CHAPTER 18.3
INPUT/OUTPUT

INPUT-OUTPUT EQUIPMENT

Input-output (I/O) equipment includes cathode-ray tubes and other display devices; printers; keyboards; character, handwriting, voice, and other recognition devices; optical scanners and facsimile equipment; speech synthesizers; process control sensors and effectors; various pointing devices such as "mice," wands, joysticks, and touch screens; card readers and punches; paper tapes; magnetic storage devices such as tape cassette units, floppy and hard disk drivers, drums, and various other types of tape drives; and several different types of optical storage devices, mainly disks.

Such equipment presents a wide range of characteristics that must be taken into account at their interface with the central processing unit (CPU) and its associated main memory. A magnetic tape unit usually operates serially by byte, writing or reading information of varying record length. Different types of printers deal with information on a character, line, or page basis, both in a coded (character) and noncoded mode. Most cathode-ray tube display devices build up the image in a bit-by-bit raster, but others handle data in the form of vectors or even directly coded as characters. Different types of I/O gear operate over widely different speed ranges, ranging from a few bytes per second to millions of bytes per second. Many I/O devices do little more than pass bits to a processor, while others have substantial internal processing power of their own.

In the many systems making use of telecommunications, an ever-growing array of I/O equipment can be attached to a variety of telecommunications links, ranging from twisted-pair copper wire to long-haul satellite and optical fiber links. Virtually any I/O device can be attached both locally and remotely to other elements of a system.

Data are entered by human operators, read from storage devices, or collected automatically from sensors or instruments.

Keyboards, wands, and character recognition equipment are typical *input devices*, while printers and displays are the most common *output devices.* Storage devices that can function in a read-write (R/W) mode naturally can fill both functions.

I/O CONFIGURATIONS

Input-output devices can be attached to other elements of the system in a wide variety of ways. When the attachment is *local*, the attachment may be an I/O bus, or a channel. Figures 18.3.1 and 18.3.2 show a typical bus attachment configuration of small systems such as personal computers. A program gains access to the I/O gear through the logic circuitry of the ALU. For example, to transfer information to I/O gear, a program might extract requested information from storage, edit and format it, add the address of the I/O device to the message, and deliver it to the bus, from which the proper I/O device can read it and deliver it to the requestor. In this simple

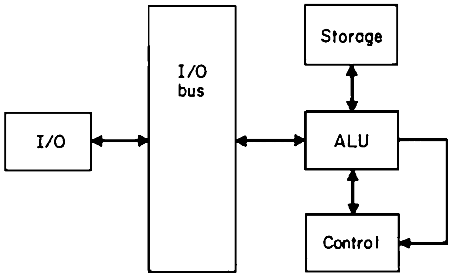

FIGURE 18.3.1 Machine organization with an I/O bus.

configuration, time must be spent by the CPU waiting for delivery of the information to the I/O gear and, if so programmed, for reception of an acknowledgement of the reception. If several types of I/O gear are used, the program or—more commonly—the operating system must take into account variations in formatting control, and speed of each type. A common method of improving the performance of the *I/O subsystem* is I/O buffering. A buffer may be part of the CPU or a segment of main memory, which accumulates information at machine speeds so that subsequent information transfers to the I/O gear can take place in an orderly fashion without holding up the CPU. In other cases, the I/O gear itself will be buffered, thus enabling, for example, synchronous input devices to operate at maximum throughput without having to repeatedly interrupt the CPU. A buffered system is desirable in all but the simplest configurations and is all but essential in any case in which the speed of the I/O device is not matched to the data rate of the receiving communications link or I/O bus.

I/O MEMORY–CHANNEL METHODS

The arrangement shown in Fig 18.3.2 involves information transfer from the I/O gear into the ALU and thence to main store. With a modularized main store, I/O data can be directly entered in, and extracted from, main storage. Direct access to storage implies control interrelationships at the interface between the I/O and the CPU to ensure coordination of accesses between the two units. The basic configuration used for direct access of I/O into storage is shown in Fig. 18.3.2. The connecting unit between the main store and the I/O gear is called a *channel.* A channel is not merely a data path but a logical device that incorporates control circuits to fulfill the relatively complex functions of timing, editing, data preparation, I/O control, and the like.

TERMINAL SYSTEMS

Terminal systems use key-entry devices with a display and/or printer for direct access to a computer on a time-shared basis. The programming system in the CPU is arranged to scan a number of such terminals through a set of ports. Each character, line, or block of data from each terminal is entered into a storage area in the computer until an appropriate action signal is received. Then by means of an internal program the computer generates such responses as error signals, requests for data, or program outputs. Terminal systems extend the power of a large computer to many users and give each user the perception of using an independent system. Terminals remotely connected to host CPUs have the same overall description, except that communication is mediated by means of various network architectures and protocols (q.v.).

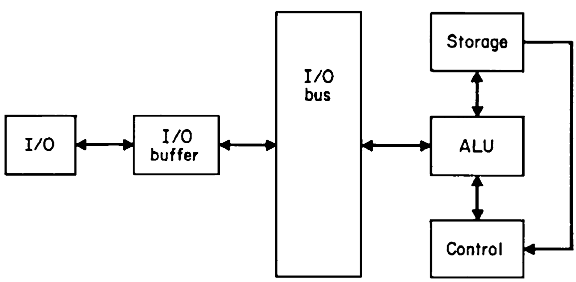

FIGURE 18.3.2 Machine organization with an I/O bus and I/O buffer.

A terminal that contains a high level of logic operational capability is often called an *intelligent terminal.* A personal computer with a communications adapter or modem is increasingly filling this description. If there are many terminals and a centralized computer or computer complex in a system, it is sometimes more economical to use simpler terminals and use the logic in the computer. More commonly, it has been the practice to attach a number of *dumb* terminals to an intelligent control unit, which, in turn, communicates with a host processor. Economic factors such as communication line cost and hardware cost, as well as such intangibles as system reliability and data security, are among those factors which influence the choice.

The proliferation of communicating personal computers has formed the basis of local and wide area networks (LANs and WANs). These networks provide both connectively between sites and local processing power.

PROCESS-CONTROL ENTRY DEVICES

In process control, e.g., chemical plants and petroleum refineries, inputs to the computer generally come from sensors (see Sec. 8) that measure such physical quantities as temperature, pressure, rate of flow, or density. These units operate with a suitable analog-to-digital converter that forms direct inputs to the computer. The CPU may in turn communicate with such devices as valves, heaters, refrigerators, pump, and so forth, on a time-shared basis, to feed the requisite control information back to the process-control system.

MAGNETIC-INK CHARACTER-RECOGNITION EQUIPMENT

A number of systems have been developed to read documents by machine. *Magnetic-ink character recognition* is one. Magnetic characters are deposited on paper or other carrier materials in patterns designed to be recognized by machines and by operators. A change in reluctance, associated with the presence of magnetic ink, is sensed by a magnetic read head. The form of the characters is selected so that each yields a characteristic signature.

OPTICAL SCANNING

Information to be subjected to data processing comes from a variety of sources, and often it is not feasible to use magnetic-ink characters. To avoid the need for retranscription of such information by key-entry methods, devices to read documents optically have been developed.

Character recognition by optical means occurs in a number of sequential steps. First, the characters to be recognized must be located and initial starting points on the appropriate characters found. Unless the characters have been intentionally constrained to be placed in well-defined locations, it is then necessary to segment the characters, that is, to determine where one character leaves off and the next one begins. Third, the characters must be scanned to generate a sequence of bits that represents the character. The resulting bit pattern must then be compared with prestored reference patterns in order to identify the pattern in question.

The scanning can be performed in a number of ways. In one technique the pattern to be recognized is transported past a linear photodetector array and the output is sampled at periodic intervals to determine the bit pattern. A second way is to scan a laser beam across the page using a rotating mirror or holographic mirror device, with the light scattered from the document being detected by an assembly of photodetectors. Other approaches that use various combinations of arrays of sources and detectors with mechanical light deflection devices are also used.

As an alternative to comparing the scanned bit pattern with stored references, a *correlation function* between the input and stored functions can be computed. Optical spatial filtering is occasionally used, too.

Another common method of data entry is through the encoding of characters into a digital format, permitting scan by a wand or, alternatively, by moving the pattern past a "stationary" scanning laser pattern. Typical of this encoding scheme is the bar code, known as the *universal product code* (UPC), which has been standardized in the supermarket industry by industry agreement to the code selection by the Uniform Grocery Product Code Council, Inc.

BATCH-PROCESSING ENTRY

One computer can be used to enter information into another. For example, it is often advantageous for a small computer system to communicate with a larger one. The smaller system may receive a high-level language problem, edit and format it, and then transmit it to a larger system for translation. The larger system, in turn, may generate machine language that is executed at the remote location when it is transmitted back.

Remote systems may receive data, operate on them, and generate local output, with portions transmitted to a second unit for filing or incorporation into summary journals. In other cases, two CPUs may operate generally on data in an independent fashion but may be so interconnected that in the event of failure one system can assume the function of the other. In other cases, systems may be interconnected to share the work load. Computer systems that operate to share CPU functions are called *multiprocessing* systems.

PRINTERS

Much of the output of computers takes the form of printed documents, and a number of *printers* have been developed to produce them. The two basic types of printer are *impact printers*, which use mechanical motion of type slugs or dot-forming hammers driven against a carbon film or ink-loaded fabric ribbon to mark paper, and *nonimpact* printers, which use various physical and chemical phenomena to produce the characters.

Printers are additionally characterized by the sequence in which characters are placed on the page. *Serial* printers print one character at a time, by means of a moving print head. After a line of characters has been printed, the paper is indexed to the next print line. *Line* or *parallel* printers are designed such that they print all the characters on a print line at (nearly) the same time, after which the paper is indexed. Finally, *page* printers are limited by the technology to marking an entire piece of paper before the output is accessible to a user. Most such printers actually generate a page one dot at a time using some type of raster scanning means.

Printers may either be *fully formed character* printers, or *matrix* (or *all-points-addressable*) printers. Matrix printers build up characters and graphics images dot by dot; some matrix-printing mechanisms allow the dots to be placed anywhere on a page, while others are constrained to printing on a fixed grid. Dot size (resolution) ranges from 0.012 in. (0.3 mm) in many wire matrix printers to 0.002 in. (0.05 mm) in some nonimpact printers. Dot spacing (addressability) ranges from as few as 60 dots per inch (24 dots per centimeter) to more than 800 dots per inch.

The speed of serial printers typically ranges from 10 to over 400 characters per second. Line printer speeds range from 200 to about 4000 lines per minute. Page printers print from 4 to several hundred pages per minute.

IMPACT PRINTING TECHNOLOGIES

Figure 18.3.3 is a sketch of a band printer, the most common high speed impact line printer. Band printers are used where price and throughput (price/performance) are the dominant criteria. In this printer, the characters are engraved on a steel band, which is driven continuously past a bank of hammers—usually, 132—at speeds of the order of 500 cm/s. A ribbon is suspended between the band and the paper, behind which hammer bank is located. This is known as *back printing*. When a desired character is located at a specific print position, that hammer is actuated without stopping the band. Several characters may be printed simultaneously, although activating too many hammers at one time must be avoided, since that would slow or stop the band and overload the hammer power supply. Timing marks (and an associated sensor) provide precise positional information for the band.

Figure 18.3.4 is a schematic sketch of a print head for a wire matrix printer. These printers are most commonly used with personal computers. The print head contains nine hammers, each of which is a metal wire. The wires are activated in accordance with information received from a character generator so that a desired character can be printed in an array of dots. The wires strike a ribbon, and the print head is positioned on a carrier. Printing is done while the carrier is in continuous motion, and it is necessary to know the position to an

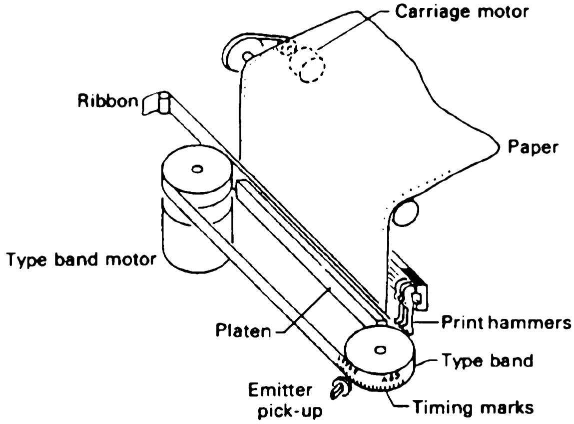

FIGURE 18.3.3 An engraved band print mechanism.

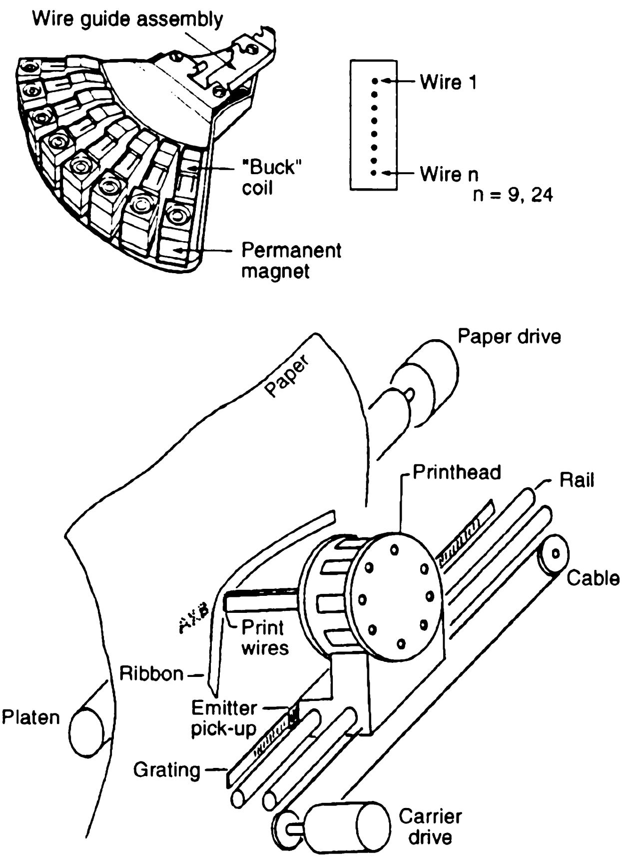

FIGURE 18.3.4 Schematic drawing of wire-matrix print head and print mechanism.

accuracy of a fraction of a dot spacing, that is, to much better than 0.1 mm. To do this, an optical grating is placed on the printer frame, and the motion of the carrier is detected by the combination of a simple light source and photodetector. Other sensing and control schemes are also used. Quite attractive printing can be generated with a wire matrix printer if the dots are more closely spaced. This can be done in the serial printer shown here by printing the vertical rows of dots in an overlapping fashion, and then indexing the page vertically only one-half or one-third the dot spacing in the basic matrix, so that the dots also overlap in the other direction. Many variations of this print mechanism are available. Among them are the use of 18 or 24 wires, and the use of multicolor ribbons to enable color printing.

NONIMPACT PRINTERS

Nonimpact printers use various mechanisms to imprint characters on a page. There are two major classes of such devices, those using *special paper* and those using *ordinary paper.* In one type of special-paper device, the paper is coated with chemicals that form a dark or colored dye when heated. Characters are formed by contact with selective heated wires or resistors that form a matrix character. Another type of special paper, known as *electro-erosion paper*, is coated with a thin aluminum metal film that can be evaporated when subjected to an intense electric spark, obtained from a suitably actuated array of wires.

Another method of nonimpact printing on special paper, called *electrography*, or, more often, *electrostatic printing*, uses a dielectric coated paper that is directly charged by ion generation. In this case an electrostatic latent image is generated by an array of suitably addressed styli. The charged paper can be toned and fixed as in *electrophotography*, also known as *xerography.* In the xerographic process, the image is developed by exposing the latent image to a toner, which may be very small solid powder particles or a suspension of solid particles or a dye in a neutral liquid. The toner particles, suitably charged, adhere either to the charged or the neutral portions of the image. This pattern is then fixed, by heat, pressure, or evaporation of the liquid carrier.

Most plain-paper nonimpact page printers combine laser and electrophotographic techniques to produce high-speed, high-quality computer printout. The printing process uses a light-sensitive photoconductive material (such as selenium, or various organic compounds) wrapped around a rotating drum. The photoconductor is electrically charged and then exposed to light images of alphanumeric or graphic (image) information). These images selectively discharge the photoconductor where there is light and leave it charged where there is no light. A powdered black toner material is then distributed over the photoconductor where it adheres to the unexposed areas and does not adhere to the exposed areas, thus forming a dry-powder image. This image is then electrostatically transferred to the paper where it is fixed by fusing it with heat or pressure. In some printers the toner is suspended in a liquid carrier, and the image is fixed by evaporation of the carrier liquid. A different choice of toner materials allows development of the unexposed rather than the exposed areas. The various elements of this process can all be quite complex and many variations are used. (See Fig. 18.3.5.)

IMAGE FORMATION

Electrophotography, transfer electrophotography, and thermal methods require that optical images of characters be generated. In some types of image generation, the images of the characters to be printed are stored. Other devices use a linear sweep arrangement with an on-off switch. The character image is generated from a digital storage device that supplies bits to piece together images. In the latter system any type of material can be printed, not just the character sets stored; the unit is said to have a *noncoded-information* (NCI) capability or *all-points-addressable* (APA) capability.

Patterns are generated either by scanning a single source of energy, usually a laser in printers and an electron beam in display devices, and selectivity modulating the output, or else by controlling an array of transducers. The array can be the actuators in a wire matrix printer, light-emitting diodes, ink-jet nozzles (see below), an array of ion sources, thermal elements, or magnetic or electrical styli.

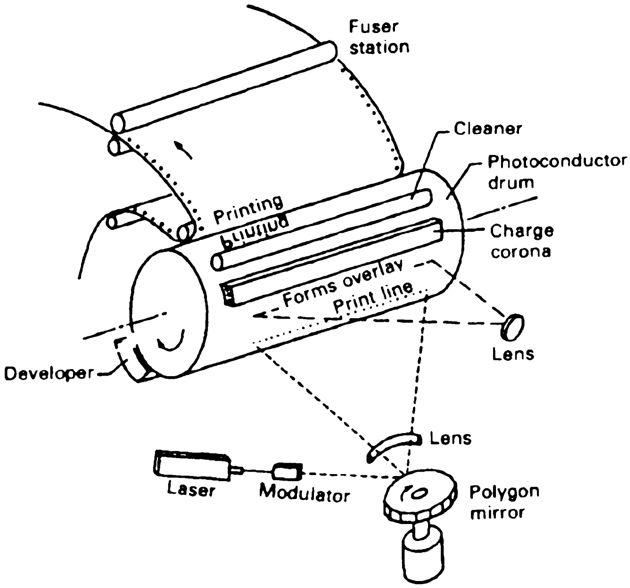

FIGURE 18.3.5 Xerographic printer mechanism.

The output terminals for digital image processing are high-resolution monochrome or color monitors for real-time displays. For off-line processing, a high-resolution printer is required. The inputs to digital image-processing requirements are generally devices such as the vidicon, flying-spot scanner, or color facsimile scanner.

INK JETS

Another method of direct character formation, usually used on untreated paper, employs ink droplets.

An example of this technique is continuous-droplet ink-jet printers, which have been developed for medium- to high-quality serial character printing and is also usable for very high speed line printing. When a stream of ink emerges from a nozzle vibrated at a suitable rate, droplets tend to form in a uniform, serial manner. Figure 18.3.6 shows droplets emerging from a nozzle and being electrostatically charged by induction as they

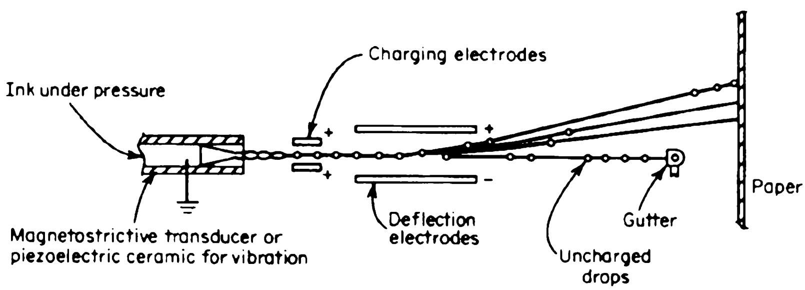

FIGURE 18.3.6 Ink-jet printing. Ink under pressure is emitted from a vibrating nozzle, producing droplets that are charged by a signal applied to the charging electrodes. After charging, each drop is deflected by a fixed field, the amount of deflection depending on the charge previously induced by the charging electrodes.

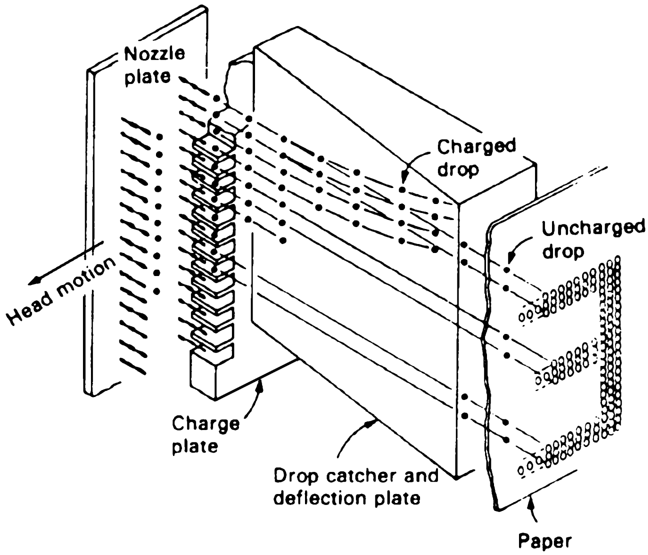

FIGURE 18.3.7 Nozzle-per-spot binary ink-jet printer.

break off from the ink stream. In subsequent flight through an electrostatic field, the droplets are displaced according to the charge they received from the charging electrode. Droplets generated at high rates (of the order of 100,000 droplets per second) are guided in one dimension and deposited upon untreated paper. The second dimension is furnished by moving the nozzle relative to the paper. Since the stream of droplets is continuous, it is necessary to dispose of most of the ink before it reaches the paper. This is usually done by selectively charging only those drops that should reach the paper; uncharged drops are intercepted by a gutter and recirculated.

An array of nozzles can be used in combination with paper displacement to deposit characters. The simplest form of such a printer has a separate nozzle for each dot position extending across the page (Fig. 18.3.7). Uncharged droplets proceed directly to the paper, while unwanted drops are charged such that they are deflected into a gutter for recirculation. Moving the paper past the array of nozzles provides full two-dimensional coverage of the paper. Such a binary nozzle-per-spot printer can print tens of thousands of lines per minute. Ink-jet systems have also been developed using very fine matrices for character generation, producing high document quality.

VISUAL-DISPLAY DEVICES

Visual-display devices associated with a computer system range from console lights that indicate the internal state of the system to *cathode-ray-tube* displays that can be used for interactive problem solving. In the cathode-ray tube, a raster scan, in conjunction with suitable bit storage, generates the output image.

For certain graphics applications, it is preferable to write the vectors making up the image directly, rather than by means of raster deflection. Such a vector mode provides higher-quality output, but can be overcome by refresh limitations when there is a large number of vectors to be displayed. Such displays are usually used with a keyboard for information entry, so that the operator and computer can operate in an interactive mode. Graphic input devices such as mice, joysticks, touch panels, and the like are also often used.

A relatively low cost, low power, and high-performance display is the *liquid-crystal display* (LCD). Liquid crystals differ from most other displays in that they depend on external light sources for visibility; i.e., they employ a *light-valve* principle. This type of display is composed of two parallel glass plates with conductive lines on their inner surfaces and a liquid-crystal compound (e.g., of the *nematic* variety) sandwiched between them. In the dynamic scattering display, the clear organic material becomes opaque and reflective when subjected to an electric field. As in other displays, the characters are built up from segments or dots.

Color LCDs are available. In one version, a dot triad is used, with thin-film color filters deposited directly on the glass. Again, the complexity and cost are such that where low power and the thin-form factors are not critical, the technology has been uncompetitive with CRTs.