The fluidized bed is heated at 13501650°C by passing an electric current between carbon electrodes immersed in the bed. The performance of the reaction system is reported to be:

	Conversion (Mole %)	Yield (Mole %)	Yield (lb/lb HCN)
C3H8	88	8890	0.600.62
NH3	86	8690	0.700.73

The reactor off-gas typically has the following molar composition: HCN = 25 percent, NH3 = 0.25 percent, H2 = 72 percent, and N2 = 3 percent. The HCN is separated from the off-gas for refining, and the hydrogen leaves the system as a by-product together with a small amount of nitrogen and unreacted ammonia. The high electrical consumption (•3 kWhr/lb HCN) makes the process attractive only in locations where the cost of electricity is extremely low. In 2001, the Fluohmic process was used in Spain, Australia, and South Africa.

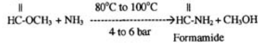
Formamide Process

The Formamide process was developed by several companies, but in 2001 it was only used by BASF. The process involves the following steps:46 (1) methyl formate synthesis, (2) formamide synthesis, and (3) HCN formation/dehydration.



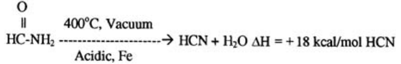
$$CO + CH_3OH \xrightarrow[20-200 \text{ bar}]{70°C, \text{ Sodium Methylate}} \underset{\text{Methyl Formate}}{HC(=O){-}OCH_3}$$

With excess methanol, the conversion to methyl formate is 95 percent with nearly 100 percent selectivity.



$$HC(=O){-}OCH_3 + NH_3 \xrightarrow[4 \text{ to } 6 \text{ bar}]{80°C \text{ to } 100°C} \underset{\text{Formamide}}{HC(=O){-}NH_2} + CH_3OH$$

The conversion of ammonia is 6085 percent.



$$HC(=O){-}NH_2 \xrightarrow[\text{Acidic, Fe}]{400°C, \text{ Vacuum}} HCN + H_2O \quad \Delta H = +18 \text{ kcal/mol HCN}$$

Process details are given in Weissermel and Arpe.46

The overall performance of the reaction system is:

	Yield (Mole %)	Consumption (lb/lb HCN)
CH3OH		0.200.30
NH3	7990	0.700.80
CO	5294	1.102.00

The reactor off-gas contains 6070 percent HCN.

BASF has obtained European Patent 1,110,913, which describes HCN production from gaseous formamide in a fluidized bed. Elements of the commercialized formamide process are included in a methyl methacrylate process that Mitsubishi Gas Chemical started up in 1997. The formamide is obtained by reacting hydroxyisobutyramide with methyl formate. Therefore, the external requirements for HCN are reduced to the replacement of yield losses.131 The process is described in U.S. Patent 6,075,162 and the catalyst is described in European Patent 1,086,744.

Methanol Process

HCN is produced when methanol reacts with ammonia and oxygen in the presence of an oxide catalyst that contains iron, antimony, phosphorous, and vanadium. The reaction occurs in the vapor phase in a fluidized bed reactor. The process and the catalyst are described in patents that were issued to Nitto Chemical (now Mitsubishi Rayon) during the late 1990s (European Patent 864,532; Japanese Patents 10-167,721, 10-251,012, 11-043,323; U.S. Patent 5,976,482).

Mitsubishi Rayon is believed to operate a commercial-scale, methanol-based HCN process and has offered to license the technology to other companies. The technology might provide a low-cost way to convert an acrylonitrile plant to HCN-only production. However, methanol is a higher cost source of carbon compared to natural gas so the methanol process probably has a higher operating cost than the Andrussow process.

Acrylonitrile Process

The Standard Oil Company of Ohio (now BP Amoco) commercialized in 1960 a fluidized bed process in which the catalytic oxidation of a mixture of propylene and ammonia produces acrylonitrile (ACRN). By-products from this reaction are HCN and acetonitrile. The yields of HCN depend on the process conditions and on the catalyst system.131 The reactions are:

$$\underset{\text{Propylene}}{CH_2{=}CHCH_3} + NH_3 + \tfrac{3}{2}O_2 \rightarrow \underset{\text{Acrylonitrile}}{CH_2{=}CHCN} + 3H_2O \qquad (29\text{-}78)$$

$$CH_2{=}CHCH_3 + \tfrac{3}{2}NH_3 + \tfrac{3}{2}O_2 \rightarrow \underset{\text{Acetonitrile}}{\tfrac{3}{2}CH_3CN} + 3H_2O \qquad (29\text{-}79)$$

$$CH_2{=}CHCH_3 + 3NH_3 + 3O_2 \rightarrow 3HCN + 6H_2O \qquad (29\text{-}80)$$

Depending on the type of ACRN process and the operating conditions, about 1024 lb HCN are obtained per 100 lb of ACRN.46 U.S. Patent 5,840,648 describes a catalyst that permits more HCN production.

During the 1990s, several patents were issued (European Patent 878,464, in U.S. Patent 5,288,473, and U.S. Patent 6,204,407) in relation to the addition of methanol to the feed stream of an ACRN plant to increase HCN production.

BP Amoco and Asahi have also been running tests related to the ammoxidation of propane to produce ACRN. This process is believed to yield a higher level of HCN than the optimized oxidation of propylene.131

Other HCN Processes

Some alternate HCN processes that have been studied during the 1990s are:

1. production of HCN from acetonitrile (see Japanese Patent 10,167,721)
2. production of HCN by use of a corona (German Patent l,054,982)
3. production of HCN by use of microwave heating (U.S. Patents 5,393,393, 5,470,541, and 5,529,669)

None of these processes are believed to be used in commerical production in 2001.

Production

In 1999, a total of 34 companies were operating 47 HCN production facilites in the United States, Western Europe, and Japan. Capacity in these three regions was about 3.6 billion lb (1.6 million tonnes). Direct production accounts for 75 percent of total capacity and the balance is co-product from ACRN production. Worldwide capacity is about 1.8 million tonnes. DuPont owns or has a financial interest in about 36 percent of this capacity while Degussa-Hüls and Rohm and Hass have about 6 percent each.

Between 1994 and 1999, HCN consumption increased to about 2.2 percent per year. Consumption is expected to increase at around 2.8 percent per year from 1999 through 2004.

Uses

HCN is usually consumed at its production site, and most consumers without on-site HCN production capability are believed to be developing plans to eliminate HCN shipments. For example, BASF in the United Kingdom produces by-product HCN in an ACRN plant and ships HCN railcars to Ineos. This HCN is used to produce acetone cyanohydrin (ACH) in a methyl methacrylate plant. To eliminate these shipments, BASF and Ineos plan to: (1) install an ACH plant on the BASF plant site and (2) transfer the ACH to Ineos via a 9 km pipeline.136

ACRN plants are highly dependent on the actions of the acrylic fiber industry, the major outlet for ACRN. As a result, HCN consumers may have to operate at reduced rates if an adequate supply of HCN is not being produced by the ACRN plant. Because of a weak ACRN market in 2000 and 2001, many consumers of HCN from ACRN plants evaluated options that would give them a more consistent supply of HCN. Some ACRN plants were shut down during this time, and market conditions plus company finances will determine whether any of these plants will be restarted.

Some of the uses of HCN are:

1. *Adiponitrile (ADN)*. ADN is made when HCN reacts with butadiene. It is a key ingredient in the production of nylon.
2. *ACH*. ACH is made when HCN reacts with acetone. It is a key ingredient in many methyl methacrylate (MMA) plants. MMA is used to make acrylic sheeting and related products.
3. *Sodium Cyanide (NaCN)*. NaCN is made when HCN reacts with sodium hydroxide (NaOH). NaCN is used in precious metals (gold and silver) extraction, metal plating, and some chemical production.
4. *Methionine*. Methionine is an essential amino acid that is used in animal feed. The key ingredients are HCN or NaCN, methyl mercaptan, acrolein, and NH4HCO3.
5. *Cyanuric Chloride (CYC)*. CYC is made when HCN reacts with chlorine and is then rearranged into a ring compound. It is used to make herbicides, pharmaceuticals, and explosives.
6. *Chelating Agents*. Chelating agents are made when HCN reacts with formaldehyde and amines to make aminocarboxylic acids, which are then saponified to produce the acid salts. Their major uses are in soaps, water treating agent, and various cleaning agents.
7. Other uses include ferrocyanides (for blue jeans), acrylates, lactic acid, pharmaceuticals, and specialty chemicals.

Table 29.32 lists the HCN consumption for different products.

TABLE 29.32 Consumption of Hydrogen Cyanide by Major Region1999 (Thousands of Tonnes)

Product	United States	Western Europe	Japan	Total
Adiponitrile	340	127		467
Acetone cyanohydrin	206	152	50	408
Sodium cyanide	72	55	16	143
Methionine and Analogs	37	32	4	73
Cyanuric chloride	20	20	1	41
Chelating agents	18	11		29
Other	35	13	13	61
Total	728	410	84	1222

29.19 Aniline

Aniline was first produced in 1826 by the dry distillation of indigo, the oldest known vat dye. Fritsche also obtained aniline from indigo by heating it with potash (K2O). Hofmann obtained aniline by reduction of nitrobenzene in 1843. Aniline is a colorless, oily, flammable liquid that is slightly soluble in cold water and infinitely soluble in alcohol and ether. Its physical properties are summarized in Table 29.33.

TABLE 29.33 Physical Properties of Aniline

Property	Value
Molecular weight	93.12
Boiling point, °C	
101.3 kPa (760 mmHg)	184.4
4.4 kPa (33 mmHg)	92
1.2 kPa (9 mmHg)	71
Melting point, °C	-6.15
Density, *d*	
at 20/4°C	1.02173
at 20/20°C	1.022
Viscosity at 20°C, mPa·sec (= cP)	4.4234.435
Dissociation constant, *pK*	
at 20°C	4.60
at 60°C	8.88
Enthalpy of dissociation, kJ/mol (kcal/mol)	21.7 (5.19)
Heat of combustion, kJ/mol (kcal/mol)	3389.72 (810.55)
Specific heat, 2025°C	0.518
Latent heat of vaporization, J/g (cal/g)	476.3 (113.9)
Flash point (closed-cup), °C	76
Autoignition temperature, °C	615
Flammable limits in air (% by volume)	
LEL	1.3
UEL	11.0
Vapor density (air = 1)	3.22
Solubility, g in 100 g water at 20°C	3.5
Odor threshold, ppm	1.1

Processes

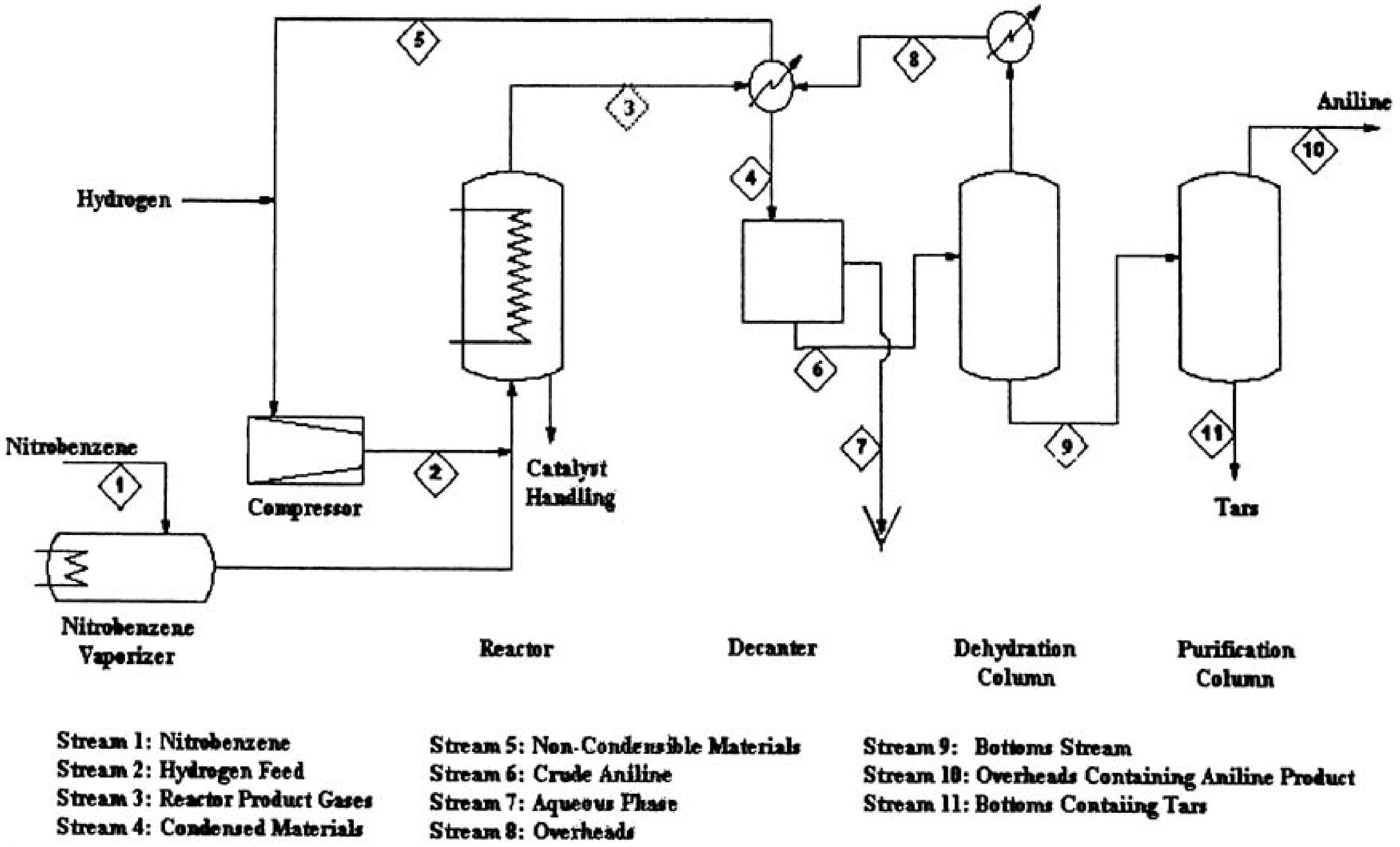
The aniline processes that are currently in use include: (1) hydrogenation of nitrobenzene, (2) nitrobenzene reduction with iron filings, and (3) ammonolysis of phenol. Almost 97 percent of the nitrobenzene produced in the United States is converted to aniline between the hydrogenation process and the iron filings process.137

Hydrogenation of Nitrobenzene

The primary aniline production process in the world is the hydrogenation of nitrobenzene. BASF, DuPont, ChemFirst and Rubicon use this process in the United States. This technology is also used by all Western European aniline producers and all but one Japanese aniline producer.

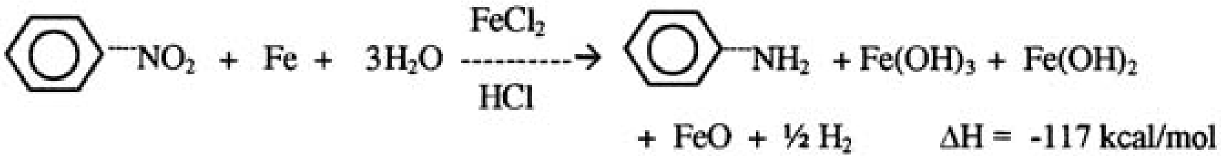
The catalytic gas-phase hydrogenation processes for nitrobenzene can be carried out using a fixed-bed or a fluidized bed reactor. Details of both processes are given in Weissermel and Arpe.46 A typical process diagram is shown in Fig. 29.31.

Fig. 29.31 Aniline production process: hydrogenation of nitrobenzene.137 Stream (1) Nitrobenzene; (2) Hydrogen feed; (3) Reactor product gases; (4) Condensed materials; (5) Non-condensible materials; (6) Crude aniline; (7) Aqueous phase; (8) Overheads; (9) Bottoms stream; (10) Overheads containing aniline product; (11) Bottoms containing tars.



Nitrobenzene Reduction with Iron Filings

The older Bechamp method for iron oxide pigment production gives aniline as a co-product and is operated by Bayer in West Virginia. Nitrobenzene is reduced by reaction with iron filings in the presence of a hydrochloric acid catalyst. The iron is oxidized to the ferrous or ferric state and the co-product aniline is separated. The yield is 9095 percent of theoretical. The reactions are represented as follows:138

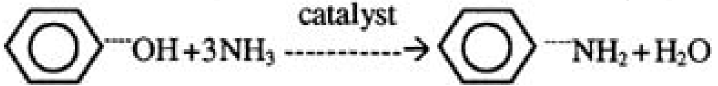


$$2\mathrm{Fe(OH)}_3 + \mathrm{Fe(OH)}_2 \rightarrow \mathrm{Fe}_3\mathrm{O}_4 + \mathrm{H}_2\mathrm{O} \qquad (29\text{-}83)$$

This process would have been replaced much earlier by more economical reduction methods if it had not been possible to obtain valuable iron oxide pigments from the iron oxide sludge. However, the increasing demand for aniline has far surpassed the market for iron pigments, so this process is no longer preferred.46

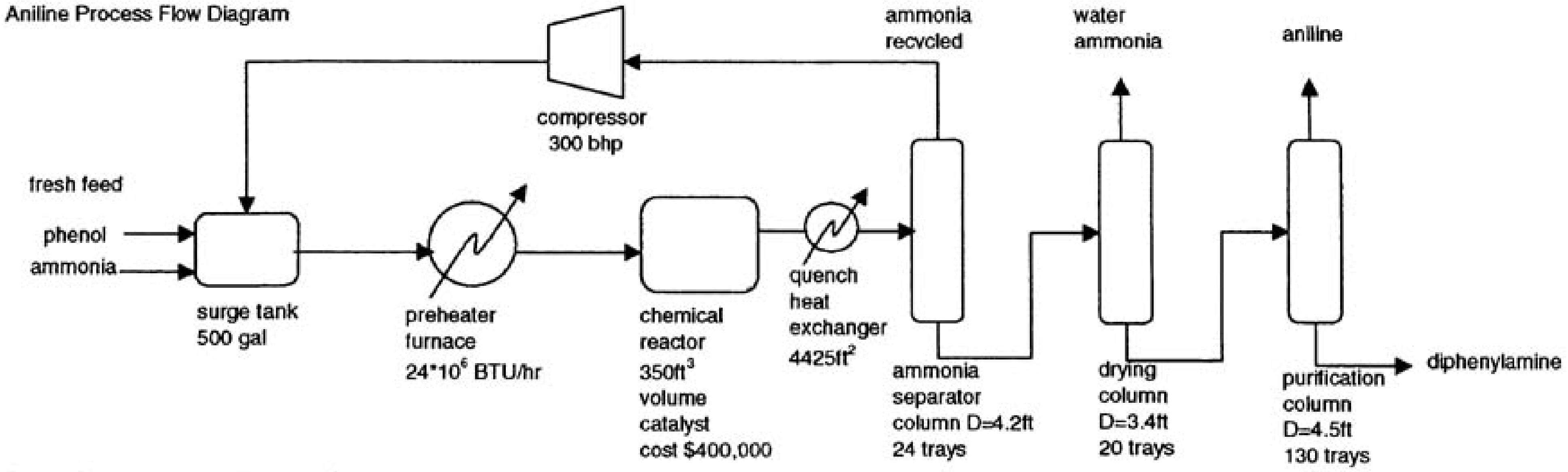
Ammonolysis of Phenol

Aniline can also be produced when phenol is subjected to gas-phase ammonolysis at 200 bar and 425°C. This is the Halcon/Scientific Design process. The chemistry is:



The catalysts are Al2O·SiO2 (possibly as zeolites) and oxide mixtures of Mg, B, Al, and Ti. These can be combined with additional co-catalysts such as Ce, V, or W. With a large excess of ammonia, the selectivity to aniline is 8790 percent at a phenol conversion of 98 percent. The by-products are diphenylamine and carbazole. This technology is used at one plant in Ohio and at another plant in Japan. The economics of this process are favorable if low-cost phenol is available, and high-purity aniline is desired. Capital costs are low because benzene nitration is avoided. A typical process sketch along with a material balance is shown in Fig. 29.32.

Fig. 29.32 Aniline production process: ammonolyis of phenol.139



Production rate = 100 million pounds per year

Utilities:

Fuel: 24*10⁶ BTU/hr

Steam: 600psig - 10.2*10⁶ BTU/hr 150psig - 10*10⁶ BTU/hr

Cooling water: 9*10⁶ BTU/hr for ΔT=20°F

Electricity to drive the motor for the 300 bhp compressor

Preliminary Material Balance on Process

					ammonia separator column		drying column		purification column	
mass flow rates (lb/hr)	fresh feed	ammonia recycle	reactor feed	reactor product	overhead	bottoms	overhead	bottoms	overhead	bottoms
phenol	13,159	0	13,159	13	0	13	0	13	13	0
ammonia	4,199	43,070	47,269	44,900	43,070	1,830	1,830	0	0	0
aniline	0	10	10	12,890	10	12,880	315	12,565	12,565	0
water	0	11	11	2,528	11	2,517	2,505	12	12	0
diphenylamine	0	0	0	118	0	118	0	118	0	118
total	17,358	43,091	60,449	60,449	43,091	17,358	4,650	12,708	12,590	118
	60449				60449		17358		12708	

			preheater		quench exchanger							
temperature(°F)	100	142	118	670	673	110	112	248	141	289	278	491
pressure(psig)	260	260	260	250	240	230	220	220	6	6	5	5

Uses

MDI (4,4′-methylenebis phenylisocyanate) accounted for almost 80 percent of the worldwide demand for aniline in 1996. MDI is used primarily to make rigid polyurethane foam and polyurethane elastomers. From 1996 to 2001, MDI growth is expected to be at least 4.0 percent per year as its use continues to increase in the construction industry (the largest user of rigid polyurethane foam) and the auto industry (the largest user of reaction-injection molding plastics).138

Other aniline uses and the percent of worldwide aniline production that goes into these uses are: rubber-processing chemicals (11%), dyes and pigments (3%), agricultural chemicals (2%), specialty fibers (1.5%), and miscellaneous, such as explosives, epoxy curing agents, and pharmaceuticals (3%).138

Production

MDI has been the driving force behind the recovery of the aniline business since 1982 when the industry had a capacity utilization rate of less than 50 percent. By 1996, capacity utilization had approached 95 percent in some regions.138

Aniline production is concentrated in the United States, Western Europe, and Japan, where aniline capacity is over 1.8 million tonnes per year.138 Aniline global production capacity in 1995 was 2.2 million tonnes, and demand in 2000 was estimated to be 2.68 million tonnes per year.140

From 1983 to 1998, aniline prices ranged from a low of $0.40 per lb to a high of $0.62 per lb. The price in 1999 was $0.50 per lb.141

The OSHA Permissible Exposure Limit (PEL) for aniline is 5 ppm for an 8 hr TWA.

Other Compounds

Several other nitrogen compounds are commercially important, including:

1. hexamethylenediamine, which is used primarily in the manufacture of nylon
2. dimethylformamide, a versatile solvent for organic and inorganic compounds and an important reaction medium for ionic and nonionic compounds
3. dimethylacetamide, an important industrial solvent for polyacrylonitrile, vinyl resins, cellulose derivatives, styrene polymers, and linear polyesters
4. isocyanates, important materials in the production of foams, resins, and rubbers
5. calcium cyanamide, which is used as a fertilizer, herbicide, insecticide, a steel-making additive, and an ore processing material. It can also be used to make thiourea, guanidine, and ferrocyanides142

29.20 References

Comyns, A. E., Named Processes in Chemical Technology, 2nd ed., CRC Press LLC, Boca Raton, 1999.
Knoblauch, K., Chem. Eng. (N.Y.), 85 (25), 87 (1978).
Suzuki, M., in Adsorption and Ion Exchange: Fundamentals and Applications, M. D. LeVan (Ed.), p. 120, American Institute of Chemical Engineers, New York, 1998.
Suzuki, M., in Adsorption and Ion Exchange: Fundamentals and Applications, M. D. LeVan (Ed.), pp. 121122, American Institute of Chemical Engineers, New York, 1998.
Eur. Chem. News, 53 (1391), 31 (1989).
Chem. Mark. Rep., 29 (1990, October 5).
Hopkins, B. S., Smith, H. R. et al., Chemistry and You, Lyons & Carnahan, Chicago, 1949.
Chem. Eng. News, 16 (1979, August 27).
Haber, F., and van Oordt, G., Z. Anorg. Allg. Chem., 43, 111 (1905).
Harding, A. J., Ammonia, Manufacture and Uses, Oxford University Press, London, 1959.
Haber, L. F., The Chemical Industry 19001930, p. 198, Clarendon Press, Oxford, 1971.
Vancini, C. A., Synthesis of Ammonia, L. Pirt (Trans.), Macmillan Press, Basingstoke, England , 1971.
Jennings, J. R., and Ward, S. A., in Catalyst Handbook, 2nd ed., p. 384, M. V. Twigg (Ed.), Wolfe Publishing, London, 1989.
Travis, T., Chem. Ind. (London), 15, 581 (1993).
Livingston, J. G., and Pinto, A., Chem. Eng. Prog., 79 (5), 62 (1983).
Chem. Eng. (Rugby, Eng.), 21 (1990).
Hydrocarbon Processing, 70 (3), 132136 (1991).
Isalski, W. H., Separation of Gases, p. 155, Clarendon Press, Oxford, 1989.
Nitrogen, 223, 25 (1996).
Chem. Eng. Prog., 48, 468 (1952).
Shearon, W. H., Jr., and Thompson, H. L., in Modern Chemical Processes, Vol. 3, p. 16, Reinhold Publishing, New York, 1954.
Chem. Week, 13 (1991, April 3).
Eur. Chem. News, 58 (1524), 42 (1992).
Eur. Chem. News, 60 (1592), 27 (1993).
Oil & Gas J., 94 (47), 37 (1996).
Chem. Eng. (N.Y.), 96 (7), 43 (1989).
Chem. Eng. (Rugby, England), 471, 21 (1990).
Hydrocarbon Processing, 70 (3), 134 (1991).
Hill, W. H., in Chemistry of Coal Utilization, Vol. 2, p. 1029, H. H. Lowry (Ed.), John Wiley & Sons, New York, 1945.
Cohen, J. M., The Life of Ludwig Mond, p. 176, Methuen & Co., London, 1956.
Scholvien, W. R., Chem. Met. Eng., 38 (2), 82 (1931).
Spitz, P. H., Petrochemicals, The Rise of an Industry, p. 84, John Wiley & Sons, New York, 1988.
Chem. Eng. News, 16 (1979, August 27).
Terra e Vita (France), 40 (22), 7879 (1999).
SRI International, Chemical Economics Handbook, Nitrogen Industry Overview, 1998.
www.fertilizer.org/STATSIND, International Fertilizer Industry Association, 2001, June.
Chem. Mark. Rep. (1999, November 29).
Chem. Mark. Rep. (2001, June 18).
Personal Communication, D. L. Mills to G. R. Maxwell, BP Nitrogen Fatality, 2001, July 10.
SRI International, Chemical Economics Handbook, Hydrogen Product Review, 1998.
Hydrogen, Airco Industrial Gases, Airco, Inc., 1991.
Green, D. S., Chemical Processing, 74, 2001, January, 74.
Alves, J., "Analysis and Design of Refinery Hydrogen Distribution Systems," PhD Thesis, UMIST (1999).
Hallale, N., and Liu, F., "Efficient Refinery Hydrogen Management for Clean Fuels Production," Presented at the AIChE Annual Meeting, Los Angeles, 2000.
Hallale, N., "Burning Bright Trends in Process Integration," Chem. Eng. Prog., 97 (7), 3041 (2001).
Weissermel, K., and Arpe, H.-J., Industrial Organic Chemistry, 3rd ed., VCH Publishers, New York, 1997.
U.S. Patent 3,827,987.
Wolff, D., "Producing Hydrogen On-Site," Chem. Eng. Prog., 97 (3), 5153 (2001).
Am. Petrol. Inst., "Selected Values of Properties of Hydrocarbons," Research Project 44, National Bureau of Standards.
Leseur, H., Czernichowski, J., and Chapelle, J., Int. J. Hydrogen Energy, 19 (2), 139 (1994).
Hydrocarbon Processing, 75 (4), 149 (1996).
PEP Yearbook, 1996, Process Economics Program, SRI Consulting.
www.efma.org/Publications, Production of Ammonia, European Fertilizer Manufacturers Association, 2001, July.
www.mbendi.co.za, Mbendi Information Services (Pty) Ltd., South Africa, 2001, July.
www.hydrocarbononlline.com, Hydrocarbon Online News, Vert Tech LLC, 1999, February 16.
www.hydrocarbononlline.com, Hydrocarbon Online News, Vert Tech LLC, 2001, July 17.
SRI International, Chemical Economics Handbook, Ammonia, 2001.
Tebo, M., "Honda Opens First Los Angeles-Area Hydrogen Fuel Station," PRNNewswire, American Honda Motor Co. Press Release, 2001, July 10.
Gronich, S., "DOE Hydrogen Program Overview," Proceedings of the 2000 U.S. DOE Hydrogen Review, 2000, May 9.
Muradov, N., "Thermocatalytic CO2-Free Production of Hydrogen from Hydrocarbon Fuels," Proceedings of the 2000 U.S. DOE Hydrogen Review, 2000, May 9.
Noble, T., "Microreactors: Smaller Can Be Better," Chemical Eng. Prog., 97 (7), 1011, 2001.
Lombard, J. F., Hydrocarbon Processing, 48 (8), 111 (1969, August).
Lundberg, W. C., Chem. Eng. Prog., 81 (6) (1979, June).
Abrardo, J. M., and Khurana, V., "Hydrogen Technologies to Meet Refiners' Future Needs," Hydrocarbon Processing, 74 (2), 4349 (1995, February).
Chem. Eng. Prog., 69 (2), 6770 (1973, February).
Chem. Eng., 67 (19), 166169 (1960, September 19).
Chem. Eng., 86 (13), 8889 (1979, December 3).
Hydrocarbon Processing, 57 (4), 145151 (1978, April).
"Amine Guard Systems in Hydrogen Production," AIChE, 86th National Meeting, Houston, TX, 1979, April 15.
www.synetix.com, Hydrogen Plant Products, Imperial Chemical Industries PLC, 2001, August 14.
Foreman, J. M., "Pre-Reformer Aids Syngas Units," Hydrocarbon Processing, 69 (12), 34B34D (1990, December).
Farnell, P. W., "Synetix's Advanced Gas Heated Reformer," AIChE Ammonia Safety Symposium, Seattle, WA, 1999, September.
Carstensen, J. H., et al., "New Developments in Shift Catalyst for Ammonia," Hydrocarbon Processing, 69 (3), 5762 (1990, March).
Bakemeier, H., Huberich, T., et al., in Ullmann's Encyclopedia of Industrial Chemistry, Vol. A2, 5th ed., Reinhold Publishing, New York, 1985.
www.haldor.dk, Synthesis Gas Technologies, Haldor Tops⊘e A/S, 2001, August 13.
www.shell.com, "Generating Energy Using Gasification," Royal Dutch/Shell Group of Companies, 2001, August 13.
Bajusz, I. G., and Goodwin, Jr., J. G., "Hydrogen and Temperature Effects on the Coverages and Activities of Surface Intermediates during Methanation on Ru/SiO2," J. Catal., 169, 157165 (1997).
Hoost, T. E., "Chemical Modification of Supported-Metal Catalysis," PhD Dissertation, University of Pittsburgh, Pittsburgh, PA, 1991.
www.itd.anl.gov/highlights/9-2/ceramic.html, Argonne National Laboratory, 2001, August 13.
www.hydrocarbononlline.com, Telepneff, M., "Hyprotech and Synetix Integrate Catalytic Reactor Models," Hydrocarbon Online News, Vert Tech LLC, 2001, August 10.
www.halliburton.com/KBR/KBRTCH, Kellogg Brown & RootA Halliburton Company, 2001, August 22.

www.synetix.com, Ammonia Plant Products, Imperial Chemical Industries PLC, 2001, August 22.
Thompson, D., "New Advances in Gold Catalysis Part II," Gold Bull., 32 (1), 1219 (1999).
Thompson, D., "Highlights from Recent Literature," Gold Bull., 33 (1), 33 (2000).
www.haldor.dk, Ammonia Synthesis, Haldor Tops⊘e A/S, 2001, August 22.
Fleshman, J. D., "Cost Efficient Revamps in Hydrogen Plants," Proceedings of NPRA Annual Meeting, New Orleans, USA , 2001, March.
Fleshman, J. D., McEvoy, L. J., et al., "New Hydrogen Plant Design Achieves Low Cost and High Efficiency," Proceedings of NPRA Annual Meeting, San Antonio, USA , 1999.
Nair, M. P. S., "Control Corrosion Factors in Ammonia and Urea Plants," Hydrocarbon Processing, 80 (1), 8593 (2001, January).
www.dakota.com, The Coal Gasification Process, Dakota Gasification Company, 2001, July 16.
www.kvaerner.com, Compact Reformer Technology, Kvaerner Process Technology, 2001, August 27.
SRI International, Chemical Economics Handbook, Nitric Acid, 2001.
International Critical Tables, Vol. III, pp. 304, 305, 309, McGraw-Hill Book Co., Inc., New York, 1928.
International Critical Tables of Numerical Data, Physics, Chemistry, and Technology, Vol. IV, p. 255, McGraw-Hill Book Co., Inc. New York, 1928.
Perry, R. H., and Chilton, C. H. (Eds.), Chemical Engineers' Handbook, 5th ed., McGraw-Hill Book Co., Inc., New York, 1973.
Rossini, F. D., and Coworkers, National Bureau of Standards, Circular 500, 53 (1952).
Encarta.msn.com, "Ostwald, Wilhelm," Microsoft Corporation, 2001.
www.efma.org/Publications, Production of Nitric Acid, European Fertilizer Manufacturers Association, 2001, September.
Sauchelli, "Fertilizer Nitrogen," ACS Monograph Series 161.
www.enviro-chem.com, Nitric Acid Plant Technology, Monsanto Enviro-Chem, 2001.
www.yet2.com, Nitric Acid Processes, yet2.com Inc., 2001.
www.chemtics.kvaerner.com, Nitric Acid Concentration, Kvaerner Pulp & Paper, 2001.
Powell, R., "Nitric Acid TechnologyRecent Developments," Noyes Development Corp., Park Ridge, NJ, 1969.
www.efma.org/Publications, Production of Ammonium Nitrate, European Fertilizer Manufacturers Association, 2001, September.
Kirk-Othmer, Encyclopedia of Chemical Technology, Vol. 17, pp. 80107, John Wiley & Sons, Inc., New York, 1997.
www.agcentral.com/imcdemo, Production of Major Nitrogen Fertilizers, IMC-Agrico Company, 2001.
www.sasol.com/smx, Ammonium Nitrate, Sasol SMX, 1998.
Chemical Week, 101 (2), 75, 1967, July 8.
Kirk-Othmer, Encyclopedia of Chemical Technology, Supplement, pp. 597621, John Wiley & Sons, Inc., New York, 1996.
Meessen, J. H., and Petersen, H. et al., in Ullmann's Encyclopedia of Industrial Chemistry, Vol. A27, 5th ed., Reinhold Publishing, New York, 1996.
www.efma.org/Publications, Production of Urea and Urea Ammonium Nitrate, European Fertilizer Manufacturers Association, September 2001.
www.agrium.com, Carseland Nitrogen Operations, Agrium, Canada , 3/01.
www.enviro-chem.com, Urea Plant Technology, Monsanto Enviro-Chem, 2001.
SRI International, Chemical Economics Handbook, Urea, 2001.
SRI International, Chemical Economics Handbook, Melamine, 1999.
"DSM Expands Melamine Production in Netherlands," European Paint and Resin News, 38 (12), 5, 2000, December.
SRI International, Chemical Economics Handbook, Alkylamines, 2001.
www.kvaerner.com, Industrial Amine Process Technologies, Kvaerner Process Technology, 2001, August 27. (May be under www.davyprotech.com as of 2001, November).
Chemical Marketing Reporter (2001, June 4).
Chemical Marketing Reporter (2000, April 24).
"Ethanolamines," Eur. Chem. News, 74 (1953), 16 (2001, April 1622).
www.owlnet.rice.edu/~ceng403/ethamsep.html, C. A. Garcia, M. Reisdorf, and L. Palombo, "Separation Unit for Ruehl Corp. Ethanolamine Plant," Rice University, 1997, October 25.
www.himtek.ru/english/technology1.shtml, Technology of Ethanolamine Production, Himtek Engineering, 2000.
Chem. Mark. Rep. (2001, June 11).
Chem. Mark. Rep. (1999, August 9).
www.chemicals-india.com, Hexamine Manufacturing Process, Kothari Phytochemicals International, 2001, October 5.
www.wrightcorp.com, Formaldehyde and Hexamine, Wright Corporation, 2001, October 5.
www.saudiform.com, Hexamethylenetetramine (HMT), Saudi Formaldehyde Chemical Company Ltd., 1999.
www.hydrazine.com, About Arch Hydrazine, Arch Hydrazine, 2001.
Kirk-Othmer, Encyclopedia of Chemical Technology, 4th ed., Vol. 7, pp. 753782, John Wiley & Sons, Inc., New York, 1993.
D. Hasenberg, "HCN Synthesis on Polycrystalline Platinum and Rhodium," Dissertation, University of Minnesota, 1984.
SRI International, Chemical Economics Handbook, Hydrogen Cyanide, 2000.
Schirmann, J. P., in Ullmann's Encyclopedia of Industrial Chemistry, Vol. A13, pp. 177190, Reinhold Publishing, New York, 1996.
Botz, M., "Cyanide Treatment Methods," Mining Environ. Manage., 9 (3), 2830 (2001, May).
"Hypergolic Propellant Hazard Response Guide, Cape Canaveral Air Force Station (CCAFS)," Draft Volume I, ICF Technology Incorporated, 1988, July 1.
www.hydrazinehydrate.com, Hydrazine Hydrate: Safety and Health, Bayer Corporation, 2001.
"Kvaerner Awarded Seal Alliance Contracts by BASF and Ineos," Hydrocarbon Processing, Int. ed., 80 (3), 41 (2001, March).
www.epa.gov/ttn/chief/le/benzene/benz.cfe.pdf., "Emissions from Major Uses of Benzene," Section 5E, EPA-454/4-98-011, Locating & Estimating Documents, Technology Transfer Network, United States EPA, pp. 5-585-61, 1998, June.
SRI International, Chemical Economics Handbook, Aniline, 1997.
www.che.LSU.edu/courses/4171/computerprojects1.htm, Aniline Process Flow Diagram, Louisiana State University, 2001, September.
"Production and Consumption of Aniline in the World," Xiandai Hyagong, 18 (6), 3436 (1998, June).
Aniline Chemical Profile, Chemical Marketing Reporter (1999, February 8).
ntp-server.niehs.nih.gov/htdocs/LT-studies/TR163.html, Bioassay of Calcium Cyanamide for Possible Carcinogenicity, 1979.
"Catalytic Processes in Nitric Acid Manufacture," Fertilizer Society Proceedings, London, 1978.

< previous page

next page >

30.1 Introduction

Scope of the Chapter

This chapter deals with the chemicals used in agriculture mainly to protect, preserve, and improve crop yields. The term "agrochemical" is used broadly. Much agrochemical research and some advanced development is directed toward the introduction of genes that may provide disease, insect, or viral resistance into plants or other organisms. Further progress is being made in improving the protein, fat, or carbohydrate composition of the plant itself. Microorganisms are being propagated and currently marketed that are insecticidal (e.g., *Bacillus thuringiensis*) fungi that are herbicidal, bacteria that are fungicidal, nematodes that are widely biocidal, and so on, are all products or candidate products for use in agriculture.

Arbitrarily excluded from discussion in this chapter are those substances that serve as fundamental nutrients, which are treated in Chapters 11 and 29 on fertilizer and nitrogen technology, respectively. Nevertheless, it is the current practice of the farmer, particularly in advanced agriculture, to integrate nutritional and plant-protection application schedules, and even provide single formulations that include both fertilizers and pesticides. Further, plant nutrition at this stage of scientific sophistication is far more complex than the older classical "N-P-K" applications alone.

Many chemicals that accelerate plant growth act as hormonal agents, modifying plant metabolic processes at some stage of development. Because these substances are manufactured and marketed by the agrochemical industry, they are included as subject matter here. Also included in this chapter are chemicals that are significant to public health. Many organisms are vectors in the dissemination of human and animal disease. Because products of the pesticide industry control the insect, the rodent, the mollusk, and so forth (the vectors), they often are the most effective and sometimes the only practical means for controlling some of the most serious health problems of humankind, especially, but not exclusively, in the underdeveloped countries. A historical analog would be the use of rodenticides in the control of plague.

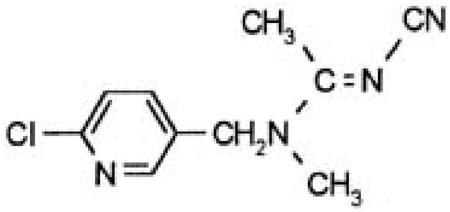
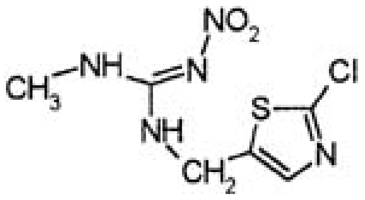
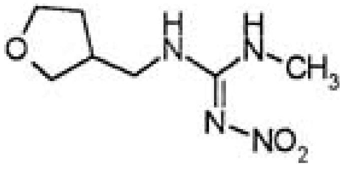
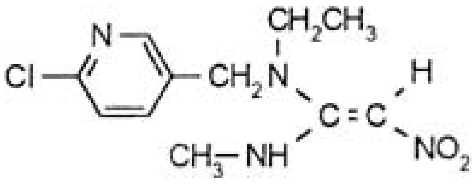
History

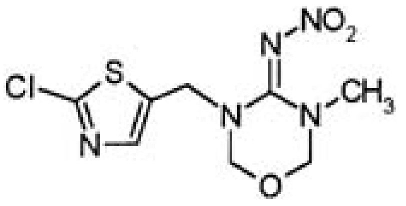
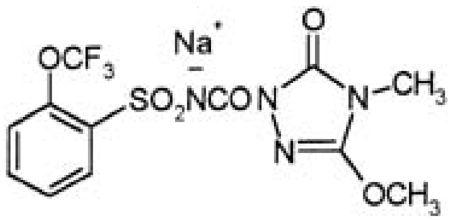
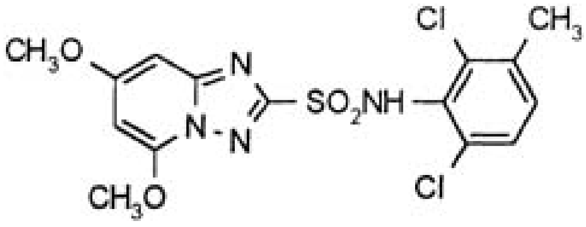
It is probable that farmers' treatment of crops with foreign substances dates back into prehistory. The Bible abounds with references to insect depredations, plant diseases, and some basic agricultural principles such as periodic withholding of land in the fallow state. Homer speaks of "pest-averting sulfur." More recently, in the nineteenth century, there was a great increase in the application of foreign chemicals to agriculture. Discovered or, more precisely, rediscovered was the usefulness of sulfur, lime sulfur (calcium polysulfides), and Bordeaux mixture (basic copper sulfates). With the exception of the organic compound formaldehyde, inorganic chemicals provided farmers with their major weapons.

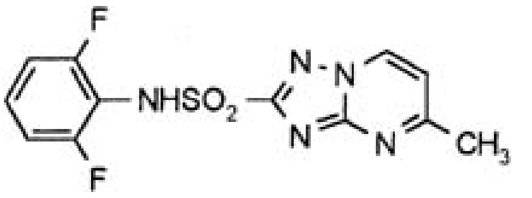
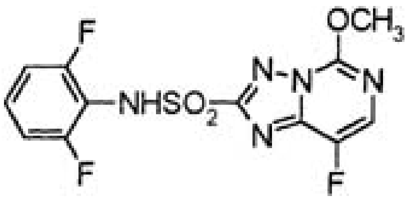
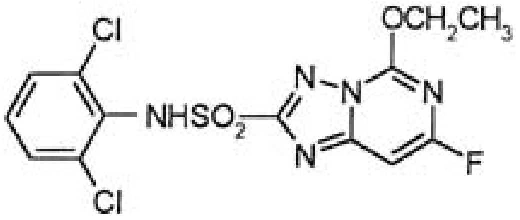
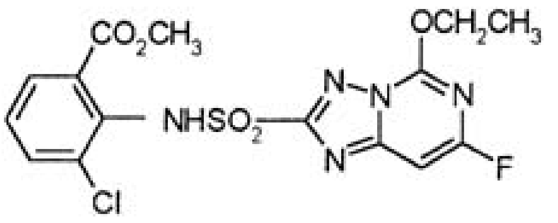
The earliest of the organic compounds generally were chemicals derived from natural products or crude mixtures of chemicals in states of very elementary refinement. Extracts of ground-up plant tissue were useful in the control of insects. Such extracts were employed in agriculture quite often before the chemist had elucidated the structure or synthesized the molecule responsible for biological activity. These extracts included the pyrethroids, rotenoids, and nicotinoids, which continue to be derived in large part from plant extracts. Crude petroleum fractions were recognized for their effectiveness in the control of mites, scale, and various fungi, as well as for their phytopathological properties.

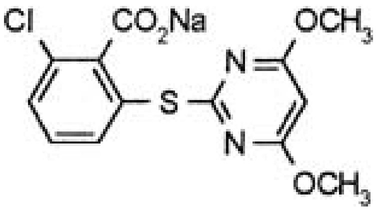
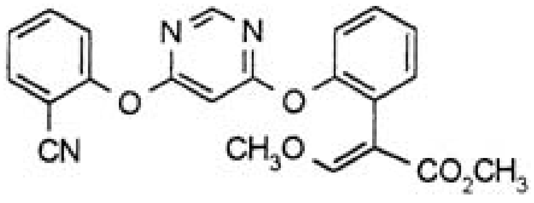
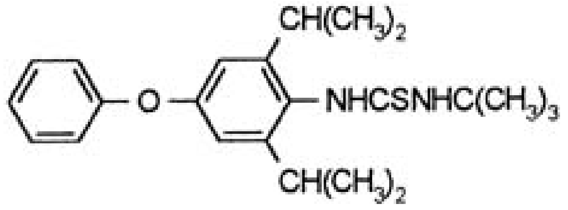
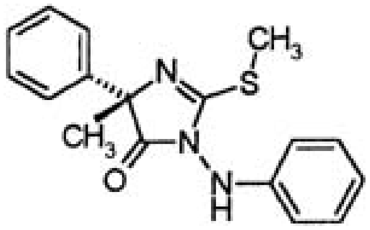
Although a few synthetic organics were already known, the great revolution in the use of organic chemicals in agriculture roughly coincides with the period of the onset of World War II. The more important of these discoveries were DDT (invented by Othmar Zeidler in late 1800s, insecticidal properties discovered by Mueller in 1939), 2.4-D (Jones patent1945), benzenehexachloride (ICI and French developmentca. 1940), and the organic phosphate esters (Schraderbegun in the late 1930s, revealed in the 1940s). These new chemicals were so enormously more potent than their predecessors intheir biological activity (frequently by orders of magnitude) that they very rapidly displaced almost all of the chemicals previously employed. The classical chemicals of today, some discovered in the 1950s and 1960s, are predominantly extensions of this almost revolutionary transition from inorganics to synthetic organics that dates from the period of World War II (Table 30.1). It is fair to say that within the United States the agrochemical industry since the late 1950s has been dominated by synthetic organic chemicals.

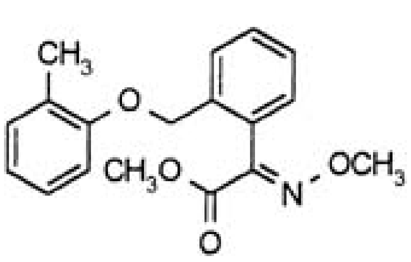
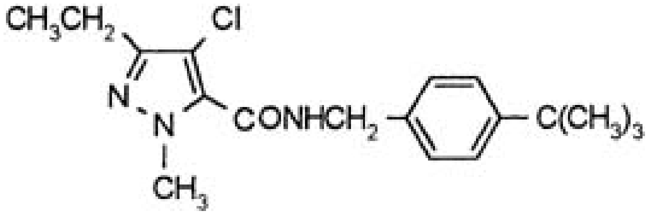
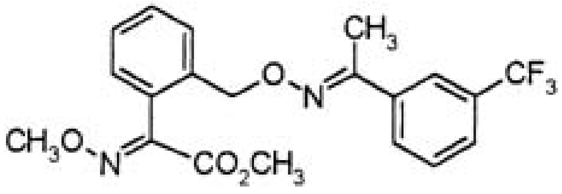
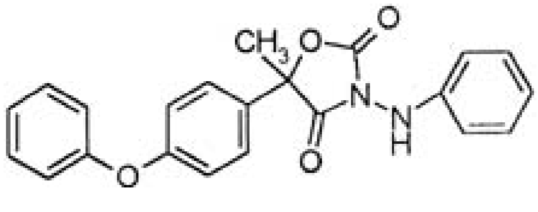
TABLE 30.1 New Chemistry

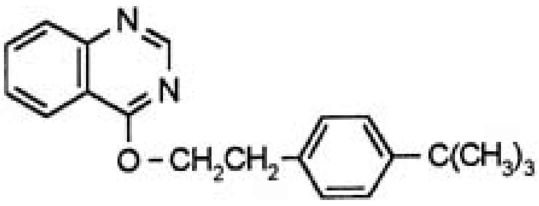
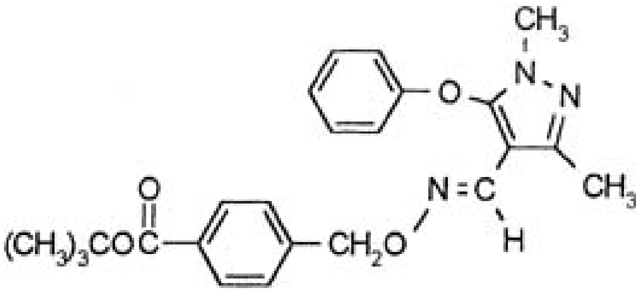
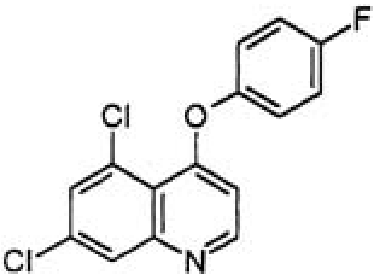
Chemical (Chemical Class) (Year of Reporting)	Company's Name	Structure	Biochemistry	Mode of Action	Known Profile of Use	Environmental Fate
Acetamiprid (Neonicotinoid) (1992)	Nippon soda		Nicotinic acetyl-choline receptor agonist	Systemic insecticide with translaminar activity and with contact and stomach action	Control of Hemiptera, Thysanoptera and Lepidoptera by soil and foliar application on a wide range of crops especially vegetables, fruit, and tea	*Plants*. Slowly degraded on or in plants, forming five identified metabolites *Soil*. DT50 in clay loam 1 day, in light clay 12 days. DT50 for total residues 1530 days
Clothianidin* (Nitromethylene neonicotinoid)	Takeda		Nicotinic acetyl-choline receptor agonist	Insecticide affecting the synapses in insects' central nervous system	Control of sucking and chewing insects by soil, foliar and seed application on rice, fruit, and vegetables	
Dinotefuran* (Nitromethylene neonicotinoid) (1998)	Mitsui		Nicotinic acetyl-choline receptor agonist	Insecticide active by ingestion and contact; also exhibits root systemic activity	For control of a range of Hemipterous and other pests	
Nitenpyram (Nitromethylene neonicotinoid (1993)	Takeda		Nicotinic acetyl-choline receptor agonist	Systemic insecticide with translaminar activity with contact and stomach action	Control of aphids, thrips, leafhoppers, whitefly and other sucking pests on rice and glass house crops	*Soil*. DT50 in soil 115 days, depending on soil type

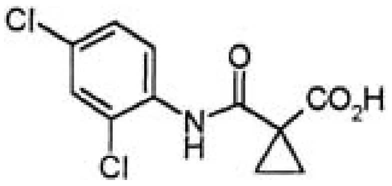
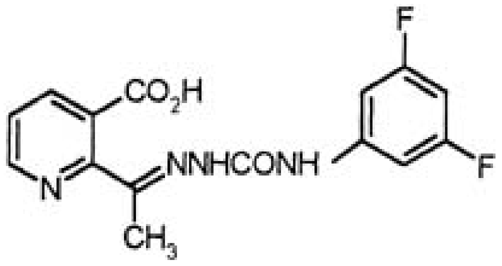
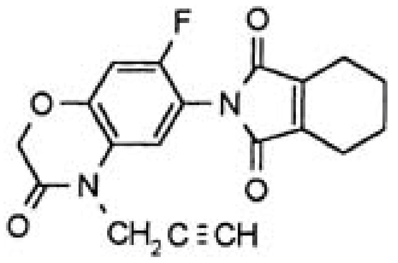
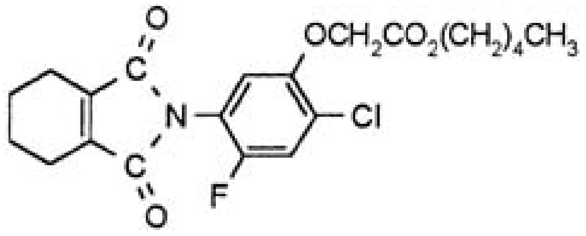
Thiamethoxam (Neonicotinoid) (1998)	Novartis		Nicotinic acetyl-choline receptor agonist	Insecticide with contact stomach and systemic activity	For the control of aphids, whitefly, thrips, ricehoppers, ricebugs, ealybugs, whitegrubs, olorado potato beetle, flea beetles, wireworms, ground beetles, leaf miners and some lepidopterous species. Foliar and soil applications: cole crops, leafy and fruity vegetables, potatoes, rice, cotton, deciduous fruits, citrus, tobacco, soybeans. For seed treatment: maize, sorghum, cereals, sugarbeet, oil seed rape, cotton, peas, beans, sunflower, rice, potatoes	*Animals*. Quickly and completely absorbed, rapidly distributed in the body and rapidly eliminated *Plants*. Degradation/ metabolism has been studied in six different crops with soil, foliar and seed treatment application. The qualitative metabolic pattern was similar for all types of applications and for all studied crops *Soil*. Soil DT50 (median) 51 days
Propoxycarbazone* (Sulfonylamino-carbonyltriazolinone) (1999)	Bayer		Amino-acid synthesis inhibitor	Herbicide translocated both acro-petally and basipetally within both xylem and phloems	Post-emergence control of annual and some perennial grasses and some broadleaf weeds in wheat, rye, triticale	*Animals*. Rapid and nearly complete (>88%) within 48 hr, primarily via faeces; 7589% unchanged parent compound in urine and faeces *Plants*. The unchanged parent compound and its 2-hydroxypropoxy metabolite are regarded as the relevant residues for plants *Soil*. Soil DT50 c. 36 days. Field dissipation DT50 c. 9 days
Flucarbazonesodium (Sulfonylamino-carbonyl-triazolinone) (1999)	Bayer		Amino-acid synthesis inhibitor	Herbicide absorbed through foliage and roots and translocated acropetally and basipetally	Wheat for post-emergence control of grass weeds especially *Avena fatua* and *Setaria viridis* and some broad-leaved weeds	*Animals.*Almost completely excreted via faeces and urine within 48 hr *Plants*. Extensively metabolized. The relevant residues are the parent compound and the *N*-desmethyl metabolite *Soil*. Average soil DT50 17 days. Not mobile in soil
Metosulam (Trizolopyridine) (1993)	Dow Agro-Sciences		Amino-acid synthesis inhibitor	Herbicide readily taken up by roots and foliage	Post-emergency control of many important broadleaf weeds including *Galium aparine*, *Stellaria media*, *Chenopodium* spp. *Amaranthus retroflexus* etc. in wheat, barley, rye, maize	*Animals*. Rapidly absorbed. DT50 <1 hr, extensively metabolized in rodents, much less in dogs, and excreted with metabolites 3-hydroxy (aliphatic oxidation) and 5-hydroxy (*O*-demethylation) in urine (DT50 5460 hr in rodents, 73 hr in dogs) *Plants*. Poorly absorbed

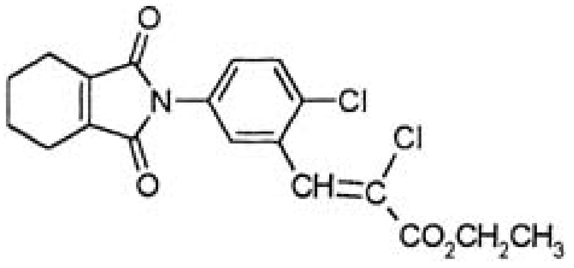
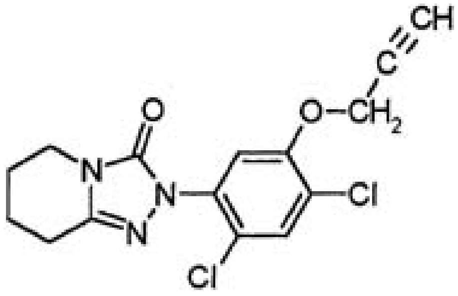
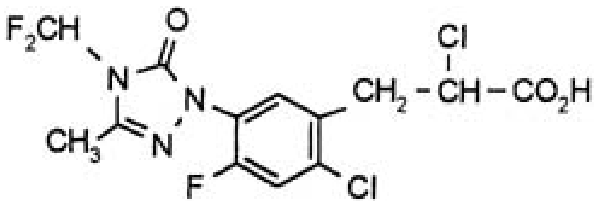
						metabolized by hydroxylation of the ring methyl, to give a 3-hydroxymethyl-metabolite and its glycoside *Soil*. Field DT50 in the 010 cm horizon has a mean value of 25 days degradation via the 5- and 7-hydroxy analogue to 5-amino-*N*-(2,6-dichloro-3-methylphenyl)-1*H*-1,2,4-triazole-3-sulfonamide and CO2. Does not have leaching potential
Flumetsulam (Triazolopyrimidine	Dow Agencies		Amino-acid synthesis inhibitor (ALS or AHAS)	Systemic Herbicide absorbed by roots and leaves of plants and translocated to growth points	Used alone and in combination with trifluralin or metolachlor for control of broad-leaved weeds and grasses in Soyabeans, field peas, maize	*Animals*. Rapidly cleared via urine and faeces with no metabolities. 5-Hydroxy metabolite found in the hen *Plants*. DT50 in maize 2 hr soya beans 18 hr, *Chenopodium* 131 hr. Metabolites depend on the species; 5-hydroxy or 5-methoxy derivatives are common S*oil*. DT50 in soil (25° C, pH 67, o.m. content 24%, 12 mos DT50 in soil (pH 67, o.m. content 24%) 12 months
Florasulam (Triazolopyriidine))	Dow Agencies (1999)		Aminoacid synthesis inhibitor (ALS or AHAS)	Herbicide taken up by both roots and shoots, and translocated in both xylem and phloem	Post-emergence control of broad-leaved weeds especially *Galium aparine, Stellaria media, Polygonum convolvutus, Matricaria* spp., and various cruciferae in cereals and maize	*Soil*. DT50 218 days; neither florasulam nor its degradates should leach
Diclosulam (Triazolopyrimidine)	Dow Agencies		Acetolactate synthase inhibitor	Herbicide taken up by roots and foliage and translocated to new growing points. Lethal amounts in meristems halting cell division and resulting in plant death	Soil applied control of broad-leaved weeds in Peanuts, soyabeans	*Animals*. Metabolized primarily by dealkylation of the ethoxy group and hydrolysis of the sulfonamide linkage *Soil*. Dissipation occurs primarily through microbial degradations oil DT50 (in a wide variety of soils) c. 3365 days
Cloransulam-methyl (Triazolo-pyrimidine (1997)	Dow Agencies		Acetolactate synthase inhibitor	Herbicide having plant meristems as the primary site of activity	Control of broad-leaved weeds in Soyabeans. Applied to the soil surface or incorporated pre- or post-emergence	*Animals*. In female rats, excreted mainly via the urine; in male rats, excreted in both urine and faeces. After 72 hr, <0.1% of the) dose was found in any tissue *Soil*. Photolysis on soil surface, DT50 3070 days (corrected for metabolism) The apparent transformation DT50 in aerobic soils 913 days

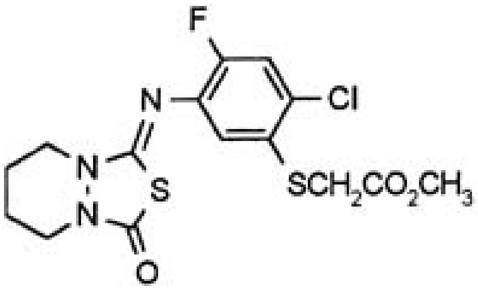
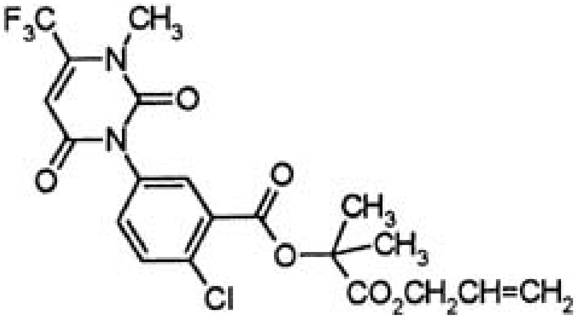
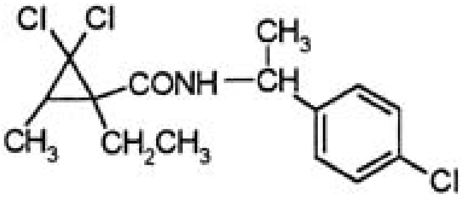
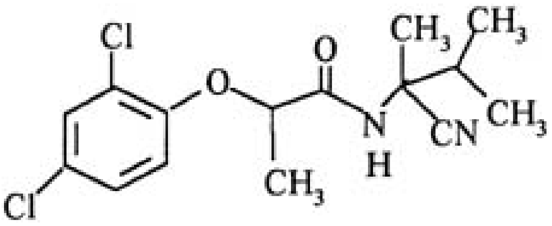
						(est.). May be mobile
Pyrithiobac-sodium (Pyrimidinyl oxybenzoic analogue (1991)	Ihara/ Kumaio		Amino-acid synthesis inhibitor (ALS or AHAS)		Pre- and post-emergence control of wide range of broad-leaved weeds in cotton	*Animals*. More than 90% excreted in urine and faeces within 48 hr; the major excreted metabolite was the *O*-desmethyl derivative. *Plants*. At 62 dat, no residues were found; major metabolites were the phenol formed by mono-demethylation, and its glucose conjugate *Soil*. Microbial and photochemical degradation play a major role in degradation t; DT50 in silty soil 60 days
Azoxystrobin (Strobilurin) (1992)	Zeneca		Inhibitor of mitochondrial respiration	Fungicide with protectant, curative, radicant, translaminar and systemic properties	Control of a number of pathogens in cereals, rice, vines, cucurbits, potato, tomato, peanuts, peach, turf, banana, pecan, citrus, coffee	*Animals*. Majority of radiolabel is excreted in the faeces. Of a large number of metabolites, only the glucuronide of azoxystrobin acid is present at >10% of the administered dose *Plants*. Metabolism was extensive, but parent azoxystrobin was the only major (>10%) residue *Soil*. In soil, in the dark, six identified metabolites were formed; over 1 year, 45% of applied radiolabel is evolved as CO2. DT50 18 weeks. Low to moderate mobility in soil; typical Koc for azoxystrobin c. 500
Diafenthiuron (Strobilurin) (1998)	Novartis		Converted by light into the corresponding carbodiimide, which is an inhibitor of mitochondrial respiration	Insecticide, Acaricide having contact and stomach action. Also shows some ovicidal action	Control of phytophagous mites, Aleyrodidae, Aphidadae, and Jassidae and some leaf feeding pests in cotton, various field and fruit crops, ornamentals, vegetables	*Animals*. Major portion excreted with the faces. Degraded to yield its corresponding carbodiimide, which, in turn forms urea and fatty acid derivatives *Plants*. Shows a complex metabolism pattern *Soil*. Diafenthiuron and its main metabolites show a strong sorptivity to soil particles. Degradation in soils proceeds rapidly: DT50 <1 hr to 1.4 days
Fenamidone* (strobilurin) (1992)	Aventis		Inhibitor of mitochondrial respiration by blocking electron transport at ubihydroquinone; cytochrome oxido reductase	Protectant and aerative fungicide	Under development for control of a range of Oomycete diseases	

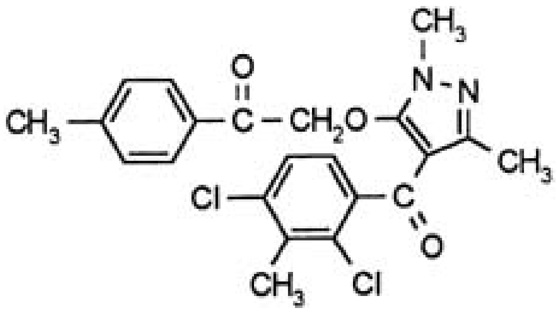
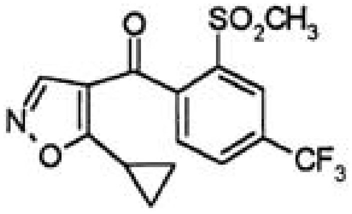
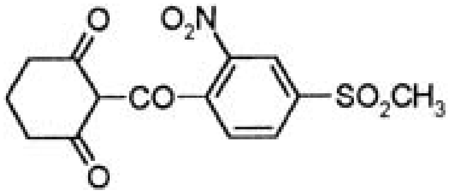
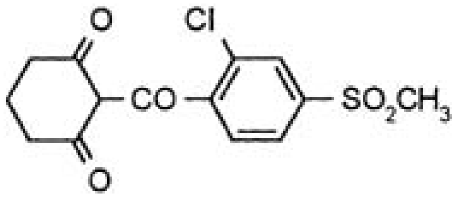
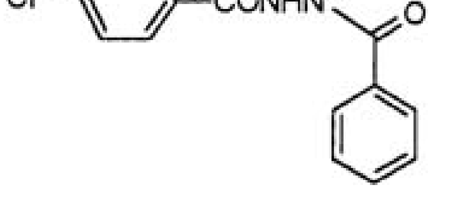
Kresoxim-methyl Strobilurin type: (Strobilurin) (1992)	BASF		Inhibitor of mitochondrial respiration by blocking electron transfer between cytochrome b and cytochrome C1	Fungicide with protective, curative, eradicative and long residual disease control	Control of scab, powdery mildew, mildew, scald, net blotch, glume blotch on apples, pears, vines, cucurbits, Sugarbeet, cereals, vegetables	*Animals*. Widely distributed and quickly eliminated; no bioaccumulation. The major routes of excretion were faeces and urine. Thirty two different metabolites were identified *Plants*. Residues in cereals and pome fruit at harvest are <0.05 mg/kg, in grapes and vegetables <1 mg/kg *Soil*. Rapidly degraded. In soil, DT90 (lab.) <3 days, the main metabolite is the corresponding acid. Very mobile in soil. However, in lysimeter studies, only low levels of kresoxim-methyl and its metabolite were found in leachates
Tebufenpyrad (Pyrazole) (1993)	Mitsubishi chemical		Mitochondrial respiration Inhibitor. Acts as an inhibitor of electron transport chain at site I	Non-systemic Acaricide active by contact and ingestion. Exhibits translaminar movement following application to leaves	Control of all stages of *Tetranychus*, *Panonychus*, *Oligonychus*, *Eotetranychus* spp., on fruit, vines, citrus, vegetables, hops, ornamentals, melons, cotton	*Animals*. Metabolite is *N*-[4-(1-hydroxymethyl-1-methyl-ethyl) benzyl]-4-chloro-3-(1-hydroxyethyl)-1-methyl-pyrazole-5-carboxamide *Plants*. As for animals *Soil*. Aerobic degradation occurs in soil, DT50 2030 days
Trifloxys-trobin Strobilurin type: (Strobilurin) (1998)	Novartis		Inhibits mitochondrial respiration by blocking electron transfer at the Q0 centre of Cytochrome bcl	Mesostemic broadspectrum fungicide with preventive and specific curative activity and displaying rain-fastness. Redistributed by superficial vapour movement and also have translaminar activity	For control of powdery mildew, leaf spots, rusts, bunch and fruit rots of cereals, Pome fruit, grapes, Peanuts, bananas, vegetables	*Animals*. Absorbed from the gastrointestinal tract, rapidly metabolised and quickly and completely eliminated from the body *Plants*. Rapidly degraded *Soil*. Dissipates rapidly. DT50 4.29.5 days. No leaching potential
Famoxadone* Strobilurin type: (Oxazolidinedione) (1996)	DuPont		Inhibits mitochondrial electron transport, by blocking ubiquinolcytochrome C oxido-reductase at complex III	Protectant translaminar and residual Fungicide	Control of mildew, potato and tomato late and early blights, wheat leaf and glume blotch and barley net blotch in grape, potato, tomoto, wheat, barley	*Animals*. Elimination is rapid. Unmetabolised famoxadone was the major component in the faeces; mono- (at 4′-phenoxyphenyl) and di- (also at 4-phenylamino) hydroxylated famoxadone were the primary faecal metabolites. In urine, products arising from cleavage of the heterocyclic ring were found. Metabolism was complex, involving hydroxylation, cleavage of the oxazolidinedione-aminophenyl linkage, cleavage of the phenoxyphenyl ether linkage and opening of the oxazolidinedione

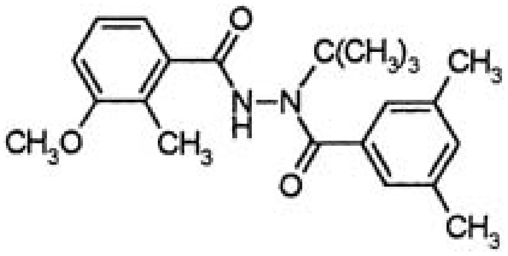
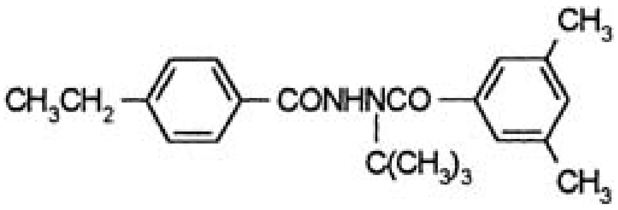
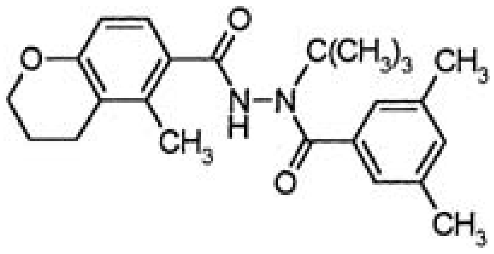
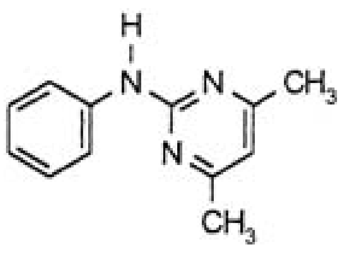
						ring *Plants*. In grapes and potatoes, famoxadone was the main residue; no residues were found in potato tubers. In wheat, famoxadone was extensively metabolised, primarily by hydroxylation, followed by conjugation *Soil*. In laboratory soil, DT50 6 days (aerobic), 28 days (anaerobic). Degradation routes include hydroxylation (at the 4′-phenoxyphenyl position), ring opening (with formation of a glycolic acid derivative), and is primarily microbial
Fenazaquin Strobilurin type: (Oxazolidinedione) (1992)	Dow Agro-sciences		Inhibitor of mitochondrial electron transport chain by binding with complex I at co-enzyme site Q	Contact Acaricide with good knockdown activity on motile foms as well as true ovicidal activity. Preventing eclosion of mite eggs	Control of *Euteranychus, Panonychus, Tetranychus, Brevipalpus phoenici* in almonds, apples, citrus, cotton, grapes, ornamentals	
Fenpyro-ximate Pyrazole (acaricide) (1990)	Nihon Nohyaku		Inhibitor of mitochondrial electron transport at complex I	Acaricide having quick knockdown activity against larvae, numphs and adults maily by contact and ingestion	Effective against some phytophagus mites. Control of Tetranychidae, Tarsonemidae, Penuipalpidae and Eriophyiade. In citrus, apple, pear, peach, grapes	*Soil*. DT50 26.349.7 days
Quinoxyfen (Quinoline) (1996)	Dow Agro-sciences		Growth signal disruptor	Mobile, Protectant fungicide acting through inhibition of appressorial development. Active through systemic acropetal and basipetal movement and by vapour transfer	Control of powdery mildew in cereals	*Plants*. Only slightly metabolized in wheat, with low residues found in the crop. Extensively photo-degraded on the wheat leaf surface, giving multiple polar degradation products. On grapes and cucumbers, the main residue was unchanged quinoxyfen *Soil*. DT50 (field) 123494 days (biphasic); non-leaching. The main metabolite in the soil (also classed as non-leaching) was formed by hydroxylation at the 3-position of the quinoline ring; a minor metabolite (DCHQ), formed by cleavage of the ether bridge, was observed, especially in acidic soil

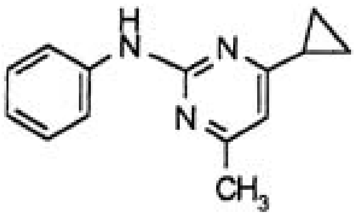
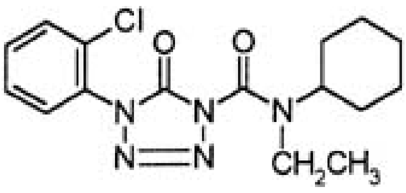
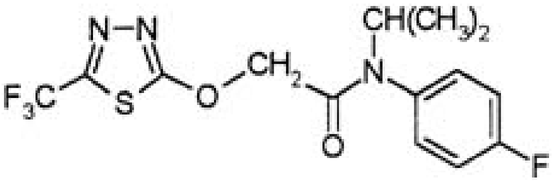
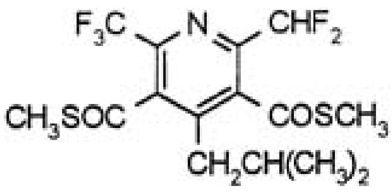
Cyclanilide Strobilurin type: (Oxazolidinedione) (1994)	Aventis		Inhibits Polar auxin transport	Plant growth regulator	Cotton and other crops	*Animals*. Rapidly excreted, primarily as unchanged cyclanilide *Plants*. Little degradation occurs in plants; cyclanilide is the major residue *Soil*. Low to moderate persistence, DT50 c. 16 days under aerobic conditions. Degrades primarily by microbial activity. Medium to low mobility
Diflufenzopyr (Semi-carbazone) (1999)	BASF		Inhibits auxin transport, apparently by binding with a carrier protein on the plasmalemma	Systemic, post-emergence herbicide	Control of annual broad leaved and perennial weeds in maize	*Animals*. Partially absorbed and rapidly eliminated; 2044% of the dose was eliminated in urine and 4979% in faeces. Total radioactive residues in tissues <3% of the administered dose. Eliminated primarily as unchanged parent compound *Soil*. Average DT50 in field soil 4.5 days. Very mobile Metabolities also very mobile. However, based upon proposed use, US EPA does not expect diflufenzopyr to reach drinking water
Flumioxazin *N*-phenyl-phthalimide)	Sumitonio		Protoporphyrinogen oxidase inhibitor	Herbicide absorbed by foliage and germinating seedlings	Control of many annual grasses, pre- and post-emergence in soyabeans, peanuts, orchards and other crops	
Flumiclorac-pentyl *N*-phenyl phthalimide (1998)	Sumitono valent		Protoporphyrinogen oxidase inhibitor	Fast acting, contact herbicide. When applied to foliage of susceptible plants. It is readily absorbed into plant	Control of problem broad leaved weeds including *Xanthiu strumarium*, *Chenopodium album*, *Ambrosia artemisifolia*, *Datura stramonium*, *Amarnthus* sp., *Sida spinosa*, *Euphorbia maculata*, *Abutilon theophrasti*, pre- and post-emergence in soyabeans and maize	*Plants*. In soya beans and maize, the major metabolite is 2-chloro-4-fluoro-5-(4-hydroxy-1, 2-cyclohexane-dicarboximido) phenoxyacetic acid formed by reduction of the tetrahydrophthaloyl double bond and hydroxylation; other metabolic pathways include cleavage of the ester, and cleavage of the imide linkage *Soil*. Degrades in soil by microbial and 0.484.4 days in loamy-sand soil (pH 7); degradates have DT50 c. 230 days. The a.i. is immobile in soil; degradates have low to medium mobility

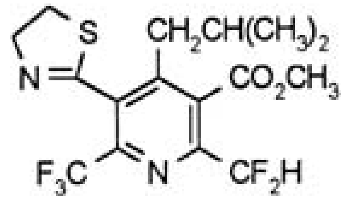
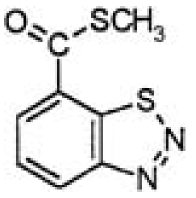
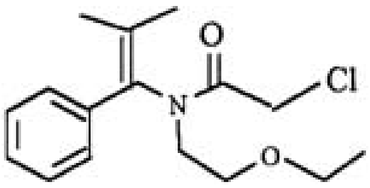
Cinidon-ethyl *N*-phenyl phthalimide (1998)	BASF		Protoporphyrinogen IX oxidase inhitor		Post-emergence control of annual broad-leaved weeds especially *Galium aparine*, *Lamium* sp. and *Veronica* sp. in winter and spring small grain cereals	*Animals*. Following limited, but rapid, absorpotion, and widespread distribution in organs and tissues, the a.i. is extensively metabolised and rapidly excreted *Plants*. The a.i. is extensively metabolised *Soil*. Readily biodegradable. Soil DT50 0.62 days (lab., aerobic conditions, 20° C); rapidly mineralized
Azafenidin (Triazolinone) (1998)	Dupont		Protoporphyrinogen oxidase inhibitor	Herbicide absorbed through roots and shoots. Improves the efficacy of other post-emergence herbicides and increases the speed of action of contact herbicides	Pre-emergence control of annual and perennial weeds in citrus, grapes, olives, sugarcane and other perennial crops	*Soil*. Degrades in soil by microbial and photolytic process. In the field, in a range of 4 soils, mean DT50 c. 25 days mean DT90 c. 169 days. There was minimal movement in soil column leaching studies
Carfentrazone-ethyl (Triazolinone) (1993)	FMC		Protoporphyrinogen oxidase inhibitor	Herbicide absorbed by foliage with limited translocation	Post-emergence control of a wide range of broad leaved weeds especially *Galium aparine*, *Abutilon theophrasti*, *Ipomoea hederacea Chenopodium album* and several mustard species in cereals	*Animals*. c. 80% is rapidly absorbed and excreted in the urine within 24 hrs. The major metabolite was the corresponding acid. Further metabolism appears to involve oxidative hydroxylation of the methyl group or dehydrochlorination to form the corresponding cinnamic acid *Plants*. Rapidly converted to the free acid, which is hydroxylated and then oxidised at the triazolinone methyl to form the dibasic acid; DT50 (carfentrazone-ethyl) <7 days, DT50 (carfentrazone) <28 days *Soil*. Broken down in the soil by microbial action. Strongly adsorbed to serile soils in non-sterile soils, rapidly converted to the free acid, which has low soil. In the laboratory, soil DT50 is a few hours, degrading to the free acid, which in turn has DT50 2.54.0 days

Fluthiacet-methyl (Thiadiazole) (1993)	Ihara/ Kumai (Also reported by Ciba-Geigy AG)		Protoporphyrinogen oxidase inhibitor	Selective, herbicide requiring light for activity	Post-emergence control of broad-leaved weeds e.g. *Abutilon theophrasti, Chenopodium album, Amaranthus retroflexus, Xanthium strumarium* in *maize* and soyabeans	*Animals*. Within 48 hr 80% is eliminated via the faeces, 14% via urine. Metabolism proceeds via hydrolysis of the methyl ester, isomerisation at the thiadiazole ring and hydroxylation of the tetrahydropyridazine moiety *Plants*. Organosoluble metabolites are similar to those in animals *Soil*. DT50 (hydrolysis, pH 7) 18 days (photolysis on soil) 21 days (UV light) 2 hr. In loam soil, DT50 1.2 days (25°C, 75% of max. water capacity
Butafenacil (Pyrimidindione) (1998)	Novartis		Protoporphyrinogen oxidase inhibitor	Non-selective contact herbicide, rapidly absorbed by the foliage. Translocation occurs only within leaves	Control of a wide range of annual and perennial broadleaved weeds in fruits, orchards, vineyards, citrus, non-crop land	*Soil*. Rapidly degraded in soil: DT50 12 days
Carpropamid (MBI: dehydrase) (1994)	Bayer		Inhibitor of melanin biosynthesis, by inhibiting the dehydration reactions from Scytalone to 1,3,8-trihydroxy naphthalene and from vermelone to 1,8-dihydroxy naphthalene	Systemic, protective fungicide	Control of *Pyricularia oryzae* as protective treatment or seed treatment	*Animals*. Readily excreted via faeces and urine. Metabolised oxidatively, mainly in the liver *Plants*. Absorbed by the roots and translocated to the shoots. The major residue in rice was carpropamid *Soil*. Metabolised oxidatively under paddy soil conditions; CO2 was the major metabolite. The calculated half-lives ranged from several weeks to several months, respy. Low mobility
Fenoxanil* (MBI: dehydrase)	BASF, Nihon Nohyaku		Melanin biosynthesis inhibitor Inhibits dehydratase enzymes, which dehydrate scytalone to trihydroxy naphthalene and vermelone to dihydroxy naphthalene	Systemic, protective fungicide with residual effects	Under development for control of rice blast by foliar or into-water application in rice	

Benzofenap (Pyrazole)	Mitsubishi/ Acentis		*p*-Hydroxy-phenyl pyruvate dioxygenase inhibitor	Systemic Herbicide, absorbed principally through root and bases of target weeds	Used in combination with pyributicarb and bromo-butide, controls annual and perennial broadleaved weeds in rice	*Plants*. No detectable residues in rice crops (detection limit 0.005 ppm). *Soil*. DT50 38 days. Non-mobile
Isoxaflutole (Isozaxole) (1995)	Aventis		*p*-Hydroxy-phenyl pyruvate dioxygenase inhibitor	Herbicide, Systemic by either root or foliar uptake.	For pre-emergence or pre-plant broad spectrum grass and broad-leaved weeds control in maize	*Animals*. Rapidly excreted *Plants*. Residue levels at harvest are very low, and comprise mainly a non-toxic metabolite *Soil*. Degradation proceeds via hydrolysis and microbial degradation, with final mineralization to CO2. Isoxaflutole and its major metabolites are non-mobile under field conditions
Mesotrione (Triketone) (1999)	Zeneca		*p*-Hydroxy-phenyl pyruvate dioxygenase inhibitor	Herbicide, Uptake is foliar and via the root, with both acropetal and basipetal translocation	Pre- and post-emergence control of broad leaved weeds such as *Xanthium strumarium*, *Ambrosia trifida*, *Abutilon theophrasti*, *chenopodium*, *Amaranthus* and *polygonum* spp. and some grass weeds in maize	*Soil*. Stable to hydrolysis under sterile conditions at pH 59, with <10% degradation after 30 days (25°C). Degradation is influenced by soil pH; DT50 31.5 days (pH 5.0% o.c. 2.0) to 4.0 d (pH 7.7% o.c. 0.9).
Sulcotrione (Triketone) (1991)	Zeneca		*p*-Hydroxy-phenyl pyruvate dioxygenase inhibitor	Herbicide Absorbed predominantly by leaves but also by roots	Post-emergence control of broad-leaved weeds and grasses in maize and sugarcane	*Animals*. Rapidly excreted in the urine, the major metabolite being 4-hydroxysulcotrione *Plants*. Deactivated by the formation of 2-chloro-4-methylsylfonylbenzoic acid *Soil*. DT50 111 days. The major metabolite is 2-chloro-4-methylsulfonylbenzoic acid
Halofenozide (Diacylhydrazine) (1997)	Rohm & Hass		Ecdysone agoinst	Systemic, ingested insecticide active by root application. Interferes with moulting affecting larval stages of insects. Also reduces fecundity in treated adults	Control of Coleoptera and Lepidoptera in turf and ornamentals	*Soil*. Soil dissipation DT50 (field) 42267 days (five sites); turf DT50 377 days

				and have some ovicidal properties		
Methoxyfenozide (Diacythydraxine) Hass (1997)	Rohm & Hass		Ecdysone agonist	Insecticide active primarily by ingestion, also with contact, ovicidal and root systemic activity	Control of lepidopterous larvae in vines tree fruits, vegetables, row crops	*Animals*. Rapidly absorbed, metabolized via phase II conjugation and eliminated *Soil*. Aerobic soil metabolism DT50 3361100 days; field DT50 23268 days
Tebufenozide (Diacylhydrazine) (1996)	Rohm & Hass		Ecdysone agonist	Insecticide Lethaly accelerates moulting process	Control of lepidopterous larvae in rice, fruit, row crops, nut-crops, vegetables, vines and forestry	*Animals*. 16 whole-molecule metabolites are formed as a result of oxidation of the alkyl substituents of the aromatic rings, primarily at the benzylic positions *Plants*. In apples, grapes, rice and sugar beet, the major component is unchanged tebufenozide. Small amounts of metabolites result from oxidation of the alkyl substituents of the aromatic ring, primarily at the benzylic positions *Soil*. Metabolic DT50 in soil 766 days; DT50 for field dissipation 453 days. No mobility below 30 cm
Chromafenozide (Diacylhydrazine) (1996)	Nippon Kayaku; Sankyo		Ecdysone agonist	Insecticide Initiating a precocious incomplete lethal moult	Control of lepidopteran larvae in rice, fruit, vegetables, tea, cotton, beans and forestry	*Animals*. Rapidly excreted with 48 hr and is not persistent in tissues and organs. The major component excreted is unchanged chromafenozide *Plants*. Many minor metabolites are detected in small amounts, but the major component is unchanged chromafenozide. *Soil*. DT50 for field dissipation 44113 days (unpland soil), 22136 days (paddy soil)
Pyrimethanil (Aminopyrimidine) (1992)	Aventis		Inhibitor of methionine biosynthesis leading to inhibition of the secretion of enzymes necessary for fungal infection	Fungicide Protectant in *Botrytis* and both protective and curative action in *Venturia*	For control of grey mould on vines, fruits, vegetables and ornamentals and of leaf scab on pome fruit	*Animals*. Rapidly absorbed, extensively metabolised and rapidly excreted. No evidence of accumulation, even on repeated dosing. Metabolism proceeds by oxidation to phenolic derivatives which are excreted as glucuronide or sulfate conjugates *Plants*. Little metabolism occurs in fruit *Soil*. Rapid degradation, DT50 754 days. Low potential for

						leaching
Cyprodinil (Anilino pyrimidine) (1994)	Novartis		Inhibitor of methionine biosynthesis and secretion of fungal hydrolytic enzymes	Systemic Fungicide with uptake into plants after foliar application and transport throughout the tissue and acropetally in the xylem. Inhibits penetration and mycelial growth both inside and on the leaf surface	Control wide range of pathogens like *Tapesia yallundae, T. acuformis*, *Erysiphe* spp., *Pyrenophora teres*, *Rhynchosporium secalis*, *Botrytis* spp., in cereals, grapes, pome fruit, stone fruit, strawberries, vegetables, field crops and ornamentals, barley *Alternaria* spp., *Venturia* spp. and *monilinia* spp.	*Animals*. Rapidly absorbed and almost completely eliminated with urine and faeces. Metabolism proceeds by 4-hydroxylation of the phenyl and 5-hydroxylation of the pyrimidine rings, followed by mono- or di-sulfation. No evidence for accumulation or retention of cyprodinil or its metabolites *Plants*. Metabolism mainly via hydroxylation of the 6-methyl group of the pyrimidine ring, as well as hydroxylation of the phenyl and pyrimidine rings *Soil*. DT50 2060 days Formation of bound residues the major route for dissipation. Immobile in soil
Fentrazamide (Tetrazolinone) (1997)	Bayer		Cell division inhibitor. Primary target site may be fatty acid metabolism	Herbicide inhibiting cell division in root and meristem	Control of barnyard grass *Echinochloo* spp. and annual sedges in rice for pre-emergence	*Animals*. The main pathway of biotransformation proceeded via hydrolytic cleavage of the parent compound *Plants*. No parent compound was detected in any plant fraction *Soil*. Thoroughly degraded and mineralized Calculated half-lives were in the range of a few days and several weeks, respectively. Immobile
Flufenacet (Oxyacetamide) (1995)	Bayer		Cell division inhibitor. Primary target site may be fatty acid metabolism	Pre- and early Post-emergence herbicide	Selective herbicide with broad spectrum grass control and control of some broad leaved weeds in maize, soybeans, sunflower, wheat, rice	*Animals*. Rapidly excreted. Metabolism takes place via cleavage of the molecule, followed by conjugation of the fluorophenyl moiety with cysteine and formation of a thiadazolone and its various conjugates *Plants*. Rapidly and extensively metabolized; no parent compound was detected, even at early sampling dates *Soil*. Rapidly degraded, immobile
Dithiopyr (Pyridine) (1994)	Rohm & Hass		Inhibits cell division by disrupting spindle microtubule formation		Pre-emergence and early post-emergence control of annual grass and broad-leaf weeds in turf	*Animals*. Rapidly absorbed, extensively metabolized and rapidly excreted *Soil*. DT50 in soil 1761 days, depending on the formulation type. The major soil metabolites are the di-acid, the normal mono-acid and the reverse mono-acid; these metabolites, themselves, dissipate

						almost completely within 1 year
Thiazopyr (Pyridine) (1994)	Rohm & Hass		Inhibits cell division by disrupting spindle microtubule formation	Herbicide causing root growth inhibition and swelling meristematic regions	Pre-emergence control of annual grass and some broad-leaved weeds in tree fruit, vines, citrus, sugarcane, pineapple alfalfa, forestry	*Animals*. Rapidly and extensively metabolized and eliminated. Oxidized by rat liver microsomes via sulfur and carbon oxidations and via oxidative de-esterification *Plants*. Initially metabolized in the dihydrothazole ring by plant oxygenases to the sulfoxide, sulfone, hydroxy derivative and thiazole, and is also de-esterified to the carboxylic acid *Soil*. Degraded by both soil microorganisms and hydrolysis. Average DT50 64 days (8150 days). Minimal mobility: The monoacid metabolite also has limited mobility
Acibenzolar-*S*-methyl Plant activator/Plant host defense induces (1995)	Novartis		Acts as a functional analogue of the natural signal molecule for systemic activated resistance, salicylic acid	Activates plants' natural defense mechanism (systemic activated resistance [SAR]). Has no intrinsic fungicidal activity	For control of fungal infections in wheat under development against a range of diseases in rice, bananas, vegetables and tobacco	*Animals*. Rapidly absorbed and also rapidly almost completely eliminated with urine and faeces. No evidence of accumulation or retention of acibenzolar-*S*-methyl or its metabolites *Plants*. The metabolism proceeds via hydrolysis with subsequent conjugation with sugars, or by oxidation of the phenyl ring followed by sugar conjugation *Soil*. Dissipates via hydrolysis: DT50 0.3 day. The product further degrades, DT50 20 days; metabolites become completely degraded and mineralized. Strong adsorption to soil, low mobility
Pethoxamid Acetamide (2001)	Tokuyama		Presumed to be acting by inhibiting fatty acid biosynthesis	Precise mode of action not yet been clarified. Absorbed by roots and young shoots after application to soil surface	Controls grass weeds including *Echinochloa cruss-galli, Digitaria sanguinalis* and *Setaria geniculata* and broad leaf weeds such as *Amaranthus retroflexus Chenopodium album, Convolvulus arvensis* and *Polygonum pericaria*	

Role of the Agrochemical Industry

The world's population, which stands as 6 billion today and will continue to grow over the years, needs food. This food needs to be grown on the arable land available today, without further destruction of forest resources, wetlands, or mangroves. Thus, improving farm productivity by reducing losses experienced during crop cycles and post harvest is a major challenge. Agrochemicals, which are selective toxicants, provide an important tool. Available agrochemicals need to be deployed judiciously and newer ones must be developed to minimize impacts on water supplies and food quality. Presently, it is believed that more than a third of global farm output is lost due to ineffective pest control. The agrochemical industry has the responsibility of using the power of science and technology for providing the world population with not only sufficient food but with food of increased nutritional value. On another, less obvious front, it needs to be noted that some 75 percent of the global population use medicinal plants for primary health care, and these plants also require crop protection. In yet another related area, public health, vector control is more dependable than are curative drugs in combating diseases such as malaria, yellow and dengue fevers.

The agrochemical industry is large, complex, and involves many facets and many players. Listing several of these, from multinational manufacturer to individuals who apply the product in the field, will help in visualizing the breadth of the industry.

The "players" include:

1. Large, multinational companies engaged in discovery, manufacture, and distribution of agrochemicals, seeds, and other products of biotechnology.
2. Large corporate entities engaged in the manufacturer and distribution of off-patent agrochemicals.
3. Companies engaged in the formulation and distribution of agrochemicals purchased from (1) or (2) above.
4. Retailers who make end use products available to growers.
5. Extension workers from governmental or non-governmental sources who provide guidance on the proper use of agrochemicals in the field.
6. Professional consultants who perform activities in (5) to individual growers.
7. Pest control operators who are professionally trained to properly apply restricted use of agrochemicals as well as other agrochemicals deployed in disease vector control and termite control.

For the year 2000, it was estimated that the global sales revenue of the agrochemical industry amounted to US$30 billion. This breaks down as follows:

Herbicides	US$14 billion	73% for cereals, maize, soybeans, fruits, vegetables
Insecticides	US$8 billion	71% for fruits, vegetables, cotton, rice
Fungicides	US$6 billion	70% for cereals, fruits, vegetables
Others	US$1 billion	

The revenues were distributed globally approximately as follows: North America27.2 percent; Far East26 percent; Europe25.5 percent; Latin America14.8 percent; and the rest of the world6.5 percent.

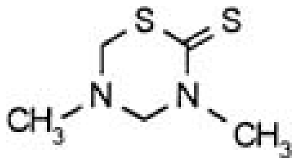
30.2 Characteristics of the Agrochemical Industry

Among the distinguishing characteristics of the agrochemical industry are: (1) the multitude of chemical agents employed, (2) a limited price range (which derives from the limited chemical complexity, in turn driven by by the economics of agricultural production), (3) a fairly rapid obsolescence of the chemicals used, and (4) a high degree of government regulation for the production, application, shipment, and use of agrochemicals.

Government Regulation

In the United States, the first state laws on insecticides were enacted in 1900 to establish standards of purity for the arsenical Paris green (copper acetoarsenite) which is no longer used in agriculture in the United States. Gradually these laws were extended to cover a wide list of inorganic compounds and plant extracts, many of them, like Paris green, extremely toxic to humans. Included in this group are such compounds as arsenic combined with copper, lead, and calcium; phosphorus pastes for ants and roaches; strychnine in rodent baits; thallium in ant and rodent baits; and selenium for plant-feeding mites. Mercury, both as a corrosive sublimate and as calomel, was used as an insect repellent and later as a seed disinfectant. Sodium fluoride was a common ant poison, and sodium cyanide, calcium cyanide, and HCN itself were general fumigants (Table 30.2). Nicotine sulfate was used generally in the garden and on the farm. These compounds, among the most toxic of any known at that time, were widely marketed without supervision under any of the early state laws. There was no provision for public health, either in regulating the amounts applied or regarding the possible danger of minute amounts (residues) remaining on the marketed produce. The need to protect the applicator, farmer, laborer, and the general public against the dangerous qualities of the insecticides, or their residues on crops, provided the motives for all the legislation that followed.

TABLE 30.2 Fumigants

Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Environmental Fate
Chloropicrin	CI3CNO2		Fumigant	
Dazomet (Methyl isothiocyanate precursor)		Non-selective inhibition of enzymes by degradation products	A pre-planting soil fumigant, acting by decomposition to methyl Eothocyanate	*Plants*. Following application to strawberries, noresidues of dazomet or of its degradation products methyl isothiocyanate, dimethyl- or monomethylthiourea were detected at >0.01 ppm in the fruit *Soil*. In the presence of moisture, undergoes degradation to methyl-(methylaminomethyl) dithiocarbamic acid, which then undergoes further degradation to methyl isothiocyanate, formaldehyde, hydrogen sulfide and methylamine
1,3-dichloropropene (Chloroalkene)			Soil fumigant nematicide	
Methyl Bromide	CH3Br		Fumigant insecticide and nematicide	*Animals/Plants*. Metabolism not totally elucidated; inorganic bromide ion is formed

Governmental concern was first related to standardization of the manufactured chemical and protection of the farmer in relation to the product that he or she purchased. This was then extended to the handling of the chemical in interstate commerce, to the protection of the consumer of raw agricultural products (e.g., apples, corn, and lettuce), and, in other legislation, to the protection of the consumer of finished goods (e.g., canned juice, margarine, cereal food, meat, and milk). Included in this legislation were provisions that protect the shipper of the chemical, the applicator of the chemical, and all personnel proximal to the application of the chemical. Legislation now regulates chemicals applied to crops or foods as protective agents-pesticides, emulsifiers, solvents, packaging materials (wax, container materials, plasticizers, antioxidants, etc.).

Toward the end of the 1960s, a new area of concern arose: the effect of the manufacture and application of pesticide chemicals *on the environment* was recognized. Concern for the environment was the subject matter of Rachel Carson's book, *Silent Spring*, which was published in 1962. In 1970, this new focus led to the establishment of the Environmental Protection Agency (EPA; aka US-EPA), which was given authority to regulate virtually all aspects of agrochemical manufacture and use in the United States. Since its inception, the principal objectives of the EPA's agrochemical activities have been to: (1) establish procedures that ensure that new pesticides will not pose unreasonable risks to human health and the environment, and (2) terminate the use of those previously registered pesticides that exceed certain risk criteria. Among the requirements called for are studies on mammalian toxicology (including lifetime animal feeding studies), environmental chemistry (persistence, mobility, etc.), and effects on fish and other wildlife. New product registrations are granted only after EPA scientists and administrators are satisfied that use of the product does not pose unreasonable risk to humans or hazard to the environment. All aspects of the environment are considered. Soil, air, and water (streams, lakes, oceans, rivers, marshes, and underground aquifers) are matters of environmental concern, as are the living organisms that reside therein. The protocols employed to ensure safety are complex and not infrequently are at the boundaries of scientific capability. The registration of any new product is a highly complicated and expensive procedure. As to the second objective, much the same criteria are used in judging whether or not to allow continued use of previously registered pesticides.

Although comprehensive regulatory legislation was developed first and most extensively in the United States, all the technically developed nations of the World now regulate the manufacture, sale, and use of agricultural chemicals. The criteria used are not unlike those which were developed over the past several years in the United States. In Europe, the European Economic Community (EEC) in 1991 adopted Directive 91/414/EEC which, included the following goals:*

1. Coordinate the overall arrangements for authorization of plant protection products within the European Union. Whereas it is intended to coordinate the process for considering the safety of particular substances at the Community level, individual Member States have responsibility for product authorization.
2. Establish a list of active substances which have been shown to be without unacceptable risk to humans or the environment.
3. Maintain an up-to-date listing (Annex I of the Directive) of active substances which have been authorized.
4. Member States can authorize the sale and use of plant protection products only if they are listed in Annex I.

*http://www.pesticides.gov.uk/

Under provisions of directive 91/414/EEC, all existing agrochemicals are being reviewed and new ones approved using common database criteria. The review process started in 1993 and is expected to be completed by 2008. Many agrochemicals will cease to be used in the European Union because of the failure of participating companies to submit full dossiers on them. Additionally, water quality directive 98/83/EC demands detectability for a given agrochemical in water below 1 ppb.

The 29 country member Organization for Economic Cooperation and Development Working Group offers a common platform to the national pesticide regulators for discussion of activities on conventional, biological, and microbial pesticides.

On a broader scale, several years ago the Food and Agriculture Organization of the United Nations adopted the International Code of Conduct on the Distribution and Use of Pesticides.† A few of the many provisions of that Code are listed below to give a general idea of its thrust.

1. Governments have the overall responsibility and should take the specific powers to regulate the distribution and use of pesticides in their countries.
2. The pesticide industry should adhere to the provisions of this Code as a standard for the manufacture, distribution, and advertising of pesticides, particularly in countries lacking appropriate legislation and advisory services.
3. Manufacturers and traders should supply only pesticides of adequate quality, packaged, and labeled as appropriate for each specific market.

†http://www.fao.org/ag/agp/agpp/pesticid/Code/References.htm

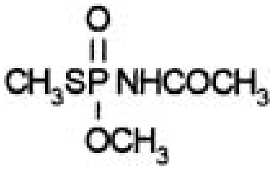
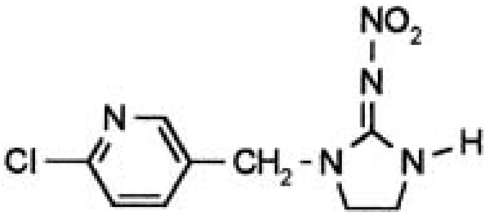
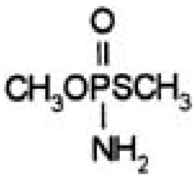
Manufacture of Agrochemicals

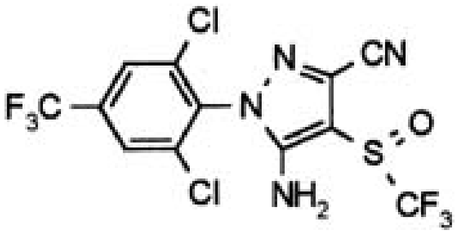
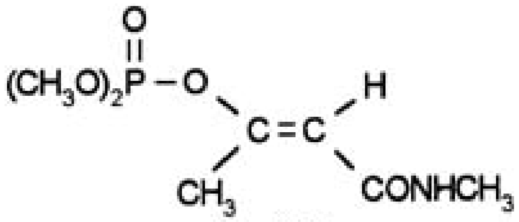
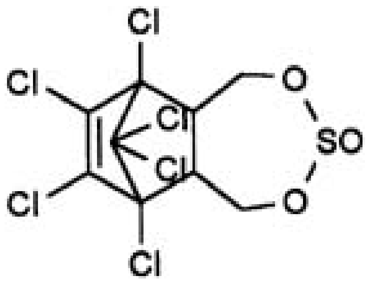
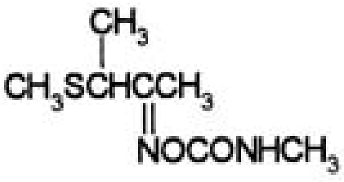
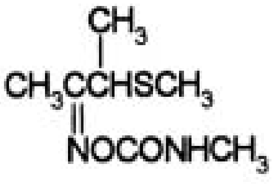
The manufacturing route to an agrochemical can have multiple options. The route chosen may depend on commercial availability of desired reagents, engineering capabilities, byproduct formation, separation techniques, and so on. This is on par with the manufacture of drugs, dyestuffs, or speciality chemicals. One feature unique to agrochemicals is a detailed and precise label preapproved by a regulatory body on the product safety, usage instructions, compatibility statements, and other statements as deemed necessary for the needs of medical professionals and general public information.

Classes of Agrochemicals

Agrochemicals that control insects by growth regulation, or by mortality through contact or stomach action are called *insecticides* (Table 30.3a, b). Those that control competing weeds (grasses, broad leaved plants, or sedges) through preplant incorporation, pre-emergence, early post- or post-emergence application with respect to the main crop are called *herbicides* (Tables 30.4ad). These interfere with chlorophyll formation and/or activity, and various other metabolic processes in the weeds at much higher efficiency than the main crop. Agrochemicals that control fungal diseases are called *fungicides* (Tables 30.5a, b), and those with plant hormone-like action are called *plant growth regulators* (Table 30.6). Agencies such as WHO (World Health Organization) and the US EPA separate agrochemicals into different classes as per agreed upon toxicity levels, as a useful guide to consumers. In many countries, field usage of more toxic agrochemicals is restricted to trained personnel only.

TABLE 30.3a Insecticides for Sucking Pests

Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Environmental Fate
Acephate (Organophosphorous)	 $CH_3S\overset{O}{\underset{OCH_3}{P}}NHCOCH_3$	Cholinesterase inhibitor	Systemic insecticide	*Animals*. Metabolized to methamidophos (*q.v.*). Plants Residual activity lasts for c. 1015 days. The major metabolite is methamidophos *(q.v.)*. *Soil*. Readily biodegraded and non-persistent; soil DT50 2 days (aerobic) to 7 days (anaerobic). Methamidophos (*q.v.*) has been identified as a soil metabolite
Imidacloprid (Neonicotinoid)		Acts as an antagonist by binding to post-synaptic nicotinic receptors in the insects' central nervous system	Systemic insecticide with translaminar activity and with contact and stomach action. Readily taken up by plant and further distributed acropetally, with good root-systemic action	*Animals*. The radioactivity was quickly and almost completely absorbed from the gastrointestinal tract and quickly eliminated (96% within 48 hr, mainly via the urine). Only c. 15% was eliminated as unchanged parent compound; the most important metabolic steps were hydroxylation at the imidazolidine ring, hydrolysis to 6-chloronicotinic acid, loss of the nitro group with formation of the guanidine and conjugation of the 6-chloronicotinic acid with glycine. All metabolites found in the edible organs and tissues of farm animals contained the 6-chloronicotinic acid moiety *Plants*. Metabolized by loss of the nitro group hydroxylation at the imidazolidine ring, hydrolysis to 6-chloronicotinic acid and formation of conjugates; all metabolites contained the 6-chloropyridinylmethylene moiety *Soil*. The most important metabolic steps were oxidation at the imidazolidine ring, reduction or loss of the nitro group, hydrolysis to 6-chloronicotinic acid and mineralisation. Medium adsorption to soil, imidacloprid and soil metabolites are to be classified as immobile
Methamidophos (Organophosphorous)	 $CH_3O\overset{O}{\underset{NH_2}{P}}SCH_3$	Cholinesterase inhibitor	Systemic insecticide with contact and stomach action	*Animals*. Absorbed rapidly and distributed uniformly among all organs and tissues. More than half of the radioactivity was rapidly eliminated from the body, mainly via urine and respiratory air. Radioactivity remaining in the animal was incorporated into endogenous compounds (carbon-1 pool) and eliminated with the natural turnover of these compounds. Metabolism in the rat was by deamination and demethylation. *Plants*. Taken up rapidly and translocated into the leaves *Soil*. Rapidly degraded in soil; field DT50 c. <2 days

Fipronil (Phenyl pyrazole)		Acts as a potent blocker of the GABA-regulated chloride channel	Broad spectrum insecticide, toxic by contact and ingestion Moderately systemic, Good to excellent residual control following foliar application	In plants, animals and the environment, fipronil is metabolized via reduction to the sulfide, oxidation to the sulfone, and hydrolysis to the amide *Animals*. Distribution and is rapid. Elimination is mainly via the faeces as fipronil and its sulfone. The two major urinary metabolites were identified as conjugates of ring-opened pyrazole products. The distribution of radioactive residues in tissues was extensive after 7 days *Plants*. Uptake of fipronil into plants was low (c. 5%). At crop maturity, the major residue components were fipronil, the sulfone, and the amide. Following foliar application to cotton, cabbage, rice and potatoes, at crop maturity, fipronil and the photodegradate were the major residue components *Soil*. Readily degraded: major degradates in soil (aerobic) are sulfone and amide, (anaerobic) are sulfide and amide. Present a low risk of downward movement in soil
Monocrotophos (Organophosphorous)	 (*E*)	Cholinesterase inhibitor	Systemic insecticide with contact and stomach action	*Animals*. In mammals, following oral administration, 6065% is excreted within 24 hr, predominantly in the urine *Soil*. Rapidly degraded in soil; DT50 (lab.) 15 days
Endosulfan (Cyclodiene organochlorine)		Antagonist of the GABA receptor-chloride channel complex	Non-systemic insecticide with contact and stomach action	*Animals*. The principal route of elimination is faeces; most of the radioactivity is excreted within the first 48 hr. Metabolised rapidly to less-toxic metabolites and to polar conjugates. *Plants*. The plant metabolites (mainly endosulfan sulfate) were also found in animals 50% of residues are lost in 37 days (depending on plant species) *Soil*. DT50 3070 days. The main metabolite was endosulfan sulfate, which is degraded more slowly. DT50 for total endosulfan (alpha- and beta- endosulfan and endosulfan sulfate) in the field is 58 months. No leaching tendency
Butocarboxim (Oxime carbamate)	 (*E*)  (*Z*)	Cholinesterase inhibitor	Systemic insecticide with contact and stomach action	*Animals*. Metabolized to butoxycarboxim, and excreted in the urine as butoxycarboxim and its degradation products *Plants/Soil*. The methylamine moiety is split off, and the sulfur atom is oxidized to sulfoxide and sulfone. DT50 in soil 18 days DT50 for metabolites 1644 days

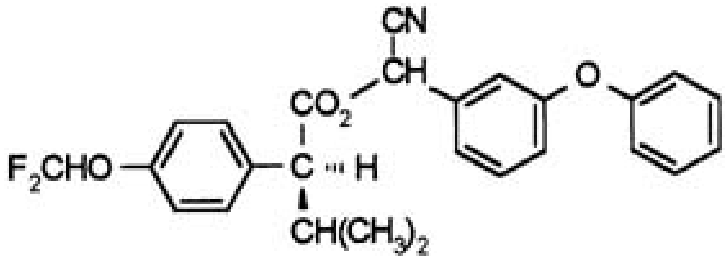
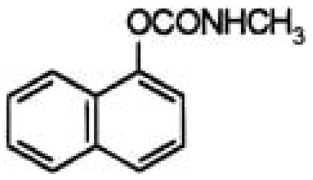
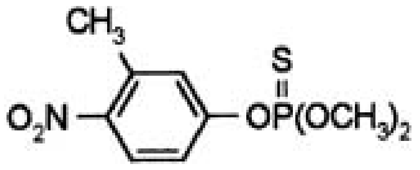
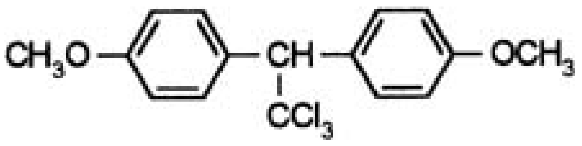
Disulfoton (Organophosphorous)	 $CH_3CH_2SCH_2CH_2S\overset{S}{\overset{\|}{P}}(OCH_2CH_3)_2$	Cholinesterase inhibitor	Systemic insecticide, absorbed by roots, with translocation to all parts of the plant	*Animals*. 14C-disulfoton is rapidly absorbed, metabolised, and the radioactivity excreted in the urine. The main metabolites are disulfoton sulfoxide and sulfone, their corresponding oxygen analogues and diethylthiophosphate *Plants*. Very rapidly metabolized. The metabolism is the same as in animals *Soil*. Very rapidly degraded. The metabolism is similar to that in animals and plants. It exhibits medium to low mobility in soil
Flucythrinate (Pyrethroid)			Non-systemic insecticide with contact and stomach action	*Animals*. 6070% is eliminated within 24 hr, and >95% within 8 days, in the faeces and urine. In the facets, the parent compound makes up most of the material excreted, but in the urine and in tissue, several metabolites are present. The major route of degradation is through hydrolysis, with subsequent hydroxylation of the hydrolysis products *Soil*. Immobile DT50 c. 2 months

TABLE 30.3b Insecticides for Chewing Pests

Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Environmental Fate
Carbaryl (Carbamate)		Weak cholinesterase inhibitor	Insecticide with contact and stomach action and slight systemic properties	*Animals*. Does not accumulate in body tissues, but is rapidly metabolised to non-toxic substances, particularly 1-naphthol. This, together with the glucuronic acid conjugate, is eliminated predominantly in the urine and faeces *Plants*. Metabolites are 4-hydroxycarbaryl, 5-hydroxycarbaryl and methylol-carbaryl *Soil*. DT50 (aerobic) 714 days in a sandy loam and 1428 days in a clay loam
Fenitrothion (organophosphorous)		Cholinesterase inhibitor	Non-systemic insecticide with contact and stomach action	*Animals*. Rapidly excreted in the urine and faeces. After 3 days c. 90% has been excreted by rats, mice and rabbits. The most important metabolites are dimethylfenitrooxon and 3-methyl-4-nitrophenol *Plants*. DT50 4 days; 7085% is degraded within 2 weeks. Major metabolites are 3-methyl-4-nitrophenol, the oxygen analogue and their decomposition products desmethylfenitrothion, dimethylphosphorothionic acid and phosphorothionic acid *Soil*. DT50 1228 days under upland conditions, 420 days under submerged conditions. The major metabolites under upland conditions are 3-methyl-4-nitrophenol and CO2, whereas, under submerged conditions, the major decomposition product is aminofenitrothion
Methoxychlor (Organochlorine)			Insecticide with contact and stomach action	*Animals*. Degradation in animals is principally by *O*-dealkylation to the corresponding phenol and diphenol, and by dehydrochlorination to 4,4′-dihydroxybenzophenone

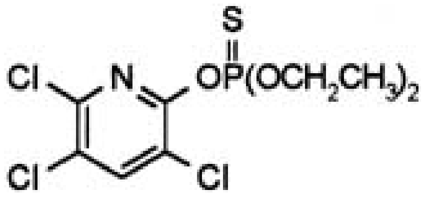
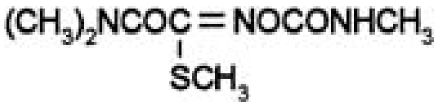
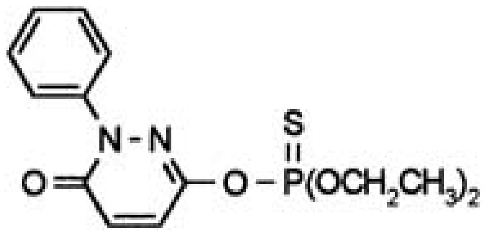
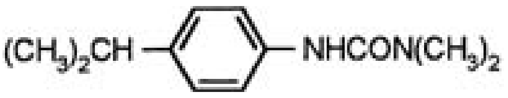
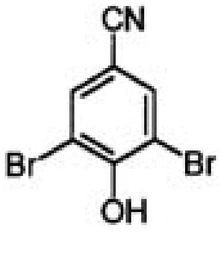
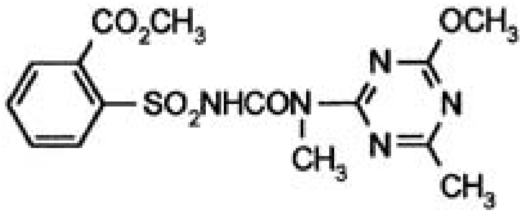
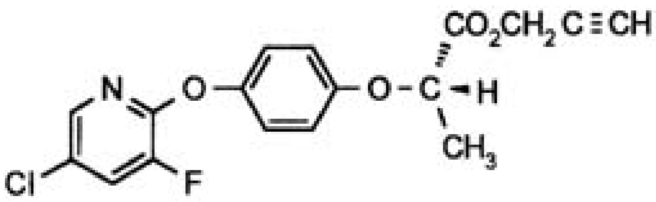
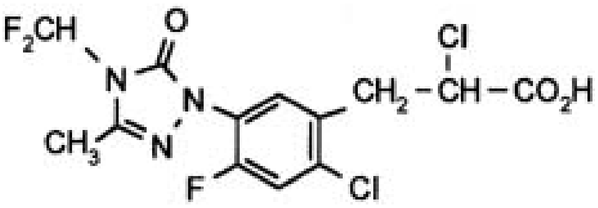
Chlorpyrifos (Organophosphorous)		Cholinesterase inhibitor	Non-systemic insecticide with contact, stomach and respiratory action	*Animals*. Rapid metabolism occurs, the principal metabolite being 3,5,6-trichloropyridin-2-ol. Excretion is principally in the urine *Plants*. Residues are metabolised to 3,5,6-trichloropyridin-2-ol which is conjugated and sequestered *Soil*. Field DT50 for soil-incorporated applications 3356 days for soilsurface applications 715 days. Primary route of degradation is transformation to 3,5,6-trichloropyridin-2-ol, which is subsequently degraded to organochlorine compounds and CO2
Oxamyl (Oxime carbamate)	 $(CH_3)_2NCO\underset{SCH_3}{C}=NOCONHCH_3$	Cholinesterase inhibitor	Contact and systemic insecticide, absorbed by foliage and roots	*Animals*. Hydrolyzed to an oximino metabolite (methyl *N*-hydroxy-*N′*,*N′*-dimethyl-1-thiooxamimidate) or converted enzymically via *N*, *N*-dimethyl-1-cyanoformamide to *N*, *N*-dimethyloxamic acid. Conjugates of the oximino compound, the acid, and their monomethyl derivatives constituted over 70% of thermetabolites excreted in the urine and faeces *Plants*. Hydrolyses to the corresponding oximino compound which, in turn, conjugates with glucose. Total breakdown into natural products has been demonstrated *Soil*. Degraded rapidly in soil, DT50 c. 7 days
Pyridaphenthion (Organophosphorous)		Cholinesterase inhibitor	Insecticide with contact and stomach action	*Animals*. In mice and rats, the parent compound, *O*-ethyl-*O*-(3-oxo-2-phenyl-2*H*-pyridazine-6-yl) phosphorothioate and the corresponding phosphate are found *Plants*. In rice, phenyl maleich ydrazide, *O*, *O*-diethyl thiophosphoric acid, and PMH glycoside are formed *Soil*. DT50 1124 days

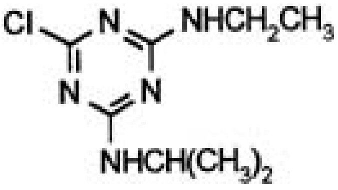
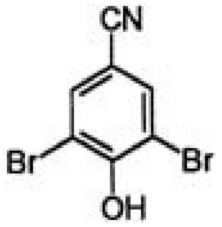
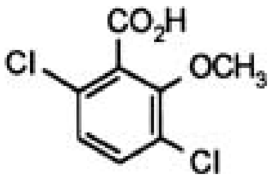
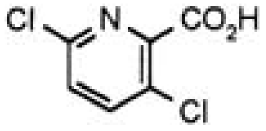
TABLE 30.4a Cereal Herbicides

Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Activity	Environmental Fate
Isoproturon (Urea)		Photosynthetic electron transport inhibitor at the photosystem II receptor site	Selective systemic herbicide, absorbed by roots and leaves, with translocation	Graminicide and broadleaf weeds controller	*Animals*. 50% is eliminated within 8 hr, predominantly in urine *Plants*. Degradation mainly via hydroxylation of the isopropyl group to 1,1-dimethyl-3-[4(2′-hydroxy-2′-propyl) phenyl] urea; *N*-dealkylation also occurs *Soil*. Undergoes enzymic and microbial demethylation at the nitrogen, and hydrolysis of the phenylurea to 4-isopropylamine DT50 628 days
Bromoxynil (Hydroxybenzonitrile)		Photosynthetic electron transport inhibitor at the photosystem II receptor site, also uncouples oxidative phosphorylation	Selective contact herbicide with some systemic activity. Absorbed by foliage with limited translocation	Broadleaf weed controller	*Animals/plants*. Metabolism by hydrolysis of the ester and nitrile groups with some debromination occurring *Soil*. DT50 *c*. 10 days. Degraded by hydrolysis and debromination to less toxic substances

					such as hydrobenzoic acid
Tribenuronmethyl (Sulfonyl urea)		Branched chain amino acid synthesis (ALS and AHAS) inhibitor. Acts by inhibiting biosynthesis of the essential amino acids valine and isoleucine, hence stopping cell division and plant growth	Rapidly absorbed by foliage and roots and translocated throughout the plant	Broad leaf weed controller	*Soil*. Degrades by hydrolysis and direct microbial degradation. Hydrolysis is faster in acidic than alkaline soils. DT50 17 days
Clodinafoppropargyl [(2-(-4-aryloxyphenoxy) propionic acid]		Fatty acid synthesis inhibitor, by inhibition of acetyl COA carboxylase (Assase)	Post-emergence, systemic herbicide	Graminicide	*Animals*. Hydrolysed to the corresponding acid *Plants*. Rapidly degraded to the acid derivative as major metabolite *Soil*. Undergoes rapid degradation to the free acid (DT50 <2 hr) and then further to phenyl and pyridine moieties which are bound to the soil and mineralised. The free acid is mobile in soil, but is further degraded with DT50 520 days; negligible leaching potential
Carfentrazoneethyl (Triazolinone)		Protoporphyrinogen oxidase inhibtor leading to membrane disruption	Absorbed by foliage with limited translocation	Broad leaf weed controller	*Animals*. About 80% excreted in urine within 24 hr. Major metabolite is corresponding acid. Further metabolism appears to involve oxidative hydroxylation of the methyl group or dehydrochlorination to form corresponding Cinnamic acid *Plants*. Converted to free acid, which is hydroxylated and oxidized at triazolinone methyl to form the dibasic acid. DT50 <7 days *Soil*. Degradation by microbial action. Strongly adsorbed to sterile soils. In non-sterile soils, rapidly converted to free acid, which has low soil binding. In laboratory soil DT50 is a few hours
Flufenacet (Oxyacetamide)		Inhibits cell division and growth. Primary target site may be fatty acid metabolism	Pre and early post-emergence herbicide	Broad spectrum herbicide controlling grasses and broad leaved weeds	*Animals*. Rapidly excreted by rats, goat and hen. Metabolism via cleavage of the molecule followed by conjugation of flurophenyl moiety with cysteine and formation of a thiadozolone and its various conjugates *Plants*. Rapidly metabolized, residues accounted based on total amount of *N*-

flurophenyl-*N*-isopropyl derived residues
Soil. Rapidly degraded in soil. No threat of leaching

TABLE 30.4b Maize Herbicides

Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Activity	Environmental Fate
Atrazine (1,3,5-triazine)		Photosynthetic electron transport inhibitor at the photosystem II receptor site	Selective systemic herbicide, absorbed principally through roots, but also through foliage, with translocation acropetally in the xylem and accumulation in the apical meristems and leaves	Cross spectrum weed controller	*Animals*. Rapidly and completely metabolized, primarily by oxidative dealkylation of the amino groups and by reaction of chlorine atom with endogenous thiols. Diaminochlorotriazine is the main primary metabolite, which readily conjugates with glutathione. More than 50% of the dose is eliminated in the urine and around 33% in faeces within 24 hr *Plants*. In tolerant plants, readily metabolized to hydroxyatrazine and amino acid conjugates, with further decomposition of hydroxyatrazine by degradation of side-chains and hydrolysis of resulting amino acids on the ring together with evolution of CO2 *Soil*. Major metabolites are desethylatrazine and hydroxyatrazine; DT50 1677 days
Bromoxynil (Hydroxybenzonitrile)		Photosynthetic electron transport inhibitor at the photosystem II receptor site, also uncouples oxidative phosphorylation	Selective contact herbicide with some systemic activity. Absorbed by foliage with limited translocation	Broadleaf weed controller	*Animals/plants*. Metabolism by hydrolysis of the ester and nitrile groups with some debromination occurring *Soil*. DT50 c. 10 days: Degraded by hydrolysis and debromination to less toxic substance such as hydrobenzoic acid
Dicamba (Benzoic acid)		Synthetic auxin (acting like indolylacetic acid)	Selective systemic herbicide, absorbed by the leaves and roots, with ready translocation throughout the plant via both symplastic and apoplastic systems	Broadleaf weed controller	*Animals*. Rapidly eliminated in the urine, partly as a glycine conjugate *Plants*. Degradation rate varies greatly with species. In wheat, the major metabolite is 5-hydroxy-2-methoxy-3,6-dichlorobenzoic acid, while 3,6-dichlorosalicylic acid is also a metabolite *Soil*. Microbial degradation occurs, the principal metabolite being 3,6-dichlorosalicylic acid. DT50, < 14 days
Clopyralid (Pyridinecarboxylic acid)		Synthetic auxin (acting like indoleacetic acid)	Selective systemic herbicide, absorbed by roots and leaves with translocation both acropetally and basipetally and accumulation in meristematic tissue	Broadleaf weed controller	*Animals*. In rats, there is rapid and almost quantitative unchanged elimination in urine *Plants*. Not metabolized in plants *Soil*. Microbial degradation occurs. Major product is CO2

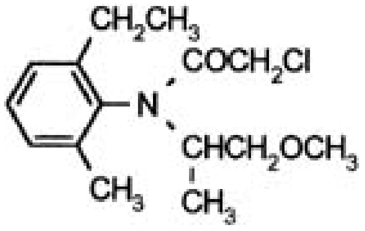
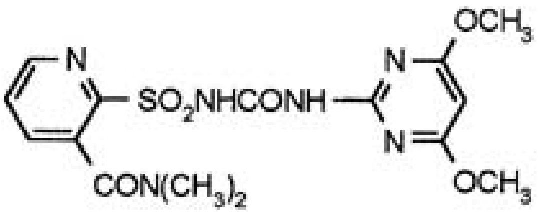
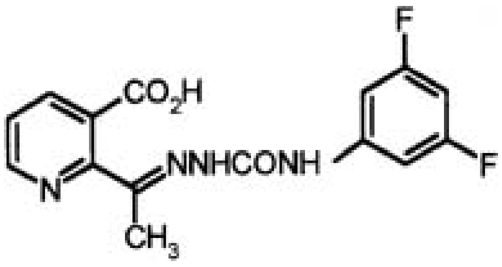
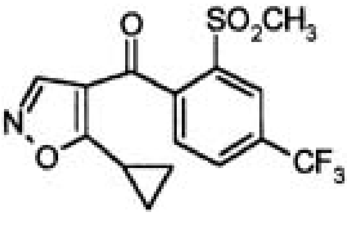
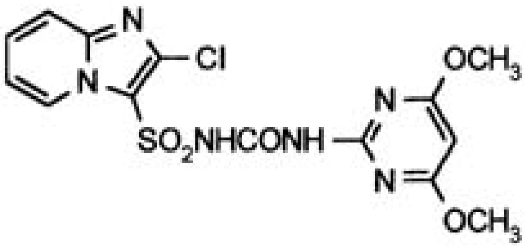
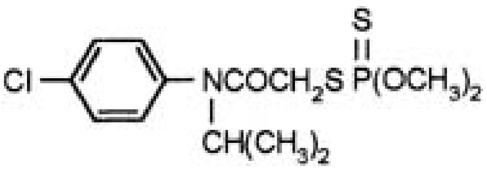
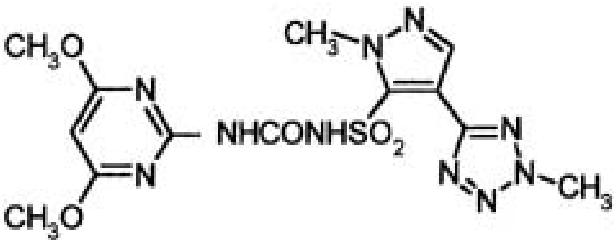
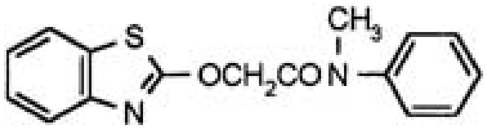
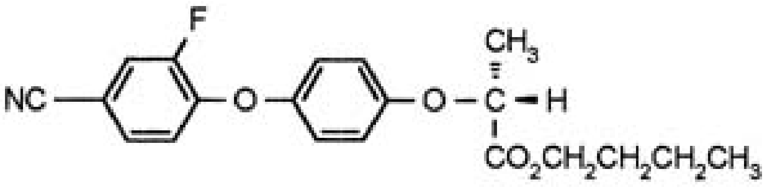
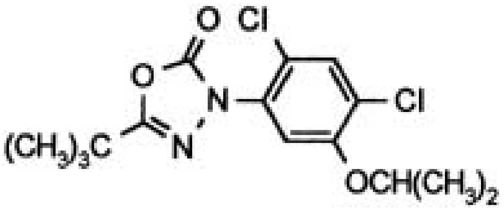
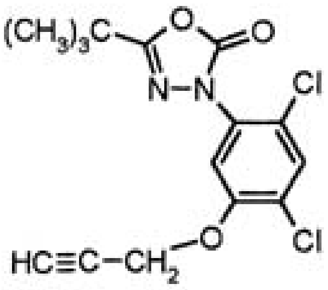
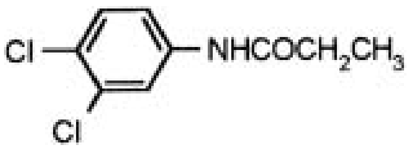
Metolachlor (Chloroacetamide)		Cell division inhibitor	Selective herbicide, absorbed predominantly by the hypocotyls and shoots. Inhibits germination	Grass weed controller	*Animals*. Rapidly oxidized by rat liver microsomal oxygenases via dechlorination, *O*-demethylation and side-chain oxidation *Plants*. Metabolism involves natural product conjugation of the chloroacetyl group and hydrolysis and sugar conjugation at the ether group. Final metabolites are polar, water-soluble and non-volatile *Soil*. Major aerobic metabolites are derivatives of oxanilic and sulfonic acids. DT50 = 20 days
Nicosulfuron (Sulfonylurea)		Branched chain amino-acid (ALS) and (AHAS) synthesis inhibitor. Acts by inhibiting biosynthesis of essential amino acids valine and isoleucine, hence stopping cell division and plant growth	Selective systemic herbicide, absorbed by roots and leaves with rapid translocation in xylem and phloem to the meristematic tissues	Graminicide and broadleaf weed controller	*Animals*. Nicrosulfuron and its metabolites do not bioaccumulate. Hydrolysis of the sulfonylurea bridge and hydroxylation were the main metabolic pathways *Plants*. Degraded rapidly. DT50 1.54.5 days main metabolic pathways were hydrolysis of the sulfonyl urea bridge to form the pyridine sulfonamide and pyrimidine amine, and hydroxylation on the pyrimidine ring *Soil*. DT50 2443 days
Diflufenzopyr (Semi-carbazone)		Inhibits auxin transport, apparently by binding with a carrier protein on the plasmalemma	Systemic, postemergence herbicide	Graminicide and broad leaf weed controller	*Animals*. 2044% of oral dose eliminated in urine and 4979% in faeces. Intravenous administration showed excretation of 6189% in urine. Elimination DT50 in urine and faeces was about 6 hr. Eliminated mainly as the parent compound *Soil*. Average DT50 in field: 4.5 days very mobile
Isoxaflutole (Isoxazole)		*p*-hydroxyphenyl pyruvate diooxygenate inhibitor	Systemic by either root or foliar uptake	Graminicide and broadleaf weed controller	*Animals*. Rapidly excreted *Plants*. Residues at harvest comprise mainly of a non-toxic metabolite *Soil*. Degradation via hydrolysis and microbial degradation with final mineralization to CO2. Mobile under simulated conditions. However, under field conditions, residues remain in the surface horizons

TABLE 30.4c Rice Herbicide

Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Activity	Environmental Fate

Imazosulfuron (Sulfonylurea)		Branched chain amino acid synthesis (ALS and AHAS) inhibitor. Acts by inhibiting biosynthesis of the essential amino acids valine and Isoleucine, hence stopping cell division and plant growth	Absorbed by plants mainly through roots and translocated throughout the plant. Inhibits shoot growth and restarts root development	Broad leaf weed and sedge controller	Unknown
Anilofos (Organophosphorous)		Inhibits cell division	Selective herbicide, absorbed through roots and to some extent, through leaves	Grass weed and sedge controller	*Soil.*Metabolizes into Chloroaniline and CO2 DT50:3045 days
Azimsulfuron (Sulfonylurea)		Branched chain amino acid synthesis (ALC or AHAS) inhibitor. Acts by inhibiting biosynthesis of essential amino acids valine and isoleucine, hence stopping cell division and plant growth	Post-emergence herbicide with mainly foliar uptake, translocated in xylem and phloem	Broad leaf weed and sedge controller and control of *Echinochloa* spp.	*Animals.* >95% was extracted within 2 days 673% in unmetabolized form. The major metabolic pathway was *O*-demethylation followed by pyrimidine ring hydroxylation and subsequent *O*-conjugation, a pyrimidine ring-cleaved guanidine was also identified *Plants.*Metabolism was rapid; little parent compound was found in any plant tissue at maturity *Soil.*The most significant mechanisms are indirect photolysis and soil metabolism, together with chemical hydrolysis
Mefenacet (Oxyacetamide)		Inhibits cell division and growth	Selecticide herbicide	Graminicide	*Animals.* Degrades to *N*-methylaniline which is subsequently demethylated, acetylated and hydroxylated to 4-aminophenol and its sulfate and glucuronide conjugates *Plants.* Besides 4-aminophenol, benzothiazolone and benzothiazoylacetic acid are found, both of which are formed by hydroxylation *Soil.* Metabolities formed are benzothiazole and benzothiazolyl-acetic

Cyhalofop-butyl (Aryloxyphenoxy propionate)		Fatty acid synthesis inhibitor, by inhibition of acetyl CoA Carboxylase (ACCase)		Graminicide	*Animals.* Metabolised by hydrolysis to acid which may further breakdown to other metabolities which in turn are rapidly excreted *Soil.* Rapidly metabolized to acid. DT50 210 hr in field
Oxadiazon (Oxadiazole)		Protoporphyrinogen oxidase inhibitor	Selective contact herbicide	Graminicide and broadleaf weed controller	*Animals.* 93% is eliminated within 72 hr, predominantly in the urine *Plants.* Rapidly metabolized. Metabolites do not accumulate *Soil.* Strongly adsorbed by soil colloids and humus with very little migration or leaching. DT50 c. 36 months
Oxadiazon (Oxadiazole)		Protoporphyrinogen oxidase inhibitor	Selective herbicide active mainly preemergence; effects being at germination. It is not absorbed by plants	Broad leaf weed and grass and annual sedge controller	*Animals.* Rapidly excreted with no accumulation *Plants.* Very low levels of residues at harvest in lemons, sunflowers and rice *Soil.* DT50 (lab, aerobic) 1872 days
Propanil (Anilide)		Photosynthetic electron transport inhibitor at the photosystem II receptor site	Selective contact herbicide with a short duration of activity	Broadleaf weed and grass controller	*Animals.* The major metabolic pathway in microsomal incubations was acylamidase hydrolysis to 3,4-dichloroaniline *Plants.* Hydrolysed by an aryl acelamidase to 3,4-dichloroaniline and propionic acid as metabolic intermediates *Soil.* Rapid microbial degradation to aniline derivative occurs. Degradation products are proportionate which is rapidly metabolized to CO2 and 3,4-dichloroaniline which is bound to soil

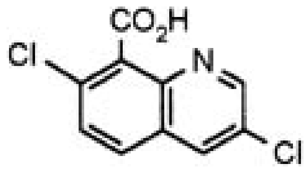
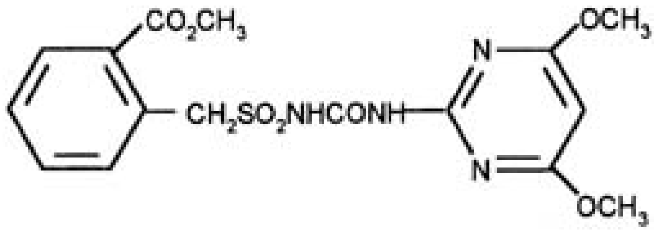
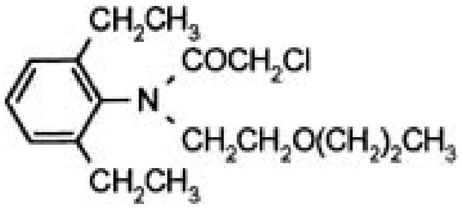
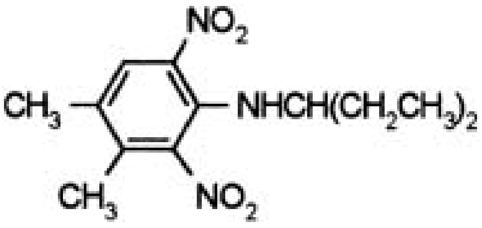
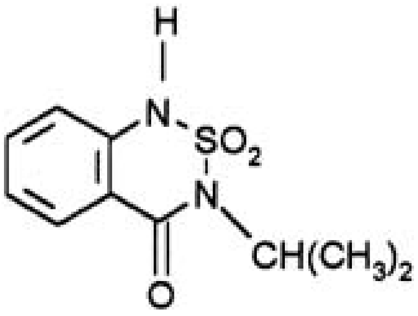
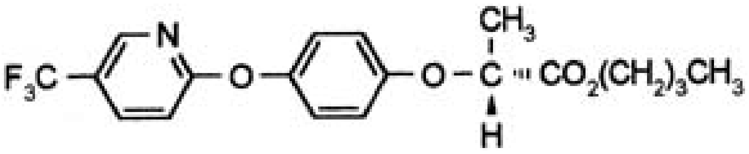
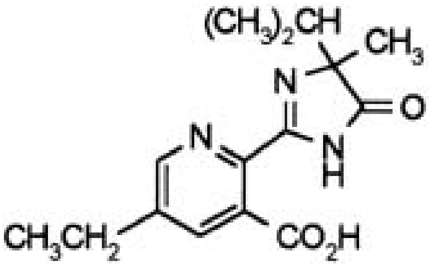
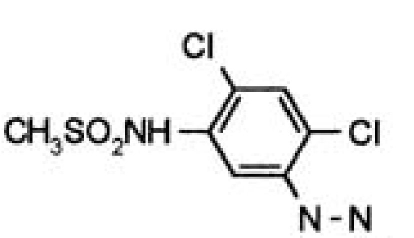
Quinclorac (quinoline-carboxylic acid)		Synthetic auxin (acting like indolyacetic acid), also inhibitor of cell wall (cellulose) biosynthesis	Rapidly absorbed through the foliage	Graminicide	*Animals.* More than 90% excreted in the urine within 5 days *Plants.* Systematically translocated to the roots and to the leaves *Soil.* Only slightly adsorbed by the soil. Depending on soil type and organic matter content, the chemical is relatively mobile. Degraded by microorganisms, 3-chloro-8-quinolinecarboxylic acid being the major metabolite
Bensulfuron-methyl (Sulfonylurea)		Branched chain amino acid synthesis (ALS and AHAS) inhibitor. Acts by inhibiting biosynthesis of the essential amino acids valine and isoleucine, hence stopping cell division and plant growth	Selective systemic herbicide, rapidly absorbed by root and foliage with rapid translocation to meristematic tissue	Broad leaf weed and sedge controller	*Animals.* Almost completely bio-transformed and rapidly excreted in urine and faeces *Plants.* After uptake by rice, converted to a non-herbicidal metabolite *Soil.* DT50 420 weeks on Planagan and Keyport silt loam soils. In rice fields, DT50 in water averages 46 days
Pretilachlor (Chloroacetamide)		Cell division inhibitor	Selective herbicide, readily taken up by hypocotyls, mesocotyls, and coleoptiles, and to a lesser extent by roots of germinating weeds	Broad leaf weed, grass and sedge controller	*Animals.* Substitution of the chlorine atom for glutathione to form a conjugate. Cleavage of the ether bond to yield an ethyl alcohol derivative. Both metabolities are susceptible to further degradation *Plants.* Substitution of chlorine atom to form a conjugate. Cleavage of the ether bond to yield an ethyl alcohol derivative. Hydrolytic and reductive removal of the chlorine atom *Soil.* Applied to paddy water, disappeared from the water by adsorption to the soil, where it is rapidly degraded under practical conditions, median DT50 (lab) 30 days. Due to strong soil adsorption, unlikely to leach

TABLE 30.4d Soyabean Herbicides

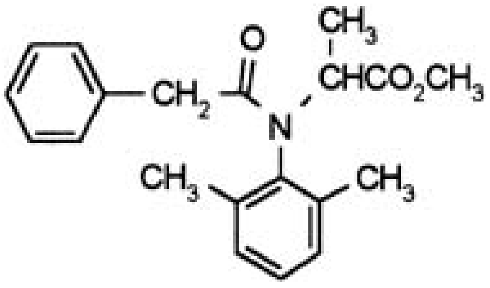
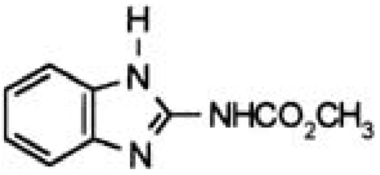
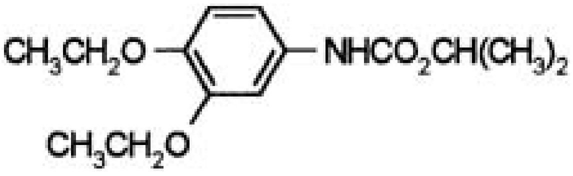
Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Activity	Environmental Fate

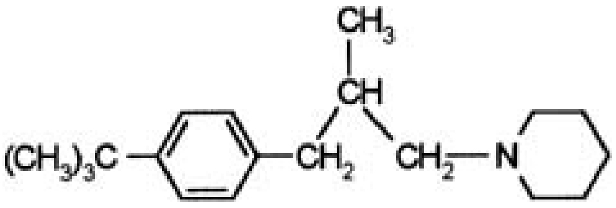
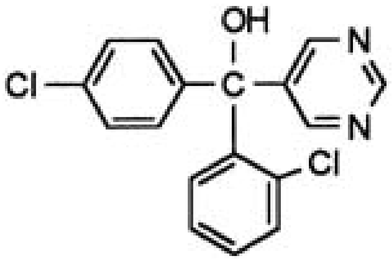
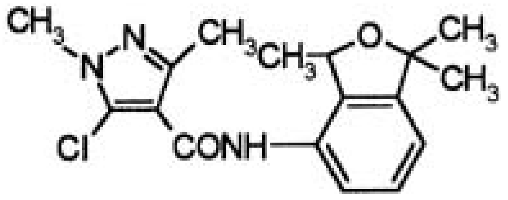
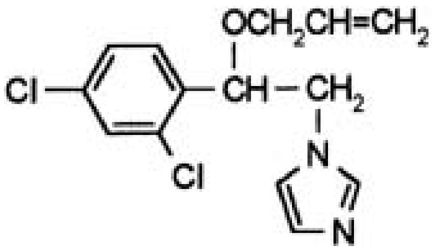
Glyphosate (Glycine derivative)	 $\mathrm{HO_2CCH_2NHCH_2\overset{O}{\overset{\|}{P}}(OH)_2}$	Inhibits 5-enolpyruvyl-shikimate-3-phosphate synthase (EPSPS), an enzyme of the aromatic acid and biosynthesis pathway. This prevents synthesis of essential aromatic amino acids needed for protein biosynthesis	Systemic herbicide, absorbed by rapid translocation throughout the plant	Non selective herbicide	*Animals.* In mammals, following oral administration, glyphosate is very rapidly excreted unchanged and does not bioaccumulate *Plants.* Slowly metabolised to aminomethylphosphonic acid ([*1066-51-9*]), which is the major plant metabolite *Soil/environment.* In soil (field), DT50 3174 days, depending on edaphic and climatic conditions. In water, DT50 varies from a few to 91 days. Photodegradation in water occurs under natural conditions, DT50 ca 28 days; no substantial photodegradation in soil was recorded over 31 days. In a lab. whole system with water and sediment, DT50 ca 14 days (aerobic), 1422 days (anaerobic). The major metabolite in soil and water is aminomethylphosphonic acid
Pendimethalin (Dinitro-aniline)		Microtubule assembly inhibitor	Selective herbicide absorbed by roots and leaves	Graminicide and broad-leaf weed controller	
Bentazone (Benzothiadiazinone)		Photosynthetic electron transport inhibitor at photosystem II receptor site	Selective, contact herbicide, absor-bed by leaves with very little translocation, but also absorbed by roots, with trans-location acropetally in the xylem	Broad leaf weed controller	*Animals.* In rats, the major metabolic routes for pendimethalin involve hydroxylation of the 4-methyl and *N*-1-ethyl groups, oxidation of these alkyl groups to carboxylic acids, nitro-reduction, cyclisation and conjugation (J. Zulian, *J. Agric. Food Chem.*, 1990, 38, 1743) *Plants.* In plants, the 4-methyl group on the benzene ring is oxidised to the carboxylic acid via the alcohol. The amino nitrogen is also oxidised. At harvest time, residues in crops are below the validated sensitivity of the analytical method (0.05 ppm) *Soil/environment.* In soil, the 4-methyl group on the benzene ring is oxidised to the carboxylic acid via the alcohol; the amino nitrogen is also oxidised. DT50 in soil is 34 months (A. Walker & W. Bond, *Pestic. Sci.*, 1977, 8, 359). Kd ranges from 2.23 (0.01% o.m., pH 6.6) to 1638 (16.9% o.m., pH 6.8) (H. J. Pedersen et al., *Pestic. Sci.*, 1995, 44, 131)

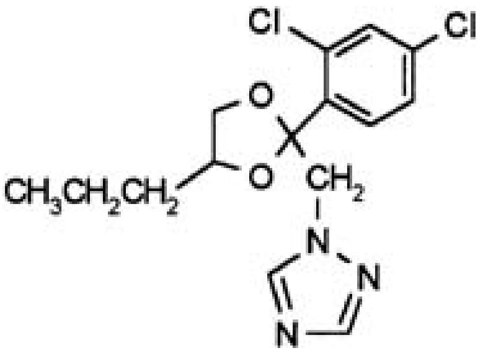
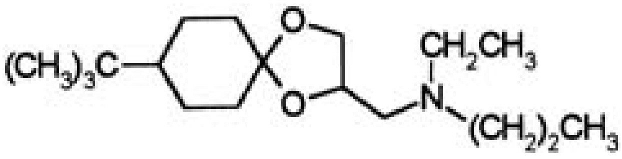
Fluazifop-*P*-butyl (Aryloxyphenoxypropionate)		Fatty acid synthesis inhibitor, by inhibition of acetyl CoA carboxylase (ACCase)	Quickly absorbed through the leaf surface, hydro-lyzed to fluazifop-*P*- and located through the phloem and xylem, accumulating in the rhizomes and stolons of perennial grasses and meristems of annual and perennial grasses	Graminicide	*Animals.* In mammals, fluazifop-*P*-butyl is metabolised to fluazifop-*P*, which is rapidly excreted *Plants.* In plants, fluazifop-*P*-butyl is rapidly hydrolysed to fluazifop-*P*, which is then partly conjugated. Ether cleavage gives the pyridone and propionic acid metabolites, which may both be further metabolised or conjugated *Soil/Environment.* c. 5800. In moist soils, rapid degradation of fluazifop-*P*-butyl occurs, DT50 <24 hr. The major degradation product is fluazifop-*P*, which is hydrolysed to 5-trifluoromethylpyrid-2-one, and 2-(4-hydroxyphenoxy) propionic acid, both of which are further degraded, ultimately to CO2 *Soil/environment.* In laboratory soil (40% MHC, pH 5.37.7), DT50 29 days (20°C). Field DT50 <4 week. Koc 3984. For degradation route, see fluazifop-*P*-butyl
Imazethapyr (Imidazolinone)		Branched chain amino acid synthesis (ALS and AHAS) inhibitor reducing levels of valine, leucine, isoleucine leading to disruption of protein and DNA synthesis	Systemic herbicide absorbed by roots and foliage with translocation in xylem and phloem and accumulation in meristetic regions	Germicide and broad leaved weed controller	*Animals.* In rats, following oral administration, 92% was excreted in the urine and 5% in the faeces within 24 hours. Residue levels in blood, liver, kidney, muscle, and fat tissues were <0.01 ppm after 48 hr *Plants.* Rapidly metabolised in non-susceptible plants; half-life in soya beans 1.6 days. The primary metabolic route in maize is oxidative hydroxylation at a carbon atom of the ethyl substituent on the pyridine ring *Soil/environment.*Half-life in soil 13 months
Sulfentrazone (Triazolinone)		Protoporphyrinogen oxidase inhibitor (Chlorophyl biosynthesis pathway)	Herbicide absorbed by roots and foliage, with translocation primarily in the apoplasm and limited movement in pholem	Graminicide and broadleaf weed controller	*Animals.* In rats, nearly all of administered sulfentrazone is excreted in the urine within 72 hr *Plants.* In soya beans, over 95% of the parent sulfentrazone is metabolised to the non-polar, ring-hydroxymethyl analogue within 12 hr. This analogue is also rapidly converted, over

the same time period, to three polar metabolites, two of which are glycosidic derivatives and one a non-glycoside metabolite
Soil/environment. Stable in soil (DT50 18 months). In water, stable to hydrolysis (pH 59), but readily undergoes photolysis (DT50 <0.5 days). Low affinity for organic matter (Koc 43), but is mobile only in soils with high sand content. Low potential to bioaccumulate

TABLE 30.5a Systemic Fungicides

Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Environmental Fate
Benalaxyl (Acylalanine)		Nucleic RNA-polymerase inhibitor	Systemic fungicide absorbed by roots, stems, leaves with translocation acropetally to all parts of the plant including subsequent growth	*Animals*. Rapidly metabolised, and eliminated in the urine (23%) and faeces (75%) within 2 days *Plants*. Slowly metabolised to glycosides in plants *Soil*. Slowly degraded by soil micro-organisms to various acidic metabolites. DT50 in silt loam soil 77 days
Carbendazim (Benzimidazole)		Inhibits betatubulin synthesis	Systemic fungicide, absorbed through roots and green tissues, with translocation acropetally. Acts by inhibiting development of the germ tubes, the formation of appressoria and the growth of mycelia	*Animals*. 66% was eliminated in the urine within 6 hr *Plants*. Readily absorbed by plants. One degradation product is 2-aminobenzimidazole *Soil*. 2-Aminobenzimidazole has been found as a minor metabolite. DT50 in soil 832 days under outdoor conditions. Mainly decomposed by microorganisms
Diethofencarb (*N*-phenyl carbamate)			Systemic fungicide, readily absorbed through leaves and roots and translocated throughout the plant	*Animals*. 98.5100% of 14C was excreted within 7 days. The major metabolic routes were deethylation of the 4-ethoxy group, cleavage of the carbamate linkage, acetylation and finally formation of the glucuronide and sulfateconjugates *Plants*. Readily degraded in plants *Soil*. Readily degraded in soil; DT50 <16 days under aerobic conditions; only very slightly degraded under anaerobic sterilized conditions

Fenpropidin (Piperidine)		Ergosterol biosynthesis inhibitor, by inhibition of steroid reduction (sterolΔ14reductase) and isomerisation (Δ8 to Δ7isomerase)	Systemic foliar fungicide with translocation acropetally in the xylem	*Animals*. Rapidly absorbed, distributed, metabolised and excreted in the urine and faeces. No bioretention potential. *Plants*. Relatively rapid and extensive degradation. Principal metabolic pathway involves hydroxylation of the piperidine ring and oxidation of the tertiary butyl group. DT50 in wheat and barley plants *c.* 411 days *Soil*. Strongly adsorbed and extensively degraded, DT50 58 (loam)95 (sandy loam). Fenpropidin and its metabolites have little or no tendency to leach
Fenarimol (Pyrimidine)		Ergosterol biosynthesis inhibitor	Systemic fungicide, translocated acropetally within the plant	*Animals*. Rapidly excreted *Plants*. Forms numerous photodegradation products. *Soil*. DT50 14130 (average 79) days
Furametpyr (Oxathiin)		Inhibitor of mitochondrial succinate oxidation	Fungicide with systemic and translaminar action	
Imazalil (Imidazole)		Steroid demethylation inhibitor	Systemic fungicide	*Animals*. 90% is eliminated in the metabolized form within 4 days *Plants*. Transformed into (-2,4-dichlorophenyl)-1*H*-imidazole-1-ethanol *Soil*. DT50 (field) 45 d; DT90 (field) 5468 days
Metalaxyl-*M* (Phenylamide: acylalanine)		Inhibits protein synthesis in fungi, by interference with the synthesis of ribosomal RNA	Systemic fungicide, absorbed through leaves, stems and roots	*Animals*. Rapidly absorbed and also rapidly and almost completely eliminated in urine and faeces. Metabolism proceeds via hydrolysis of the ester bond, oxidation of the 2-(6)-methyl group and of the phenyl ring and *N*-dealkylation. Residues in tissues were generally low and there was no evidence for accumulation or retention of metalaxyl-*M* or its metabolites *Plants*. Metabolized by more than four types of phase I reaction (oxidation of the phenyl ring, oxidation of the methyl group, cleavage of the methyl ester and *N*-dealkylation) to form eight metabolites; at phase II, most of the metabolites are sugar conjugated *Soil*. DT50 in soil 21 days (realistic range 530 days).

Propiconazole (Triazole)		Steroid demethylation inhibitor	Systemic foliar fungicide, translocated acropetaly in the xylem	*Animals*. Rapidly absorbed and also rapidly and almost completely eliminated with urine and faeces. Residues in tissues were generally low and there was no evidence for accumulation or retention of propiconazole or its metabolites. The major sites of enzymic attack are the propyl side-chain and the cleavage of the dioxolane ring, together with some attack at the 2,4-dichlorophenyl and 1,2,4-triazole rings *Plants*. Degradation through hydroxylation of the *n*-propyl side-chain and deketalisation of the dioxolan ring. After cleavage of triazole, triazolealanine is formed as the main metabolite. Metabolites are conjugated mostly as glucosides *Soil*. DT50 in aerobic soils (25° C) 4070 days. The main degradation pathways are hydroxylation of the propyl side-chain and the dioxolane ring, and finally formation of 1,2,4-triazole. Immobile in soil
Spiroxamine (Spiroketalamine)		Sterol biosynthesis inhibitor acting mainly by inhibition of Δ14reductase	Systemic fungicide, which readily penetrates into the leaf tissue followed by acropetal translocation to the leaf tip	*Animals*. Highly absorbed followed by fast elimination from the body (>97% within 48 hr). The radioactivity was readily distributed from the plasma into peripheral compartments. The main metabolite was the compound oxidized to the carboxylic acid in the *t*-butyl moiety. Metabolism proceeds either via oxidation of the *t*-butyl moiety to yield the carboxylic acid compound or via des-alkylation of the amino group resulting in the des-ethyl and des-propyl derivatives of spiroxamine *Plants*. Extensively metabolized by oxidation, desalkylation and cleavage of the ketal structure; the resulting metabolites bearing a hydroxylated *t*-butyl group or an aminodiol were further conjugated *Soil*. Readily degraded, ultimately to CO2; oxidation on the *t*-butyl moiety and des-alkylation of the amine are the primary reaction steps. The des-alkylated compounds were either further oxidised to the corresponding acids or further degraded to a ketone metabolite. Soil DT50 (lab. and field) in the range 3564 days. Bound rapidly to the sediment

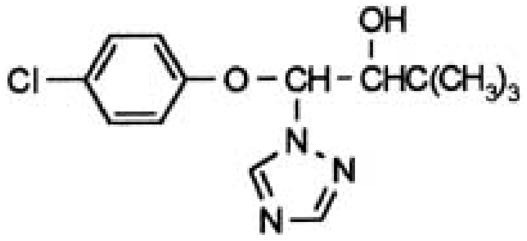
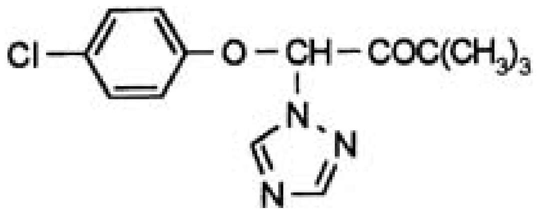
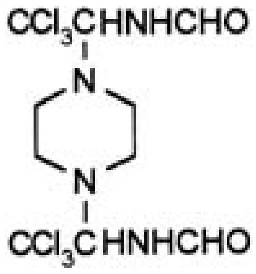
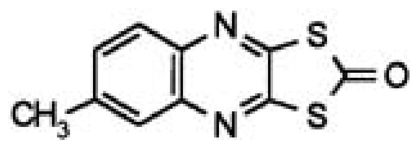
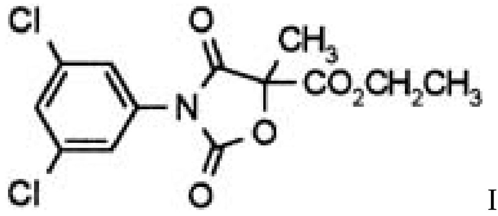
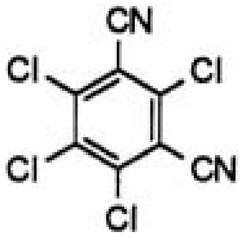
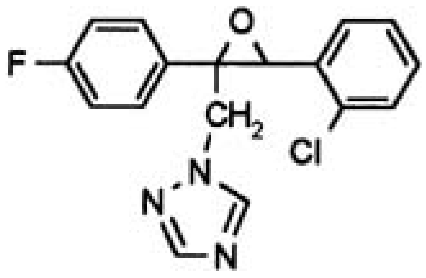
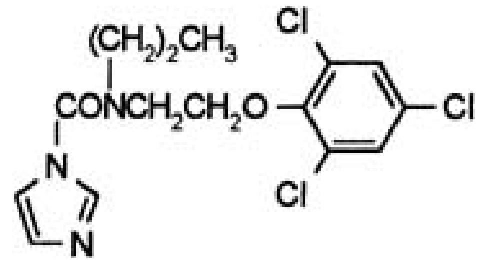
Triadimenol (Triazole)		Inhibits gibberellin and ergosterol biosynthesis and hence the rate of cell division	Systemic fungicide absorbed by roots and and leaves with ready translocation in young growing tissues but less ready translocation in older, woody tissues	*Animals*. Metabolized mainly by oxidation of the *tert*-butyl moiety to the corresponding alcohol and then to carboxylic acid. A small fraction of these compounds was conjugated *Plants*. The most important breakdown reactions are conjugation with various sugar compounds (especially hexose) and oxidation at the *tert*-butyl moiety. The resulting primary alcohol is likewise partly conjugated *Soil*. Triadimenol is a degradation product of triadimefon (*q.v.*). Degradation involving hydrolytic cleavage leads to the formation of 4-chlorophenol. DT50 (sandy loam) 110375 days; (loam) 240270 days
Triadimefon (Triazole)		Steroid demethylation (ergosterol biosynthesis) inhibitor	Systemic fungicide, absorbed by leaves and leaves with ready translocation in young growing tissues but less ready translocation in older, woody tissues	*Animals*. 8396% is excreted unchanged in the urine and faeces within 23 days. However, metabolism occurs in the liver, mostly to triadimenol (*q.v.*) and its glucuronic acid conjugates. Half-life in blood plasma is c. 2.5 hr *Plants*. In plants, the carbonyl group is reduced to a hydroxyl group, with the formation of triadimenol (*q.v.*) *Soil*. In soil, the carbonyl group is reduced to a hydroxyl group, with the formation of triadimenol (*q.v.*). DT50 of triadimefon in sandy loam *c.* 18 days, in loam c. 6 days
Triforine (Piperazine)		Ergosterol biosynthesis inhibitor	Systemic fungicide, absorbed by leaves and roots with translocation acropetally	*Soil*. A range of non-fungitoxic metabolic end-products are formed, presumably including piperazine. DT50 in soil c. 3 weeks. Does not accumulate in the environment

TABLE 30.5b Contact Fungicides

Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Environmental Fate
Bordeaux mixture (inorganic)		Cu++ is taken up by the spores during germination and accumulates until a sufficiently high concentration is achieved to kill the spore cell	Foliar fungicide	*Animals*. Copper is an essential element and is under homeostatic control in mammals
Chinomethionat (Quinoxaline)			Selective, non-systemic contact fungicide	*Animals*. Rapidly metabolized, and *c.* 90% is eliminated within 3 days in the faeces and urine. The main metabolite is chinomethionat acid (dimethylmercaptoquinoxaline-6-carboxylic acid), which also occurs in the conjugated form *Plants*. No penetration of the a.i. or metabolites in the fruit pulp was observed. The only metabolite detected was dihydromethylquinoxalinedithiol *Soil*. DT50 in standard soil land 2: 13 days

Chlozolinate (Dicarboximide)		Inhibitor of Lipid Peroxidation in mitochondrial membranes	Contact fungicide	*Animals*. Readily absorbed, metabolized and excreted. Metabolites identified in urine are: 3-(3,5-dichlorophenyl)-5-methyloxazolidin-2,4-dione, *N*-(3,5-dichlorophenyl)-2-hydroxypropionamide, *O*-1-carboxyethyl-*N*-3,5-dichloro-phenyl carbamate, and *N*-(3,5-dichloro-2(or 4)-hydroxyphenyl)-2-hydroxypropionamide and its sulfate and glucuronide conjugates *Plants*. Undergoes hydrolysis and decarboxylation processes, giving the same metabolites as those identified in animals. *Soil*. In silt-loam, sandy loam and clay loam soils, hydrolysis and decarboxylation occur; aerobic DT50 <7 hr
Chlorothalonil (Chloronitrile)		Conjugation with, and depletion of thiols (particularly glutathione) from germinating fungal cells, leading to disruption of glycolysis and energy production, fungistasis and fungicidal action	Non-systemic foliar fungicide	*Animals*. Not well absorbed following oral dosing. It reacts with glutathione in the gut lumen, or immediately on absorption into the body, to give mono-, di- or tri- glutathione conjugates. These may be excreted through urine or faeces, or subject to further metabolism resulting in thiol or mercapturic acid derivatives. In ruminants, the 4-hydroxy metabolite may also be present *Plants*. The majority of the residue remains as parent compound. The most abundant metabolite, 4-hydroxy-2,5,6-trichloroisophthalonitrile, is generally <10% of applied parent *Soil*. Low mobility to immobile. In aerobic and anaerobic soil studies, DT50 is 536 days. Degradation is faster in biotic aquatic systems, typical DT50 (aerobic) <8 hr, (anaerobic) <10 days
Epoxiconazole (Triazole)		Inhibitor of C-14-demthylase in sterol biosynthesis	Preventive and curative fungicide	*Animals*. Readily excreted via faeces. No major metabolites, but a high number of minor metabolites was identified. The important metabolic reactions were cleavage of the oxirane ring, hydroxylation of the phenyl rings and conjugation *Plants*. There is extensive degradation *Soil*. Degradation is by microbial activity, DT50 *c.* 23 months
Prochloraz (Imidazole)		Steriod demethylation (ergosterol biosynthesis) inhibitor		*Animals*. Rapidly metabolised initially by cleavage of the imidazole ring and quantitatively eliminated from the body. Whilst absorption following dermal exposure is low, residues in plasma and tissues are rapidly eliminated from the body *Plants*. The primary metabolite, *N*-formyl-*N*′-1-propyl-*N*-(2-(2,4,6-trichlorophenoxy)ethyl)urea, is formed from cleavage of the imidazole ring. This is degraded to *N*-propyl-*N*-(2-(2,4,6-trichlorophenoxy) ethyl)urea, which occurs in both free and conjugated forms. Other metabolites include 2-(2,4,6-trichlorophenoxy)ethanol, 2-(2,4,6-trichlorophenoxy) acetic acid, traces of 2,4,6-trichlorophenol and conjugates of the above. Little unchanged prochloraz is present *Soil*. Degrades in the soil to a range of mainly volatile metabolites. Well adsorbed onto soil particles, and is not readily leached DT50 537 days
Cozeb Alkylenebis (dithiocarbamate)	[-SCSNHCH2CH2NHCSSMn-]x Zny	Non-specific thiol reactant, inhibiting respiration	Fungicide with protective action	*Plants*. Extensively metabolized, forming ethylenethiourea, ethylenethiuram monosulfide, ethylenethiuram disulfide, and sulfur as transitory intermediates. Terminal metabolites are natural products, especially those derived from glycine *Soil*. Rapidly degraded by hydrolysis, oxidation, photolysis, and metabolism. DT50 in soil *c.* 615 days

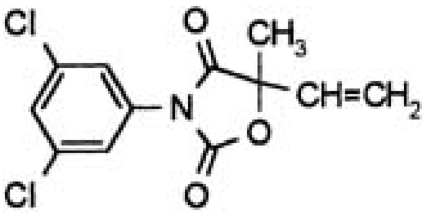
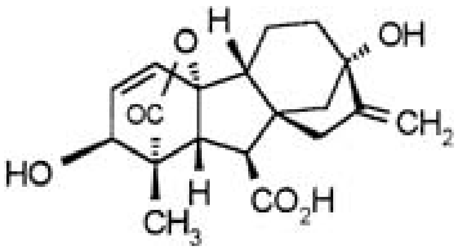
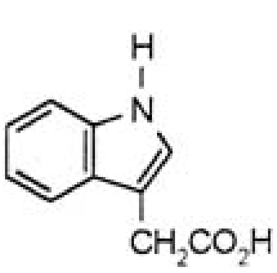
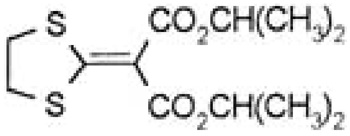
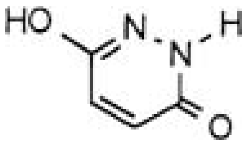
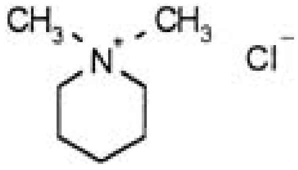
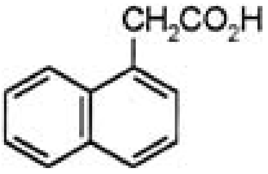
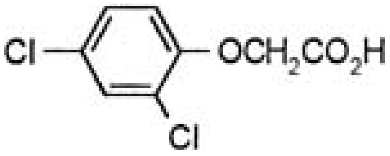
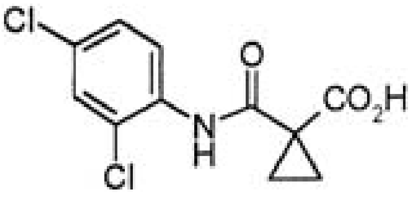
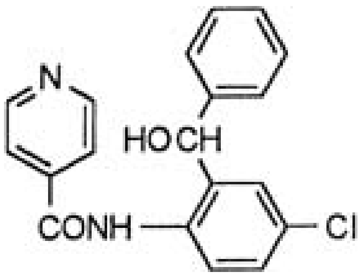
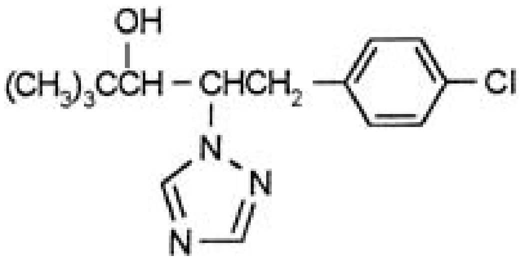
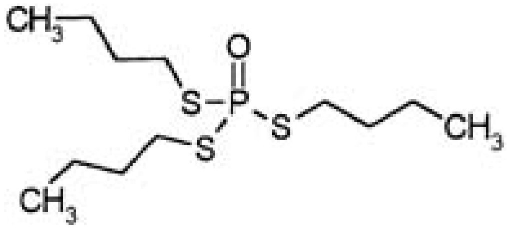
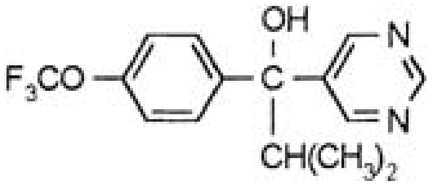
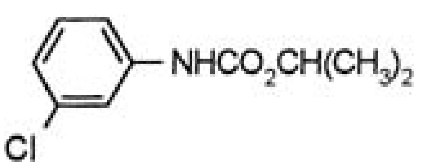
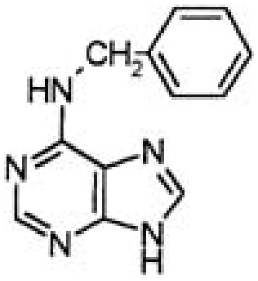
Sulfur (inorganic)		Nonpspectic thiol reactant, inhibiting respiration	Non-systemic fungicide	*Plants*. Degradation proceeds primarily by microbial reduction *Soil*. Slight oxidation to the volatile oxides
Vinclozolin (Dicarfboximide)			Non-systemic fungicide	*Animals*. The major metabolic routes are epoxidation of the vinyl group, followed by hydration of the intermediate epoxide, and by hydrolytic cleavage of the heterocyclic ring. Eliminated in approximately equal proportions in the urine and faeces, with the principal metabolite being *N*-(3,5-dichlorophenyl)-2-methyl-2,3,4-trihydroxybutanamide *Plants*. The primary metabolites are (1-carboxy-1-methyl) allyl 3,5-dichlorophenylcarbamate and *N*-(3,5-dichlorophenyl)-2-hydroxy-2-methyl-3-butenamide. Alkaline hydrolysis leads to loss of 3,5-dichloroaniline from vinclozolin and its metabolites. The metabolites exist as conjugates *Soil*. Metabolism occurs by loss of the vinyl group, cleavage of the 5-membered ring and eventual formation of 3,5,-dichloroaniline. Soil degradation takes place with half-lives of several weeks, and mainly leads to the formation of bound residues
Ziram (Dimethyldithio-carbamate	[(CH3)2NCS2]2Zn	Inhibitor of enzymes containing copper ions or sulfonyl groups	Basic contact, Foliar fungicide	*Animals*. Mostly eliminated within 12 days leaving 12% of the dose in the tissue and carcass after 7 days *Plants*. The major metabolite is dimethylamine salt of dimethyldithiocarbamic acid; tetramethylthiourea, carbon disulfide and sulfur can also be formed. Dimethyldithiocarbamic acid can be present as the free acid or as the metabolic conversion products. *Soil*. A. aerobic DT50 42 hr. Unlikely to leach

TABLE 30.6 Plant Growth Regulators

Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Environmental Fate
Chlormequat chloride (Quaternary ammonium)	ClCH2CH2N+(CH3)3 Cl-	Gibberelin biosynthesis inhibitor	Inhibits cell elongation. Also influences developmental cycle, leading to increased flowering and harvest. May also increase chlorophyll formation and root development	*Animals*. 97% is eliminated within 24 hr, principally unchanged *Plants*. Converted to choline chloride *Soil*. Rapidly degraded by microbial activity. DT50 in 4 soils averaged 32 days at 10°C; 128 days at 22°C. Low to medium mobility
Ethephon (Ethylene generator)	 $ClCH_2CH_2\overset{O}{\overset{\parallel}{P}}(OH)_2$		Plant growth regulator with systemic properties. Penetrates into the plant tissue and is decomposed to ethylene, which affects growth processes	*Animals*. Rapidly excreted intact via the urine, and as ethylene via the expired air *Plants*. Rapidly undergoes degradation to ethylene *Soil*. Rapidly degraded and strongly adsorbed; unlikely to leach
Gibberellic acid (Gibberellins)			Shows physiological and morphological effects on the plant parts above soil surface at extremely low concentrations. Translocated Plant Growth Regulator	
Indol-3-ylacetic acid (Auxin)			Affects cell division and cell elongation	*Soil*. Rapidly degraded in soil
Isoprothiolane (Phosphorothiolate)		Inhibits penetration and elongation of infection hyphae, by inhibiting formation of infection peg or cellulose secretion	Absorbed by leaves and roots with translocation acropetally and basipetally	

Maleic hydrazide		Inhibits cell division in the meristematic regions.	Absorbed by leaves and roots with translocation in xylem and phloem	*Animals*. 4362% of the dose excreted unchanged within 48 hr *Plants*. Various acids, e.g., succinic, fumaric, and maleic, are found as metabolites *Soil*. DT50 c. 11 hr
Mepiquat chloride (Quaternary ammonium)		Inhibits biosynthesis of gibberellic acid	Plant Growth regulator, absorbed and translocated throughout the plant	*Animals*. c. 48% is excreted in the urine *c.* 38% in the faeces, with <1% remaining in the tissues. The unmetabolised material constitutes c. 90% in each case *Soil*. DT50 1097 days at 20 ± 2°C and 40% of maximum
1-naphthyl acetic acid (Synthetic auxin)			Plant Growth Regulator with auxin like activity	
2,4-Dichlorophenoxy acetic acid (aryloxyalkanoic acid)		Synthetic auxin, acting like indol-lacetic acid	Salts are readily absorbed by the roots while esters are readily absorbed by foliage. Translocation occurs, with accumulation principally at the meristematic regions of shoots and roots. Acts as a growth inhibitor	*Animals*. Elimination is rapid, and mainly as the unchanged substance. Following single doses of up to 10 mg/kg, excretion is almost complete after 24 hr, although, with higher doses, complete elimination takes longer. The maximum concentration in organs is reached after c. 12 hr *Plants*. Metabolism involves hydroxylation, decarboxylation, cleavage of the acid side-chain, and ring opening *Soil*. Microbial degradation involves hydroxylation, decarboxylation, cleavage of the acid side-chain, and ring opening. Half-life in soil <7 days. Rapid degradation in the soil prevents significant downward movement
Cyclanilide (Anilide)		Inhibits polar auxin transport		*Animals*. Rapidly excreted, primarily as unchanged cyclanilide *Plants*. Little degradation occurs in plants; cyclanilide is the major residue *Soil*. Low to moderate persistence, DT50 c. 16 days under aerobic conditions. Degrades primarily by microbial activity. Medium to low mobility
Inabenfide (Pyridine)		Inhibits gibberellin biosynthesis	Plant Growth Regulator which shortens lower internodes and upper leaf blades	*Animals*. The major urinary metabolite is 4-hydroxyinabenfide *Plants*. Metabolized to inabenfide ketone *Soil*. Half-life under Japanese paddy field conditions, c. 4 months

Thidiazuron (Phenylurea)		Cytokinin activity	Plant Growth Regulator absorbed by the leaves which stimulates formation of an abscission layer between the plant stem and leaf petioles, causing dropping of the entire green leaves	*Animals*. Metabolism involves hydroxylation of the phenyl group, followed by formation of water-soluble conjugates. Following oral administration, the compound is excreted in the urine and faeces within 96 hr *Plants*. Only small amounts of residue (normally <0.1 mg/kg) are likely in cottonseed *Soil*. Strongly absorbed by soil. DT50 in soil c. 26144 days (aerobic), 28 days (anaerobic)
Paclobutrazol (triazole)		Inhibits gibberellin and sterol biosynthesis and hence the rate of cell division	Plant Growth Regulator taken up into xylem through the leaves, stems or roots and translocated to the growing sub-apical meristems. Produces more compact plants and enhances flowering and fruiting	*Soil*. Soil DT50 0.51.0 years in general; in calcareous clay loam (pH 8.8, 14% o.m.), DT50 <42 days; in coarse sandy loam (pH 6.8, 4% o.m.), DT50 >140 days
Tribufos (Phosphorothioate)			Plant Growth Regulator absorbed by leaves. Stimulates foration of an abscission layer between the plant stem and the leaf petioles, causing the dropping of entire green leaves	*Animals*. Rapidly absorbed and metabolized; 96% of the administered radioactivity was excreted within 72 hr. Metabolism proceeds by hydrolysis followed by methylation and successive oxidation of butylmercaptan, yielding the main metabolite (3-hydroxy)-butylmethylsulfone *Plants*. Unmetabolised tribufos is the primary residue in treated cotton *Soil*. Very strongly adsorbed, leaching is extremely unlikely. The half-life under field conditions is 27 weeks. The main metabolite is 1-butane sulfonic acid
Flurprimidol (Pyrimidinyl carbanol)		Gibberellin synthesis inhibitor	Plant Growth Regulator absorbed by the leaves and roots with translocation in the xylem and ploem. Reduces internode elongation	*Animals*. In mammals, the skin forms is significant barrier to absorption. Following oral administration, excretion follows in the urine and faeces within 48 hr, and more than 30 metabolites have been identified. No accumulation potential *Soil*. Degradation in soil under aerobic conditions leads to more than 30 metabolites
Chlorpropham (Carbamate (mi))		Mitosis inhibitor (Microtubule organization)	Absorbed by roots and coleoptile and readily translocated acropetally	*Animals*. The principal metabolic route is by hydroxylation at the para position and conjugation of the resultant 4-hydroxychloropropham with sulfate. There is also some hydroxylation of the isopropyl residue *Plants*. Three major metabolites have been identified, isopropyl *N*-4-hydroxy-3-chlorophenylcarbamate, isopropyl *N*-5-chloro-2-hydroxyphenylcarbamate and 1-hydroxy-2-propyl-3′-chlorocarbanilate, These aglycones are found in plants as water-soluble conjugates of glucose or other plant components *Soil*. Microbial degradation leads to the production of 3-chloroaniline by an enzymic hydrolysis reaction, with

				liberation of CO2. DT50 in soil c. 65 days (15°C), 30 days (29°C)
6-Benzylaminopurine (Cytokinin)		Stimulates RNA, RuDP-carboxylase, NADP-gleceraldehyde-3-phosphate-dehydrogenase, protein synthesis	Synthetic cytokinin, little translocated	*Animals*. Almost all of administered 14C was excreted in urine and faeces. Three metabolites were identified *Plants*. More than nine metabolites were identified. Urea is an end product *Soil*. 16 Days after application to soil at 22°C, 6-benzylaminopurine had degraded to 5.3% (sandy loam) and 7.85% (clay loam soil) of applied dose. Other studies indicate DT50 79 weeks

Delivery Systems of Agrochemicals

Agrochemicals are generally formulated with surface-active agents in dry forms (WP, WDG) or wet forms (EC, SC) for efficient delivery at the site of action. This might include sticking to the foliage, translaminar action in the leaves or through uptake by the root system. An efficient delivery system plays an important role in optimizing dosages per hectare of a given agrochemical in order to achieve maximum efficiency of pest control while reducing risk to personnel and livestock. Selective uses of appropriate sprayers, nozzles are other aids for the same purpose. Good farm practices during spray application and clean-up controls wasteful non-target dissipation.

Obsolescence of Agrochemicals

Resistance to agrochemicals occurs through the natural selection. Fungi with short life cycles exhibit the most pronounced resistance development. Insects are in an intermediate category with respect to life cycle, while weeds (plants) take much longer. Resistance management is brought through rotational or combination uses of agrochemicals with differing modes of action. Sub-lethal dosages are strictly to be avoided, as are extreme over-applications. These measures improve the functional life-time of a product. These practices are very pertinent since major breakthroughs with a new class of agrochemicals usually occur only once in 25 years or so. Availability of safer, more efficacious, cost effective agrochemicals tends to render prevailing agrochemicals obsolete.

Even with all of the advances to date, it would be wrong to state that all issues of plant protection are adequately addressed. There are still issues defying solutions, and these will continue to attract R&D efforts. As higher and higher farm productivities are achieved, coupled with more judicious distribution of food, the growing global population should enjoy higher levels of nutrition. Agrochemicals will continue to play an important role in health care, protection of farm produce, production of medicinal plants of value, and offer recyclable feedstock to produce chemicals currently derived solely from diminishing petroleum feed stocks. Hence, the value of the agrochemical industry should be judged more on its multiple impacts on the improved human condition rather than its modest size.

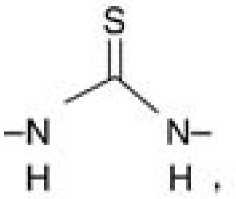
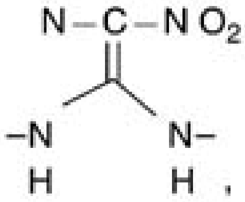
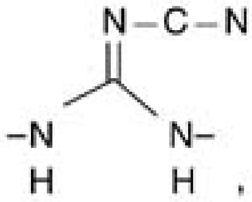
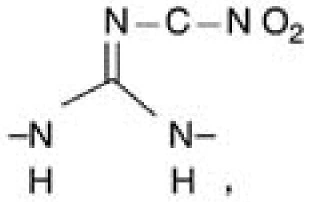
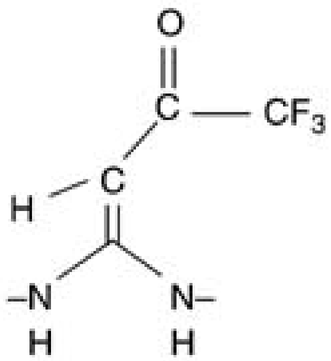
30.3 Products of the Agrochemical Industry

Many of the chemical structures of Agrochemicals demonstrating similar target (enzyme) specific biological activity can be rationalized under the term *BIOISOSTERISM*. Bioisosterism is a phenomenon where molecules possessing related structure have similar or antagonistic properties. Bioisosterism is the biological analog of isosterism, which is the close physical similarity of molecules or ions having the same number of atoms and valence electrons, such as CO and N2. This similarity is thought to explain certain analogies amongst the physical constants of molecules.

No two substituents are exactly alike. Any substitution impacts size, shape, electronic distribution, lipophilicity, pKa, chemical reactivity, susceptibility to metabolism, and the like. The bioisosteric approach is the total change induced by substituent replacement on the potency, selectivity, duration of action, bioavailability and toxicity, of an agrochemical. The following groups are examples of bioisosterism:

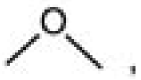
(A) CO2H, SO2NHR, SO3H, PO(OH)2, PO(OH)NH2, CONHCN

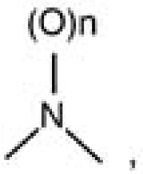
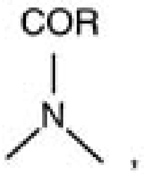
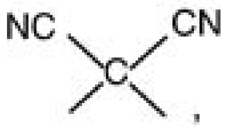
(B) F, Cl, Br, CF3, CN, SCN, N(CN)2, C(CN)3

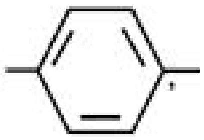
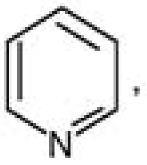
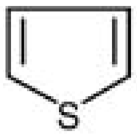
(C) OH, NHCOR, NHSO2R, CH2OH, NHCONH2, NHCN, CH(CN)2

(D)



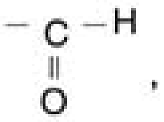




(E) $-(CH=CH)_n$,

, and so on

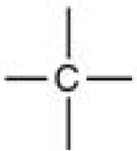
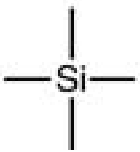
(F)



$-\underset{\underset{O}{\|}}{C}-H$,

$-C \equiv N$

(G)

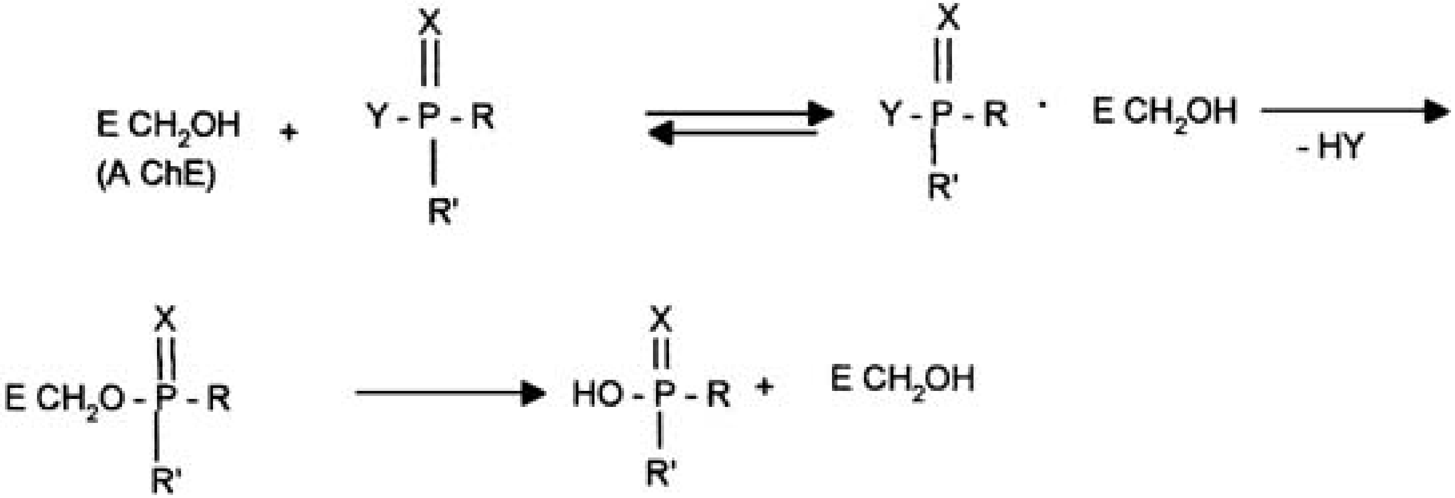
The concept of bioisosterism has been used to theoretically evaluate structural variation within the lead structures of synthetic or natural origin prior to and during the preparation of molecules of specified efficacy, safety, stability and so on.

30.4 Structural Basis of Agrochemicals

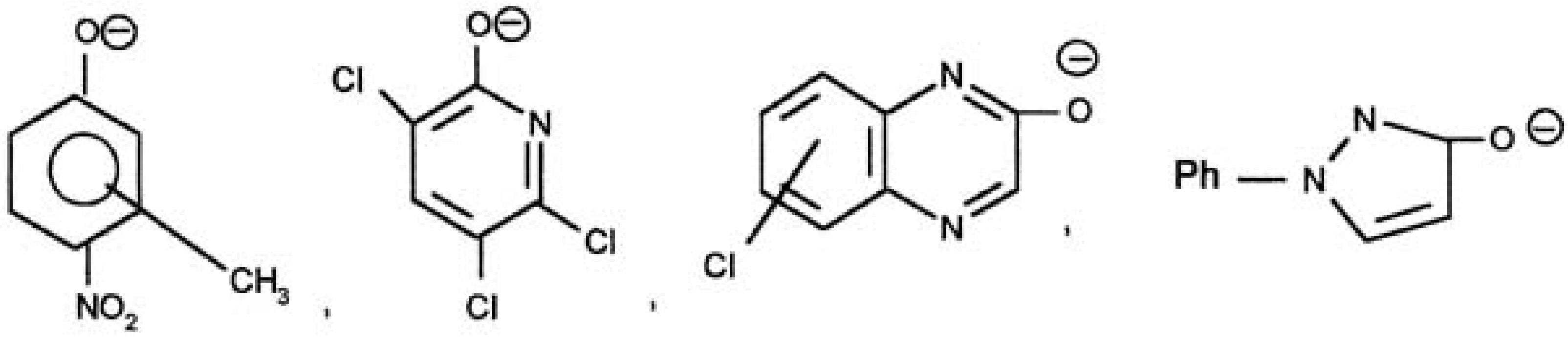
Organophosphorous Agrochemicals

Insecticides (Tables 30.3a, b)

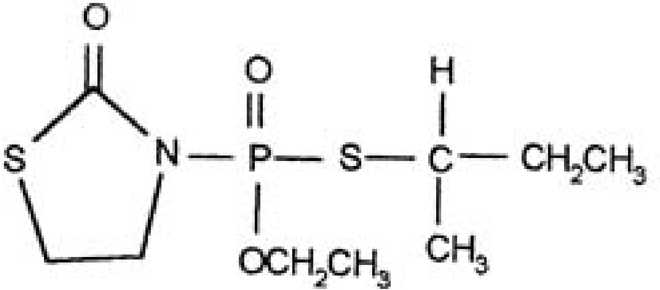
organophorous insecticides bind to Acetylcholinesterase (AchE), that is,



where R and R′ are lower alkyl alkoxy, alkylthio, or substituted amino groups; X is oxygen or sulfur; and Y is a good leaving group, for example aryloxy groups substituted with electron with drawing substituents.



Y can also be a mixed (C, N, S) heterocycle as in the case of the nematicide Fostiazate.

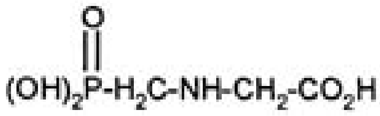


Nematicide
(Fosthiazate)

Herbicides (not numerous) (Tables 30.4ad)

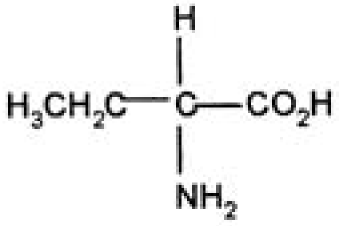
Organophosphorous herbicides are derived from amino acids, for example,

$H_2N\text{-}CH_2\text{-}CO_2H$

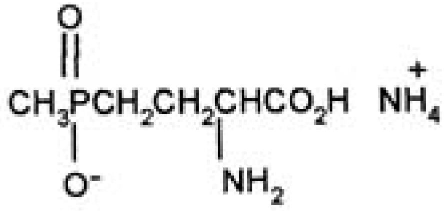


$(OH)_2\overset{\overset{O}{\|}}{P}\text{-}H_2C\text{-}NH\text{-}CH_2\text{-}CO_2H$ Glyphosphate

Glycine derivative *N*-(phosphonomethyl) glycine



$H_3CH_2C\text{—}\overset{\overset{H}{|}}{\underset{\underset{NH_2}{|}}{C}}\text{—}CO_2H$

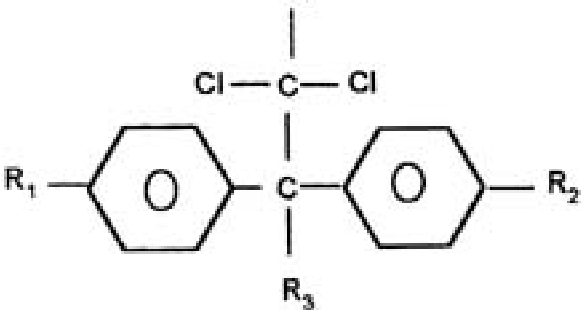


$CH_3\underset{\underset{O^-}{|}}{\overset{\overset{O}{\|}}{P}}CH_2CH_2\underset{\underset{NH_2}{|}}{C}HCO_2H \quad \overset{+}{N}H_4$ Gluphosinate

Homoalanine derivate Ammonium 4[Hydroxy(methyl)phosphonoyl]homoalalinate

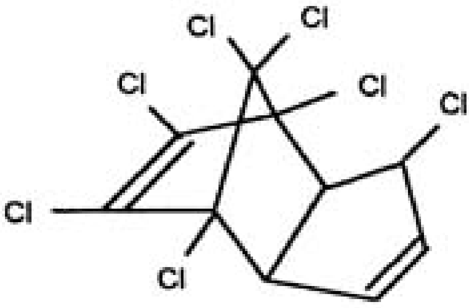
Others are: fosamine (herbicide used on non-crop areas, in meadows and pastures), piperofos (used in rice), anilofos (used on rice).

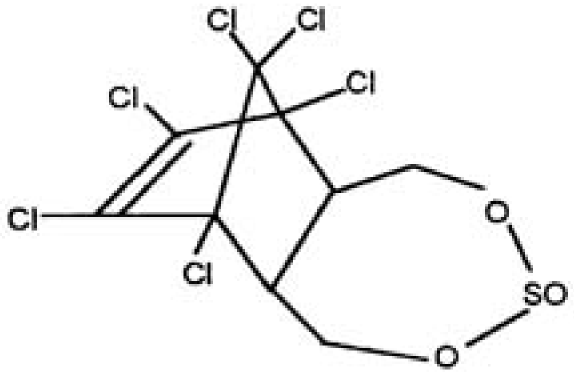
Organochlorine



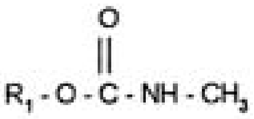
R1 = R2 = Cl, R3 = H-DDT
R1 = R2 = Cl, R3 = OH-Dicofol.

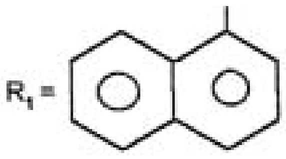
OrganochlorineCyclodiene

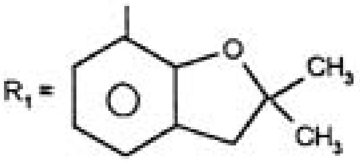

Heptachlor

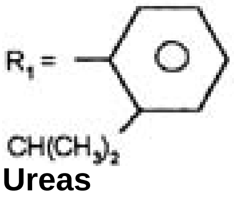

Endosulfan

Carbamates

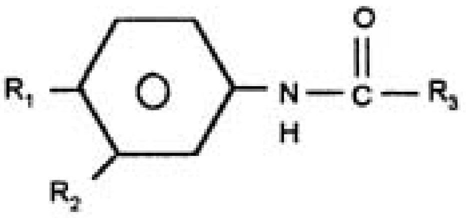

$R_1\text{-}O\text{-}\overset{\overset{\displaystyle O}{||}}{C}\text{-}NH\text{-}CH_3$


Carbaryl

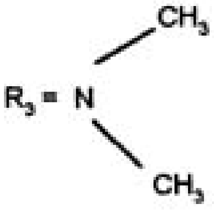

Carbofuran


Isoprocarb

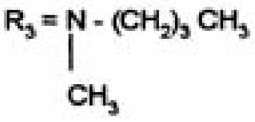
Ureas



$R_1 = CH(CH_3)_2$, $R_2 = H$,

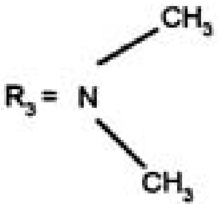

Isoproturon

$R_1 = R_2 = Cl$,

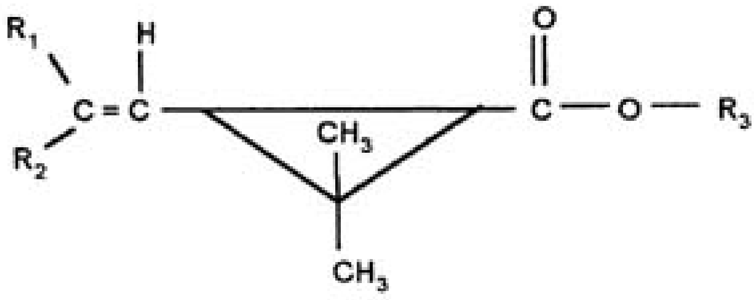

$R_3 = \underset{\underset{\displaystyle CH_3}{|}}{N}\text{-}(CH_2)_3\,CH_3$

Neburon

$R_1 = R_2 = Cl$,

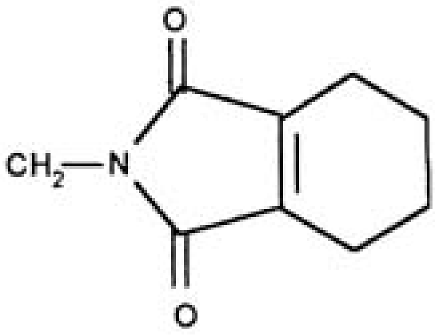

Diuron

Pyrethroids



(a)

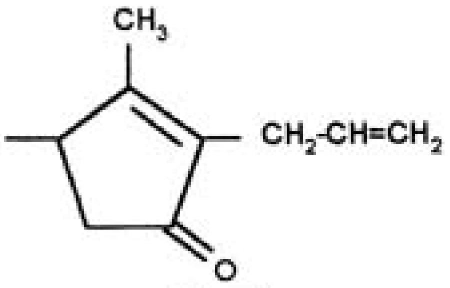
$R_1 = R_2 = CH_3$, R_3 =



Tetramethrin

(b)

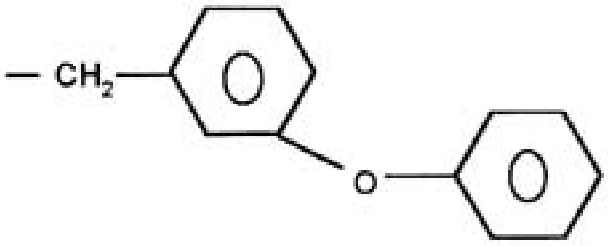
$R_1 = R_2 = CH_3$, R_3=



Bioallethrin

(c)

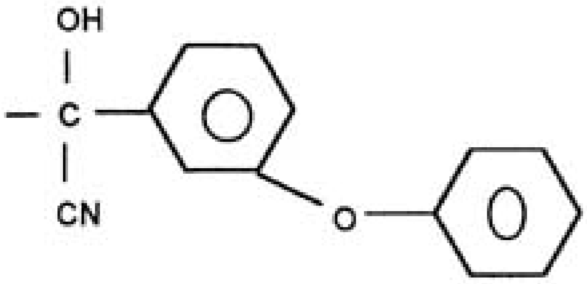
$R_1 = R_2 = Cl$, R_3 =



Permethrin

(d)

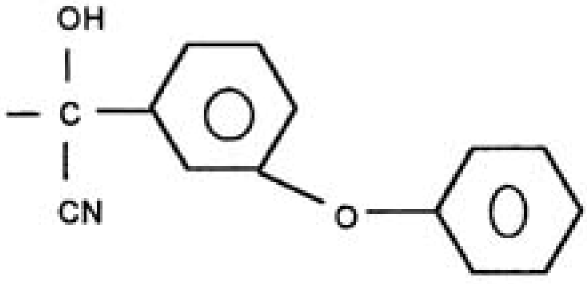
$R_1 = R_2 = Cl$, R_3 =



Cypermethrin

(e)

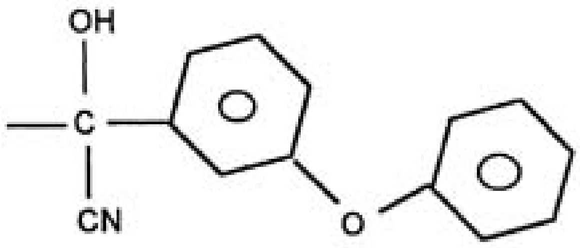
$R_1 = R_2 = Br$, R_3 =



Deltamethrin

(f)

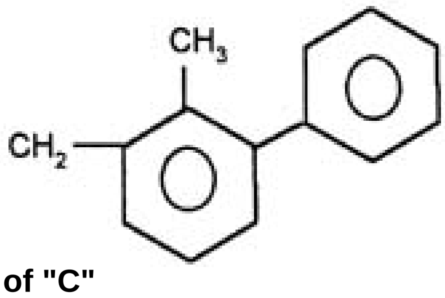
$R_1 = Cl$, $R_2 = CF_3$, R_3=



Lamdacyhalothrin

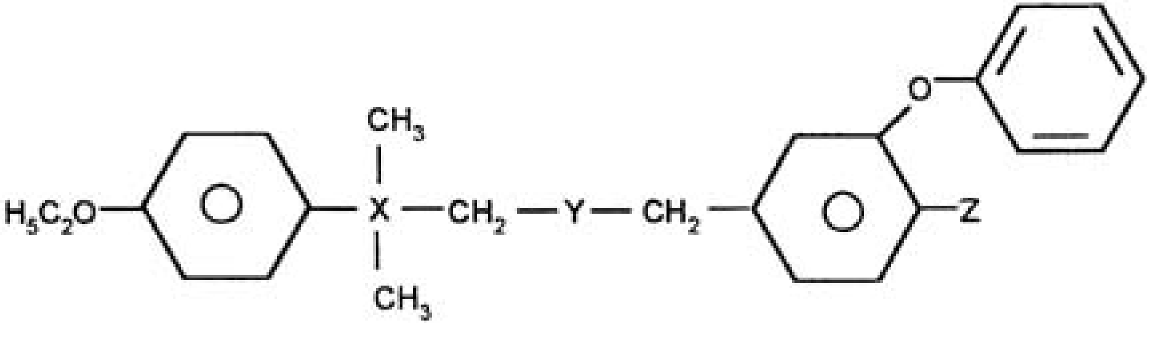
(g)

$R_1 = Cl$, $R_2 = CF_3$, R_3 =



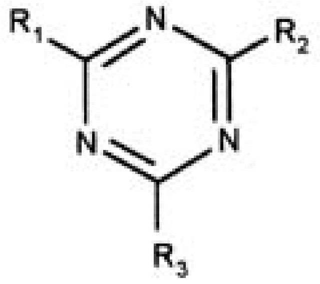
Bifenthrin

Non-ester pyrethroids/"Si" in lieu of "C"



X = C, Y = O, Z = H Ethofenprox
X = Si, Y = CH2, Z = F Silafluofen

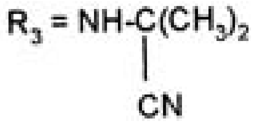
Triazenes



R_1 = Cl, $R_2 = NHCH_2CH_3$, $R_3 = NHCH_2CH_3$ Simazine

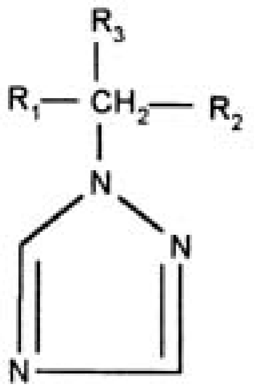
R_1 = Cl, $R_2 = NHCH_2CH_3$, $R_3 = NHCH(CH_3)_2$ Atrazine

R_1 = Cl, $R_2 = NHCH_2CH_3$,

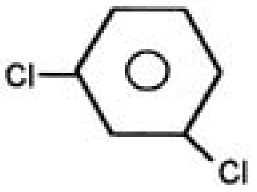


$R_3 = NH\text{-}\underset{\displaystyle CN}{\underset{|}{C}}(CH_3)_2$ Cyanazine

Azoles

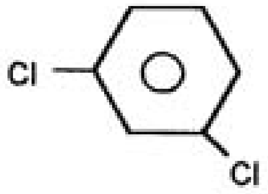


R_1 =

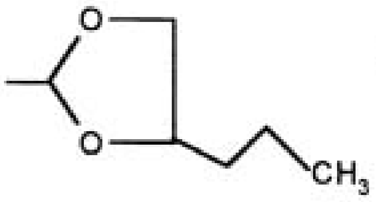


, R_2 = $(CH_2)_3CH_3$, R_3 = OH Hexaconazole

R_1 =

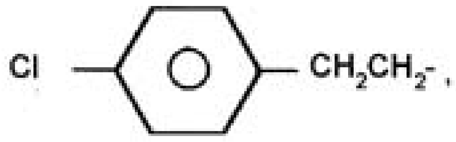


, $R_2 + R_3$ =



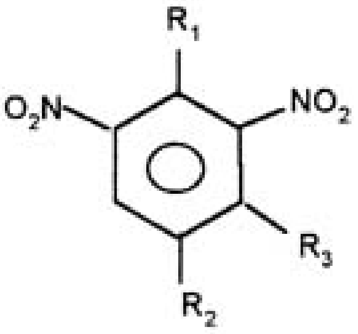
Propiconazole

R_1 =



R_2 = $-C(CH_3)_3$, R_3 = OH Tebuconazole

Dinitroanilines

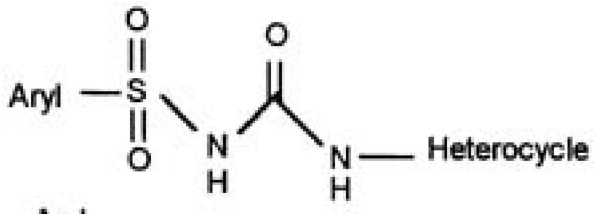


$R_1 = NHCH(CH_2CH_3)_2$, $R_2 = CH_3$, $R_3 = CH_3$ Pendimethalin

$R_1 = N(CH_2CH_2CH_3)_2$, $R_2 = CF_3$, R_3 = H Trifluralin

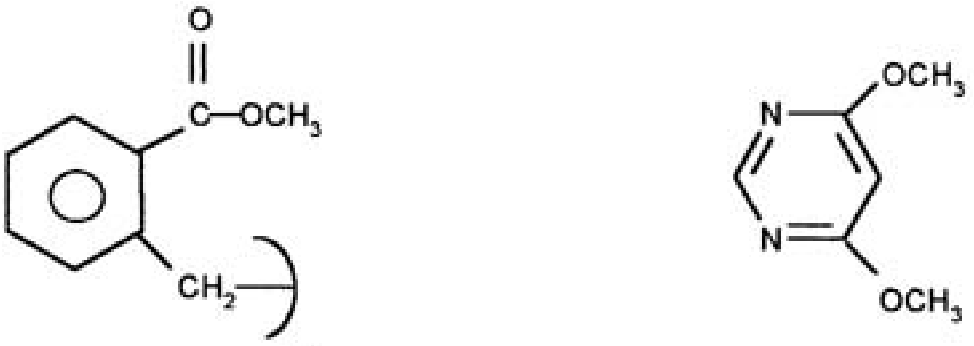
Chloracetanilides

Sulfonyl Ureas

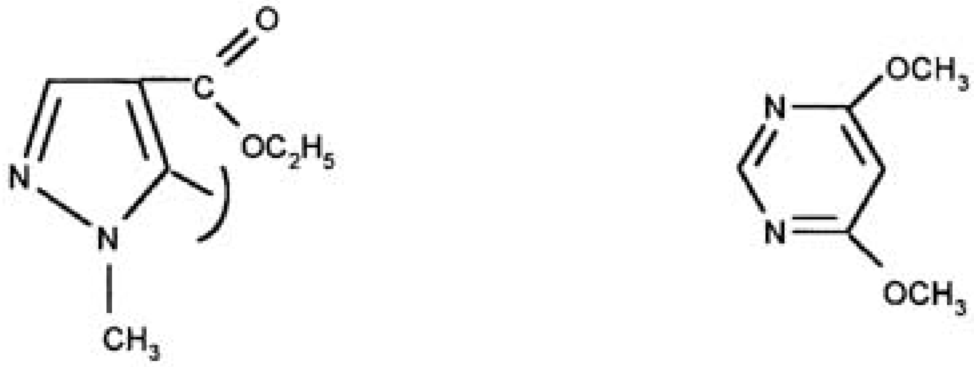


Aryl

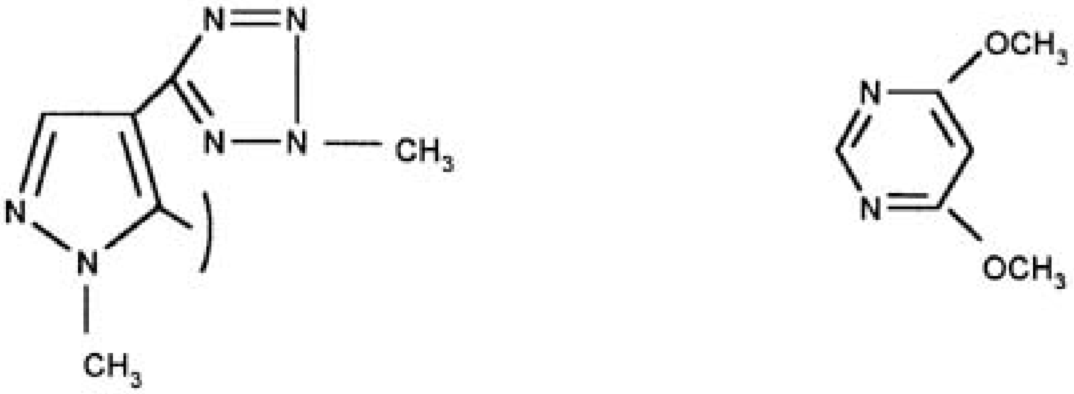
Heterocycle



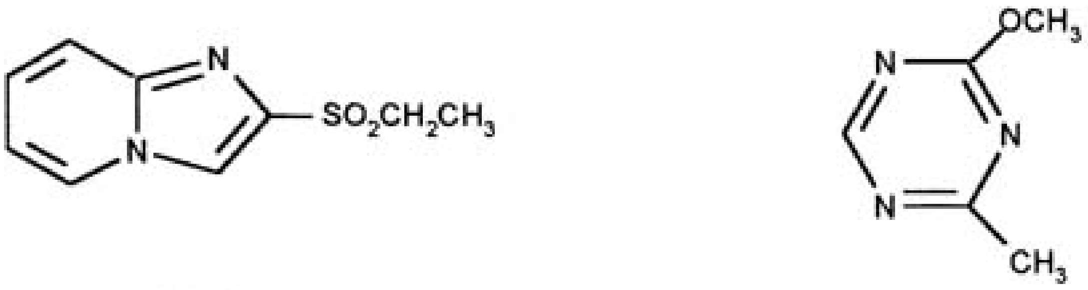
Rice Herbicide
Bensulfuron



Rice Herbicide
Pyrazosulfuron

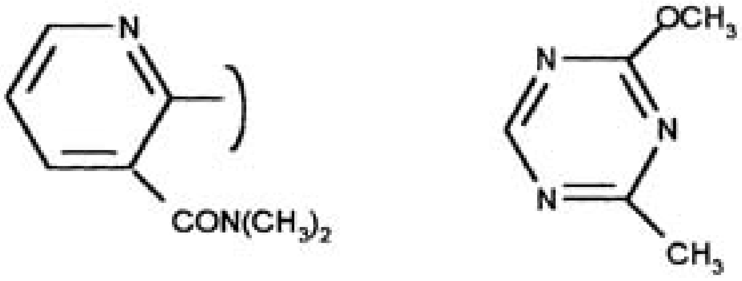


Rice Herbicide
Azimsulfuron

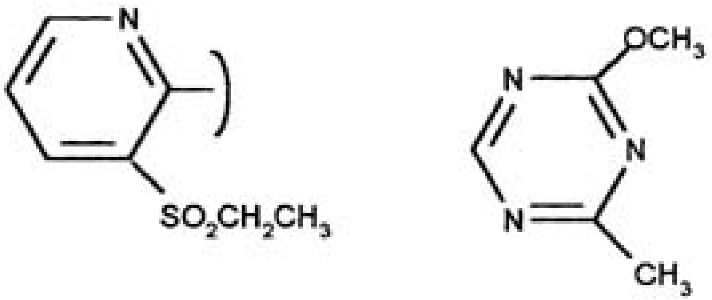


Wheat Herbicide
Sulfosulfuron

Wheat Herbicide
Metsulfuron

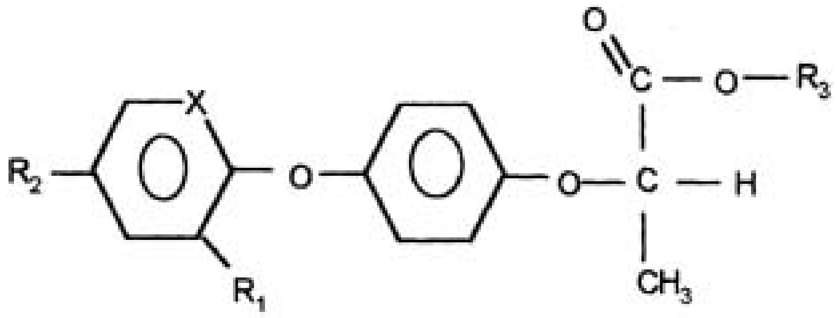


Corn Herbicide
Nicosulfuron



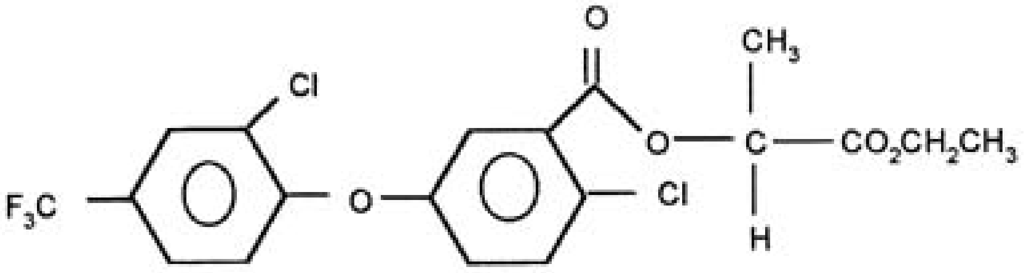
Corn Herbicide
Rimsulfuron

Aryloxyphenoxypropionic Acids (Cereal Herbicides)
Clodino/cyhalo/diclo/fenoxa/fluazi/haloxypropaquiza-fops

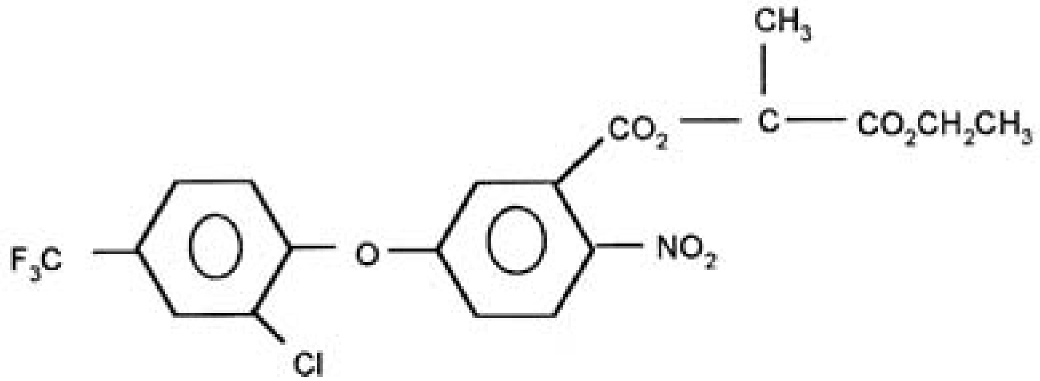


1. X = H, R = H, R2 = H, R3 = MethylDiclofop
2. X = H, R1 = F, R2 = CN, R3 = *n*-ButylCyhalofop
3. X = N, R2 = Cl, R1 = F, R3 = CH2C≡CHClodinofop
4. X = N, R1 = H, R2 = CF3, R3 = *n*-Butyl = Fluazifop 4.5

Diarylether Carboxylic Acids (Cereal Herbicides)

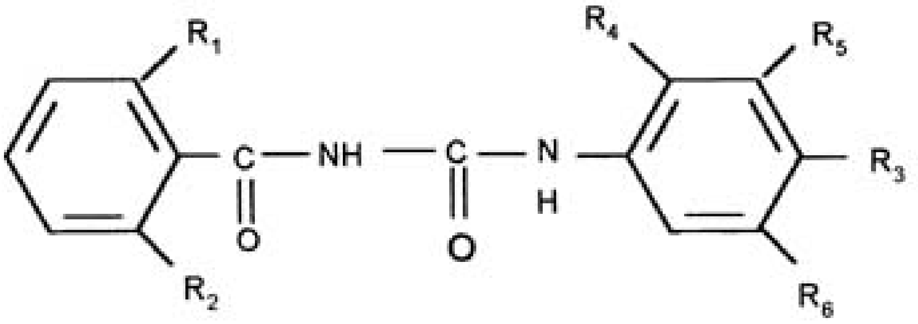


HC-252



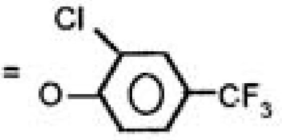
Fluoroglycofen

Benzoyl Ureas (Insect Growth Regulators) (Table 30.7)



$R_1 = R_2 = F$, $R_3 = Cl$, $R_4 = R_5 = R_6 = H$ Diflubenzuron

$R_1 = R_2 = F$, $R_3 =$

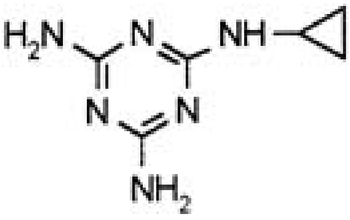
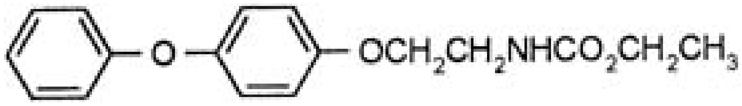
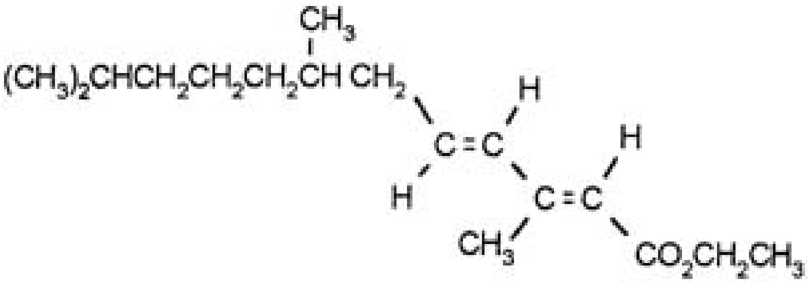
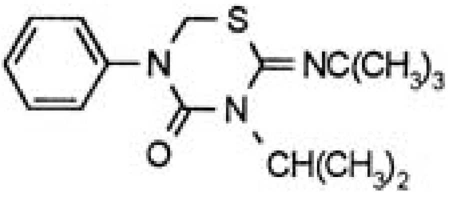


$R_4 = F$, $R_5 = R_6 = H$ Flufenoxuron

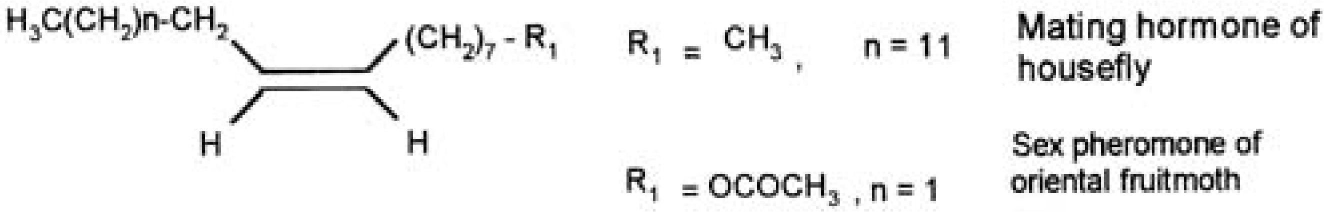
$R_1 = R_2 = F$, $R_3 = OCF_2CHF_2$, $R_4 = H$, $R_5 = R_6 = Cl$ Hexaflumuron

TABLE 30.7 Insect Growth Regulators

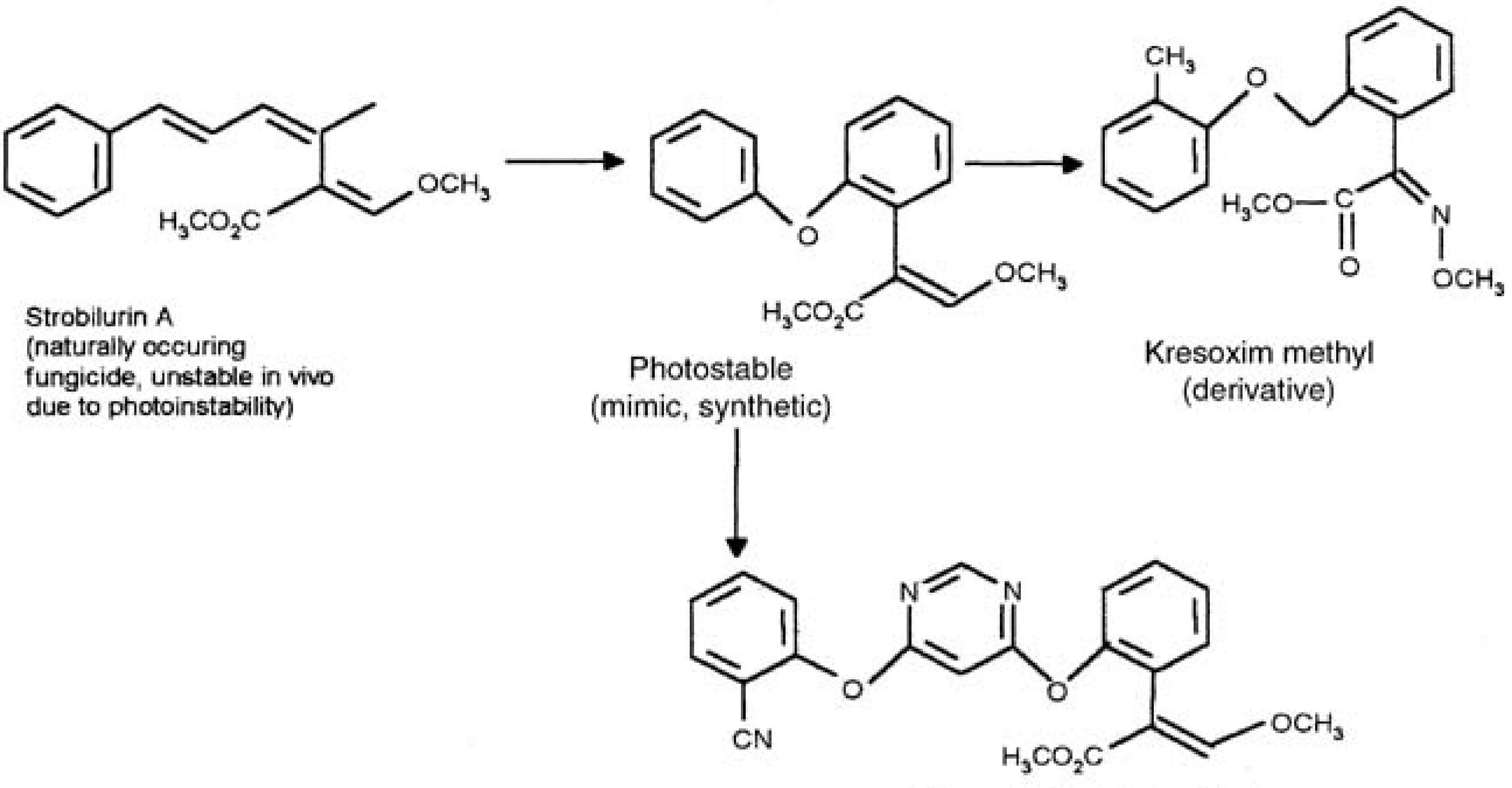
Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Environmental Fate
Diflubenzuron (Benzoylurea)		Chitin synthesis inhibitor and so interferes with the formation of insect cuticle	Non-systemic insect growth regulator with contact and stomach action. Acts at the time of insect moulting or at hatching of eggs	*Animals*. Elimination is partly as the unchanged parent compound in the faeces, partly as hydroxylated metabolites (for c. 80%) and as 4-chlorophenylurea plus 2,6-difluorobenzoic acid (for c. 20%). The intestinal absorption is strongly related to the dosage administeredthe higher the dosage, the more (relatively) is excreted unchanged in the faeces. *Plants*. Non-systemic. Not metabolized on plants *Soil*. Strongly absorbed by soil/humic acid

				complex and is virtually immobile in soil. Rapidly degraded in soil, with a half-life of <7 days. The principal degradation products are 4-chlorophenylurea and 2,6-difluorobenzoic acid
Cyromazine (Triazine)		Interfere with moulting and pupation	Insect Growth Regulator with contact action. When used on plants action is systemic. If applied to leaves, it exhibits a strong translaminar effect, applied to soil it is translocated acropetally after absorption by roots	*Animals*. Efficiently excreted, mainly as the parent compound *Plants*. Rapidly metabolized. The principal metabolite is melamine *Soil*. Cyromazine and its main metabolite melamine are moderately mobile. Efficiently degraded by biological mechanisms
Fenoxycarb (Carbamate)			Non neurotoxic insect growth regulator with contact and stomach action. Exhibits a strong juvenile hormone activey, inhibiting metamorphosis to the adult stage and interfere with the moulting of early instar larvae	*Animals*. The major metabolic path is ring hydroxylation to form ethyl [2-[*p*-(*p*-hydroxyphenoxy) phenoxy]ethyl]carbamate *Plants*. Rapidly degraded in plants *Soil*. Low mobility in soil, no bioaccumulation. Relatively fast degradation.: DT50 1.72.5 months (lab.), few to 31 days (field)
Hydroprene (Juvenile hormone mimic)			Prevents metamorphosis to viable adults when applied to larval stage	*Plants*. Degradation principally involves ester hydrolysis, *O*-demethylation, and oxidative splitting of the double bond *Soil*. Rapidly decomposed, DT50 is only a few days
Buprofezin (Cyclic urea)		Probable Chitin synthesis and prostaglandin inhibitor. Hormone disturbing effect, leading to suppression of ecdysis	Contact and stomach action, not translocated in the plant. Inhibits moulting of nymphs and larvae, leading to death. Also suppresses oviposition by adults, treated insects lay sterile eggs	*Animals*. Low residues were found in nearly all ruminant and poultry tissues. Extensive metabolism was observed, with a large number of minor metabolites being produced *Plants*. Limited metabolism in most plant species; minor metabolites indicate a pathway involving hydroxylation or oxidative loss of the *tert*-butyl group, followed by opening of the heterocyclic ring *Soil*. DT50 (25°C) 104 days (flooded conditions, silty clay loam, o.c. 3.8%, pH ≥ 6.4), 80 days (upland conditions, sandy loam, o.c. 2.4%, pH 7.0)

Pheromones



Strobilurins (Fungicide)



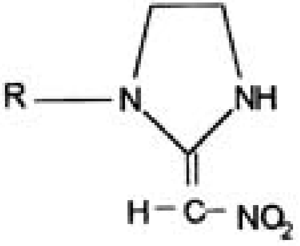
Strobilurin A (naturally occuring fungicide, unstable in vivo due to photoinstability)

Photostable (mimic, synthetic)

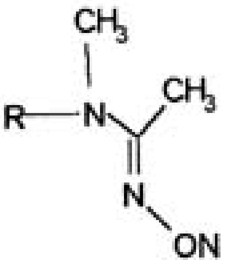
Kresoxim methyl (derivative)

Azoxystrobin (derivative)

Neonicotinoids (Sucking pest control insecticides)



(I)



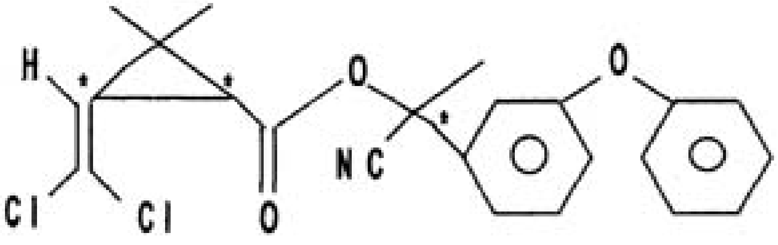
(II)

LC90 (Against Rice Plant hopper)
R =
1. 4-chlobenzyl40 ppm
2. 3-pyridinobenzyl08 ppmimprovedspectrum of activity
3. 6-chloropyridino-3-benzyl0.32 ppm (I)Imidacloprid
4. 6-chloropyridino-3-benzyl-(II)Acetamiprid

a……d having isosteric relationships. Distance between "N" in pyridine to acidic proton in (I) crucial to activity.

30.5 Role of Chirality

Whenever a chemical structure has one or more asymmetric centers or double bonds, either diastereoisomerism or geometric isomerism is possible. Usually only one of the specific stereoisomers is responsible for all or most of the biological activity. An example is cypermethrin,

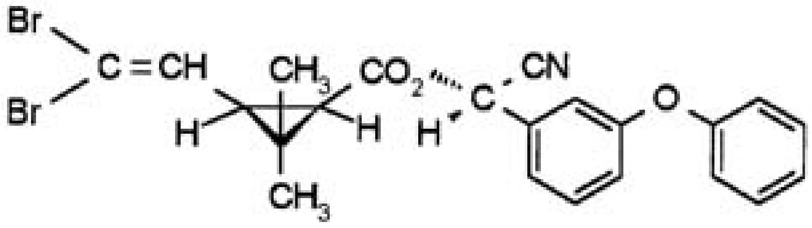
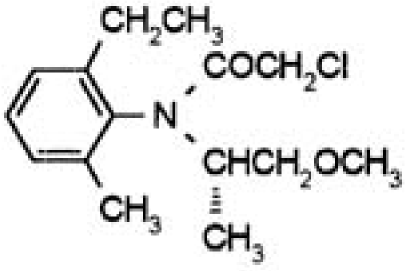
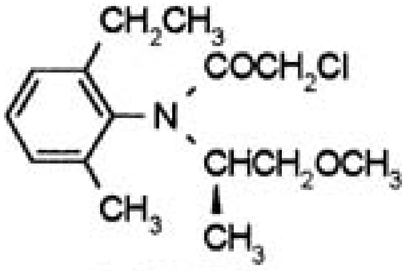


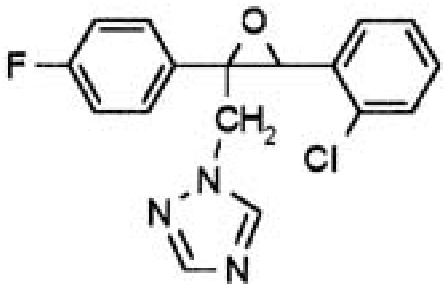
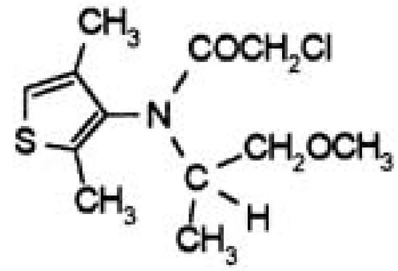
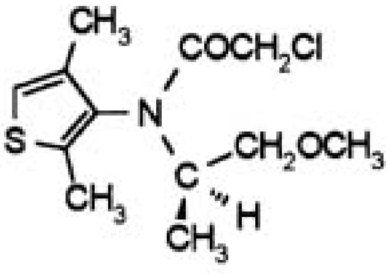
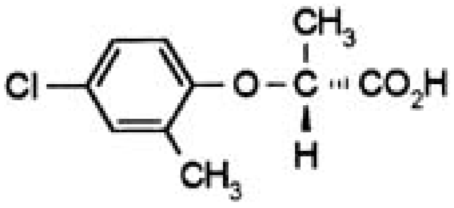
Cypermethrin

cypermethrin, which possesses three asymmetric centers and hence eight possible isomers. Of eight possible isomers, the 1R-*cis*-S isomer is the most biologically active isomer, and the bromo analogue is deltamethrin, commercialized in 1984. In the search for low dose agrochemicals, chiral synthesis plays an important role. For example, one methodology, as used for deltamethrin, is the separation of *cis* and *trans* acids, resolution of the *cis* acid using a chiral amine to prepare the 1R-*cis*-acid. Condensation of the 1R-*cis*-acid chloride with *in situ* generated cyanohydrin, derived from metaphenoxybenzaldehyde, offers a mixture of "1R *cis* ∝ S" and "1R *cis* ∝ R" deltamethrin. The crystallization of the mixture under conditions of epimerization allows isolation of high purity 1R *cis* ∝ S deltamethrin. Today, stereoisomer separations are also relatively facile using chiral phases in preparing HPLC.

Table 30.8 lists examples of chiral agrochemicals.

TABLE 30.8 Chiral Chemistry

Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Environmental Fate
Deltamethrin (Pyrethroid)		Prevents sodium channels from functioning so that no transmission of nerve impulses can take place	Non-systemic insecticide with contact and stomach action	*Animals*. Eliminates within 24 days. The phenyl ring is hydroxylated, the ester bond hydrolyzed, and the acid moiety is eliminated as the glucoside and glycine conjugates *Soil*. Undergoes microbial degradation within 12 weeks DT50 in field <23 days. Soil photolysis DT50 9 days. No risk of leaching
S-Metolachlor (Chloroacetamide)	 (a*RS*,1*S*)-  (a*RS*,1*R*)-	Cell division inhibitor	Selective herbicide, absorbed predominantly by the hypocotyls and shoots; inhibits germination	*Animals*. Rapidly oxidized by rat liver. Oxygenases via dechlorination, *O*-demethylation and side-chain oxidation, conjugation by glyutathione *S*-transferases *Plants*. Metabolism involves dechlorination and conjugation to glutathione-*S*-transferases, followed by further degradation to polar, water soluble, non-volatile metabolites *Soil*. Major aerobic metabolites are derivatives of oxalic and sulfonic acids; DT50 (field): 1130 days. DT90 (field): 3690 days

Epoxiconazole (Azole)		Inhibits C14 demethylase in sterol biosynthesis	Preventive and curative fungicide	*Animals*. Readily excreted via faeces. Metabolic reactions are cleavage of the oxirane ring, hydroxylation of the phenyl rings and conjugation. No major metabolites; high number of minor metabolites *Soil*. Degradation by microbial activity. DT50 a. 23 months
Dimethenamid (Chloroacetamide)	 **dimethenamid**  **dimethenamid-P**	Cell division inhibitor	Herbicide absorbed by Coleoptile	*Animals*. Metabolites in rat, goat and hen include glutathione, cysteine and thioglycolic acid *Plants*. Metabolism in maize leads to thiolactic acid besides above *Soil*. Rapidly degraded in soil, probably through microbial action, with DT50: 843 days, depending upon soil type and weather conditions. Photolysis DT50 on soil a. 7.8 days
Mecoprop-*P* (Aryloxyalkanoic acid)		Systemic auxin, acting like indolylacetic acid	Selective herbicide, absorbed by leaves, with translocation to the roots	*Animals*. In mammals predominantly eliminated as conjugates in urine *Plants*. Hydroxylated at the methyl group with formation of 2-hydroxy methyl-4-chlorophenoxy propionic acid *Soil*. Degraded predominantly by micro-organisms to 4-chloro-2-methyl phenol, followed by ring hydroxylation at the 6-position and ring opening. DT50 (aerobic) : 313 days

Basis of Chemistry Used in Synthesis of Agrochemicals

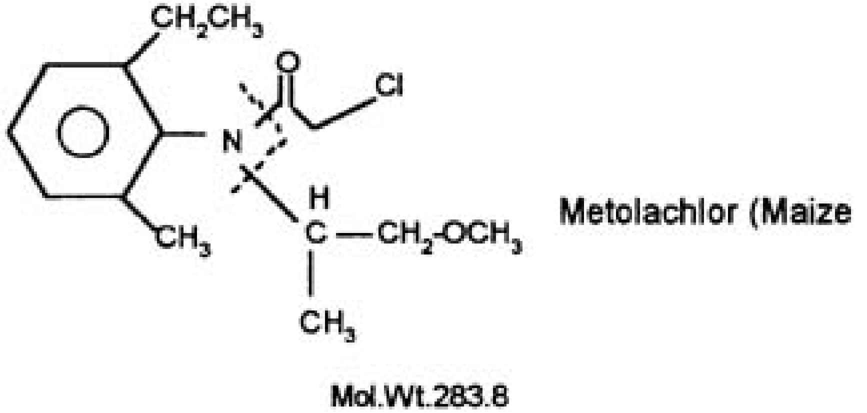
Agrochemicals belong to a variety of chemical classes. Each chemical class demands certain key building blocks. Some examples are: cyanuric chloride (triazines), iminodiacetic acid (glyphospate), 2,6-disubstituted anilines (chloracetanilides), dialkoxythiophosphoryl chlorides (organophosphorous group), metaphenoxybenzaldehyde/alcohol (pyrethroid), isocyanatesalkyl/aryl (carbamates / ureas), sulfonyliso-cyanates (sulphonyl ureas), dialkyl/dihalovinyl-dimethyl-cyclopropane carboxylic acid chloride (pyrethyroids), dihalophenols (phenoxyherbicides), chlorocresols (dichlorophen, etc.), ethylenediamine (dithiocarbamates).

During the synthesis of specific building blocks (key intermediates, penultimate intermediates, or target-active ingredients) a whole range of chemical reactions are deployed. Examples obviously include aromatic electrophilic and nucleophilic substitution, Diels Alder cycloadditions, telemerizations, "N," or "0" alkylations, and so on. New and safer reagents are often employed, for example diphosgene/triphosgene/diarylcarbamates, in lieu of phosphene in the manufacture of carbamates and sulfonyl ureas. Toxic reagents are generated for immediate chemical reaction (consumption) avoiding hazards of storage through leakage. As an example, the standard production of alkyl isocyanates for production of carbonatesfrom alkalimetal cyanates and alkyl halides or sulfatescan be replaced with the in-situ generation of arylisocyanates from arylamines and urea to produce target aryl alkyl ureas.

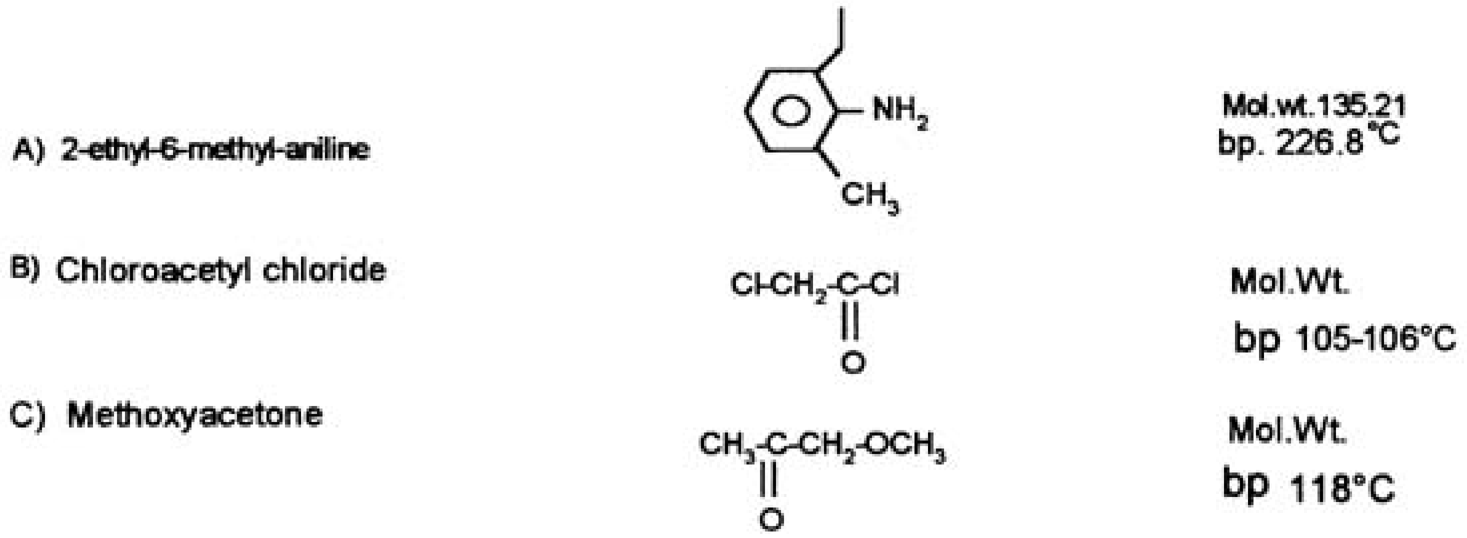
Case StudyChemistry and Manufacture of Metolochor

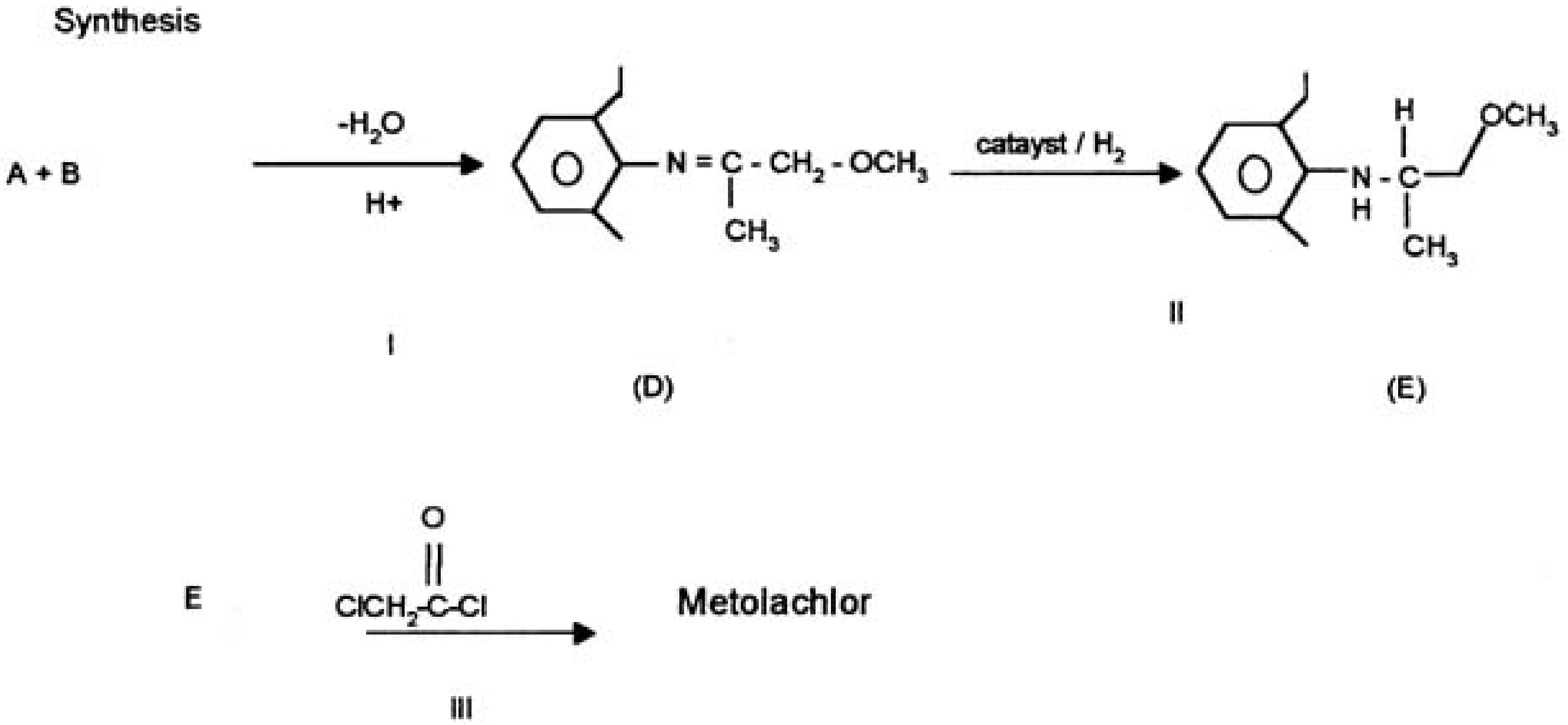
A typical example of agrochemical manufacture is provided by the maize herbicide (Table 30.2b), *metolachlor*.

The chemistry of metolochlor synthesis is shown below, followed by a description of the manufacturing process.



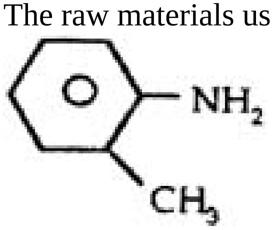
Building blocks for synthesis of Metolachlor are:



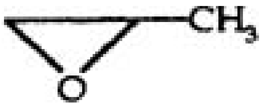


The nature of the catalyst employed during the hydrogenation of (D) offers (E) as a racemate of 1S and 1R isomers, or enriched in 1S isomer which leads into *racemic (R/S) metolachlor* or *S-metolachlor*. A chiral catalyst allows the 1S isomer enrichment to be favored.

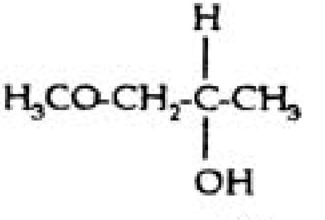
The raw materials used in the commercial synthesis of metolachlor include 2-ethyl 6-methyl-aniline, which can be built up from



ortho-toluidine and ethylene (ex ethanol if required) at high temperature and pressure over metallic aluminum. Methyoxyacetone is built up from



propylene oxide and its reaction with sodium methoxide in methanol, which favors opening from the least hindered side and results in



$$H_3CO\text{-}CH_2\text{-}\overset{H}{\underset{OH}{C}}\text{-}CH_3$$

(F)

the methylether of 2,3-propanediol. Due operational care is necessary in handling propylene oxide as it is extremely flammable and poisonous.

The methyl ether of 2,3-propanediol can be oxidized to methoxyacetone in the vapor or liquid phase using a suitable catalyst or oxidant, respectively. To achieve production of high purity metolachlor, which is a liquid (mp. 62.1°), it is necessary to control the quality of all input materials, monitor reaction progress, terminate each reaction step effectively, and use solvents which can be effectively eliminated to very low levels without volatilization losses of metolachlor (100°C/0.001 mmHg).

Reaction step I is conveniently carried out in a stainless steel reactor using a hydrocarbon solvent to expel water of the reaction as an azeotrope. Completion of the reaction can be determined by measuring the water of reaction or through disappearance of both the 2-ethyl-6-methyl aniline and methoxyacetone by gas chromatography (GC) or high performance liquid chromatography (HPLC).

In the reaction step II, the energetics of saturation of an olefinic Schiff base linkage are favorable and quantitative over most hydrogenation catalysts. 1S isomer enrichment can be achieved through use of a chiral catalyst and the enrichment may be monitored through chiral HPLC. A stainless steel reactor is considered adequate for this operation.

Reaction step III is "N-acylation." It is conveniently carried out in a refluxing chlorinated aliphatic hydrocarbon; the hydrogen chloride gas which is evolved should be scrubbed in aqueous caustic soda. Precautions are necessary when handling the highly corrosive chloroacetyl chloride. This step should be carried out in a glass lined reactor.

Effluents from the reaction are the water of reaction in step I, and hydrogen chloride gas in step III.

Usage of Agrochemicals

Agro climatic conditions, soil type, irrigation, as opposed to rain as the source of water, and other factors all govern the nature of crop cultivation and also the nature of pest attack. Soil preparation, seed dressing, pre-sowing, pre-plant incorporation, pre-emergence, early-post-, or late post-emergence, right through harvest are all various stages during which agrochemicals find use. Protection of farm produce or seeds from pest attack is dependant on the use of agrochemicals. Ectoparasite control of livestock and poultry is another area of agrochemical usage. Disease vector control from flies, mosquitoes, and cockroaches on one hand and protection of dwellings from termites and other wood borers on the other are both significant consumers of agrochemicals.

Different insecticides are used to control insects during the different stages of insect life cycle. The habitat of the insects, whether dwelling above or below the soil surface, and the different feeding habits of individual groups of insects dictate the use of specific insecticides. Use of herbicides depends upon the weed spectrum as per crop and climatic conditions. The competing growth stages of the weeds, as compared to the crop, as well as nature of weeds (e.g., grasses, broad leaf weeds, or sedges) and the efficacy characteristics of individual herbicides need to be synchronized for effective control. Biannual versus annual, or perennial weeds need be treated with different specific herbicides.

The various fungicides have either specific or broad spectrum efficacy. These can be deployed for either prophylactic or curative action. Beyond the farm crops and horticulture, maintenance of turf and lawns free of insects and weeds, and the preservation of natural and agricultural forestry are significant markets for insecticides and herbicides.

All agrochemicals must be used at well-researched recommended dosages. Sub-lethal dosages as well as single agrochemicals used repeatedly leads to resistance development. It is therefore important to use agrochemicals with an independent mode of action in rotation or in well-conceived mixtures to decrease resistance development. Different formulations of a given agrochemical are deployed when the intention is for soil application for absorption by root system or by foliar uptake. Different sprayers with specific nozzles allow better efficacy. Aerial spraying demands yet different formulations.

Differential selective or broad spectrum agents include nematacicides (Table 30.9), acaracides, ovicides, molluscicides (Table 30.10), algicides (Table 30.11), bactericides (Table 30.12), rodenticides (Table 30.13), which control damage from nematodes, mites, eggs, snails/slugs, algae, bacteria, and rodents, respectively.

TABLE 30.9 Nematicides

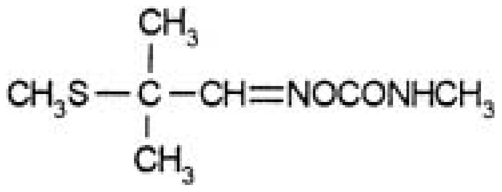
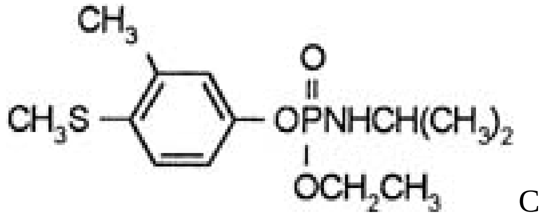
Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Environmental Fate
Aldicarb (Oxime Carbamate)		Cholinesterase inhibitor. Metabolically activated to Aldicarb sulfoxide	Systemic nematicide with contact and stomach action	*Animals*. Absorbed rapidly and completely; >80% is excreted in the urine within 24 hr, >96% within 34 days. Aldicarb is oxidized to the sulfoxide and sulfone, which undergo further metabolism *Plants*. The sulfur atom is oxidised to sulfoxide and sulfone groups. Further degradation leads to the formation of oximes, nitriles, amides, acids, and alcohols which are present in the plant only in conjugated form *Soil*. Sulfur atom is oxidised to sulfoxide and sulfone groups. Further degradation leads to the formation of oximes, nitriles, amides, acids, and alcohols
Fenamiphos (Organophosphorous)		Cholinesterase inhibitor	Systemic nematicide with contact action. Absorbed by roots with translocation to the leaves	*Animals*. In mammals, following oral administration, there is rapid metabolism involving oxidation to the sulfoxide and sulfone analogues, followed by subsequent hydrolysis, conjugation and execretion via the urine, some *N*-dealkylation also occurs *Plants*. Degradation is by thiooxidation and hydrolysis. The major metabolites are fenamiphos sulphoxide and fenamiphos sulfone *Soil*. Degradable on soil surfaces. Duration of activity in soil is c. 4 months. Compound with low mobility. Soil DT50 (aerobic and anaerobic) several weeks, The major degradation products are fenamiphos sulfoxide and fenamiphos sulfone and their phenols
1,3-dichloropropene (Chloroalkenes)			Soil fumigant nematicide.	

TABLE 30.10 Molluscicides

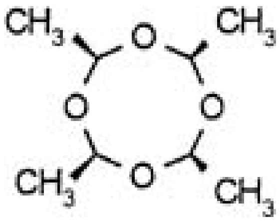
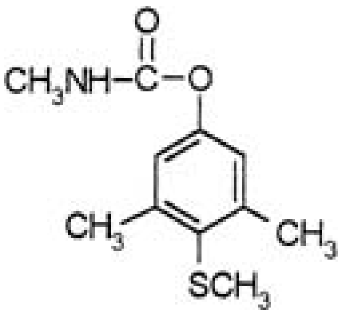
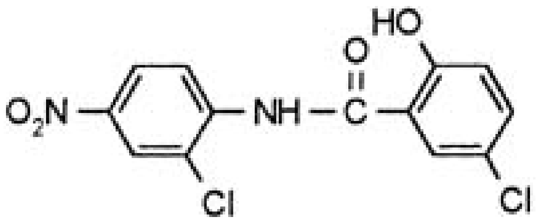
Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Environmental Fate
Metaldehyde			Molluscicide with contact and stomach action. Poisoned slugs secrete large quantities of slime, desiccate and die. Their mucus cells are irreversibly destroyed	*Soil*. Aerobic and anaerobic microorganisms in soil decompose metaldehyde to CO2 and water
Methiocarb (Carbamate)			Molluscicide with neurotoxic toxic effect	*Animals*. Rapidly absorbed and excreted, principally in the urine, with only a small proportion in the faeces. Metabolism involves hydrolysis, oxidation, and hydroxylation, followed by excretion in free or conjugated form. There is a continuous decrease of activity in all organs. *Plants*. Methylthio group is oxidised to sulfoxide and sulfone, with hydrolysis to the corresponding thiophenol, methylsulfoxide-phenol, and methylsulfonyl-phenol *Soil*. Degradation is rapid. The importance metabolites are methylsulfinylphenol and methylsulfonylpehnol
Niclosamide (Anilide)			Molluscicide with respiratory and stomach	*Animals*. The major metabolite in the urine was the reduced compound 2′,5-dichloro-4′-aminosalicylanilide. Several labile conjugates were also detected. The major constituent in the faeces was unchanged niclosamide, although considerable amounts of 2′,5-dichloro-4′-aminosalicylanilide were also present *Soil*. Degradation followed pseudo-first order kinetics, DT500.3 days

TABLE 30.11 Algicides

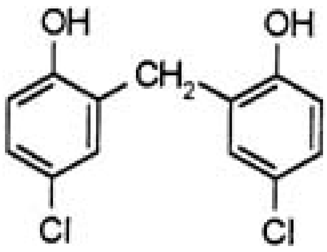
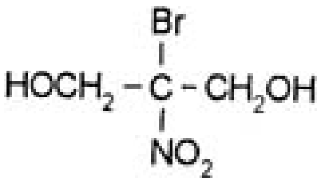
Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Environmental Fate
Nabam (Multi-site alkylenebis (dithiocarbamate)	NaSCSNHCH2CH2NHCSSNa	Non-specific thiol reactant, inhibiting respiration		*Plants*. The principal metabolite is ethylenethiourea. Other metabolites include ethylenethiuram monosulfide, ethylenethiuram disulfide, and sulfur
Dichlorophen (Chlorophenol)			Contact action	

TABLE 30.12 Bactericides

Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Environmental Fate
Bronopol		Oxidation of mercapto group of bacterial enzymes. Inhibition of dehydrogenase activity leads to irreversible membrane damage		*Animals*. Rapidly absorbed and rapidly excreted, mainly in the urine. The major metabolite is 2-nitroprophane-1,3-diol *Plants*. Biochemical degradation leads to the metabolite 2-nitropropan-1,3-diol

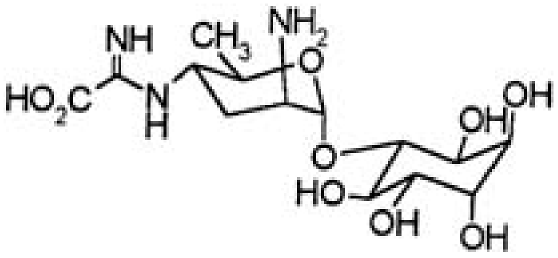
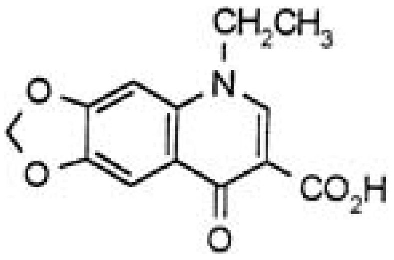
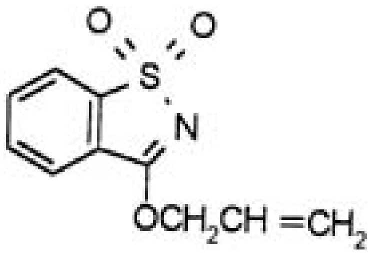
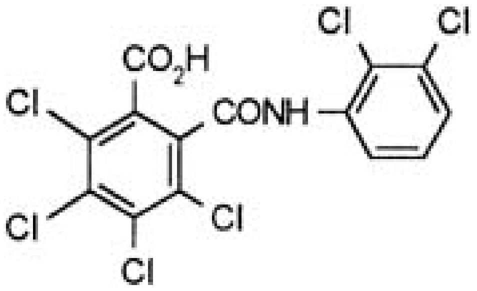
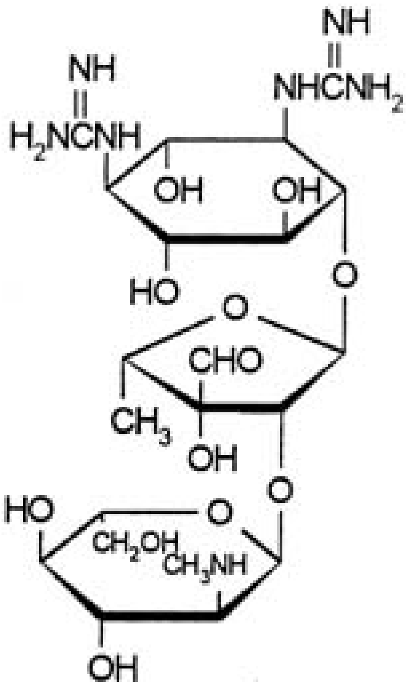
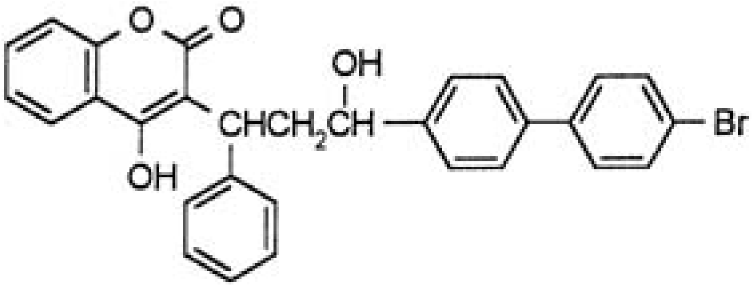
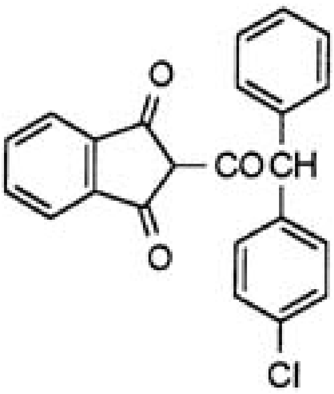
Kasugamycin (Antibiotic)		Protein synthesis inhibitor, inhibits binding of Met-RNA to the mRNA-3OS complex thereby preventing amino acid incorporation	Systemic bactericide	*Animals*. Mostly excreted in the urine within 24 hr. After oral administration to rats at 200 mg/kg, 96% of administered dose remained in the digestive tract 1 hr after administration *Plants/soil*. Degraded to kasugamycinic acid and kasuganobiosamine; finally degraded to ammonia, oxalic acid, CO2 and water
Oxolinic acid (Pyridone)			Systemic bactericide	
Probenazole			Systemic bactericide, absorbed by roots and translocated acropetally	*Soil*. Half-life in soil <24 hr (alluvial or volcanic soil)
Tecloftalam (Benzoic acid)			Bacteriostat	*Soil/Environment*. DT50 410 days undergoing loss of chlorine from the benzoic acid ring
Streptomycin (antibiotic)		Inhibits protein synthesis by binding to the 3OS ribosomal sub-unit, causing a mis-reading of the genetic code	Bactericide with systemic action	

TABLE 30.13 Rodenticides

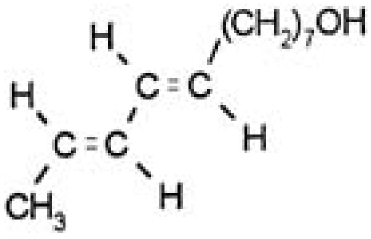
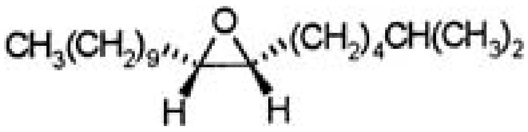
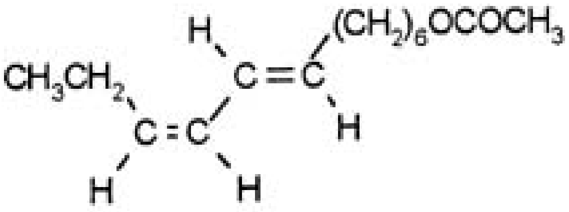
Chemical (Chemical Class) Structure	Biochemistry	Mode of Action	Environmental Fate

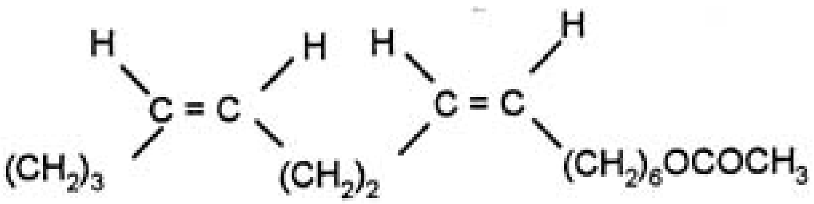
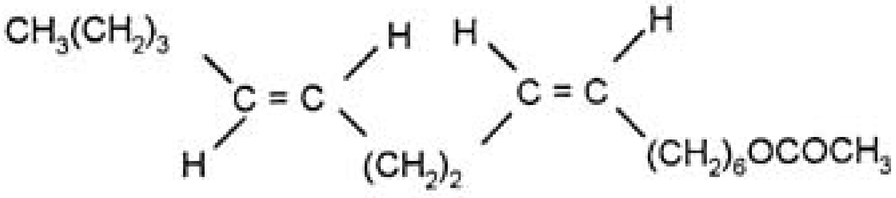
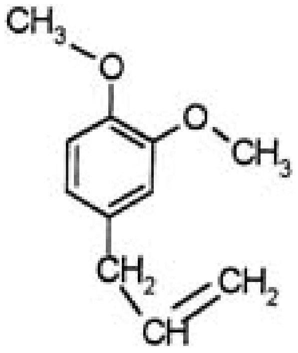
Bromadiolone (Coumarin anticoagulant		Second generation anticoagulant rodenticide which also blocks prothrombin formation		*Soil*. Leaching behaviour is inversely related to clay and organic matter content of soils. In soil column and soil layer studies, 97% was remaining in the top soil, with 0.1% in leachate
Phosphine[a]	PH_3		Respiratory, metabolic and nerve poison. Evolves non-flammable mixture of phosphine, ammonia and carbon dioxide	*Animals*. Phosphine is probably metabolised to non-toxic phosphates *Plants*. In stored products, phosphine undergoes oxidation to phosphoric acid
Chlorophacinone (Indandione anticoagulant)		Blocks prothrombin formation and uncouples oxidative phosphorylation		*Animals*. 90% is eliminated in the faces within 48 hr in the form of metabolites

[a] Includes Aluminium phosphide, Zinc phosphide, Magnesium phosphide.

Insecticides control insects with differential feeding habits, which have a reasonably predictable pattern per crop and agro climatic condition. Systemic insecticides (Table 30.3a) find use to control insects which feed by sucking plant juices. Insecticides with contact and stomach action control insects feeding (Table 30.3b) on foliage, stems, and fruits. Insect growth regulators (Table 30.7) find a complementary role to make insects more prone to control. Plant growth regulators (Table 30.6) direct nutrition to fruit formation by controlling vegetative growth. Herbicides (Tables 30.4ad) are used for farm preparation, pre-plant incorporation, pre-early post, and post-emergence application to control competing weeds. Fungicides (Tables 30.5a, b) are used for seed dressing (control from soil fungi), and for the prophylactic and curative protection from fungal attack. Economic threshold levels, use of pheromone traps (Table 30.14) to measure insect population and the like are tools to use agrochemicals judiciously. Agrochemicals, when used rotationally or in mixtures of two with differential modes of action (e.g., mixture of an insecticide with insect growth regulator, mixture of herbicides or fungicides with different modes of action) help tackle the problem of resistance development.

TABLE 30.14 Insect Pheromones

Chemical (Chemical Class)	Structure	Biochemistry	Mode of Action	Environmental Fate
Codlemone (Pheromone)			Acts by disruption of mating, either in trapping or in disorientation mode	*Plants*. Rapidly oxidised on the surface of leaves *Soil*. Rapidly degraded in soil to CO_2
Disparlure (Pheromone)			Acts both as an attractant and as a mating disruptant	
Dodeca-7,9-dienyl Acetate (Pheromone)	 (7E,9Z)- isomer		Acts as an attractant and by disruption of mating in the disorientation mode	

Gossyplure (Pheromone)	 (Z,Z) -  (Z,E) -	Acts as an attractant and by disruption of mating
Methyl eugenol		Synthetic attractant

30.6 Biotechnology for Farm Productivity

Biotechnology using recombinant DNA has impacted farm productivity in some areas and may become more significant over time. A full discussion of this interesting subject is beyond the present scope of this chapter, but mention of some aspects may be appropriate.

The determination of DNA sequences of genes for resistance to pest and pathogens, cloned from a number of crop species, suggest the existence of certain domains in the protein products. These are shared among plant species and are shared among genes, which can provide resistance to viruses, fungi, bacteria, nematodes, and insects. It may be possible to design durable resistance genes that have a broad spectrum of activity. Modification of genes controlling key steps in signal-transduction pathways and result in the generation of defense responses may influence the speed and type of plant response. This may then lead to pest and disease resistance. Herbicide tolerance is already widely used in the United States. Genetic modification of secondary plant metabolism through the synthesis of novel toxins or increased levels of existing toxins and through synthesis of molecules which regulate pest or pathogens colonization, development or morphogenesis may enhance crop resistance to pests and pathogens. Knowledge of plant signaling processes and the role of secondary plant metabolites in regulating insect behaviour offer opportunities to manipulate relevant biosynthetic pathways. This, in turn, may increase the ability of plants to deter pest colonization, inhibit pest development and reproduction, and attract natural enemies of its pests through semiochemical production.

Genes of microbial, animal or plant origin, coding for toxins, inhibitors of pathogenicity factors, such as the lecithins, or degradative enzymes, such as chitinases and proteases, may enhance resistance to pests or pathogens when expressed in transgenic plants (e.g., genetically modified (GM) crops expressing *B. thuringiensis* (Bt) toxins) (Table 30.15). However, the effect of GM crops on predator and parasites, as well as on pollinators, need to be monitored. Also, development of resistance to GM plants by target insects requires careful monitoring. Genetically modified predators and parasites for enhanced adaptability or baculoviruses with modified genes for toxin production for enhanced efficacy of viruses as biological control agents are currently being investigated. Virus resistant squash and papaya are already available in U.S. markets. Good levels of antibody expression in plants have not yet been achieved due to technical problems. Discovery of new essential processes of pests and pathogens and their genetic make-up/gene function could lead to novel chemistries having such processes as target systems. Combination of genomics and biochemical screening procedures may lead to a range of new chemical control methods for agrochemical industry while increasing the screening rates of the molecules. Genetic processes are also used to understand pesticide resistance within pest populations, leading to high-resolution diagnostics for resistant alleles, especially for pests with multiple resistance mechanisms. Molecular techniques, used in ecological research to elucidate the structure of populations and to estimate gene flow between populations occupying discrete habitat patches, help in the formation of ecological modification strategies for maintaining biodiversity as farmland promoting sustainable pest management. The risk assessment of the release of GM organisms in the environment may be aided by assessing the gene flow among populations in agricultural systems.

TABLE 30.15 Biological Agents

Biological Agents	Usage
Amblyseius spp. (a number of predatory mite species)	Control of thrips
Anagrus otomus	A wasp, parasite of leafhoppers
Aphelinus abdominatis	A wasp, parasite of aphids used in protected crops
Aphidius colemani	A wasp, parasite of aphids used in protected crops
Aphidoletes aphidimyza	Predatory midge, consumer of aphids in protected crops
Chrysoperla carnea	Entomophagus lacewing larva
Cryptolaemus montrouzieri	Beetle, consumer of mealybugs used in orchards, vines, andprotected crops
Dacnusa sibirica	A wasp, parasite of leaf miners used in protected crops
Diglyphus isaea	A wasp, parasite of leaf miners
Encarsia formosa	A wasp, parasite of glasshouse whiteflymost widely used
Hippodamia convergens	Ladybirdconsumer of aphids and other pests
Leptomastix dactylopii	A wasp, parasite of mealybugs used in horticultural and fruit crops
Metaphycus holvodus	A wasp, parasite of soft scales used in orchards and in protected crops
Orius spp.	Predatory bug. A number of species are used for control of thrips
Phytoseiulus persimillis	Spider mite consuming mite, used in protected crops
Trichogramma spp.	A number of species of wasp, parasitic of Lepidotera in protected crops
Bacillus sphaericus	Used against mosquito larvae
Beauveria bassiama	Entomopathogenic fungused under development for use on control of a wide range of coleopteran, homopteran, and heteropteran pests
Helicoverpa zea NPV	Nuclear polyhedrosis virus used for control of heliothis and Helicovenpa in cotton and tobacco
Heterorhabditis bacteriophora and*H. megidis*	Insect parasitic nematodes used for control of Japanese beetles, black vine weevils, etc.
Mamestra brassicae NPV	NPV used for control of lepidoptera
Metarhizium anisopliae	Entomopathogenic fungus under development for control of locusts
Spodoptera exigua NPV	For control of beef armyworm in various crops
Steinernema spp.	Parasitic nematode which searches for enters and kills target pests
Ampelomyces quisqualis	Hyperparasite of the Erysiphaceae genus, which causes powdery mildew diseases
Bacillus subtilis	Seed treatment for control of *Rizoctonia solani*, *Fusarium* spp., *Alternaria* spp., *Aspergillus* spp., etc. in cotton, legumes of other crops
Candida oleophila	Selective fungal antagonist for control of post harvest diease in citrus and pome fruit
Eiliocladium spp.	Used for the control of foliar pathogens in seedlings and as a post-harvest treatment. *G. vorens* is used for the control of soil diseases
Streptomyces griseoviridis	Bacterium used for control of *Fasaurium* and other pathogens
Trichoderma spp.	Mitosporic fungi used for control of range of soil/foliar pathogens
Microlarinus lareynii	Attacks seedheads of puncture vine
Urophora sirunasova	Gallfly attacks seedhead of yellow star thistle
Bangastemus orientalis	Weevilattacks seedhead of yellow star thistle
Coleophora parthenica	Attacks stem of Russian thistle

Regulatory Scene Regarding Products of Biotechnology

The health effects of eating genetically modified foodstuffs, especially the allergencity, are being investigated by the FAO/WHO Codex Alimentarius Commission, and its relevant subsidiary bodies have reflected the result of these studies. An Intergovernmental Task Force of Foods Derived from Biotechnology is being established to develop standards, guidelines, and recommendations as deemed to be appropriate for foods developed through biotechnology.

In the United States, the U.S. Department of Agriculture (USDA), EPA and FDA are involved in the regulation of products of biotechnology. State and Federal statutes and standards must be satisfied prior to registration. At present, no requirements are in place for the varietal registration of new crops. The following table gives a picture of the present regulatory review process.

New Trait/Organism	Regulatory Review Conducted by	Reviewed for
Viral resistance in food crop	USDA	Safe to grow
	EPA	Safe for the environment
	FDA	Safe to eat
Herbicide tolerance in food crop	USDA	Safe to grow
	EPA	New use of companion herbicide
	FDA	Safe to eat

Herbicide tolerance in ornamental crop	USDA	Safe to grow
	EPA	New use of companion herbicide
Modified oil content in food crop	USDA	Safe to grow
	FDA	Safe to eat
Modified flower colour ornamental crop	USDA	Safe to grow
Modified soil bacteria	EPA	Safe for environment
Degrades pollutants		

< previous page

next page >

30.7 Future Directions

Allelopathy

Herbicides comprise 6070 percent of total usage of pesticides in developed countries. Plants develop resistance to herbicides which in turn demands altered management strategies through improved herbicides and/or herbicide mixtures. This may involve different modes of action, for short-term efficacy and long-term control to be effective.

Allelopathy is the action of secondary metabolites in plants, algae, bacteria or fungi, which influences the growth and development of other species. This may help in overcoming such problems of resistance through development of crop varieties having greater ability to smother weeds, use of natural phytotoxins from plants or microbes as herbicides, and use of synthetic derivatives of natural products as herbicides. Another way allelopathy may be used in agriculture is through isolation, identification, and synthesis of the active compounds from an allelopathic plant or a microorganisms species. Knowledge of the structural features of such secondary metabolites will act as important leads in the development features of such secondary metabolites will act as important leads to develop future herbicides to tackle resistance phenomena.

Biorational Approach to Chemical Synthesis

Knowledge of pest biochemistry will open multiple target sites to which agrochemicals of appropriate structure could be directed. This is to say that new modes of action will emerge and may overcome some of the presently unmet needs. These include such items as:

1. need for insecticides with *multiple modes of action*
2. need for *higher selectivity* herbicides to control weeds that resemble the main crop
3. fungicides and insecticides of *high safety* to preserve agricultural produce
4. control agents for plant intake to *minimize damage* from soil fungi, soil insects, and nematodes
5. *safer and more effective* rodenticides and bird repellants
6. antivirals and antibacterials

30.8 References

Future Research, Development and Technology Transfer Needs for UK Crop Protection, A report prepared by the British Crop Protection Council, 1997, August.

Evaluation of Allergenicity of Genetically Modified Foods. Report of a joint FAO/WHO expert consultation on Allergenicity of Foods derived from Biotechnology, 2001, January 2225.

Biotechnology in Crop Protection: A BCPC Appraisal of Progress and Prospects. A report prepared by British Crop Protection Council, 1999, November.

Acreage, Released by National Agricultural Statistics service (NASS), Agricultural Statistics Board, US Department of Agriculture, 2001, June.

Regulatory information downloaded from various US, EEC and Australian sites.

Pesticide Manual, 12th ed., CDs Tomlin (Ed.), 2001.

Synthesis and Chemistry of Agrochemicals IV, ACS Symposium Series 584, D. R. Baker, J. G. Fenyes, and G. S. Basarab (Eds.), 1995.

KirkOthmer Encyclopedia of Chemical Technology, 4th ed., J. I. Kroschwitz and M. Howe-Grant (Eds.), 1992.

31.1 Introduction

The average citizen in today's world gives little thought to the important role that commercial explosives play in our lives and how their use is linked to our standard of living and our way of life. Explosives provide the energy required to give us access to the vast resources of the earth for the *advancement of civilization*.

To maintain our standard of living in the United States, every day 187,000 tons of concrete are mixed, 35 million paper clips are purchased, 21 million photographs are taken, using large quantities of silver, 80 pounds of gold are used to fill 500,000 cavities, and 3.6 million light bulbs are purchased. It takes more than 40 different minerals to make a telephone, and 35 to make a color television. Even everyday products such as talcum powder, toothpaste, cosmetics, and medicines contain minerals, all of which must be mined using chemical explosives.1,2

Without explosives the steel industry and our entire transportation system would not be possible. The generation of electricity has been largely dependent on coal, and coal mining today is still the largest consumer of industrial explosives. Rock quarrying for road building, and excavations for skyscrapers, tunnels, roads, pipelines, and utilities are direct beneficiaries of the labor saving use of explosives.

31.2 Commercial Explosives Market

The use of commercial explosives in the United States continued to expand during the 1990s, reaching a peak in 1998. Figure 31.1 shows the usage by year as reported by the U.S. Geological Survey (USGS) and the Institute of Makers of Explosives (IME) from 1991 to 2000.2,3 Figure 31.1 also separates the volumes by industry use. The open-pit coal mining industry continued to be the largest user throughout the 1990s. Table 31.1 shows the commercial explosive usage by seven leading states for 1999 and 2000.2 In the year 2000, the five states consuming the most explosives (in decreasing order) were: Kentucky, Virginia, Wyoming, West Virginia, and Pennsylvania, all coal mining states. The coal market is also slowly making a shift from the eastern states to the western states with lower BTU, but more importantly, lower sulfur coal. Through the 1990s the sparsely populated state of Nevada ranked in the top ten states using commercial explosives. This reflected the growth of large volume gold mining operations in North America. Many of the smaller underground gold mines were transformed into large open-pit operations using efficiencies of scale to overcome the overall lower grade of ore, the same transformations that other metal mines had made decades before. Along with this change in mining style came a conversion from small diameter packaged explosives like dynamite to large, bulk explosive loading systems using emulsions and ammonium nitrate/fuel oil (ANFO).

Fig. 31.1 Sales for consumption of U.S. industrial explosives from 1991 to 2000. (*Kramer, Deborah A., United States Geological Survey, Explosives Statistics and Information, Minerals Yearbook 2000, and the Institute of Makers of Explosives*.)

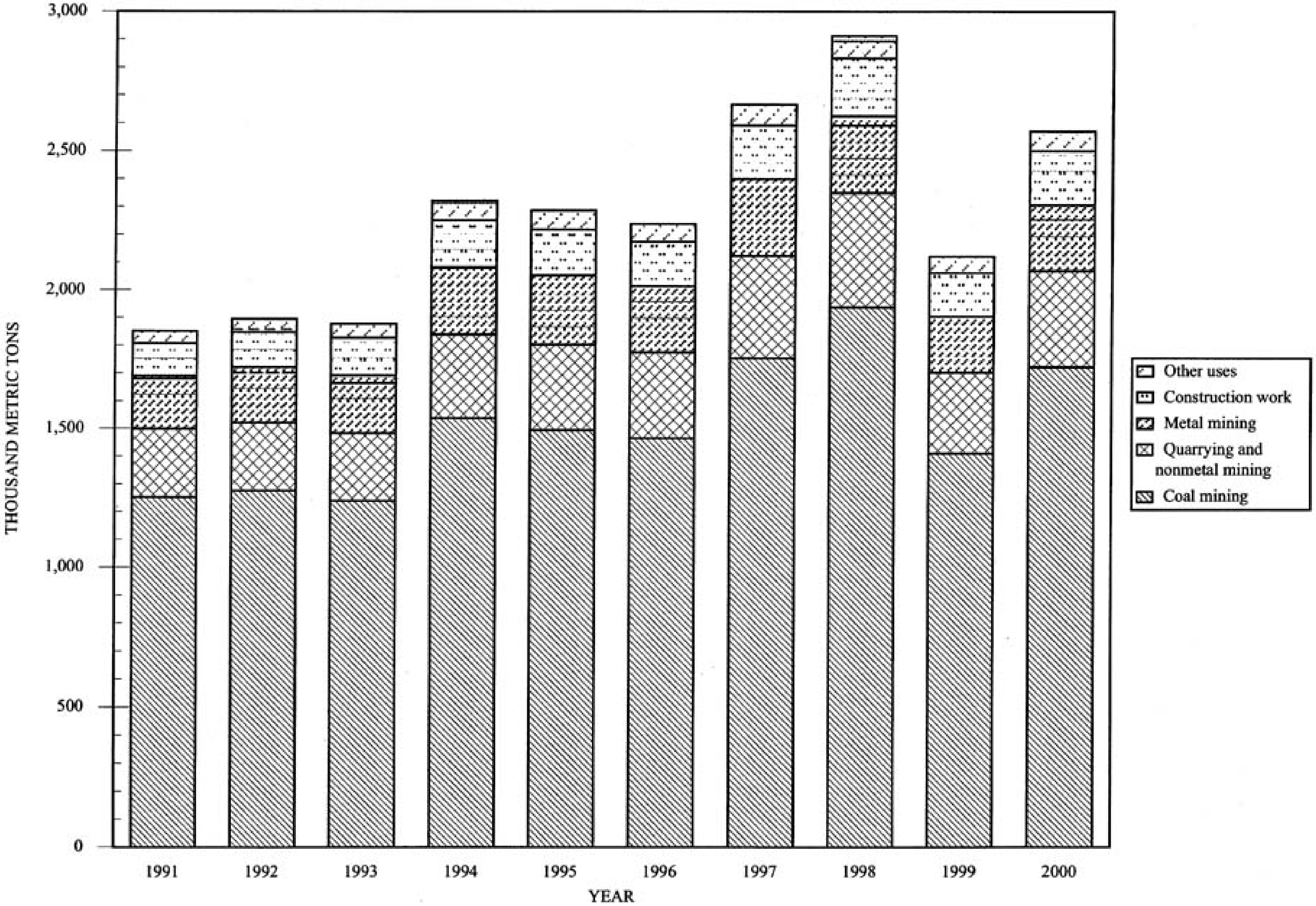


TABLE 31.1 Industrial Explosives and Blasting Agents Sold in the United States by State and Class (Metric Tons)

	Fixed High Explosives							
	1999				2000			
State	Permissibles	Other High Explosives	Blasting Agents and Oxidizers	Total	Permissibles	Other High Explosives	Blasting Agents and Oxidizers	Total
Arizona	0	414	109,000	109,000	20	548	112,000	112,000
Kentucky	975	2,460	408,000	411,000	806	2,720	499,000	502,000
Missouri	2	1,490	45,500	47,000	2	1,280	116,000	117,000
Pennsylvania	98	3,280	110,000	114,000	88	1,980	139,000	141,000
Virginia	250	1,700	194,000	196,000	206	2,690	231,000	234,000
West Virginia	121	609	214,000	214,000	117	1,240	220,000	221,000
Wyoming	4	233	81,400	81,700	44	635	223,000	224,000
Others	350	20,914	928,100	947,300	247	22,807	990,000	1,019,000
Total	1,800	31,100	2,090,000	2,120,000	1,530	33,900	2,530,000	2,570,000

Note: Data rounded; may not add up to total shown.
Source: IME.

31.3 Chemistry of Combustion and Explosion

For a simple understanding of explosives it is helpful to compare an explosive reaction with the more familiar combustion or burning reaction. Three elements are needed to have a fire: fuel, oxygen, and a source of ignition. The process of combustion is basically an oxidationreduction (redox) reaction between the fuel and oxygen from the air. Once initiated, this reaction becomes self-sustaining and produces large volumes of gases and heat. The heat given off further expands the gases and provides the stimulus for the reaction to continue.

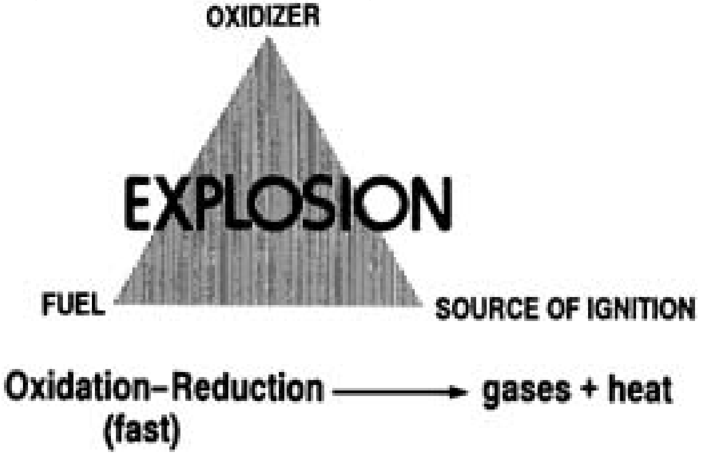
The burning reaction is a relatively slow process, depending upon how finely divided the fuel is, that is, the intimacy of contact between the fuel and the oxygen in the air. Because burning is diffusion-controlled, the more intimately the fuel and oxygen are mixed, the faster they can react. Obviously, the smaller the particles of fuel, the faster the combustion can occur.

Another result of the fineness or particle size of the fuel is the completeness or efficiency of the reaction. In a complete combustion all the fuel elements are oxidized to their highest oxidation state. Thus, wood, being mainly cellulose, and gasoline, being generally a hydrocarbon (e.g., octane), produces primarily carbon dioxide and water vapor upon complete combustion. Once initiated these burning reactions give off heat energy, which sustains the reactions. Heat is released because the oxidized products of the reaction are lower in energy (more stable) than the reactants. The maximum potential energy release can be calculated from the respective heats of formation of the products and reactants. Actual heats of combustion can be measured experimentally by causing the reaction to occur in a bomb calorimeter. The calculated energy values for the above reactions are -3,857 cal/g of cellulose and -10,704 cal/g of octane, respectively.

In the case of an inefficient burn, some less stable or higher-energy products are formed so that the resultant heat energy given off is lower than that for complete combustion. In the above examples inefficient combustion could result for lack of oxygen accessibility, producing carbon monoxide or even carbon particles instead of carbon dioxide. A smoky flame is evidence of unburned carbon particles and results from inefficient combustion where fuel particles are so large or so dense that oxygen cannot diffuse to the burning surface fast enough. If this inefficiency is great enough, insufficient heat is given off to keep the reaction going, and the fire will die out.

All explosive reactions involve similar redox reactions; so the above principles of combustion can help illustrate, in a very basic way, the chemistry involved in explosions. As in a fire, three elements (fuel, oxidizer, ignition source) are needed for an explosion. Figure 31.2 shows an explosion triangle, which is similar to the fire triangle. In general, the products of an explosion are gases and heat although some solid oxidation products may be produced, depending upon the explosive composition. As in normal combustion, the gases produced usually include carbon dioxide and water vapor plus other gases such as nitrogen, again depending upon the composition of the explosive.

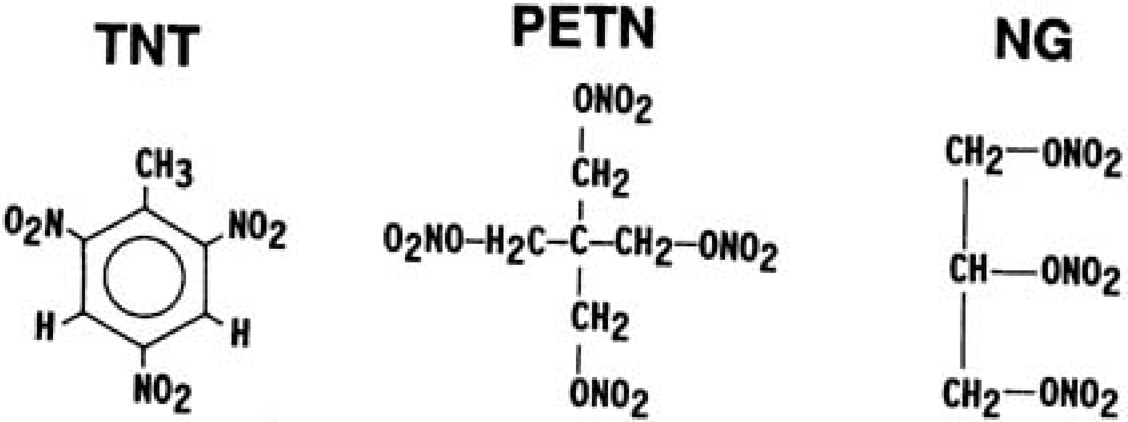
Fig. 31.2 An explosion triangle.



It should be noted that an explosion differs from ordinary combustion in two very significant ways. First, oxygen from the air is not a major reactant in the redox reactions of most explosives. The source of oxygen (or other reducible species) needed for reaction with the fuelthe oxidizermay be part of the same molecule as the fuel or a separate intermixed material. Thus an explosive may be thought of as merely an intimate mixture of oxidizer and fuel. This degree of intimacy contributes to the second significant difference between an explosion and normal combustionthe speed with which the reaction occurs.

Explosives in which the oxidizer and fuel portions are part of the same molecule are called molecular explosives. Classical examples of molecular explosives are 2,4,6- trinitrotoluene (TNT), pentaerythritol tetranitrate (PETN), and nitroglycerin (NG) or, more precisely glycerol trinitrate. The chemical structures of these explosives are shown in Fig. 31.3. As can be seen in the structures, the oxidizer portions of the explosives are the nitro (NO2) groups in TNT and the nitrate (ONO2) groups in PETN and NG. The fuel portions of all three explosives are the carbon and hydrogen (C and H) atoms. Comparison of the ratios of carbon to oxygen in these explosives (i.e., approximately 1:1 for TNT, approximately 1:2 for PETN, and 1:3 for NG) shows that TNT and PETN are deficient in oxygen; that is, there is insufficient oxygen present in the molecule to fully oxidize the carbon and hydrogen. Consequently, products such as carbon monoxide, solid carbon (soot), and hydrogen are produced, as well as carbon dioxide and water vapor. Prediction of the exact products of explosion is complex, especially for the oxygen-deficient explosives, because the ratios of CO2, CO, H2O, and H2 will vary, depending upon reaction conditions (explosive density, degree of confinement of the explosive, etc.).4,5 The following equations show typical ideal reaction products along with calculated heats of reaction for these molecular explosives:

Fig. 31.3 Chemical structure of three molecular explosives.



$$\underset{\text{TNT}}{C_7H_5N_3O_6} \rightarrow 1.5CO_2 + 0.5CO + 2.5H_2O + 1.5N_2 + 5C + 1290\,\text{cal/g}$$

$$\underset{\text{PETN}}{C_5H_8N_4O_{12}} \rightarrow 4CO_2 + 4H_2O + 2N_2 + C + 1510\,\text{cal/g}$$

$$\underset{\text{NG}}{C_3H_5N_3O_9} \rightarrow 3CO_2 + 2.5H_2O + 1.5N_2 + 0.25O_2 + 1480\,\text{cal/g}$$

Explosives in which the oxidizer and fuel portions come from different molecules are called composite explosives because they are a mixture of two or more chemicals. A classic industrial example is a mixture of solid ammonium nitrate (AN) and liquid fuel oil (FO). The common designation for this explosive is the acronym, ANFO. The oil used (typically #2 diesel fuel) is added in sufficient quantity to react with the available oxygen from the nitrate portion of AN. The redox reaction of ANFO is as follows:

$$\underset{\text{AN}}{3\text{NH}_4\text{NO}_3} + \underset{\text{FO}}{—\text{CH}_2—}$$

$$\rightarrow \text{CO}_2 + 7\text{H}_2\text{O} + 3\text{N}_2 + 880\,\text{cal/g}$$

"Oxygen balance" (O.B.) is the term applied to quantify either the excess oxygen in an explosive compound or mixture (beyond what is needed for complete combustion of the fuel elements) or oxygen deficiency (compared to the amount required for complete combustion). It is expressed as either a percentage or a decimal fraction of the molecular weight of the oxygen in excess (+) or deficiency (-) divided by the molecular weight of the explosive or the ingredient being considered. Individual components of an explosive mixture have O.B. values that may be summed for the mixture. Shown below are the O.B. calculations for AN and FO:

$$\underset{\substack{\text{AN}\\ \text{Mol. wt.} = 80}}{\text{NH}_4\text{NO}_3} \rightarrow 2\text{H}_2\text{O} + \text{N}_2 + \underset{\text{Mol. wt.} = 32}{\tfrac{1}{2}\text{O}_2}$$

$$\text{O.B} = \frac{+(1/2)(32)}{80} = +0.20$$

$$\underset{\substack{\text{FO}\\ \text{Mol. wt.} = \sim 14n}}{—(\text{CH}_2)_n—} + \frac{3n}{2}\text{O}_2 \rightarrow nCO_2 + n\text{H}_2\text{O}$$

$$\text{O.B} = \frac{-(3n/2)(32)}{14n} = -3.43$$

From the O.B. values, one can readily determine the ratio of ingredients to give a zero O.B. mixture for optimum efficiency and energy. Thus the weight ratio for ANFO is 94.5 parts of AN and 5.5 parts of FO (94.5 × 0.20 = 5.5 × 3.43).

For the molecular explosives shown previously, the respective oxygen balances are: TNT, -0.74; PETN, -0.10; and NG, +0.04. Thus, NG is nearly perfectly oxygen-balanced; PETN is only slightly negative; but TNT is very negative, meaning significantly deficient in oxygen. Therefore, combinations of TNT and AN have been employed to provide additional oxygen for the excess fuel, as, for example, in the Amatols developed by the British in World War I.6

Modern commercial explosives react in a very rapid and characteristic manner referred to as a detonation. Detonation has been defined as a process in which a shock-induced supersonic combustion wave propagates through a reactive mixture or compound. This high pressure shock wave compresses the reactive material in contact with it resulting in rapid heating of the material, initiation of chemical reaction, and liberation of energy. This energy, in turn, continues to drive the shock wave. Pressure in a detonation shock wave may reach millions of pounds per square inch. The sudden pressure pulse shocks the explosive material as it passes through, causing a nearly instantaneous chemical reaction in the body of the explosive. Once initiated, molecular explosives tend to reach a steady-state reaction with a characteristic detonation velocity. Composite or mixture explosives also have steady-state detonation velocities, but these velocities are more variable than those of molecular explosives and are influenced by such factors as diameter of the charge, temperature, and confinement.

31.4 Historical Development

The first known explosive material was black powder, a mixture of potassium nitrate (saltpeter), charcoal, and sulfur. As such it is a composite explosive whose properties are dependent upon how finely divided each of the ingredients is, and how intimately they are mixed. The exact origins of black powder are lost in antiquity. Publications referring to it seem about equally divided between those that attribute its origin to third- or fourth-century China7,8 and those that place it closer to the 13th century, at about the time of Roger Bacon's written description in 1242.913 Nevertheless, its use did not become very popular until the invention of the gun by Berthol Schwartz in the early 1300s; and its first recorded use in mining did not occur for over 300 years after that. First used for blasting in 1627, the production and application of black powder played a critical role in the rapid expansion of the United States in the early 19th century as canals were dug and railroads built to span the continent.

For over 200 years black powder was the only blasting agent known, but the 1800s brought a number of rapid developments that led to its demise, replacing it with safer and more powerful explosives. Table 31.2 presents a chronological summary of some of the significant discoveries of the 1800s. Credit for the first preparation of NG is generally ascribed to Ascanio Sobrero in Italy in 1846.

TABLE 31.2 Nineteenth Century Explosive Discoveries

1800	Mercury fulminate
1846	Nitrocellulose
1846	Nitroglycerin
1847	Hexanitromannite
1862	Commercial production of nitroglycerin
1867	Dynamite
1867	Blasting cap
1867	Ammonium nitrate explosive patented
1875	Blasting gelatin and gelatin dynamite
1884	Smokeless powder
1886	Picric acid
1891	TNT

Swedish inventors Emmanuel Nobel and his son Alfred took an interest in this powerful liquid explosive and produced it commercially in 1862. However, its transportation and its handling were very hazardous, and eventually Alfred Nobel discovered that NG absorbed into a granular type of material (kieselguhr) was still explosive, but was much safer to handle and use than the liquid. This new invention, called "dynamite," was difficult to ignite by the usual methods used for pure NG. Therefore, also in 1867, Alfred Nobel devised the blasting cap using mercury fulminate. With this development dynamite became the foundation of the commercial explosives industry.

For military and gun applications black powder continued to be the only explosive of choice as a propellant or bursting charge until the inventions of the late 1800s, when smokeless powder, based on nitrocellulose, proved to be a cleaner, safer, and more effective propellant than black powder. The synthesis of picric acid (2,4,6-trinitrophenol) followed by TNT and PETN gave solid, powerful, molecular explosives of more uniform performance for use in bombs and artillery shells. The main explosives used in World War I were TNT, Tetryl (2,4,6-trinitrophenylmethylnitramine), and Hexyl (hexanitrodiphenylamine), and in World War II they were TNT, PETN, and RDX (1,3,5-trinitro-1,3,5-triazacyclohexane).14

In the industrial arena the production of black powder in the United States dropped precipitously after reaching a peak of 277 million lb in 1917.15 By the mid-1960s it had ceased to be of commercial significance, but during the same time period dynamite production rose from 300 million lb to 600 million lb.

In 1947 a spectacular accident of catastrophic proportions ushered in the next revolution in explosives. Fertilizer-grade AN, in the form of prills (small spherical particles coated with paraffin to prevent caking), was being loaded into ships in Galveston Bay, Texas. Along with other cargo, one of these ships, the partially loaded *SS Grandcamp*, contained 2300 tons of this material. On the morning of April 16, soon after loading was resumed aboard the *Grandcamp*, a fire was discovered in one of the holds containing AN. Efforts to extinguish the fire were unsuccessful, and an hour later the bulk of the coated fertilizer detonated, killing 600 people and injuring 3000.16

This tragedy, along with several other large-scale accidents involving AN explosions, finally led researchers to the conclusion that inexpensive, readily available, fertilizer-grade AN could be used as the basis for modern industrial explosives.

Soon after the advent of porous AN prills, introduced in the early 1950s, investigators realized that these prills readily absorbed just the right amount of FO to produce an oxygen-balanced mixture that was both an inexpensive and effective blasting agent, in addition to being safe and simple to manufacture. This technology was widely adopted and soon constituted 85 percent of the industrial explosives produced in the United States.17 With ANFO's cost and safety characteristics, it became practical for surface miners to drill larger boreholes and to utilize bulk ANFO delivery systems. Nevertheless ANFO had two significant limitations: AN is very water soluble, so wet boreholes readily deactivated the explosive; and ANFO's low density of 0.85 g/cc limited its bulk explosive strength. Cook18 hit upon the idea of dissolving the AN in a small amount of hot water, mixing in fuels such as aluminum powder, sulfur, or charcoal, and adding a thickening agent to gel the mixture and hold the slurried ingredients in place. As this mixture cooled down, the AN salt crystals would precipitate, but the gel would preserve the close contact between the oxidizer and the fuels, resulting in a detonable explosive. Other oxidizers also could be added, and the density could be adjusted with chemical foaming agents to vary the bulk explosive strength of the product. With the addition of a cross-linking agent, the slurry or water gel could be converted to a semisolid material having significant water resistance. The latest significant development in industrial explosives actually was invented only a few years after slurries.19a, b Water-in-oil emulsion explosives involve essentially the same ingredients that slurry composite explosives do, but in a different physical form. Emulsion explosives are discussed fully under the section titled "Explosives Manufacture and Use."

The main developments in military types of explosives since World War II have been trends toward the use of plastic bonded explosives (PBXs) and the development of insensitive high explosives. Driving these trends are desires for increased safety and improved economics in the process of replacing aging TNT-based munitions and bomb fills. PBXs involve the coating of fine particles of molecular explosives such as RDX and HMX (1,3,5,7-tetranitro-1,3,5,7-tetrazacyclooctane) with polymeric binders and then pressing the resultant powder under vacuum to give a solid mass with the desired density. The final form or shape usually is obtained by machining. Explosives such as triaminotrinitrobenzene (TATB), nitroguanidine,20 and hexanitrostilbene (HNS)21 are of interest because of their high levels of shock insensitivity and thermal stability. The synthesis of many new, potentially explosive compounds is a very active and ongoing area of research,22,23 but recently interest also has focused on composite explosives similar to those used by industry. Examples are EAK, a eutectic mixture of ethylenediamine dinitrate, AN, and potassium nitrate,24 and nonaqueous hardened or cast emulsion-based mixtures.25

31.5 Classification of Explosives

The original classification of explosives separated them into two very general types: low and high, referring to the relative speeds of their chemical reactions and the relative pressures produced by these reactions. This classification still is used but is of limited utility because the only low explosives of any significance are black powder and smokeless powder. All other commercial and military explosives are high explosives.

High explosives are classified further according to their sensitivity level or ease of initiation. Actually sensitivity is more of a continuum than a series of discrete levels, but it is convenient to speak of primary, secondary, and tertiary high explosives. Primary explosives are the most sensitive, being readily initiated by heat, friction impact, or spark. They are used only in very small quantities and usually as an initiator as part of an explosive train involving less sensitive materials, such as in a blasting cap. They are very dangerous materials to handle and must be manufactured with the utmost care, generally involving only remotely controlled operations. Mercury fulminate, used in Nobel's first blasting cap, is in this category, as is the more commonly used lead azide. On the other end of the spectrum are the tertiary explosives that are so insensitive that they generally are not considered explosive.

By far the largest grouping is secondary explosives, which includes all of the major military and industrial explosives. They are much less easily brought to detonation than primary explosives and are less hazardous to manufacture. Beyond that, however, generalizations are difficult because their sensitivity to initiation covers a very wide range. Generally the military products tend to be more sensitive and the industrial products less sensitive, but all are potentially hazardous and should be handled and stored as prescribed by law. Table 31.3 lists some of the more prominent explosives of each type, along with a few of their properties.

TABLE 31.3 Some Properties of Common Explosives

Common Name	Symbol	Composition	Molecular Weight	Density (g/cc)	Detonation Velocity (km/sec)	Detonation Pressure (kilobars)	Explosive Energy (cal/g)
Primary Explosives							
Mercury fulminate		$Hg(CNO)_2$	284.7	3.6	4.7	220	428
Leadazide		$Pb(N_3)_2$	291.3	4.0	5.1	250	366
Silver azide		AgN_3	149.9	5.1	6.8		452
Lead styphnate		$C_6H(NO_2)_3O_2Pb$	468.3	2.5	4.8	150	368
Mannitol hexanitrate (Nitromannite)	MHN	$C_6H_8(ONO_2)_6$	452.2	1.7	8.3	300	1,420
Diazodinitrophenol	DDNP	$C_6H_2N_4O_5$	210.1	1.5	6.6	160	820
Tetrazene		$C_2H_8N_{10}O$	188.2	1.5			658
Secondary Explosives							
Nitroglycerin	NG	$C_3H_5(ONO_2)_3$	227.1	1.6	7.6	253	1,480
Pentaerythritol tetranitrate	PETN	$C(CH_2ONO_2)_4$	316.2	1.6	7.9	300	1,510
Trinitrotoluene	TNT	$CH_3C_6H_2(NO_2)_3$	227.0	1.6	6.9	190	900
Ethyleneglycol dinitrate	EGDN	$C_2H_4(ONO_2)_2$	152.1	1.5	7.4		1,430
Cyclotrimethylenetrinitramine (Hexogen or Cyclonite)	RDX	$C_3H_6N_3(NO_2)_3$	222.1	1.6	8.0	347	1,320
Cyclotetramethylenetetranitramine (Octogen)	HMX	$C_4H_8N_4(NO_2)_4$	296.2	1.9	9.1	393	1,350
Trinitrophenylmethylnitramine (Tetryl)		$(NO_2)_3C_6H_2N(CH_3)NO_2$	287.2	1.4	7.6	251	950
Nitroguanidine	NQ	$CH_4N_3NO_2$	104.1	1.6	7.6	256	721
Nitromethane	NM	CH_3NO_2	61.0	1.1	6.2	125	1,188
Nitrocellulose	NC	Variable		1.4	6.4	210	950
Triaminotrinitrobenzene	TATB	$C_6H_6N_3(NO_2)_3$	258.2	1.8	7.9	315	829
Diaminotrinitrobenzene	DATB	$C_6H_5N_2(NO_2)_3$	243.2	1.6	7.5	259	993
Ethylenediamine dinitrate	EDDN	$C_2H_{10}N_4O_6$	186.1	1.5	6.8		948
Ethylenedinitramine (Haleite)	EDNA	$C_2H_6N_2(NO_2)_2$	150.1	1.5	7.6	266	1,080
Picric acid		$C_6H_3O(NO_2)_3$	229.1	1.7	7.4	265	1,000
Ammonium picrate (Explosive D)		$C_6H_6NO(NO_2)_3$	246.1	1.6	6.9		800
Picramide		$C_6H_4N(NO_2)_3$	228.1	1.7	7.3		1,070
Hexanitrostilbene	HNS	$[C_6H_3C(NO_2)_3]_2$	450.2	1.7	7.1	200	1,005
TACOT-Z		$C_{12}H_4N_8O_8$	388.2	1.6	7.2	181	980
Azobishexanitrobiphenyl	ABH	$C_{24}H_6N_{14}O_{24}$	874.4	1.8	7.6		
Dinitrotoluene	DNT	$CH_3C_6H_3(NO_2)_2$	182.1	1.5	5.0		700
Composition B		49/50/1 TNT/RDX/wax		1.7	8.0	294	1,100
Pentolite		50/50 TNT/PETN		1.6	7.7	245	1,100
Amatol		50/50 TNT/AN		1.6	6.5		950
Dynamite		Variable NG and various oxidizers and fuels		0.81.6	1.87.6	30160	6751,090
Prilled ANFuel Oil	ANFO	94/6 AN/FO		0.80.9	1.54.0 (depends on diameter)		880
Slurries or water gels		Variable mixtures of oxidizers, fuels, and water		0.91.4	3.55.0		6001,200
Emulsions		Variable solutions of oxidizers in water and fuels		0.91.4	4.56.0		7001,100
Heavy ANFO		5075% AN with 5025% emulsion		1.11.3	4.04.5		755815
Tertiary Explosives							
Mononitrotoluene	MNT	$CH_3C_6H_4NO_2$	137.1	1.2			
Ammonium perchlorate	AP	NH_4ClO_4	117.5	1.9	3.4	187	488
Ammonium nitrate	AN	NH_4NO_3	80.1	1.4	3.2		346

For industrial applications, secondary explosives are subdivided according to their initiation sensitivity into two classes: Class 1.1 and Class 1.5. Class 1.1 explosives are sensitive to initiation by a blasting cap and usually are used in relatively small-diameter applications of 13-in. boreholes. Class 1.5 (blasting agents) are high explosives that are not initiated by a Standard # 8 electric blasting cap under test conditions defined by the U.S. Department of Transportation (DOT), and that pass other defined tests designed to show that the explosive is "so insensitive that there is very little probability of accidental initiation to explosion or of the transition from deflagration to detonation."26 Being less sensitive, blasting agents are generally used in medium- and large-diameter boreholes and in bulk applications. Dynamites are always Class 1.1, but other composite explosives made from mixtures of oxidizers and fuels can be made either Class 1.5, depending upon the formulation and the density. Density plays a significant role in the performance of most explosives, and this is especially true for slurry and emulsion explosives where the density may be adjusted by air incorporation, foaming agents, or physical bulking agents, irrespective of the formulation. Blasting agent classification is of interest because regulations governing transportation, use, and storage are less stringent for blasting agents than for Class A explosives. (Propellants and fireworks are classified by the DOT as Class 1.2 or 1.3 explosives, and blasting caps and detonating cord as Class 1.4.)

31.6 Structural Characteristics of Explosives

The number of potentially explosive compounds is virtually unlimited. A listing by the U.S. Bureau of Alcohol, Tobacco and Firearms of explosive materials under federal regulation27 numbered 225, and many of the items listed were broad, general categories. The ten-volume *Encyclopedia of Explosives and Related Items* compiled by the U.S. Army Picatinny Arsenal over a 25-year period contains several thousand entries. New organic molecular explosives are being synthesized continually; composite explosives, such as current commercial products that are mixtures of oxidizers and fuels, present an infinite number of possible combinations. The complexity of trying to comprehensively list the chemical structures of explosives is shown by a 1977 reference that listed 13 separate categories just for primary explosives.28 However, the majority of the most important explosives can be grouped into a few classes sharing common structural features that are of value to researchers in understanding and predicting explosive properties.

The following seven categories,29a, b updated to include the relatively recent fluoroderivatives,22 appear to be the most encompassing. Many explosives may contain more than one category, but not every compound that contains one of these groups is necessarily an explosive.

1. N=O
2. NN, N=N, and N≡N
3. C≡N and C≡N
4. C≡C
5. Cl=O
6. NX, where X = Cl, F, I
7. OO

Category 1 is by far the largest. It includes nitro groups, both aliphatic and aromatic; nitrate esters; nitrate salts; nitramines; and nitrosamines. Nearly all of the explosives listed in Table 31.3 fall into this category. Prominent examples are: nitromethane, an aliphatic nitro compound; TNT, an aromatic nitro compound; NG and PETN, nitrate esters; EDDN and ammonium nitrate, nitrate salts; and RDX and HMX, nitramines. Category 2 represents the hydrazine, azo, diazo, and azide compounds, both organic and inorganic. Hydrazine, tetrazene, and lead azide are examples of this group. Category 3 is represented by the explosives mercury fulminate and cyanogen, respectively. Acetylene and metallic acetylide salts constitute category 4. Category 5 consists mainly of inorganic and organic ammonium salts of chloric and perchloric acid, but would also include various chlorine oxides. Category 6 is generalized to include most of the amine halogens, nitrogen triiodide being a classic example. Also, considerable new synthetic work has focused on inserting the energetic difluoroamine groups into various organic molecules to form explosives that fall into this category. Category 7 includes organic peroxides and ozonides as well as hydrogen peroxide itself.

Commercial industrial explosives such as dynamites, slurries, and emulsions are included in these categories because their major components, nitrate esters and nitrate and perchlorate salts, are listed. However, mixtures of fuels and oxygen or other gases that may be explosive at certain ratios are not covered, including the liquid oxygen explosives that saw limited application earlier in the 20th century.

31.7 Explosives Manufacturing and Use

Details of the synthesis and larger-scale production of a number of molecular explosives including dynamites are given in the fourvolume series by Urbanski (*Chemistry and Technology of Explosives*, Pergamon Press. 196484) and in various military books such as *Engineering Design Handbook: Explosive Series*.6 Formulations of commercial slurries and emulsions generally are considered proprietary and are described mainly in the patent literature. Some specific examples of prominent explosives with general preparation methods are given below.

TNT (2,4,6-Trinitrotoluene)

TNT is no longer manufactured commercially in the United States, but is manufactured in significant quantities at several government plants because it is still an important military explosive. It is produced commercially in Canada and other countries and is imported into the United States for use in cast boosters to initiate industrial blasting agents. In a relatively straightforward process, TNT is made by the direct trinitration of toluene with nitric acid. Most modern processes are set up for continuous production in a series of nitrators and separators with the nitrating acid flowing countercurrently. This procedure avoids having to isolate the intermediate mono- and dinitration products and may also employ continuous purification and crystallization, being carried out simultaneously with production.

Mixed nitric and sulfuric acids sometimes are used with the addition of SO_3 or oleum. The sulfuric acid or oleum helps drive the reaction to completion by removing the water produced by nitration and by dehydrating nitric acid to form the more reactive nitronium ion (NO_2^+). Because toluene is not very soluble in the acid, powerful agitation is required. The spent acid is removed in successive separation steps, and the sulfuric acid is reused after the addition of more nitric acid. The molten TNT product is purified with multiple water and sodium sulfate washes, which produce significant quantities of "yellow water" and "red water" waste streams, respectively, that must be properly handled to avoid environmental problems. The low melting point of TNT (8082°C) is ideal for melt casting, and TNT usually is employed as a mixture with other higher-melting explosives such as PETN, RDX, HMX, and tetryl. This feature and the excellent chemical stability of TNT have made it the most popular and widely used military explosive in the world.

RDX and HMX

Both RDX and HMX are cyclic nitramines made by nitrolysis of hexamethylene tetramine (HMT). Their good thermal stabilities, high melting points (>200° C), and high energy properties make these crystalline compounds popular as projectile and bomb fills and for use in cast boosters and flexible, sheet explosives. HMX has superior detonation properties and a higher melting point than RDX, but it is more difficult and more expensive to manufacture.

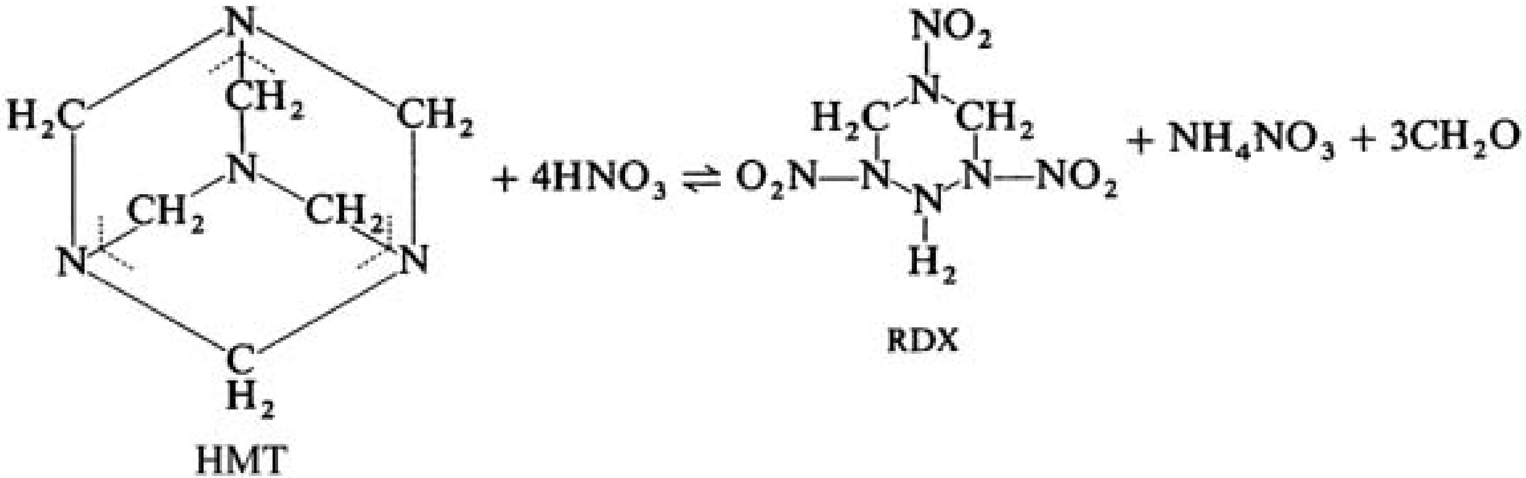
Reaction 1 shows the formation of RDX by the action of nitric acid on HMT. Schematically, RDX formation can be pictured as nitration of the three "outside" nitrogen atoms of HMT (in more accurate, three-dimensional representations all four nitrogens are equivalent) with removal of the "inside" nitrogen and methylene (CH_2) groups. AN (NH_4NO_3) and formaldehyde (CH_2O) are produced as by-products but can be used to form more RDX with the addition of acetic anhydride, as shown in Reaction 2. In actual practice these two reactions are run simultaneously, as shown in the combined reaction to produce approximately two moles of RDX for each mole of HMT.

HMX was discovered as an impurity produced in the RDX reaction. It is composed of an eight-membered ring rather than the six-membered ring of RDX. The latter is more readily formed than the eight-membered ring, but with adjustment of reaction conditions (lower temperature and different ingredient ratios), HMX formation can be favored. Schematically its formation can be pictured by nitration of all four nitrogens in hexamethylene tetramine and removal of two methylene groups as indicated in Reaction 3. To obtain pure HMX the RDX "impurity" must be removed by alkaline hydrolysis or by differential solubility in acetone.

HNS (2,2′,4,4′,6,6′-Hexanitrostilbene)

This is a relatively new explosive having been prepared unequivocally for the first time in the early 1960s.35a, b It is of interest primarily for two reasons: (1) its high melting point (316°C) and excellent thermal stability, and (2) its unique crystal-habit-modifying effects on cast TNT. The former makes HNS useful in certain military and space applications as well as in hot, very deep wells, and the second property is used to improve TNT castings. It can be manufactured continuously by oxidative coupling of TNT as shown below.

Reaction 1



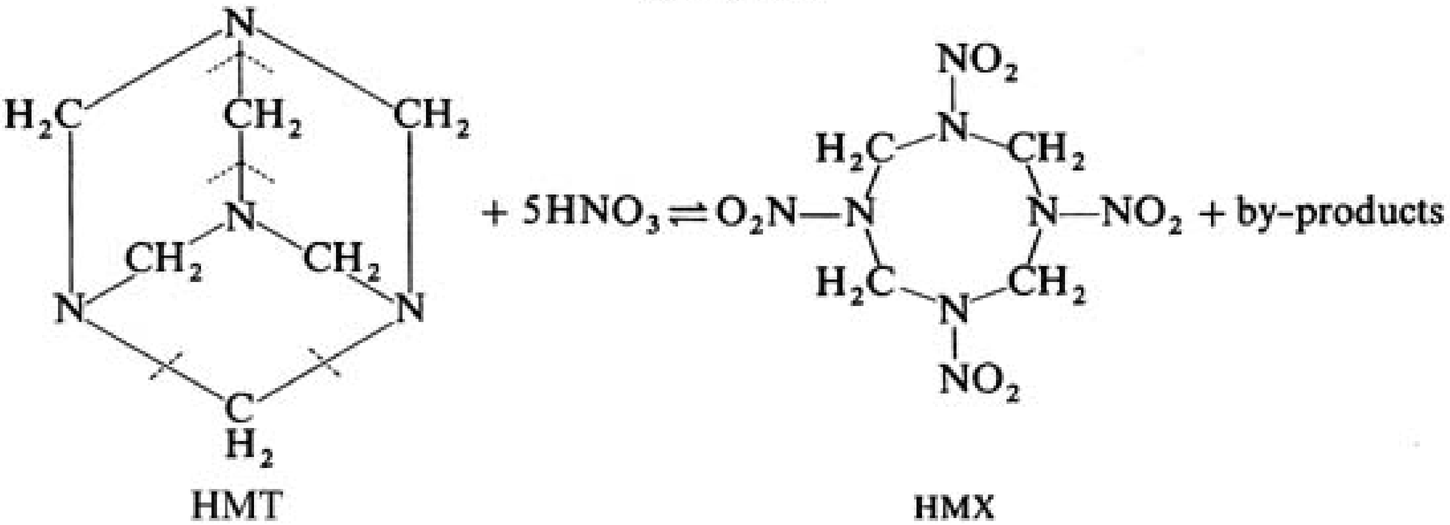
Reaction 2

$$\underset{\text{Formaldehyde}}{3CH_2O} + \underset{\text{AN}}{3NH_4NO_3} + \underset{\text{Acetic anhydride}}{6(CH_3CO)_2O} \rightleftharpoons \underset{\text{RDX}}{(CH_2 \cdot N \cdot NO_2)_3} + \underset{\text{Acetic acid}}{12CH_3COOH}$$

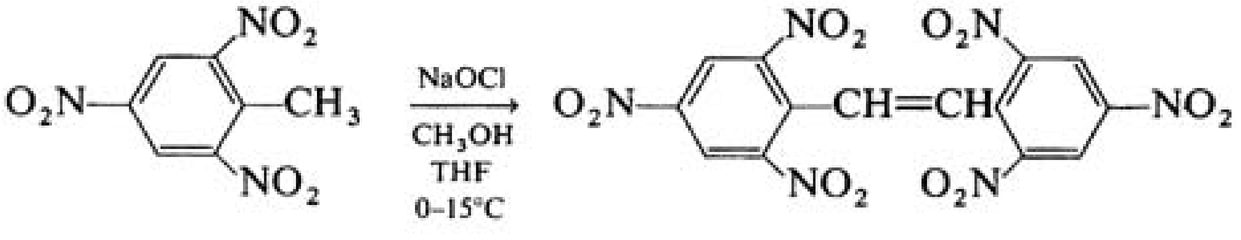
Combined Reaction

$$(CH_2)_6N_4 + 4HNO_3 + 2NH_4NO_3 + 6(CH_3CO)_2 \rightleftharpoons 2(CH_2 \cdot N \cdot NO_2)_3 + 12CH_3COOH$$

Reaction 3



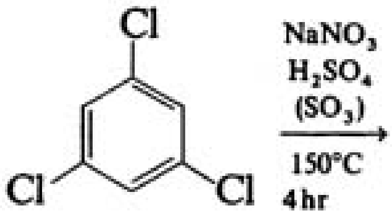
This relatively simple process from readily available TNT and household bleach (5% NaOCl solution) has been shown to involve a series of intermediate steps that give HNS in only low to moderate yields (3045%) with many by-products. Although it also involves the use of expensive organic solvents that must be recovered, this synthesis is used commercially.36,37 Studies to improve this process constitute an active area of research.



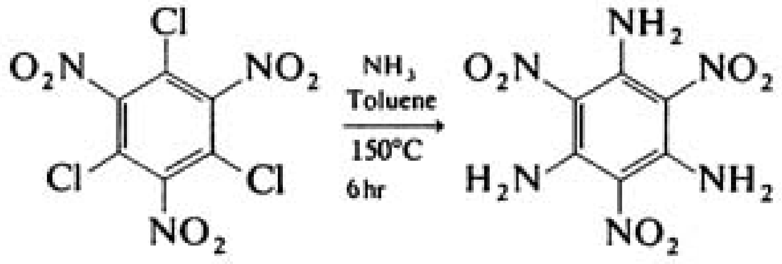
TNT HNS

TATB (1,3,5-Triamino-2,4,6trinitrobenzene)

This highly symmetrical explosive molecule has even higher thermal stability than HNS (greater than 400°C) and has become of special interest in the last two decades because of its extreme insensitivity.38,39a, b Because its accidental initiation is highly unlikely, TATB has been used in nuclear warheads and is being explored in plastic bonded systems for a number of military and space applications.19 Currently it is manufactured in large-scale batch processes that are little change from its original synthesis over 100 years ago. The two-step process involves trinitration of trichlorobenzene followed by amination to displace the chlorine groups as shown below.



Trichlorobenzene



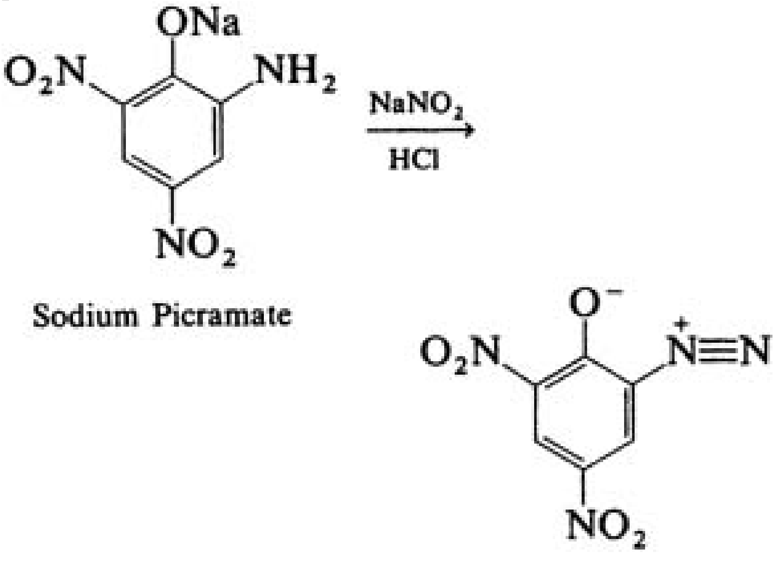
Trinitrotrichlorobenzene TATB

Both steps require high temperature and considerable reaction time but give 8090 percent yields. The major problem areas are chloride impurities in the final product and the excessively fine particle size of the final product. Because TATB is highly insoluble in most solvents, it is difficult to purify the product or to change its particle size by recrystallization. Also the starting material is expensive and not very readily available.

Recently a similar synthetic procedure starting with 3,5-dichloranisole was reported.40

DDNP (2-Diazo-4,6-dinitrophenol)

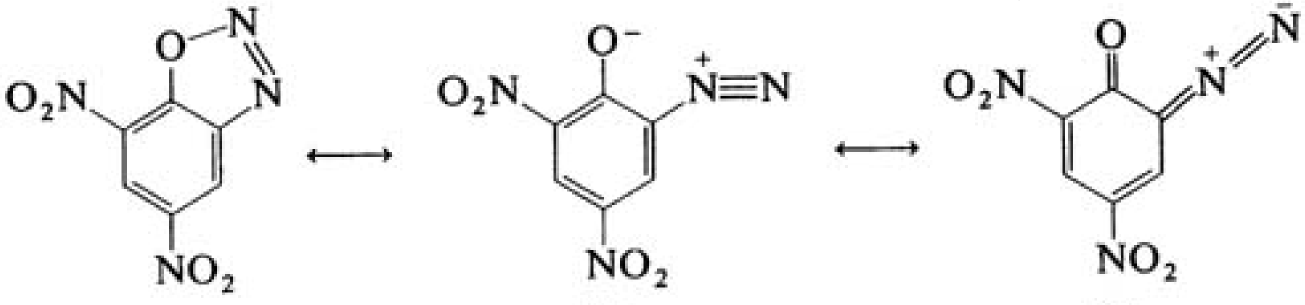
This yellow-to-brown crystalline material (melting point 188°C) is a primary explosive used as the initiator charge in electric blasting caps as an alternative to lead azide. It is less stable than lead azide but much more stable than lead styphnate, and is a stronger explosive than either of them because it does not contain any metal atoms. DDNP is also characterized as not being subject to dead pressing (tested at pressures as high as 130,000 psi). It was the first diazo compound discovered (1858) and was commercially prepared in 1928. It is manufactured in a single-step, batch process by diazotizing a slurry of sodium picramate in water.



Sodium Picramate

DDNP

The structure shown is convenient for visualization purposes, but DDNP actually exists in several tautomeric forms with form (2) apparently predominating.

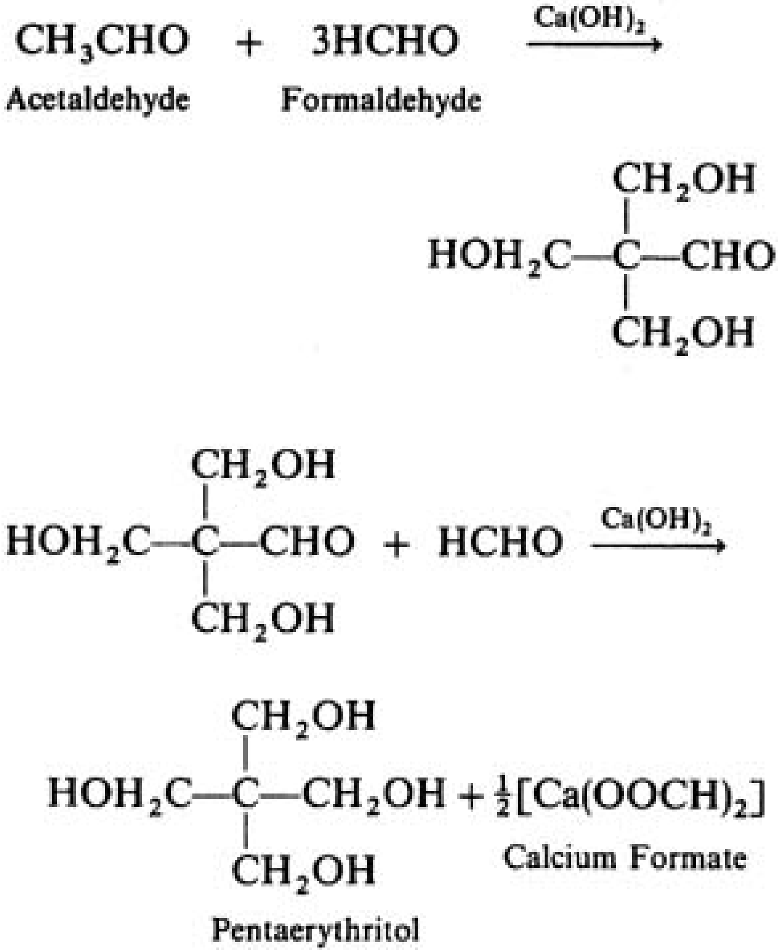


(3) (1) (2)

The sodium picramate starting material is itself explosive, but is commercially available as a chemical intermediate. It can be made by the reduction of picric acid with reducing agents such as sodium sulfide. The key to making useful DDNP is to control the rate of diazotization so that relatively large, rounded crystals are formed instead of needles or platelets that do not flow or pack well.

PETN (Pentaerythritol Tetranitrate)

Although known as an explosive since 1894, PETN was used very little until after World War I when the ingredients to makethe starting material became commercially available. The symmetrical, solid alcohol starting material, pentaerythritol, is made from acetaldehyde and formaldehyde, which react by aldol condensation under basic catalysis followed by a crossed Cannizzaro disproportionation to produce the alcohol and format salt. Although the reaction takes place in a single mixture, it is shown below in two steps for clarity.



For PETN manufacture the pentaerythritol starting material can be readily purchased as a commodity chemical from commercial suppliers. The nitration is relatively simple, involving only nitric acid (9698%) and the solid alcohol added slowly with mixing and cooling. PETN is not very soluble in nitric acid or water and is readily filtered directly from the acid or after dilution of the acid with water. Water washing and recrystallization from acetonwater mixtures give the desired particle size ranges and the desired purity. PETN can be made either batchwise or continuously for large-scale production.

Pure PETN is a white, crystalline solid with a melting point of 141.3°C. Because of its symmetry it is said to have higher chemical stability than all other nitrate esters.41 Relatively insensitive to friction or spark initiation, PETN is easily initiated by an explosive shock and has been described overall as one of the most sensitive, noninitiating, military explosives.42 As with most explosives, the detonation velocity of PETN varies with the bulk density of the explosive. Most military applications of PETN have been converted to RDX because of its greater thermal stability. However, in industry PETN is widely used as the major component in cast boosters for initiating blasting agents, as the explosive core in detonating cord, and as the base load in detonators and blasting caps. For safety in handling, PETN is shipped in cloth bags immersed in wateralcohol mixtures and dried just before use.

NG (Nitroglycerin or Glyercol Trinitrate)

This nitrate ester is one of only a very few liquid molecular explosives that are manufactured commercially. It is a clear, oily liquid that freezes when pure at 13°C. As seen in the historical section, the first practical use of NG was in dynamites, where it is still used today more than 100 years later. It also is used as a component in multibased propellants and as a medicine to treat certain coronary ailments. This latter usage is attributed to NG's ability to be rapidly absorbed by skin contact or inhalation into the blood, where it acts as a vasodilator. (At high exposure levels such as in dynamite manufacture and handling, this property is responsible for the infamous powder headache.) NG is undoubtedly the most sensitive explosive manufactured in relatively large quantities. Its sensitivity to initiation by shock, friction, and impact is very close to that of primary explosives, and extreme safety precautions are taken during manufacture. Pure glycerin is nitrated in very concentrated nitric and sulfuric acid mixtures (typically a 40/60 ratio), separated from excess acid, and washed with water, sodium carbonate solution, and water again until free from traces of acid or base. Pure NG is stable below 50°C, but storage is not recommended. It is transported over short distances only as an emulsion in water or dissolved in an organic solvent such as acetone. Traditionally it has been made in large batch processes, but safety improvements have led to the use of several types of continuous nitrators that minimize the reaction times and quantities of explosives involved. Because of its sensitivity, NG is utilized only when desensitized with other liquids or absorbent solids or compounded with nitrocellulose.

31.8 Ammonium Nitrate and ANFO

AN continues to be the most widely used component of commercial explosives. It is used in nearly all of the packaged and bulk explosives on the market. The manufacturing process is described in Chapter 29.

Ammonia is basically the main raw material needed to manufacture AN. Some of the AN manufacturers make their own ammonia and some purchase it on the open market. It is obvious that the cost of manufacturing AN will depend on the price of ammonia and, even more basically, natural gas from which it is made. The volatility of ammonia prices is shown by the fact that in 1992 it cost $95 per ton and in 1995 the cost was $207 per ton.30

There are many producers of AN in North America making both AN solution and explosive grade AN prills. The AN solution is used in the manufacture of packaged and bulk emulsion and water gel explosives, and explosive-grade AN prills are used to make ANFO. ANFO, the acronym for a mixture of AN and FO, is, the single most commonly used chemical explosive. (ANFO is an example of a *composite* explosive as described in an earlier section, "Chemistry of Combustion and Explosions".) These low density AN prills are made by a specialized process, in which internal voids are created making the prills porous able to absorb the required 5.56 percent FO.

ANFO alone represents about three fourths of the current volume of commercial explosives in use today around the world. Because of this, ANFO is commonly used as a reference when defining and comparing explosive properties. Some of these important explosive properties include density, detonation velocity, and energy release.

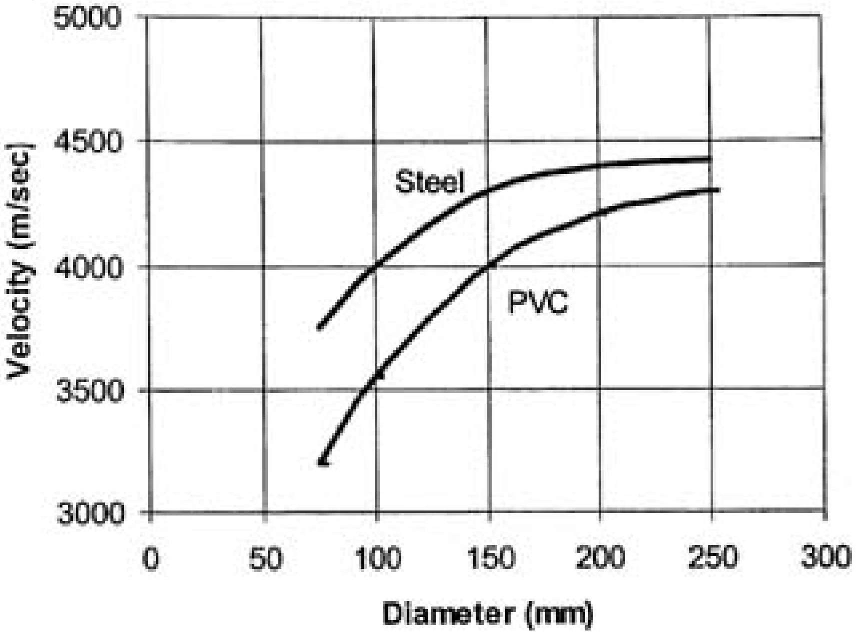
The crystal density of AN is about 1.72 g/cc, and the particle density of explosive-grade AN prills ranges from 1.40 to 1.45 g/cc depending upon the manufacturing process. This difference in crystal and particle density reveals the volume of pores or voids created by the specialized prilling process. The porosity of AN prills is the property desired in the manufacture of ANFO, since this determines how much FO can be absorbed. This intimate mixture of AN with FO is critical in the efficient detonability of ANFO. The AN prill particle density and inherent void-space value also become important when predicting and calculating the densities of ANFO blends with water-gel and emulsion explosives.

The bulk density of ANFO ranges from 0.80 to 0.87 g/cc. So, clearly about half of the ANFO is air or void space. All explosives require a certain amount of entrained void space in order to detonate properly. These void spaces also play a major role in the detonation reaction by creating "hot spots" under adiabatic compression in the detonation front.31 The amount of void space in any given explosive and the resultant change in density have a significant impact on the detonation properties like detonation velocity, sensitivity, and even energy release.

Generally speaking, the detonation velocity of an explosive will increase with density until a failure point is reached. This failure point is commonly referred to as the critical density of that particular explosive. The density at that point is so high and the void space so low that the detonation cannot be sustained and failure occurs.

Other important parameters that affect the detonation velocity and performance of ANFO are charge diameter and confinement. The detonation velocity of ANFO will increase by about 300 m/s when the charge confinement is changed from a PVC tube to a Schedule 40 steel pipe. A summary of test data on ANFO velocity versus confinement and diameter is shown in Fig. 31.4.

Fig. 31.4 Detonation velocity of ANFO versus diameter and confinement.



The basic chemical reaction of ANFO can be described with the following equation:

$$3NH_4NO_3 + CH_2 \rightarrow CO_2 + 7H_2O + 3N_2 + 880\,\mathrm{cal/g}$$

Using CH2 to represent FO is generally accepted, but it really is an oversimplification, since it is a mixture of hydrocarbons. The heat energy release of 880 cal/g is the theoretical maximum value based upon the heats of formation of the reactants and products. Of course. All of the products of detonation are gases at the detonation temperature of about 2700 K.

The theoretical work energy that is released from an explosive reaction can be calculated using a variety of equations of state and computer programs.32 Explosive energy can also be measured by a variety of techniques including underwater detonation of limited size charges with concurrent measurements of the shock and bubble energies.33 Each explosive manufacturer has an energy measurement and equation of state that is used to calculate and report their product properties. This often leads to confusion and controversy when explosive consumers try to compare product lines when given only technical information sheets. Since theoretical calculations must of necessity be based on a number of assumptions, the only valid comparisons are done in the field with product testing and detailed evaluation of results.

31.9 Bulk Emulsions

During the 1990s the commercial use of bulk emulsion explosives continued to increase. Bulk emulsion products began to significantly replace packaged products in underground mining and in quarry operations. Also, bulk emulsion and ANFO blends became very popular in large volume open-pit mining operations. Emulsions are made by combining an oxidizer solution and a fuel solution using a high-shear mixing process. The oxidizer solution is normally 9095 percent by weight of the emulsion. It contains AN, water, and sometimes a second oxidizer salt, that is, sodium or calcium nitrate. The solution must be kept quite hot, since the water is minimized for increased energy, and the crystallization temperature is typically 5070°C. The fuel solution contains liquid organic fuels, such as FO and/or mineral oils, and one or more emulsifiers. An emulsifier is a surface active chemical that has both polar and nonpolar ends of the molecule. In the high-shear manufacturing process, the oxidizer solution is broken up into small droplets, each of which is coated with a layer of fuel solution. The droplets in this metastable water-in-oil emulsion are basically held together with the emulsifier molecules, which migrate to the surface of the dispersed droplets. In today's explosives industry, much of the research work is directed towards developing better and more efficient emulsifier molecules that will improve the storage life and handling characteristics of the bulk and packaged emulsions. The emulsifiers currently used in commercial explosives range from relatively simple fatty acid esters with molecular weights of 300400 to the more complex polymeric emulsifiers having molecular weights in excess of 2000.

Figure 31.5 shows a photomicrograph of an emulsion explosive at 400 power with the typical distribution of the fuel-coated oxidizer solution droplets (normally 15 µm in diameter). Figure 31.6 shows a bulk emulsion exiting a loading hose and displaying the soft ice creamlike texture typical of bulk emulsions. The viscosities of bulk emulsions can range from nearly as thin as 90 weight oil to as thick as honey, depending upon the application requirements. Emulsion viscosity increases with product cooling, but most emulsions continue to remain stable at temperatures below 0°C, which is considerably below the crystallization temperature of the oxidizer solution. The oxidizer solution droplets in the emulsion are therefore held in a supersaturated state. Over time, the surface layers created by the emulsifier molecules can be broken, and the oxidizer solution droplets are free to form crystals. At this point the emulsion begins to "break down" and lose some of the desirable properties. For this reason the emulsion composition must be optimized for a particular application in terms of its product stability and usable storage life.

Fig. 31.5 Photomicrograph of a bulk emulsion.

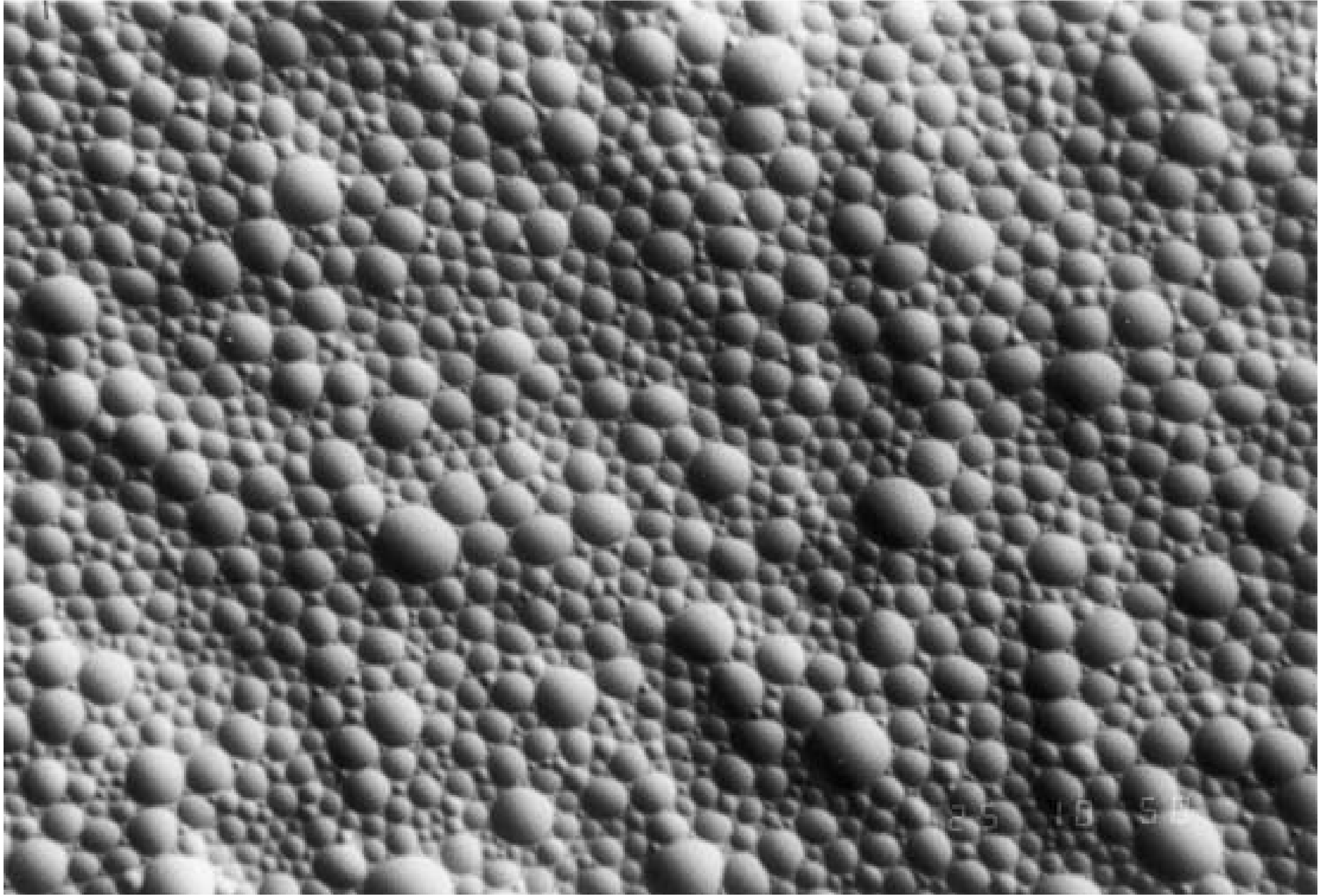


Fig. 31.6 Bulk emulsion exiting a loading hose.

The intimate mixing of oxidizer and fuel in emulsions give these explosives much higher detonation velocities when compared to ANFO. For example, in 150 mm diameter PVC ANFO has a velocity of about 4000 m/sec, and a sensitized emulsion would have a velocity closer to 6000 m/sec at a density of 1.201.25 g/cc. Also, the layer of oil surrounding each oxidizer solution droplet protects the emulsion from extraneous water intrusion and subsequent deterioration of the explosive. Many studies have shown that when mining operations use emulsion explosives rather than ANFO, which has basically no water resistance, the amount of nitrate salts in mine ground water is reduced considerably. This can be a very important factor in today's environmentally conscious mining and explosives industry.

Bulk emulsions are generally nondetonable per se and must be sensitized with some type of density control medium to become usable blasting agents. That is, voids, creating "hot spots," are required to sustain the detonation front. The two most commonly used density control methods are hollow solid microspheres or gas bubbles created by an in-situ chemical reaction. Both glass and plastic hollow microspheres are commercially available and used by explosives manufacturers. The in-situ chemical gassing techniques require considerably more expertise and generally utilize proprietary technology.

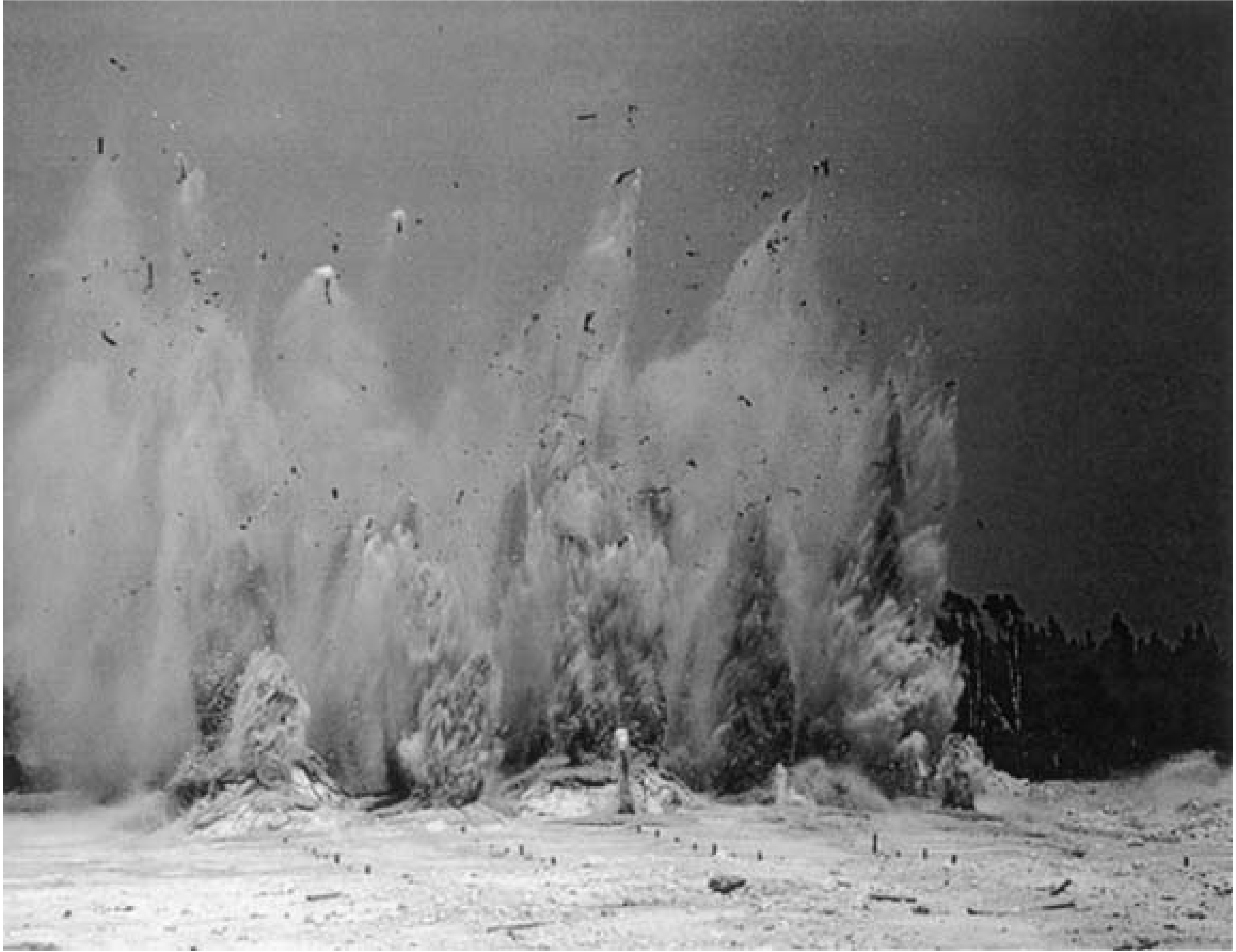
In the past decade the use of sensitized bulk emulsions has increased considerably in underground mining. Much of this has been due to the development of innovative loading equipment and techniques. One example of this is given in Fig. 31.7 which shows a small-volume pressure vessel that can be used for development and tunnel rounds utilizing horizontally drilled boreholes. The bulk emulsion blasting agent is pressurized inside the vessel and literally squeezed through the loading hose into the boreholes. A continuous column of explosives is assured by inserting the loading hose to the back of the hole and extracting it as the product is loaded. Much more complex underground loading units are available for loading bulk emulsion into boreholes drilled at any angle to the horizontal from straight up to straight down. The emulsion explosives used in these specialized loading units were specifically designed for underground use over twenty years ago, and it has been successfully used in underground mining operations around the world. The fuel and oxidizer contents are carefully balanced, and this, combined with the excellent water resistance and detonation efficiency, results in the near elimination of after-blast toxic fumes, such as CO, NO, and NO2. The fume characteristics of this product have been shown to be considerably superior to either dynamite or ANFO. For example, a series of tests in an underground chamber in Sweden compared the after-blast fumes of this emulsion to ANFO. The CO was reduced from 11 to less than 6 L/kg of explosive, and the NO plus NO2 was reduced from about 7 to less than 1 L/kg of explosive.34

Fig. 31.7 Underground pressure vessel loader for bulk emulsion. (*Courtesy Dyno Nobel*)

Many open-pit quarries also use bulk emulsions for their blasting operations. As the size of the quarry increases, the size of the explosive loading trucks also must increase. Truck payloads can range from 5000 to 30,000 lb of product. Fig. 31.8 shows an emulsion pumper truck in a quarry in south Florida. These particular trucks, with a payload of about 20,000 lb, are specially designed for a site-mixed system, in which each truck is an emulsion manufacturing unit. Combining nonexplosive raw materials directly on the truck maximizes safety and minimizes requirement for explosive storage. This particular bulk emulsion is manufactured at a rate between 300 and 500 lb/min and sensitized to the desired density with a chemical gassing system as it is loaded into the boreholes. Figure 31.9 shows a Florida blast in progress. Note the ejection of cardboard tubes from some holes. These tubes must be used in most areas to keep the boreholes from collapsing in the layered, coral limestone formation.

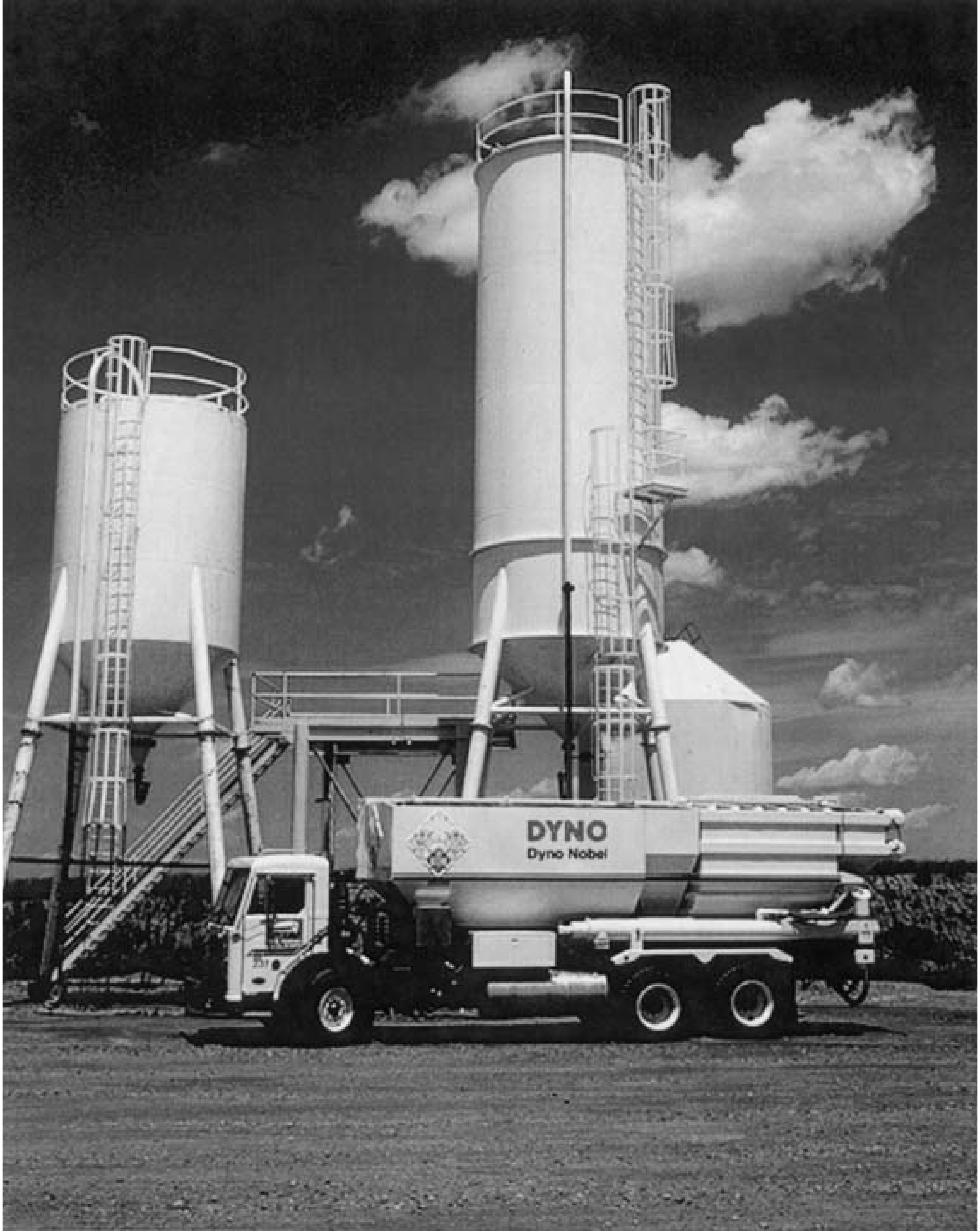
Fig. 31.8 Bulk emulsion loading truck in a Florida quarry.

Fig. 31.9 Florida quarry blast in progress.



Many of the large volume metal and coal mining operations around the world have both bulk emulsion and AN prills stored either on site or nearby so that any combination of these two products can be used. Figure 31.10 shows a typical explosive staging area in a large open-pit coal mine in Wyoming. The explosive truck in the foreground has compartments on board for emulsion, AN prills, and FO, so any combination of products ranging from straight emulsion to straight ANFO can be loaded. The truck has a capacity ot about 50,000 lb and can deliver product to the boreholes at up to a ton per minute. Each borehole can contain as much as 5 tons of explosive, and some of the blast patterns can contain as much as 10 million total pounds.

Fig. 31.10 Typical bulk explosives staging area in a large open-pit mine.



The emulsion/ANFO explosive blend selection to be used in any given mining application depends upon many factors. Typically, ANFO is the least expensive product, but it also has the lowest density and no water resistance. As emulsion is added to ANFO it begins to fit into the interstitial voids between the solid particles and coat the AN prills increasing the density, detonation velocity, and water resistance. The density increases nearly linearly with percent emulsion from about 0.85 g/cc with ANFO to about 1.32 g/cc with a 50/50 blend. This range of emulsion/ANFO blends is commonly referred to as Heavy ANFO. As the density increases the amount of explosive that can be loaded into each borehole increases, and either drill patterns based on ANFO can be spread out or better blasting results can be obtained.

It is commonly accepted in the explosive and mining industry that at least 4050 percent emulsion is required to protect the Heavy ANFO blend from borehole water intrusion. Pumped explosive blends with 6080 percent emulsion can even be used when severe water conditions are encountered. These products can be pumped through a loading hose, which can be lowered to the bottom of the borehole and displace the water during loading. Trucks similar to that shown in Fig. 31.8 can be used for these products. Most Heavy ANFO products are more simply mixed and loaded through an auger into the top of boreholes. For Heavy ANFO products the holes must be either dry or dewatered using pumps.

The basic chemical composition of a typical all-AN oxidizer emulsion explosive would be: AN plus about 15 percent water plus about 5 percent fuels. The fuels may contain fuel oil, mineral oils, and emulsifiers, the majority of which can generally be described as CH2 hydrocarbon chains. Therefore, a very simplified chemical reaction for a basic emulsion can be written as follows:

$$3NH_4NO_3 + CH_2 + 2.5H_2O + CO_2$$

$$+ 9.5H_2O + 3N_2 + 680\,\text{cal}/\text{g}$$

So, by adding 15 percent water to the ANFO reaction described earlier, the theoretical heat energy release is reduced from 880 to 680 cal/g. The difference is the energy price paid for using water and converting it to steam in the detonation reaction. The advantages and disadvantages of using ANFO or emulsions begin to become clear. ANFO is easily mixed and is probably the least expensive form of explosive energy, but it has no water resistance and has a relatively low loading density. Emulsions are considerably more complicated to formulate and manufacture, but they have excellent water resistance and more flexibility in terms of density, velocity, and energy to match rock types and blasting applications.

31.10 Dynamite

Dynamite is not a single molecular compound but a mixture of explosive and nonexplosive materials formulated in cylindrical paper or cardboard cartridges for a number of different blasting applications. Originally Nobel simply absorbed NG into kieselguhr, an inert diatomaceous material, but later he replaced that with active ingredientsfinely divided fuels and oxidizers called dopes. Thus, energy is derived not only from the NG, but also from the reaction of oxidizers such as sodium nitrate with the combustibles.

The manufacture of dynamite involves mixing carefully weighed proportions of NG and various dopes to the desired consistency and then loading preformed paper shells through automatic equipment. Because dynamites represent the most sensitive commercial products produced today, stringent safety precautions such as the use of nonsparking and very-little-metal equipment, good housekeeping practices, limited personnel exposure, and barricaded separations between processing stations are necessary. Today, the "NG" used in dynamite is actually a mixture of EGDN and NG (formed by nitrating mixtures of the two alcohols), in which NG is usually the minor component. Table 31.4 lists the common general types of dynamites with their distinguishing features. The straight dynamites and gelatins largely have been replaced by the ammonia dynamites and ammonia gelatins for better economy and safety characteristics.

TABLE 31.4 General Types of Dynamite

1. Straight dynamite	Granular texture with NG as the major source ofenergy.
2. Ammonia dynamite ("extra" dynamite)	AN replacing part of the NG and sodium nitrate of the straight dynamite.
3. Straight gelatin dynamite	Small amount of nitrocellulose added to produce soft to tough rubbery gel.
4. Ammonia gelatin dynamite ("extra" gelatin)	AN replacing part of the NG and sodium nitrate of the straight gelatin.
5. Semigelatin dynamite	Combination of types 2 and 4 with in-between properties.
6. Permissible dynamite	Ammonia dynamite or gelatin with added flameretardant.

31.11 Packaged Emulsions

The use of NG-based dynamite continued to decline during the 1990's throughout the world. For example, by 1995 there was only one dynamite manufacturing plant left in North America, and in 2000 the dynamite production at this plant had dropped to less than half the amount produced in 1990. The reasons for the declining use of dynamite are its unpopular properties of sensitivity to accidental initiation and the headache-causing fumes. Both bulk and packaged emulsions have been slowly replacing dynamite since about 1980.

Packaged emulsions are basically made with the same manufacturing equipment as the bulk emulsions. The fuel component usually contains waxes and other thickeners to give the emulsions a thick, putty-like consistency, and the oxidizer solution often contains both AN and a second oxidizer salt to produce optimum after-blast fumes. After manufacture, the thick emulsion is extruded into packaging material, normally a plastic film. The final product is then clipped together with metal clips forming firm, sausage-like chubs. Some packaged emulsions are also available in paper cartridges, designed to simulate dynamite packaging. Fig. 31.11 shows some commercial packaged emulsion cartridges in both plastic and paper wrappings.

Fig. 31.11 Commercial packaged emulsion cartridges.

To obtain reliable detonability in small diameters, the density of packaged emulsions must be maintained at a relatively low value, typically 1.101.20 g/cc. On the other hand, some dynamites are available with densities in excess of 1.40 g/cc. These higher density and energy dynamites have been the most difficult to replace with emulsions and are the primary dynamite products currently in production.

31.12 Initiation Systems

The first reliable initiation system for commercial explosives could probably be traced back to Alfred Nobel's invention of the blasting cap in 1864. This, combined with his subsequent invention of dynamite about three years later, basically started the modern era of blasting. In the century that followed, the initiation systems became more and more sophisticated and safe. Short and long-period delay electric blasting caps were perfected and detonating cord was developed. Detonating cord is a flexible cord made of cloth or plastic with a core load of high explosives, usually PETN. Strings or circuits of detonating cord could be used to initiate several explosive charges with only one blasting cap.

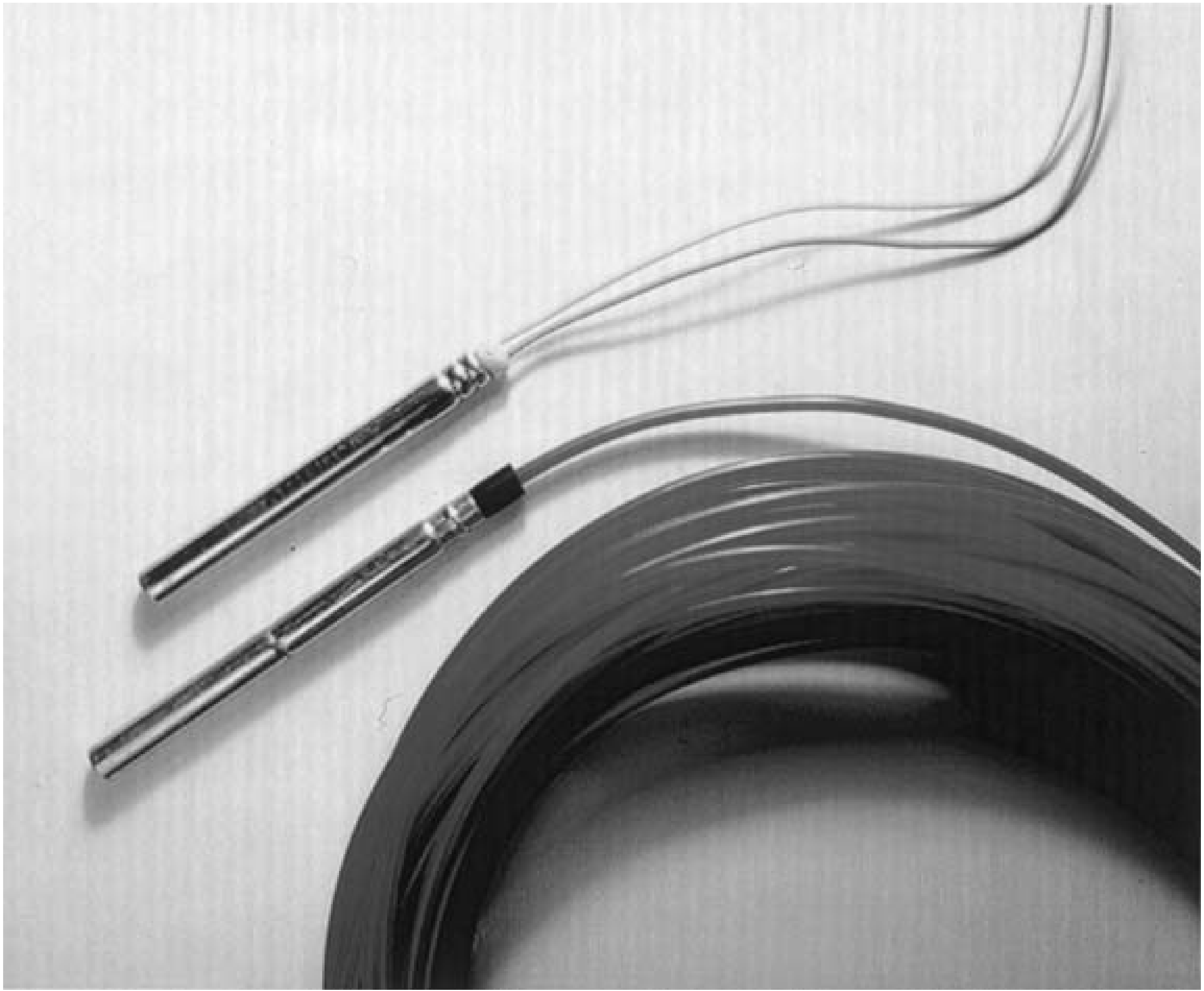
Prior to about 1950 most of the commercial explosives in the market were reliably detonable with just a blasting cap or detonating cord as the initiator. Then came the advent of ANFO and later water-gels, invented by Melvin A. Cook in 1957. These explosive products were considerably less sensitive than dynamites and required larger "booster" charges for reliable detonation. At first, a high density and high velocity dynamite was used as the booster charge, and later TNT-based cast boosters came into the market. These cast boosters continue to be used today in nearly all large mining operations. Cast TNT by itself is not reliably detonable with a blasting cap or detonating cord, and so 4060 percent PETN is normally added to the TNT melt and subsequent cast. The combination of TNT and PETN is called Pentolite. TNT has a melting point of about 80°C, which makes it an excellent base explosive for casting into forms. The military has used this concept for decades for filling bomb casings. Once the TNT has melted, other material can be added to give the final cast explosive composition the desired properties. In the case of Pentolite cast boosters the added material is finely divided PETN. Commercial cast boosters are available in a variety of sizes from about 10800 g as shown in Fig. 31.12.

Fig. 31.12 Some commercial cast boosters.

This very brief history relates the development in the commercial explosives industry of an explosive loading and initiation system that emphasized safety. An entire pattern of boreholes can now be loaded with an insensitive blasting agent primed with cast boosters on detonating cord down lines. The down lines can then later be tied to surface detonating cord line, and the entire blast initiated with just one blasting cap after the blast pattern has been completely cleared of personnel. Delay elements were also developed that could be placed between holes to control the borehole firing sequence for maximum blast movement and rock fragmentation.

In 1967 Per-Anders Persson ot Nitro Nobel AB in Sweden (now part of Dyno Nobel Europe) invented a non-electric initiation system, designated Nonel®43 that has revolutionized this aspect of the explosives Industry. The Nonel system consists of an extruded hollow plastic tube that contains an Internal coating of a mixture of powdered molecular explosive and aluminum. The plastic tube is inserted into and attached to a specially designed detonator or blasting cap. The Nonel tubing can be initiated by a number of starter devices, one of which uses a shotgun shell primer. The explosive/aluminum mixture explodes down the inside of the tube at about 2000 m/sec and initiates the blasting cap. The tubing is about 3 mm outside diameter and 1 mm inside diameter, and the explosive core load is only about 18 mg/m, not even enough to rupture the tubing. The Nonel product is not susceptible to the hazard associated with electric blasting caps wherein premature initiation by extraneous electric sources can occur. Figure 31.13 is a photograph of both an electric blasting cap with the two electrical wires and a typical Nonel unit with the plastic tubing.

Fig. 31.13 An electric blasting cap showing electrical wires, typical Nonel unit with plastic tubing.



The 1990s saw a large increase in the use of Nonel products around the world to replace both electric blasting caps and detonating cord down lines. It has long been known that detonating cord down lines disrupt and partially react with blasting agents causing some degree of energy loss. Also, the use of surface detonating cords to initiate blasts can lead to noise complaints. As a result, long-lead Nonels were developed to replace the detonating cord in boreholes. As delay elements were perfected for the Nonel blasting caps, their application and use grew even further and especially in underground mining where a large percentage of blasting caps are used.

Detonator manufacturers are now perfecting initiation systems using electronic blasting caps containing programmable delay circuitry and remote initiation features. The manufacturing cost of these units is currently relatively high, but as the science progresses these electronic detonators will likely be a wave of the future.

PART II. ROCKET PROPELLANTS

A rocket is a device that uses the expulsion of internally generated gases as a source of motive power. The gases used for propelling the rocket are generated by chemical reaction of a fuel and an oxidizer. The force that acts against a rocket as gases are expelled is called thrust. Because rockets carry their own fuel and oxidizer and do not rely on air, the thrust from reaction (combustion) of the propellant chemicals will act in a vacuum. Thus, rockets, unlike internal combustion engines, are capable of providing power in space as well as in the earth's atmosphere.

The use of rockets has been traced to 13th century China, but it was not until the development of the liquid-fueled V-2 in Germany during World War II that a practical long-range missile using rocket propulsion was achieved. Work during the early 20th century by such pioneers as Konstantin Tsiolkovsky in Russia, Robert Goddard in the United States, and Hermann Oberth in Germany provided the basis for the successful German effort and the spectacular space exploration studies that followed. The launching into orbit of the *Sputnik* satellite from the Soviet Union in 1957 was the initial event in a huge expansion of rocket development efforts in recent years. These developments have resulted in rockets used for three principal applications:

1. space exploration and satellite launching
2. strategic missiles
3. tactical missiles

Space exploration efforts have been very well publicized in recent years. These efforts have included such notable developments, in addition to Sputnik, as the launching of manned rockets (with the first astronauts Yuri Gagarin in the Soviet Union and Alan Shepard in the United Sates), the Apollo missions to the moon (with Neil Armstrong's momentous first step), the Russian, American, and International space stations, the U.S. space shuttle program, and the exploration of the solar system by such spacecraft as the Russian Venera and the U.S. Pioneer, Mariner, and Voyager. Perhaps less well publicized, but of great commercial and strategic importance, has been the launching of satellites for purposes of communication, mapping, and surveillance. Launch vehicles for the U.S. space program have included the Atlas Agena, Delta, Juno, Saturn, Scout, Thor, and Titan rockets.

The space exploration efforts were paralleled in the United States and the Soviet Union by the development of rocket-powered missiles for strategic military use. Such U.S. systems as the Air Force Minuteman, Peacekeeper, and Small ICBM and the Navy submarine-launched Polaris and Trident are widely deployed.

The use of missiles for tactical military applications has also been an area of major development since World War II. Among the first such applications were the JATO (rocket assisted takeoff) units used to provide power to boost launching of airplanes. Tactical missiles have become an important component of weaponry and include U.S. rockets such as the Navy Sidewinder and Standard Missile, the Army Hawk and Hellfire, and the Air Force Sparrow, AMRAAM, and Phoenix.

31.13 Principles of Rocket Propulsion

The flight of rockets is based on the thrust achieved by expelling gases from the aft end of the missile; this provides a forward impetus. A schematic diagram of a simple rocket is shown in Fig. 31.14. Combustion of the propellant causes pressurization of the chamber by hot gases; the pressure from the gases is counterbalanced by the strength of the chamber. At a narrow opening, the throat, gases are allowed to escape, providing thrust. If there were no expansion cones, and gases were expelled at the throat, the force, *F*, acting to propel the rocket would be:

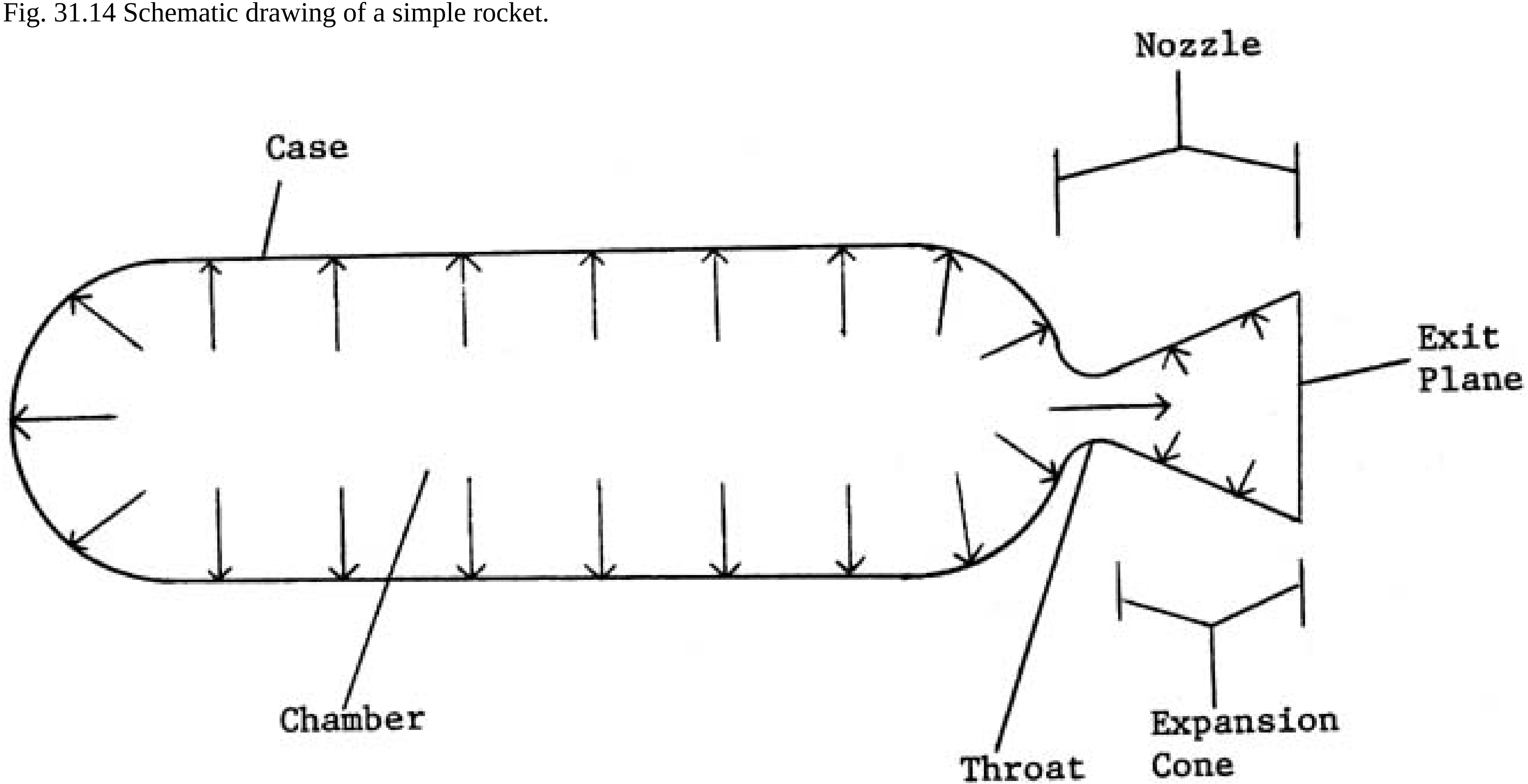
$$F = A_t P_c$$

where *A*t is the area of the throat and *P*c is the chamber pressure. When an expansion cone is present, a new term called the thrust coefficient, *C*f, enters the equation:

$$F = A_t P_c C_f$$

The value of *C*f depends on the ratio of the chamber pressure to the pressure at the exit plane, and on the ratio of the throat area to the exit plane area. The optimum performance of a rocket results when the pressure at the exit plane, *P*e, is equal to the pressure of the surrounding atmosphere (which is one atmosphere for firings at sea level and zero atmospheres in space).

Fig. 31.14 Schematic drawing of a simple rocket.



Because thrust is dependent on motor design and the rate of propellant combustion, it is not a convenient measure of propellant effectiveness. A parameter that is used to compare effectiveness is the specific impulse, *I*sp, which is equivalent to the force divided by the mass flow rate of the propellant:

$$I_{\mathrm{sp}} = \frac{Ft}{W} = \int \frac{F}{\dot{w}}$$

where $\dot{w}$ is the weight flow rate of the propellant, *W* is the total weight of the propellant, and *t* is the time. Because the impulse is dependent on a variety of parameters, it is customary to use the standard specific impulse, *I*sps0, which is the value of the specific impulse for an ideal rocket motor fired at 1000 psi, exhausting to 14.7 psi, with no heat loss, and with a nozzle of 00 half-angle. Frequently, measured or delivered impulse, *I*spd values from motor firings will be converted to *I*sps0 for comparison with previous firings and with expectations. The ratio of delivered to predicted impulse is termed the efficiency. In engineering units, specific impulse is given in (pounds-force * seconds)/pound-mass. A thorough yet succinct discussion of the physics and thermodynamics of rocket propulsion is found in Sutton.48

The prediction of rocket propellant specific impulse, as well as impulse under other conditions, may be reliably accomplished by calculation using as input the chemical composition, the heat of formation, and the density of the component propellant chemicals. Not only impulse but also the composition of exhaust species (and of species in the combustion chamber and the throat) may be calculated if the thermodynamic properties of the chemical species involved are known or can be estimated. The present standard computer code for such calculations is that described by Gordon and McBride.49 Theoretical performance predictions using such programs are widely used to guide propellant formulation efforts and to predict rocket propellant performance; however, verification of actual performance is necessary.

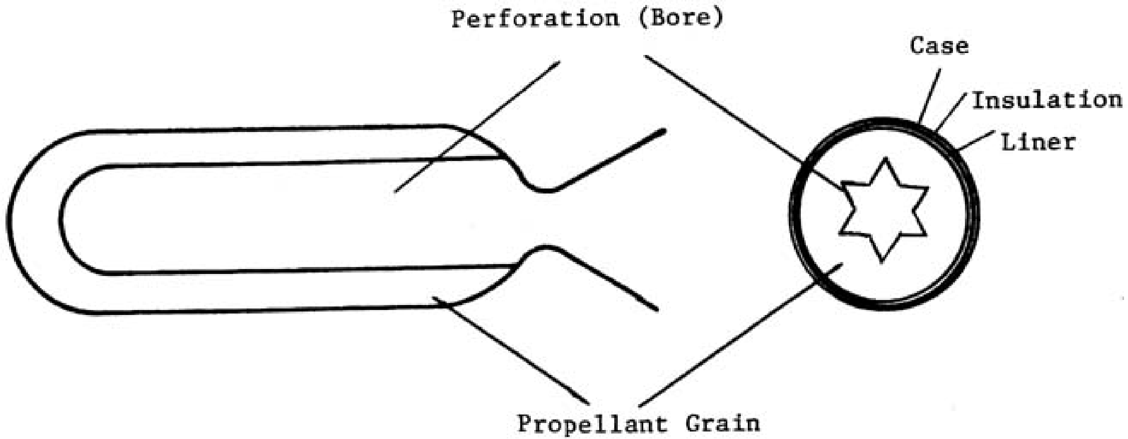
31.14 Types of Propellants

The two principal types of rocket propellant in general use are solid propellants and liquid propellants. Solid propellants are chemical compositions that burn on exposed surfaces to produce gases for rocket power. Liquid propellants rely on pumping or pressurized flow of stored liquids to the combustion chamber. The choice between solid and liquid propellants for a specific application depends on a variety of considerations; to date, many of the strategic missiles and most of the tactical missiles rely on solid propellants because of the lower cost of the rocket and the greater storability of the propellant. On the other hand, large space vehicles and rockets fired for maneuvering in space use liquid propellants, in part because of the ready controllability of liquid systems. Less widely used are hybrid rockets, which use solid fuel and liquid or gas-phase oxidizer. If a single chemical compound (e.g., nitromethane) containing both oxidizing and reducing functions in the same molecule is employed to power a rocket, it is called a monopropellant. If two chemicals combine to provide the propulsion, they form a bipropellant system.

31.15 Solid Propellants

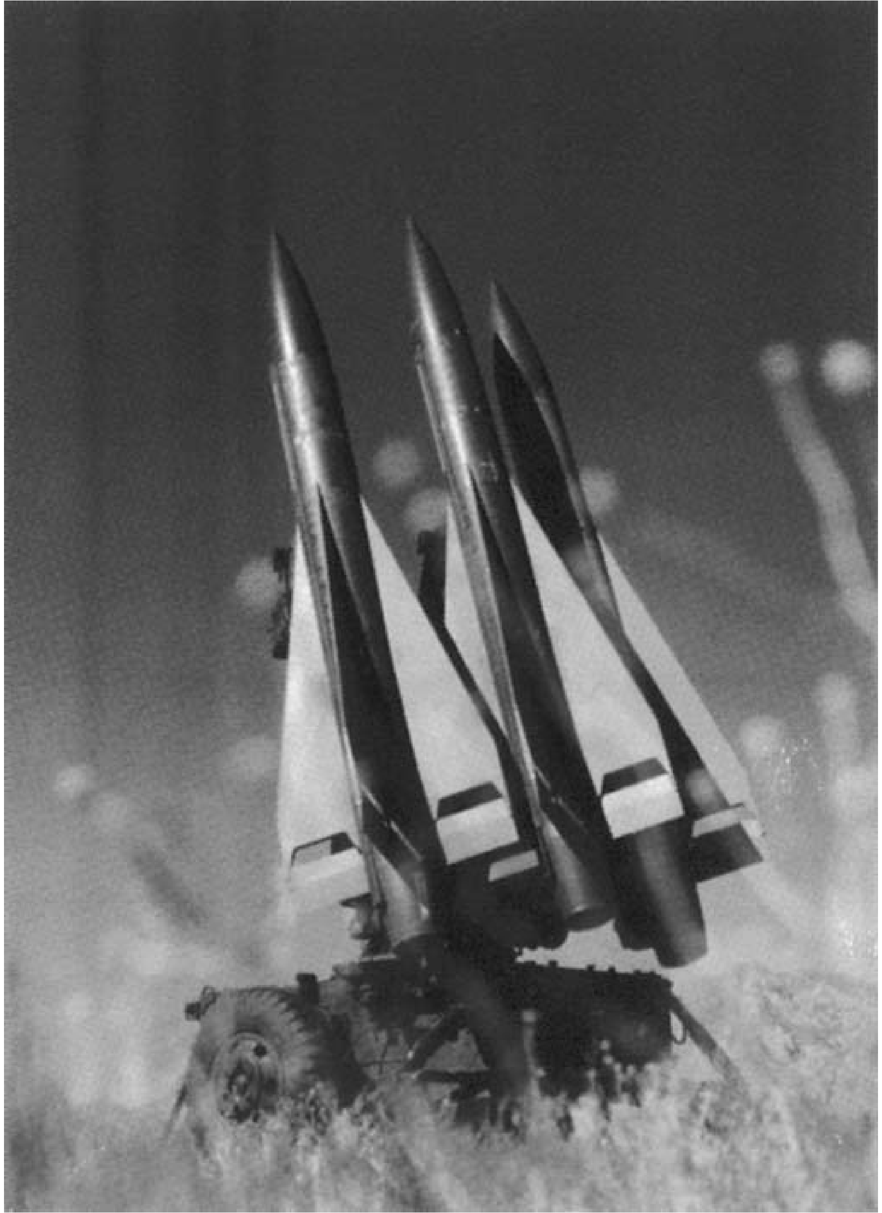
A solid propellant rocket motor is quite simple in concept, although in practice a complete motor is more complex. As shown in Fig. 31.15, the rocket propellant is contained within a case, which may be metal or a reinforced high-performance composite. Frequently, the case is internally shielded by a bonded layer of insulation. The insulation is coated with a liner that bonds the propellant to the insulation. The integrity of the propellant-to-liner bond is of utmost importance; failure at this interface during a motor firing can result in a sudden increase in the area of propellant surface exposed to combustion, with potentially catastrophic results.

Fig. 31.15 A solid propellant rocket motor.



The bore or perforation of the propellant grain is a major factor in determining the ballistic performance of a rocket. In the simplest instance, the grain has no perforation, and the burning is restricted to the end of the grain; the resulting end-burning rockets have a relatively long burning duration with low thrust. More commonly, a perforation extends through the grain (center-perforated) and may have a cylindrical, star, cross-shaped, wagon-wheel, or other more complex profile. The configuration of the grain is used to control the burning behavior of the propellant; the more surface area there is exposed, the more rapidly propellant will be consumed, affecting performance over time. The length of time a rocket motor will burn is governed not only by the perforation geometry, but also by the web thickness of the propellant (distance from perforation to liner), the burning rate of the particular rocket propellant, and the throat diameter of the nozzle. The pressuretime curve resulting from motor burning may be neutral (a single pressure is achieved and maintained throughout the burn), progressive, or regressive. Progressive burning leads to acceleration, whereas regressive burning gives lower pressure as the firing progresses. Tactical solid propellant motors frequently are manufactured with two types of propellant: a rapid-burning boost propellant to provide initial acceleration and a slower-burning sustain propellant to complete the desired flight profile. The trajectory of missiles as a function of propellant properties, grain configuration, and missile design may be reliably predicted or simulated by sophisticated computer calculations. A typical tactical solid propellant-based motor is shown in Fig. 31.16.

Fig. 31.16 Solid propellant-based HAWK tactical missile.



Single and Double-Base Propellants

Early solid rocket propellants were based on processing methods similar to that used in the rubber industry, with propellants extruded into the desired grain configuration. Propellants that have been successfully manufactured by this technique include nitrocellulose (single-base) and nitrocellulose-nitratoester (double-base) materials.* Double-base propellants contain both nitrocellulose and NG as the principal components; additionally, chemicals such as stabilizers, plasticizers, and burning rate modifiers may be added as appropriate. Other nitratoesters also may be in double-base systems. Extruded propellants usually are limited to small grain diameters (≪12 in.) by the size of the equipment required for extrusion, or by the difficulties of solvent removal if a solvent-based process is employed.

*Triple-base propellants have also been produced, having double-base composition plus nitroguanidine, added as a combustion coolant or as a ballistic modifier.

A processing advantage is achieved with the castable double-base systems; a rocket chamber is filled with particulate nitrocellulose (casting powder), which then is treated with NG-containing lacquers or a mixture of nitratoesters. The nitrocellulose is swollen by the nitratoester to give the final propellant, a tough material with relatively low elasticity.

Double-base propellants may be formulated to include fuels such as aluminum metal, oxidizers such as ammonium perchlorate (AP), or energetic materials such as the high energy-density nitramines RDX or HMX. The resulting compositions are termed *compositemodified double-base (CMDB) propellants*. A further modification, the addition of a polymer that is curable with a low molecular weight curing agent, allows formulation of propellants with much improved mechanical properties over the temperature range of usage. Such propellants are termed elastomer-modified composite double-base (EMCDB) propellants, and are currently among the most energetic and highest-performance propellant formulations.

Composite Propellants

In recent years, the great majority of solid rockets have utilized composite propellants. Most composite propellants are based on a solid oxidizer and a curable liquid polymeric binder. The binder also serves as fuel. Optionally, metallic fuels such as aluminum or boron may also be used.50 The propellant components are mixed together, and then the binder is cured to give the tough, flexible, elastomeric (rubbery thermoset) solid propellant required for modern missile use. An excellent source of information on the formulation of solid propellants is a report by Oberth.51

Oxidizer

The major component, by weight and volume, of composite solid propellants is the oxidizer. By far, the most important oxidizer used is AP, a crystalline solid material ground to exacting particle size distributions. This chemical possesses the desirable properties of high density, good thermal stability, and oxygen availability, and relatively low reactivity and cost. Properties of AP and several other materials that are used as oxidizers are summarized in Table 31.5.

TABLE 31.5 Properties of Oxidizers Used in Solid Propellants

Material	Formula	Molecular Weight	Density g/cm^3	Heat of Formation ΔH°_f at 298 K, cal/100 g	g Atom Oxygen/100 g
AP	NH_4ClO_4	117.4	1.95	-60.21	3.404
AN	$N_2H_4O_3$	80.0	1.725	-109.12	3.748
NaN	$NaNO_3$	85.0	2.26	-131.23	3.530
KP	$KClO_4$	138.6	2.53	-74.49	2.887
RDX	$C_3H_6N_6O_6$	222.1	1.82	+6.61	2.701
HMX	$C_4H_8N_8O_8$	296.2	1.90	+6.05	2.701

AN has been considered as an oxidizer for many applications; but its principal use to date is in gas generator propellants, where generation of gases is required to provide initial motor pressurization or to power turbines. Widespread use of AN has been hindered by a propensity of the compound to undergo numerous crystalline phase transitions, some involving a large (4%) volume change of the oxidizer, with concomitant oxidizer particle and binder degradation, upon warming or cooling. Phase-stabilized AN (PSAN), which avoids this difficulty, has been developed but for various reasons has not yet found widespread use. An area of potential application for AN is insensitive minimum smoke propellants, which better meet military criteria of handling and storage safety. Another alternative solid propellant oxygen source, sodium nitrate, has been demonstrated as a cooxidizerscavenger propellants. These propellants contain sufficient alkali metal (sodium) in the formulation to react with the chlorine generated during combustion of the major oxidizer, AP. Sodium chloride, rather than hydrochloric acid, is produced in the exhaust stream as the final chloride-containing reaction product. The benefit of scavenging the chloride ion is lowered exhaust toxicity; however, a substantial loss of propellant impulse results.

The energetic nitramines, RDX and HMX, provide excellent impulse and nonsmoky exhaust, but their use gives propellants that may detonate when subjected to shock or impact. The possibility of unwanted violent burning or detonation of propellants during transport or storage has resulted in recent emphasis on the development of insensitive (reduced hazard) rocket propellants for tactical applications. Special attention are required in using these energetic materials.52

A number of other materials have received attention as potential oxidizers for propellant use, but to date have found little actual use. They include hydroxylammonium nitrate (HAN), hydroxylammonium perchlorate (HAP), hydrazinium nitrate (HN), hydrazinium perchlorate (HP), and hexanitroisowurtzitane (CL-20).

Metallic fuels

In rocket applications where exhaust smoke is not a major concern, the use of metallic fuels adds considerable impulse to the composition. By far, the most common metal in use as a solid propellant fuel is finely divided aluminum, because of a combination of several desirable properties:

1. low equivalent weight
2. high heat of formation of its oxide
3. low reactivity
4. relatively high density
5. low volatility
6. low cost

7. low exhaust product toxicity

An interesting comparison of some properties of metals with respect to their use as solid propellant fuels is available.53

Although theoretical considerations indicate that boron and beryllium might be preferred to aluminum, practical considerations dictate otherwise. Some of the theoretical advantage of boron is lost because of the volatility of the oxide, and because boron is oxidized to a mixture of oxidation states, not cleanly to the trivalent oxide. The use of beryllium is, in general, not possible today because of the high toxicity of the metal and its exhaust products. Aluminum is preferred to magnesium because of its lower equivalent weight and reactivity; aluminum metal powder normally has a thin oxide coating that diminishes its reactivity until combustion temperatures are reached.

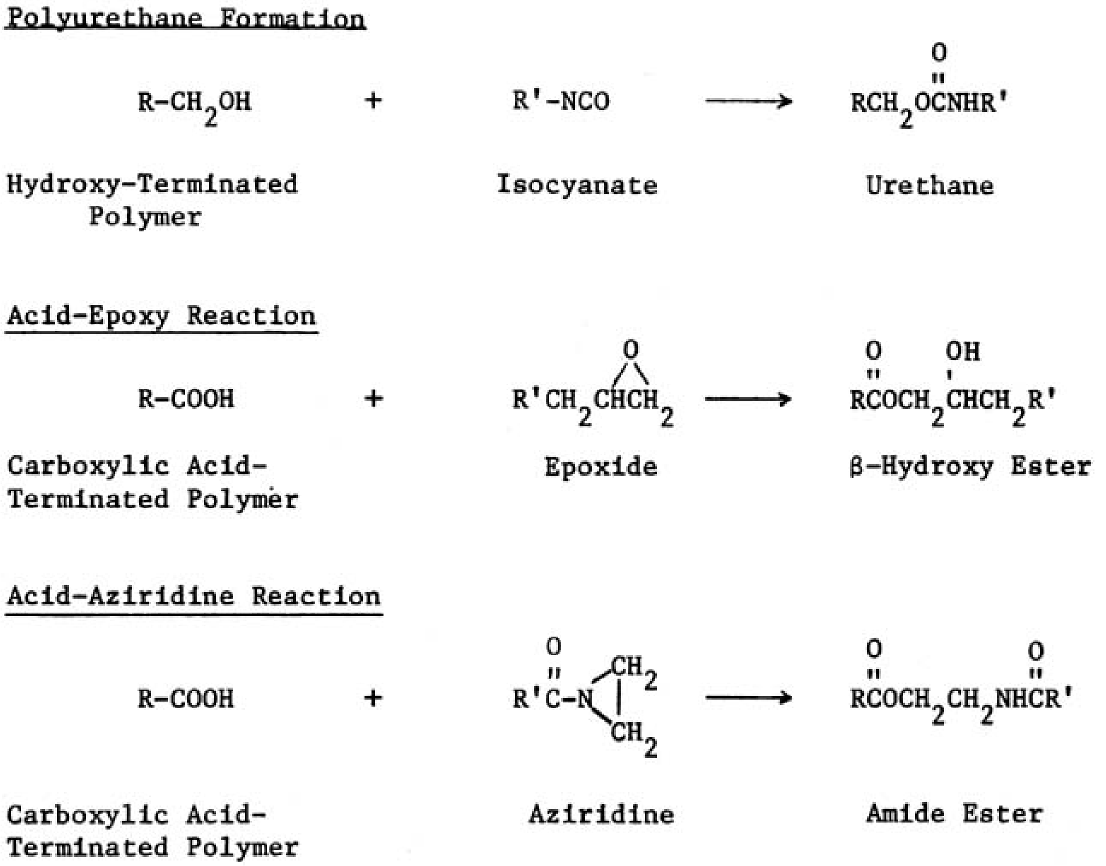
Binder

The binder of a composite solid propellant serves the dual function of providing a matrix to hold the oxidizer and the metal fuel, and of serving as a fuel itselfalthough its total makeup in modern formulations may only be 810 wt percent. The binder of the propellant usually is considered to consist of the polymer, the curing agent, and the plasticizerand can arguably include soluble stabilizers, as well.

In early years, natural rubber, asphalt, polysulfide-based organic polymers (Thiokol rubber), and acrylate polymers were employed as binders, but polymers based on polybutadiene, polypropylene, polyethylene, or polyesters have largely supplanted them. The two most important types of prepolymers used in present propellants are those terminated with cure-reactive carboxyl or with hydroxyl functional groups. These functional groups are used to react with curing agents (cross-linkers), as shown in Fig. 31.17, to provide the high molecular weight polymer networks that function as propellant binders. In general, the lower molecular weight prepolymer introduced into the propellant formulation is di- or polyfunctional so that the polymer resulting from reaction with a curing agent has a degree of cross-linking sufficient to lend a desirable degree of rigidity to the flexible propellant. The most common prepolymers in recent propellant use are the hydroxy-terminated polybutadienes; in the previous generation, *carboxy*-terminated polybutadienes were employed. For example, the binder used in the space shuttle solid booster propellant is based on PBAN, a carboxy-terminated terpolymer of butadiene, acrylic acid, and acrylonitrile. Hydroxy-terminated polyethers also are used, particularly in high-performance propellants. Hydroxy-terminated prepolymers, which are cured with linear or branched isocyanates to give polyurethanes, offer advantages over carboxylic acid-terminated prepolymers:

1. lower mix viscosity
2. faster, lower temperature cure
3. lower susceptibility to side reactions
4. higher oxidizer and metal fuel solids loading

Fig. 31.17 Curing reactions used in present propellants. Each reactant is di- or polyfunctional, so that high molecular weight polymers are formed as the propellant binder.



New prepolymers based on 3,3-bis (azidomethyl)oxetane (BAMO) and 3-nitratomethyl-3-methyloxetane (NMMO) are used in advanced pintle-controllable solid rocket motor applications. These polymers yield favorable propellant energy with the combination of nitrato esters and ammonium nitrate, and allow solid propellant designs to compete with liquid propellants in arenas of energy management, approaching true startstoprestart operation.

Plasticizer

In general, propellant formulations include plasticizers, which are nonreactive liquid diluents used to improve processing and mechanical properties (particularly the low temperature properties) of the propellant. Plasticizers such as high-boiling esters (e.g., dioctyl adipate) or low molecular weight isobutene oligomers are frequently used. In energetic formulations, nitratoester plasticizer blends, such as NG with BTTN, DEGDN, EGDN, and others, are used not only to improve processing and low temperature properties but also to improve impulse and to serve as oxidizers. In such formulations, high plasticizer/polymer weight ratios (231) are frequently used. The nitratoester plasticizers are not miscible with the butadiene-based polymers, hence use of telechelic hydroxy-prepolymers based on ethylene-oxide or propylene-oxide repeat units are required. Because of this, high solids loadings much above 7580 wt percent are not practical. By comparison, propellants based on (poly) butadiene prepolymers and their appropriate plasticizers can achieve solids loadings of 90 wt percent or higher, while simultaneously retaining high mechanical integrity for over 20 years' service life expectation, a wondrous achievement!

Other propellant chemicals

In addition to the binder, oxidizer, and fuel, a solid propellant may have a variety of other chemicals added (usually in small amounts) for specific purposes. These include:

1. aging stabilizers or sequestrants
2. processing aids
3. bonding agents
4. cross-linking agents
5. burning rate modifiers
6. signature-modifying agents
7. cure catalysts
8. cure catalyst scavengers
9. combustion stability enhancers

The final propellant composition is a result of the interaction of a considerable number of chemicals, each of which is important and is selected either for one characteristic or for several reasons. The ultimate purpose of the formulation is to give a propellant whose properties are reproducible from batch to batch and from motor to motor, and are adequate for the intended use as shown, for example, in Fig. 31.18.

Fig. 31.18 The solid booster rockets for the space shuttle are one of the most widely publicized applications of solid rocket propellants.

Classification of Propellents based on Exhaust Properties

Based on the exhaust properties, solid propellants can be classified as smoky, reduced smoke, minimum smoke, or minimum signature propellants. The description of these categories are listed as follows:

Smoky

Propellants containing metals (such as aluminum) give exhaust with particulate matter (such as aluminum oxide) which is visible in the exhaust stream as smoke. Solid exhaust products such as aluminum oxide are called primary smoke. Smoky propellants formulated to reduce hydrogen chloride (HCl) emissions to less than one percent of the exhaust gas mixture are termed clean propellants.

Reduced smoke

Propellants without metals or primary smoke, but containing oxidizers such as AP which gives HCl gas as a principal combustion product, are called reduced smoke propellants. If HCl is exhausted in atmospheres of high or moderate humidity, water droplets will coalesce about the HCl molecules, resulting in a visible exhaust trail of what is called secondary smoke. In atmospheres of low humidity, the exhaust plume of reduced smoke propellants is not visible.

Minimum smoke

Propellants with no metals, and having exhaust free of nucleating species such as HCl, are termed minimum smoke propellants.

Minimum signature

Propellants whose exhaust characteristics are tailored to give not only minimum smoke properties, but also to have low visible, ultraviolet, or infrared emissions are termed minimum signature propellants. Minimum signature propellants are of interest from the standpoints of launch site and missile detectability and from considerations of through-plume guidance.

Propellant Use Criteria

To function properly in its intended use, a propellant must satisfy a large number of criteria, as discussed in the following paragraphs.

Performance

The composition must have adequate specific impulse and volumetric impulse to perform its mission. Volumetric impulse is the product of impulse *and* density (or density raised to a fractional power).

Mechanical properties

The important propellant properties include the tensile strength, strain capability (elongation), modulus of elasticity, and strain endurance. The propellant is formulated so that it will be sufficiently flexible to withstand the stresses of acceleration and temperature changes without cracking, yet be sufficiently rigid so as not to slump or deform upon standing or undergoing temperature changes. Bonding agents, which improve the interaction between polymer and filler (oxidizer), are frequently employed to improve mechanical properties.

Bond properties

The strength of the propellant-to-liner-to-insulation bond must be sufficient to maintain its integrity under the stresses mentioned above.

Ballistic properties

Important parameters in this regard are the burning rate, the pressure dependence of the burning rate, and the temperature dependence of the burning rate. The burning rate is adjusted using the oxidizer particle size and combustion-modifying additives as variables. With AP oxidizer, finer oxidizer particles give faster burning rates. Finely divided iron oxide is a catalyst frequently employed to accelerate the burning rate of AP propellants. In general, low sensitivity of the burning rate to changes in pressure or temperature is mandatory. Burning rates and ballistic properties are measured in progressively larger motor firings as development of a propellant proceeds.

Combustion stability

Although burning without acoustic oscillations is partly a motor design concern, the propellant may be modified by the addition of refractory particles to dampen such vibrations. If uncorrected, pressure oscillations from combustion instability may be large enough to destroy a motor during firing.

Aging and service life

Propellants must be storable for at least as long as the intended service life of the missile system without undergoing degradation or change of ballistic or mechanical properties. Usually the service life is estimated by extrapolation from properties measured for sample aged at elevated temperatures. This accelerated aging is presumed to speed processes that would take place at lower (storage) temperatures. Stabilizers selected to enhance the aging capability are usual components of propellant compositions; the stabilizer choice depends on the polymer, plasticizer, and oxidizer types.

Processibility and castability

In order to be introduced into a rocket motor, a castable propellant composition must be blended until all components are evenly dispersed. The resulting composition must be sufficiently fluid that it may be cast into the motor without creating voids or bubbles in the propellant and yet be sufficiently viscous that dense particles (oxidizer, aluminum) do not settle, or less dense materials (polymer) rise to the surface. Certain chemicals can significantly improve the processibility of propellants when added in small amounts.

Potlife and cure

In addition to achieving a castable viscosity, it is necessary that the liquid (uncured) propellant remain fluid for a sufficient amount of time to be transported, cast, and cleaned up, with allowance for possible delays, before it solidifies appreciably. Following cast, the composition must be cured to a solid, preferably at temperatures close to the intended storage temperature. The balancing of potlife and cure rate frequently requires careful adjustment of cure catalyst levels, catalyst scavengers, mix temperature, and cure temperature.

Hazard properties

It must be verified that the propellant is sufficiently insensitive to shock, electrostatic discharge, friction, thermal decomposition, or self-heating (in larger quantities) that it does not represent an unwarranted hazard in its intended use. Rocket propellants are energetic compositions and must be formulated so that chance stimulus will not initiate violent reaction.

Ignitability

Conditions for ignition in the desired application must be defined, and the propellant formulated so that it may be reliably ignited under these conditions.

All of the above factors must be carefully studied and optimized before a solid propellant can be considered adequate for its intended use.

Composite Propellant Manufacture

In a typical batch processing sequence for a modern polyurethane-cure composite propellant, a submix is prepared first. The submix contains the liquid prepolymer, plasticizers, liquid or powder-form stabilizers, and usually a (liquid) bonding agent; it is slurried with aluminum powder to give a premix. The premix is added to a moderate-shear slurry mixer, and AP oxidizer is added in several portions with intervening mixing which may include heating rate profiles and interim vacuum degas steps. Following addition of the oxidizer, the composition is mixed under high vacuum for a defined period. Vacuum is released, and the curing agent and cure catalyst(s) are added. After a final vacuum mix, the propellant is ready for casting. Frequently, the propellant is cast into an evacuated motor; the use of vacuum increases the casting rate and lessens the possibility of air entrapment and possible void formation in the cured grain. The cast motor then is placed in a cure oven (typically held at a temperature in the range of 110160°F) until the propellant slurry has cured to the desired hardness. At many points in the process, samples may be taken and analyzed to ensure motor quality and integrity.

Although most solid propellants are manufactured in a vertical mixer batch process, a continuous mixing process has been used successfully in the production of the first stage A-3 Polaris propellant and in the NASA 260 inch demonstration motor program. The use of a continuous mixing process, in which propellant chemicals are metered into a helical kneader, offers considerable benefit in terms of safety and cost for large-volume propellant production.

Liquid Propellants

Energetic materials which support rocket propulsion via chemical and thermodynamic changes in *engines*, as opposed to rocket *motors*, distinguish the liquid propellant mission versus applications of solid propellants. Liquid propellants encompass liquid-phase materials in the unreacted state, and are controlled by pressure- or pump-fed liquid rocket engine (LRE) components. Most commonly, liquid propellant rocket systems derive their propulsive energy from the combustion of a liquid fuel and a liquid oxidizer in a combustion chamber. Additionally, new liquid fuel or oxidizer blends, fuel and oxidizer gels, plus solid/liquid propellant hybrid systems are of increasing interest when combined with advances in high strength, low weight, high temperature materials, fast actuation components, or reduced toxicity requirements.

Liquid propellants can be categorized by their type of storage (cryogenic propellants vs. storable propellants) or by their function in the chemical propulsion system (oxidizers, fuels, bipropellants, or monopropellants). In common, is their physical stateusually liquid phasefrom tankage to the injector within the combustion chamber. In addition, selected liquid propellants, either oxidizer, fuel, or both may be gelled as a neat material or as a heterogeneous gel mixture. These gelled propellants may contain suspended solid material such as metal fuel powder, together with polymeric or other particulate gel additives, for rheology tailoring or performance enhancement, usually with military applications in mind.

Chemical *bipropellants* include conventional fuel + oxidizer LRE designs that use either *cryogenic* (less than -150°C/-238°F) propellant fluids or *storable* fuels and oxidizers, the latter not requiring extensive facility cooling or refrigeration launch support. Typical *bipropellants* include LO2/LH2 (cryogenic) or dinitrogen tetroxide/hydrazine (storable). *Monopropellants* may be classified separately as either fuels or oxidizers. Their decomposition via heterogeneous catalysts (such as iridium on alumina support in the case of hydrazine monopropellant, or silver-plated catalyst screens for hydrogen peroxide decomposition) either provides the propulsive thrust or functions as gas generators.

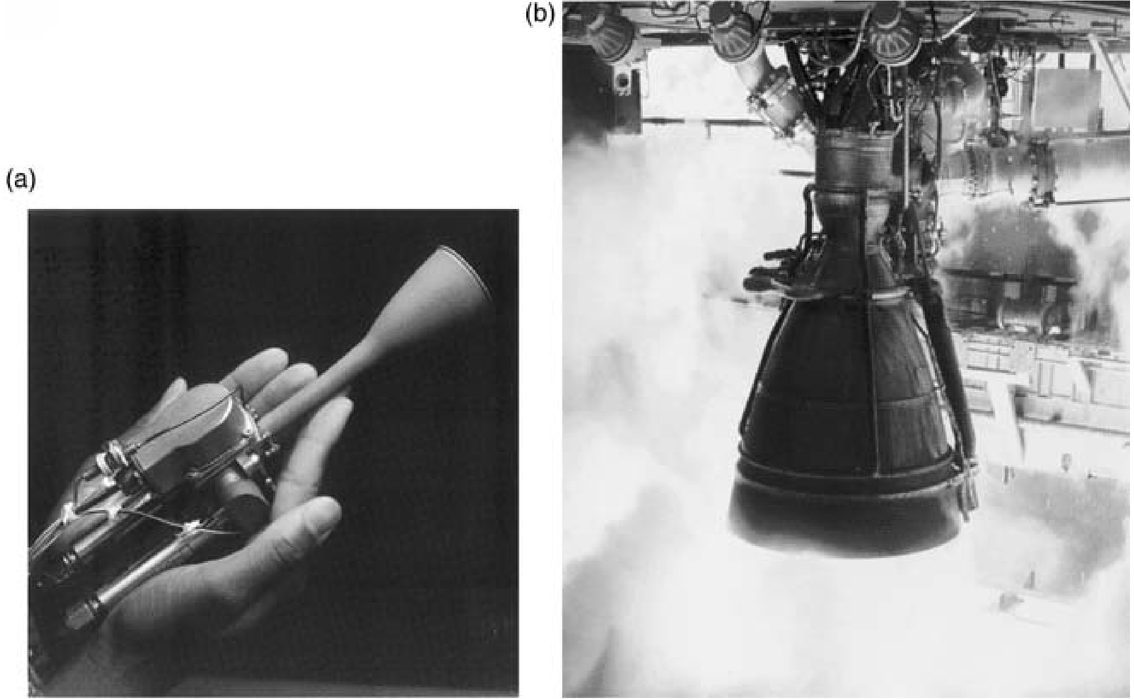
Gel propellants can be considered as a separate category of liquid propellant technology. They provide a unique application of solid propellant processing techniques and materials to the liquid propellant LRE designs, which maximize on some of the advantages of both liquid and solid propellant systems. Gel propellants are thixotropic (shear thinning) liquids by nature of various gelling additives, either polymeric in nature, or as high surface area powders, or both, and flow readily under pressure. A key requirement of gel bipropellants is *rheology matching* the viscosity vs. shear rate and flow dependencies of both the fuel and oxidizer gel, such that optimum combustion requirements are met over the entire temperature range of operationno easy feat. Of benefit in gel systems are their improved safety in handling, for spills of either gelled fuels or oxidizers are easily contained. Advanced applications in gel propellants utilize energetic, insensitive, nanoparticulate gellants and fillers together with reduced toxicity fuel or oxidizer liquids. The application of these combined technologies expands the capability and performance of many conventional currently deployed solid- and liquid-propellant military and civil applications.

Of current intense study is the category of "green", or environmentally benign, propellantsmost notably liquid hydrocarbon fuels with liquid/gaseous oxygen or highly concentrated hydrogen peroxidein studies in this country and abroad. The economic, environmental, and toxicological advantages abound when considering long-term effects of manufacturing, transporting, storing, deploying, or disposing highly toxic storable propellants based on hydrazine and nitric acid chemistry. Increased use of high performance "green" propellant alternatives can only benefit the efficient utilization of space, while at the same time improving the quality of land, water, and air resources in the ever-shrinking global community.

Applications

To date, the liquid propellant systems used in chemical propulsion range from a small trajectory control thruster with only 0.2 lbf (0.89 N) thrust for orbital station-keeping to large booster rocket engines with over 1.0 million lbf (4.44 MN) thrust. Bipropellant propulsion systems are the most extensively used type today for applications in main combustion chambers and gas generators as shown in Fig. 31.19. The monopropellant propulsion system is widely used in low temperature gas generators and auxiliary rockets for trajectory or orbital adjustment.

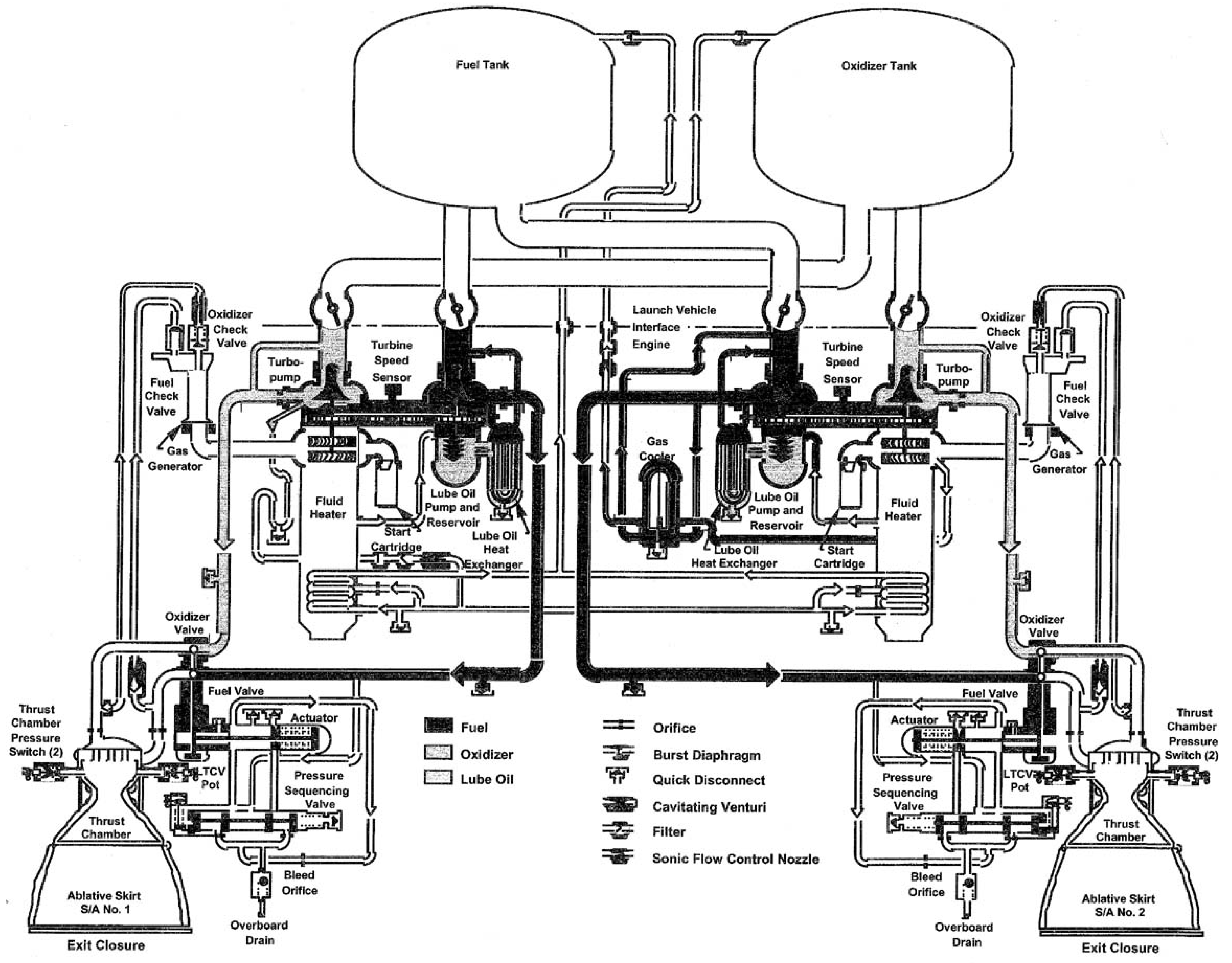
Fig. 31.19 (a) Small thruster used in Milstar and (b) high thrust-to-weight liquid oxygen/kerosene engines for commercial launch vehicles. (*Courtesy Aerojet Propulsion*)



A major difference between liquid propellants and solid propellants used in chemical propulsion systems is the ease of use or controllability. The solid propellants are cast as a solid propellant grain. The burning rate is dependent on the propellant formulation and the configuration design of the solid propellant grain in addition to the chamber pressure and grain temperature. The combustion process is continuous, and a quench and reignition combustion process may be difficult. For the liquid bipropellant system, the liquid fuel and the liquid oxidizer are stored in separate tanks and fed separately to the combustion chamber. The propellants are fed either by means of pumps or by pressurization with an inert gas. A controller generally is used to control the flow rate of the liquid propellants in the system. Ignition and reignition combustion generally is employed to fit the mission requirements and objectives.

Small-orifice injectors are used to atomize and mix the liquid propellants in appropriate proportions. The propellants enter the thrust chamber through the injection manifold and burn inside the thrust chamber. A typical liquid bipropellant rocket engine is shown in Fig. 31.20.

Fig. 31.20 Engine schematic for the Titan IV Stage I (LR87-AJ-II). (*Courtesy Aerojet Propulsion*)



Physical Properties

General physical properties, including freezing point, normal boiling point, critical temperature, critical pressure, specific gravity, heat of formation, and heat of vaporization, of some conventional or promising liquid propellants are listed in Table 31.6. Of the temperature-dependent physical properties, such as heat capacity, thermal conductivity, viscosity, and specific gravity, only the specific gravity is included in Table 31.6. Generally, cryogenic propellants are listed at their normal boiling point, whereas the storable propellants are evaluated at 68°F (293 K). Detailed information on the physical properties of the liquid propellants can be found in Vander Wall et al.54 To obtain a wide operating range and a large payload capacity, the desired physical properties are:

1. low freezing point
2. low temperature variability
3. low vapor pressure
4. high specific gravity
5. high heats of formation and vaporization

TABLE 31.6 Properties of Common Liquid Propellants

Propellant	Molecular Weight	Freezing Point (°F)	Boiling Point (°F)	Critical Temp. (°F)	Critical Press. (psia)	Specific Gravity	Heat of Formation (cal/mole at 298.16 K)	Heat of Vapor. (BTU/lb: NBP)
LO2	32.0	-362	-297	-182.0	730.6	1.14a	-2,896	91.62
F2	38.0	-365	-307	-201	808.5	1.50a	-3,056	71.5
N_2O_4	92.011	11.75	70.4	316.8	1,441.3	1.45b	-4.7	178.2
ClF_5	130.445	-153.4	7.3	289.4	771	1.795b	-60,500	76.04
H_2O_2	34.016	31.2	302.4	855	3,146	1.38b	-44,750	76.04
H_2	2.016	-434.8	-423.3	-399.9	188	0.071a	-1,895	195.3
N_2H_4	32.045	34.75	237.6	716	2,131	1.008b	12,054	583
MMH	46.072	-62.3	189.8	594	1,195	0.879b	13,106	377
UDMH	60.099	-70.94	144.18	482	867	0.785b	12,339	250.6
A-50	41.802	22.0	158	633	1,731	0.905b	12,310	346.5
(50% $N_2H_4$50% UDMH)								
RP-1(H/C = 2.0)	172.0	-55	422	758	340	0.807b	-6,222	125
H_2O	18.016	32	212	705.4	3,206.2	1.0b	-68,317	970.3

aEvaluated at NBP; bEvaluated at 68°F (293.4 K).

Because liquid propellants may also be used to cool the thrust chamber assembly, good heat transfer properties, such as high heat of vaporization, high thermal conductivity, high specific heat, and high boiling point, are desirable.

Liquid Oxidizers

Liquid oxidizers can be categorized as either storable or cryogenic. Many different types of liquid oxidizers have been used in chemical propulsion systems. In general, cryogenics such as liquid oxygen, fluorine, or fluorinated compounds give a high specific impulse. Several storable oxidizers such as nitrogen tetroxide (N_2O_4), inhibited red fuming nitric acid (IRFNA), or chlorine pentafluoride (ClF_5) have been used for their advantages in storage. A brief description of the commonly used oxidizers is given below.

Cryogenics

Liquid oxygen, the most important and extensively used liquid oxidizer, is used primarily with liquid hydrogen to give a very high specific impulse (usually near or over 400 lbf-sec/lbm). Major applications of the liquid oxygen and liquid hydrogen bipropellant system include the space shuttle main engine and the Saturn second stage engine (J-2). Liquid oxygen is also widely used with hydrocarbon fuels in the booster engine of heavy lift launch vehicles such as the Russian Energia. Although liquid oxygen can be used with storable fuels, such as hydrazine or monomethylhydrazine (MMH), this bipropellant combination engine is still in the development stage, mainly because of the difficulty of chamber cooling and combustion stability. Both the freezing point (-362°F, 54.5 K) and the boiling point (-297°F, 90 K) of liquid oxygen are low, permitting a wide range of operation. Liquid oxygen is highly reactive with most organic materials in a rapidly pressurized confinement region of rocket combustion chambers. Except for the relatively high evaporation rate, the handling and storage problems for liquid oxygen are minimal. Although liquid oxygen is not considered corrosive and toxic, the surfaces that will be in contact with the liquid oxygen must be kept extremely free of any contamination because of its reactivity. The low boiling point also can cause problems due to low temperature embrittlement. In order to minimize oxidation problems, metals such as copper are used. Storage tanks and transfer lines of liquid oxygen systems must be well insulated to prevent the condensation of moisture or air with subsequent ice formation on the outside. Vacuum jackets, formed plastics, and alternate layers of aluminum foil and glass-fiber mats have been used successfully.

Liquid fluorine and fluorine compounds (F_2, OF_2, or NF_3) are also cryogenic oxidizers. Although fluorine offers specific impulse and density advantages, extreme toxicity and corrosiveness have prevented the practical application of fluorine and fluorinated compounds in chemical rocket design. In addition, with fluorine the handling problems are significant because fluorine has the highest electronegativity, hence reactivity, of any of the elementsplus a very low boiling point (85 K), necessitating extensive facility support requirements.

Storable

Nitrogen tetroxide and IRFNA are the most important and most extensively used storable liquid oxidizers. A high density yellow-brown liquid, nitrogen tetroxide is very stable at room temperature, existing as an equilibrium with NO_2 ($N_2O_4 \leftarrow \leftarrow 2NO_2$), with the degree of dissociation varying directly with temperature and indirectly with pressure. At atmospheric pressure and room temperature N_2O_4 contains approximately 15 wt. % NO_2, but at 303°F (423 K) it is essentially completely dissociated into nitrogen dioxide. Upon cooling, nitrogen tetroxide dimer is re-formed. It is used as the oxidizer for the Titan first and second stage rocket engines, Delta second stage rocket engine, for the orbital maneuvering engines of the space shuttle, plus finds use in divert and attitude control (DACS) military and civil applications.

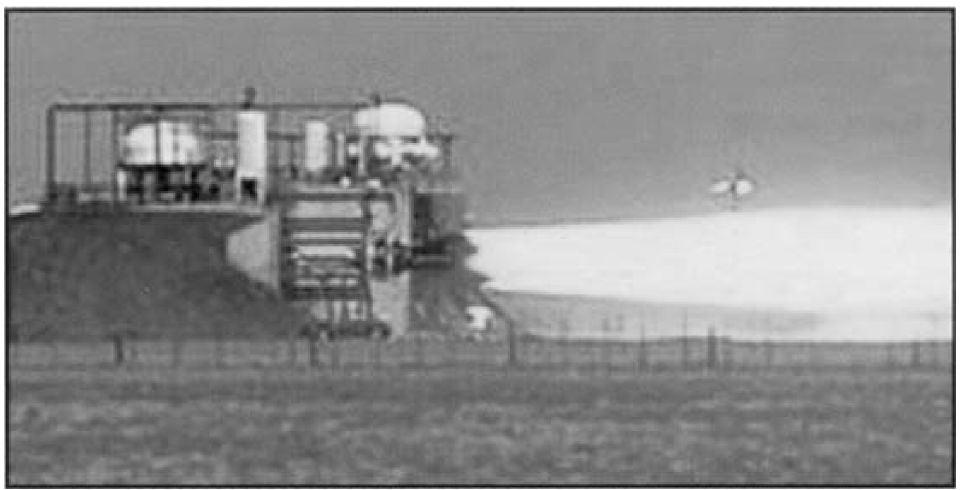
The freezing temperature of nitrogen tetroxide (11.75°F, 262 K) is relatively high, so care must be taken to avoid freezing (causing flow blockage) whenever it is used. For this reason, nitric oxide (NO) may be added (applications with as much as 30 wt. % NO, as mixed oxides of nitrogen (MON), have been noted) to the nitrogen tetroxide to depress the freezing point, improving space-storability of this oxidizer. Additionally, with 13 wt. % NO, storability in titanium tanks is greatly improved, due to reduced stress corrosion over time. Nitrogen tetroxide itself is not corrosive if pure, that is, if the moisture content is very low. Carbon steels, aluminum, stainless steel, nickel, and Inconel can be used with it. However, when the moisture content increases above about 0.2 wt. %, the nitrogen tetroxide becomes increasingly corrosive, and 300-series stainless steel should be used.

IRFNA consists of concentrated nitric acid (HNO_3) that contains dissolved nitrogen dioxide and a small amount of fluoride ion, as hydrofluric acid (HF).

The addition of HF provides the fluoride ion, which reacts with the metal containers, forming a metal fluoride coating and reducing the corrosiveness significantly. As with nitrogen tetroxide, IRFNA is mainly used with hydrazine-type fuels in bipropellant systems. Because nitric acid ignites spontaneously with aniline and amines (hypergolicity), care must be taken in handling and storing the IRFNA. However, this is an advantage in systems where extinguishment and reignition are desired. MON and high density acid (HDA) are variants of storable oxidizers based on dinitrogen tetroxide. The MON propellants consist of N_2O_4 and NO, as noted earlier, and HDA is a mixture of nitric acid and N_2O_4. Both HDA and MON behave similarly to IRFNA.

Other liquid oxidizers, such as concentrated (usually >70 wt. %) hydrogen peroxide (H_2O_2), chlorine trifluoride (ClF_3), and ClF_5 are used as rocket propellants for special applications. ClF_3 and ClF_5 can be highly corrosive to metals. Other interhalogen compounds such as ClO_3F, ClF_3O, or BrF_5 can also be considered as alternatives. Rocket-grade hydrogen peroxide, recently available in quantity at concentrations over 98 wt. % via the anthroquinone process, is gaining widespread respect as a "green" propellant option. However, hydrogen peroxide has serious shortcomings in its incompatibility with a wide range of possible contaminants, which have resulted in catastrophic failure of lines and tanks in extended storage scenarios. Booster applications have demonstrated, in static test firings, capabilities for high-performance heavy lift launch vehicles using H_2O_2/kerosene propellants as shown in Fig. 31.21.

Fig. 31.21 A 200-f (60-m) long flame shoots out from the Beal Aerospace second stage engine during a test firing March 4, 2000. The engine is supposed to produce 810,000 lb of vacuum thrust, using hydrogen peroxide and kerosene propellants. (*Photo courtesy of Beal Aerospace.*)



Liquid Fuels

Cryogenic

Liquid hydrogen, the most important and widely used liquid fuel, is used mainly with liquid oxygen to give high performance, as mentioned above. Liquid hydrogen has excellent heat transfer characteristics: high heat of vaporization, high specific heat, and high thermal conductivity. It is a very good choice to cool the thrust chamber assembly when it is used in a regeneratively cooled system. However, the low boiling point (-434.8°F, 14 K) and low density of liquid hydrogen make it difficult to handle and store. The low fuel density means that a very large and bulky fuel tank is required; this is considered a disadvantage when a high ratio of payload to vehicle dry weight is desired. Liquid hydrogen has been found to react with metals, causing embrittlement of such metals as nickel. Copper is the best and most widely used material for liquid hydrogen in rocket engine applications.

As with liquid oxygen, storage tanks and transfer lines of liquid hydrogen systems must be well insulated, to prevent the evaporation of hydrogen or condensation of moisture or air with subsequent ice formation on the outside. Vacuum jackets, formed plastics, or glass-fiber mats mixed with aluminum foil can minimize the problems.

Molecular hydrogen exists in two forms: *ortho*-hydrogen (nuclei of the two atoms spinning in the same direction) and *para*-hydrogen (nuclei of the two atoms spinning in opposite directions). These two forms are in equilibrium with each other, and at room temperature the equilibrium mixture contains 75 percent of the *ortho* form and 25 percent of the *para* form. Upon cooling to the normal boiling point (-425°F, 20.4 K), the equilibrium is shifted. At this temperature, the *ortho* form will convert slowly to the *para* form. The equilibrium concentration of *para*-hydrogen at this temperature is 99.8 percent. (The shift of the orthopara equilibrium produces energy that causes a liquid hydrogen storage problem.) Therefore, *para*-hydrogen generally is used as liquid fuel for rocket engine applications. Modern liquefiers can produce liquid hydrogen that is more than 99 percent *para*-hydrogen.55

Liquid methane (CH_4), another kind of cryogenic fuel used in testing and experiment, generally is used with liquid oxygen in a bipropellant system. It has the advantage of a considerably higher density than that of hydrogen (the specific gravity for methane is 0.4507 at (-258.7°F, 111.7 k) the normal boiling point). To date no operational rocket engine utilizes liquid methane as the fuel; all liquid oxygen/liquid methane engines are still in the development stage.

Storable

Together with liquid hydrocarbons, hydrazine-type fuels are the most important storable liquid propellants. They include hydrazine (N_2H_4), MMH, unsymmetrical dimethylhydrazine (UDMH), Aerozine-50 (50% N_2H_4 and 50% UDMH), and various blends of these fuels with other amine-based components. All hydrazine-type fuels are toxic to some degree, as are their breakdown products in the environment (especially, as in the case of dilute UDMH with nitrates and nitrites, forming carcinogenic and highly water soluble *nitrosodimethylamine*, or *NDMA*). The most notable hydrocarbon storable fuels include kerosene-based liquid propellants (RP-1, JP-8, and others).

A colorless liquid, hydrazine is stable to shock, heat, and cold. The freezing point of hydrazine (34.75°F, 274.9 K) is the highest of commonly used hydrazine-type fuels. Because it starts to decompose at 320°F (433 K) with no catalysts present, it is undesirable for use as the coolant for regenerative cooling of the thrust chamber. Different blends of hydrazine and MMH have been tested to improve heat transfer properties. Hydrazine is generally compatible with stainless steel, nickel, or aluminum. (See Chapter 28 for more information on hydrazine.)

UDMH is neither shock- nor heat-sensitive, and it is a stable liquid even at high temperature. A key advantage is its low freezing point (-71°F). Furthermore, it is compatible with nickel, Monel, and stainless steel. UDMH is often used as rocket propellant mixed with hydrazine in various proportions. A 50/50 mixture with hydrazine, Aerozine-50, is used for the current Titan IV engines.

MMH fuel is generally used with nitrogen tetroxide (N_2O_4) oxidizer in small spacecraft rocket engines such as orbital maneuvering or altitude control engines. Compared to hydrazine, MMH has better shock resistance and better heat transfer properties as a coolant. However, the specific impulse for MMH/N_2O_4 engines is slightly lower than that for hydrazine/N_2O_4 bipropellant engines. Like hydrazine, MMH is compatible with stainless steel, nickel, aluminum, Teflon, and Kel-F.

In general, the hydrazine-type fuels do not have very good heat transfer properties. Therefore, in the latest development of a high-pressure bipropellant system using N_2O_4 and hydrazine-type fuels, the oxidizer N_2O_4 has been used as a regenerative coolant instead of the fuel itself.

Monopropellants

Simplicity and low cost are the major reasons why monopropellant rocket engines are considered for development and deployment. The specific impulse for monopropellant engines generally is much lower than that for bipropellant engines (in the range of 200 lbf-sec/lbm for monopropellant vs. 280400 lbf-sec/lbm for bipropellant). Hydrazine is the most important and widely used monopropellant in small trajectory correction or altitude control rockets. In an effort to lower the freezing point for improved storability, many hydrazine blends have been studied.56

Hydrogen peroxide has been used as a monopropellant, especially in various-concentration solutions with water. It was used as a rocket propellant in the X-15 research aircraft. Current use of rocket-grade hydrogen peroxide, or "high-test" peroxide (HTPgenerally greater than 90% concentration in water) as a monopropellant is in the developmental stages, mostly for station-keeping on orbit, or as an oxidizer-compatible pressurant gas source.57,58 Long-term storability of highly concentrated hydrogen peroxide is still problematic, although recent advances in oriented crystalline polymers and other novel composites show great promise in lightweight yet compatible tankage materials.

Gelled Propellants

As with powdered metallic fuels in solid composite propellants, early interest in gelled liquid propellants focused on gelation as a means of incorporating high-energy solids into liquid propellants to achieve high specific impulse and density. Storable hydrazine-type fuels such as hydrazine and MMH have received most of the attention in gelation development. During the development of Titan IIA*, a stable suspension of aluminum powder in gelled hydrazine (called Alumazine) was developed and tested extensively. Later the needs for such a vehicle disappeared, leaving the gelation technology to languish for lack of other applications.

*Titan I utilized kerosene and LOX as propellant; the Titan II ICBM development required earth-storable propellant selection.

In the early 1980s, an increasing focus on improving the safety and handling characteristics of storable liquid propellants revived the interest in developing gelled liquid propellants. Although gelation technology for both cryogenic and storable liquid propellants has been developed, most of the activities have concentrated on storable propellants, especially in liquid fuels. Comprehensive reviews covering all gelled fuels development activities can be found in references 59 and 60. A brief description of the gelled fuels and oxidizers is given below.

Gelled fuel

Gelled fuels generally have consisted of metallic powders such as aluminum or boron, or carbon black plus other polymeric gelants, suspended in MMH or an MMH-blend fuel. These gels are typically applied to tactical missile applications. The gelled fuels without metal additives have drawn added interest as propellants with minimum or reduced exhaust signatures and relatively high specific impulse.

Gelled fuels have essentially the same toxicities as their ungelled counterparts because gelation does not significantly change their compositions or equilibrium vapor pressure. However, the rates of vaporization are decreased significantly, reducing the toxic exposure hazards. Gelled fuels, such as gelled MMH, are hypergolic but can burn only upon direct contact of the fuel gel with oxidizer; the fire hazards are greatly reduced with respect to both intensity and extent. Compatibility of the gelled fuels generally is the same as those of the ungelled fuels as regards materials of tankage.61 There is evidence, however, that selected fillers or gel loadings in MMH have caused inordinate offgas and pressure increase over time with these tactical fuel gels; ongoing studies are underway to identify and minimize this potential incompatibility in storage.

Gelled oxidizers

Development work on gelled oxidizers was started during the 1960s, but virtually no effort was expended on them during the 1970s. Development work was revived in the 1980s as interest in gels was awakened because of their improved safety and handling characteristics. Gelled oxidizers include nitrogen tetroxide, MON, red fuming nitric acids (RFNA), and IRFNA. The gelling agents used for the oxidizers include powdered lithium nitrate ($LiNO_3$), lithium fluoride (LiF), plastisol nitrocellulose (PNC), and colloidal silica. The early work on these gels can be found in references 62 and 63. For tactical applications with minimum exhaust signatures, gelled IRFNA and ClFs are the primary gel oxidizers.

Like the gelled fuels, the toxicity of gelled oxidizers is not reduced, but the handling and safety hazards of storage are greatly improved. Gelled IRFNA is considered to be compatible with aluminum. Recent efforts have shown that it can be stored for at least as long as the non-gelled IFRNA. Studies at UAH in this country (sponsored by the U.S. Army Missile Command) and University of Nottingham (sponsored by the United Kingdom's Royal Ordnance) have identified mechanisms and additives to increase storability of gelled IRFNA and analogs; their deployment, however, is yet to be seen.

31.16 References

Sudweeks, W. B., "Chemical Explosives," Chapter 30, in Riegel's Handbook of Industrial Chemistry, 9th ed., J. A. Kent (Ed.), Van Nostrand Reinhold, New York, 1992.
Institute of Makers of Explosives, Publication on the Commercial Explosives Industry.
Kramer, Deborah, A., United States Geological Survey, Explosives Statistics and Information, Minerals Yearbook 2000, and the Institute of Makers of Explosives.
Ornellas, J. Phys. Chem., 72, 2390 (1968).
Kaye, S. M., Encyclopedia of Explosives and Related Items, PATR 2700, Vol. 8, pp. 99100, U.S. Army Armament Research and Development Command, 1978.
Engineering Design Handbook: Explosives Series, Army Material Command Pamphlet 706177, AD 764, 340, p. 12, distributed by NTIS, 1971, Jan.
Fedoroff, B. T., and Sheffield, O. E., Encyclopedia of Explosives and Related Items, PATR 2700, Vol. 2, p. B165, Picatinny Arsenal, Dover, NJ, 1962.
Butler, A. R., ChemTech, p. 202 (Apr. 1990).
Van Gelder, A. P., and Schlatter, H., History of the Explosives Industry in America, Arno Press, New York (1972). (Reprint of 1927 edition).
Marshall, A., Explosives, Vol. I, History and Manufacture, P. Blakiston's Son & Co., Philadelphia, PA, 1917.
Gregory, C. E., Explosives for North American Engineers, Trans Tech Publications, Cleveland, OH, 1973.
"Black Powder," Explosives and Pyrotechnics, 16 (7) (1983, July).
Explosives and Rock Blasting, Atlas Powder Co., Dallas, TX, 1987.
Johansson, C. H., and Persson, P. A., Detonics of High Explosives, Academic Press, New York, 1970.
Blasters Handbook, 16th ed., pp. 14 and 109, Du Pont, 1980.
Kintz, G. M. et al., Report of Investigations 4245, U.S. Dept. of Interior, Bureau of Mines (1948, February).
Chem. Mark. Rep., p. 6 (Sept. 19, 1983).
Cook, M. A., The Science of Industrial Explosives, p. 2, IRECO Chemicals, Salt Lake City, UT, 1974.
Egly, R. S., and Neckar, A. E., U.S. Patent 3,161,551, Dec. 15, 1964.
Bluhm, H. F., U.S. Patent 3,447,978, June 3, 1969.
Bower, J. K., et al., I & EC Prod. Res. Develop., p. 326 (Sept. 1980).
Mohan, V. K., and Field, J. E., Combust. Flame, 56, 269 (1984).
Spear, R. J., and Wilson, W. S., J. Energ. Mater., 2, 61 (1984).
Coburn, M. D., et al., Ind. Eng. Chem. Prod. Res. Dev., 25, 68 (1986).
Dobratz, B. M., LA-9732-H, VC-45, Los Alamos National Laboratory, Los Alamos, NM (May 1983).
Cranney, D. H., and Hales, R. H., Proceedings of the Fourteenth Symposium on Explosives and Pyrotechnics, Franklin Research Center, 1990, Feb.
Code of Federal Regulations, Title 49 Transportation, Part 173.114a, p. 272, Oct. 1, 1979 Revision.
Federal Register, 55 (9), 1306 (Jan. 12, 1990).
Tarver, C. M. et al., "Structure/Property Correlations in Primary Explosives," Final Report, 76-2, Stanford Research Institute, Menlo Park, CA (Feb. 4, 1977).
Brunswig, H., Explosivstoffe, p. 17, Barth, Braunschweig, 1909.
Plets, V., Zh. Obshch. Khim, 5, 173 (1953).
The Sheet 22/00-November 8, 2000, Blue Johnson and Associates, Inc.
Cook, M. A., The Science of High Explosives, pp. 178183, American Chemical Society Monograph Series No. 139, 1958.
Ibid., Appendix II.
Cole, R. H., Underwater Explosions, Princeton University Press, Princeton, NJ, 1948.
Hanto, K., Internal Report on Fumes from Emulsion and ANFO, Dyno Nobel Europe, 1998.
Shipp, K. G., J. Org. Chem., 29, 2620 (1964).
U.S. Patent 3,505,413, April 7, 1970.
Kayser, E. G., J. Energ. Mater., 1, 325 (1983).
Gallo, A. E., and Tench, N., J. Hazard. Mater., 9, 5 (1984).
Delistraty, J., and Brandt, H., Propell., Explos., Pyrotech., 7, 113 (1982).
Rizzo, H. F., et al., Propell. Explos., 6, 27 (1981).
Kolb, J. R., and Rizzo, H. F., Propell. Explos., 4, 10 (1979).
Ott, D. G., and Benziger, T. M., J. Energ. Mater., 5, 343 (1987).
Urbanski, T., Chemistry and Technology of Explosives, Vol. II, p. 181, Pergamon Press, New York, 1965.
Military Explosives, Department of the Army Technical Manual TM 9-1300-214, pp. 732, Washington, DC, 1967.
Nonel is a Registered Trademark of Nitro Nobel AB, Gyttorp, Sweden (Part of Dyno Nobel Europe).
Sudweeks, W. B., I & EC Prod. Res. Develop., 24, 432 (1985).
Work Injury & Illness Rates, pp. 6 and 8, National Safety Council, Chicago, IL, 1988.
Sheahan, R. M., and Minchinton, A., The Australian IMM Explosives in Mining Workshop, p. 17, Melbourne, 1988, Nov.
Seltzer, R. J., Chem. Eng. News, p. 7, Aug. 8, 1988.
Sutton, G. P., and Biblarz, O., Rocket Propulsion Elements, 7th ed., John Wiley & Sons, Inc., New York, 2001.
Gordon, S., and McBride, B. J., Computer Program for Calculation of Complex Chemical Equilibrium Compositions. Rocket Performance, Incident and Reflected Shocks, and Chapman-Jouget Detonations, NASA SP-273, Interim Revision, NASA Lewis Research Center, (March 1981).
Kuo, K. K., and Pein, R., "Combustion of Boron-Based Solid Propellants and Solid Fuels," Begell House, ISBN 084939919X, 1993, Jan.
Oberth, A. E., Principles of Solid Propellant Development, CPIA Publication 469, Chemical Propulsion Information Agency, Baltimore, MD, 1987.
D'Andrea, B., and Lillo, F., "Industrial Constraints for Developing Solid Propellants with Energetic Materials," J. Propul. Power, 15 (5), 713718, 1999.
Dekker, A. O., "Solid Propellants," J. Chem. Ed., 37, 597.
Cadwallader, A., Liquid Propellant Manual, Chemical Propulsion Information Agency, 1969.
Von Doehern, J., Propellant Handbook, AFRPL-TR-66-4, Air Force Rocket Propulsion Laboratory, 1966.
Schmidt, E. W., Hydrazine and Its Derivatives, 2nd Ed., Vols. I and II, Wiley-Interscience, New York, 2001.
Wernimont, E. J., and Mullens, P., (General Kinetics, Aliso Viejo, CA), Recent Developments in Hydrogen Peroxide Monopropellant Devices, AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, 35th, Los Angeles, CA, 1999, June 2024.
Murphy, T. H., Coffman, P. E., Jensen, J. J., and Hoffman, C. S. (Boeing Co., Canoga Park, CA), Storable Liquid Propellant UsageTrends in Non-toxic Propulsion, IAF, International Astronautical Congress, 49th, Melbourne, Australia, 1998, Sep. 28Oct. 2.
Vander Wall, E. M., Andersen, R. E., and Cabeal, J. A., "Preparation and Characterization of Gelled Hydrazine Based Fuels Containing Aluminum," Final Report, AFRPL-TR-80-40, Air Force Rocket Propellant Laboratory (1980, July).
Haun, D. V., Ryder, D. D., and Graham, P. H., "Gelled Bipropellants for Advanced Missile Propulsion Systems, I. Fuel Component Initial Characterization," CPIA Publication 425, Vol. VI, pp. 215224, JANNAF Propulsion Meeting, 1985, April.
Thompson, D. M., U.S. Patent 6,210,504; Azidoalkyl-substituted Tertiary Amines as Fuel Components for Liquid- or Gel-based Rocket Propellants, Assigned 1999.
Vander Wall, E. M., Anderson, R. E., Beegle, R. L., Jr., and Cabeal, J. A., "Gelation of Nz04 and MON-25," Final Report, AFRPL-TR-82-49, Air Force Rocket Propulsion Laboratory (Aug. 1982).
Allan, B. D., "A Gelled Oxidizer for Tactical Missiles," CPIA Publication 370, Vol. V, pp. 1120, JANNAF Propulsion Meeting, 1983, Feb.