CHAPTER 21.3
DIGITAL VIDEO RECORDING SYSTEMS

Peter H. N. de With, M. van der Schaar

INTRODUCTION TO MAGNETIC RECORDING

Historical Development

Recording of video signals has established itself as a continuously growing and indispensable technique in video communication services and the video entertainment area. The currently available video recorders for professional and consumer applications have evolved from the early experiments in magnetic tape recording, which were focused on analog voice and audio recording. In the 50s, a number of companies experimented and succeeded to record video signals for broadcasting purposes, however, using rather diverging technical approaches, as the optimal concept for video recording was still being explored. The discrepancies in the applied techniques were due to a new technical hurdle that had to be taken: the required video bandwidth is roughly 300 times the bandwidth of audio signals. It is therefore not surprising that a number of system proposals were based on extending the traditional longitudinal audio recording technique by using a plurality of longitudinal tracks that are registered simultaneously, and, in order to further cope with the demanding bandwidth, a very high tape speed.

In 1956, Ampex announced a system (Quadruplex, or simply Quad) that enabled a substantially lower tape speed, by applying a rotary drum containing a head wheel that turns around at a much higher speed than the tape is transported. Another technique, which had a major impact on further developments in video recording, was the frequency modulation of the video signal prior to storage. This principle, in which a carrier having a relatively low frequency, e.g., 5 MHz, is frequency modulated by the wideband video signal, is still being applied in the current analog video recorders for registration of the luminance signal. The major drawback of the Quad system was its segmented storage of television fields on tape caused by relatively short tracks, which demanded accurate time-base correction for picture reconstruction. The short tracks resulted from the applied transversal recording technique, in which the tracks were written perpendicular to the tape length, so that several rotations, and thus tracks, were required to cover one complete television field.

The segmentation problem was solved a few years later (1959) by other companies, which realized a helical-scan video tape recorder (VTR) that formed the basis of the currently available VHS format. In the 1960s, emphasis was put on the implementation of consumer video cassette recorders (VCRs)—a logical successor of audio cassette recording—that allowed a much simpler tape loading mechanism. One of the first VCRs for professional use is the U-matic system, introduced by Sony in 1970. Since then, many systems have been realized and passed through the consumer's horizon,[1] among which a VCR by Philips in 1970, called the N-1500 (the first consumer VCR), Betamax by Sony in 1975, the VHS system by JVC in 1976, V-2000 again by Philips in 1979, and 8 mm by a company consortium in 1984.

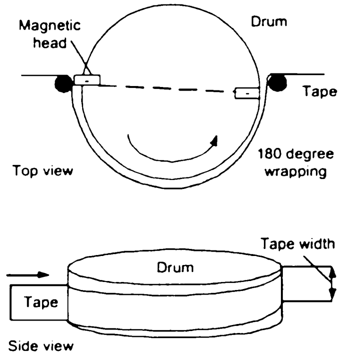

FIGURE 21.3.1 Helical-scan tape recorder with two heads.

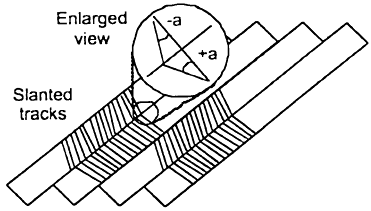

FIGURE 21.3.2 Azimuth recording track pattern and reading with a magnetic head having corresponding azimuth.

As a result of the fast progress, video recording has even started to influence audio recording by means of the introduction in 1987 of the digital R-DAT format, which is fundamentally a miniaturized video recorder.[2] The generation of video recorders from the past two decades is based on helical-scan recording, which is depicted in Fig. 21.3.1. In such a system,[3] the magnetic heads are mounted on a rotary head wheel inside a cylindrical drum, which has the tape wrapped helically around it. The tracks written on tape have a shallow angle with respect to the tape direction and are gradually crossing the tape width, so that a single track can be long enough to store an entire television field.

The continuous drive for higher quality and the demand for more compact mechanics have been satisfied by an increased recording density with each new generation of video recorders. The enormous augmentation in information density is mainly a result of better recording media and heads.[4] The linear-density improvement resulted from better magnetic tapes, offering higher signal-to-noise ratio (SNR) and more high-frequency output, thereby enabling more flux reversals per unit of length. This is expressed by the decrease in the minimum wavelength.[5] The track density improves with the higher SNR and better tracking systems, enabling a further reduction of the track pitch, i.e., the distance between the centers of two adjacent tracks.

Another technique, introduced by Sony in the Betamax system (1975), which significantly contributed to a higher track density, is azimuth recording, in which the tracks are recorded without guardband (the empty space between tracks) and with alternating azimuth (Fig. 21.3.2). Azimuth is the angle between the actual gapline of the magnetic head and the direction perpendicular to the track length. Azimuth recording, which reduces the effects of "sidereading" on adjacent tracks, is now a universally applied technique in consumer video recorders.

The areal recording density improves considerably from higher track densities than from higher linear densities. Although for the consumer the volume density is the most important—a parameter in which magnetic recording clearly outperforms optical disk recording, one of its key parameters, the tape thickness, has not decreased very rapidly. In the future, it is therefore expected that tape manufacturers will work toward tapes with thinner base films, which further contributes to a higher volume density. In recent years, the mass production of optical disks, which was initiated by the success of the compact disc for audio and the CD-ROM for data, has become the leading technology for video recording systems. This has resulted in the DVD recorder and playback-only systems, which are available for various home and personal computer systems.

Analog Consumer Video Recording

From the systems mentioned above, two analog recorders are still in use in the home today: VHS and the 8 mm system. From both standards, camcorders are available as well. The 8 mm cassette is smaller than that of the VHS system, therefore a special compact VHS-C cassette was later introduced for camcorder use only.

The small cassette needs as adaptor for playback in the desktop VHS recorder. With respect to the tape format, the following characteristics apply.

Video. Video fields originating from the odd and even lines of complete pictures (frames) are recorded line by line in individual tracks, thus one field per track. This holds both for most professional systems and the consumer VHS and 8 mm system. This recording allows access to each field individually, for trick modes, such as still picture or fast search (also called picture shuttle).

Audio. VHS has two audio tracks in the direction of the tape travel, while 8 mm uses audio recording inside the slanted tracks.

Mechanics. All home recorders use 180° wrap angle with two video heads diametrically opposed, such as in Fig. 21.3.3. The heads are mounted on a wheel inside a rotating drum. The scan speed, i.e., the actual speed of reading or writing over the tape surface varies between 25 m/s for professional recorders and about 4 m/s for consumer recorders. The tape travel speed is only a fraction of this. Furthermore, the systems are usually cassette based. Some professional recorders use an omega wrap that covers nearly a full circle.

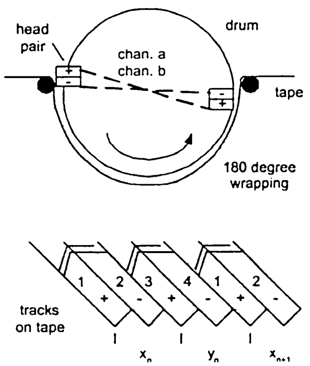

FIGURE 21.3.3 Headwheel with two head pairs and 180° wrap angle used in consumer recorders.

Recording density. The consumer recorders apply relatively narrow tracks of, say, 20 μm. The consumer recorders apply azimuth recording for high recording density, whereas professional recorders use guard bands between two succeeding slanted tracks. The tracks in professional systems are less narrow.

Tape format. Figure 21.3.4 shows a typical footprint of a tape format. VHS uses longitudinal tracks for the audio signal. The video tracks are slanted. The 8 mm system uses an audio slanted track with an edit gap between the video and the audio part, thereby enabling separate editing of video and audio. The control track at the bottom allows the insertion of markers and time code on tape for programming and fast searching to a specific point.The place and location of control and audio tracks may vary from standard to standard, especially for professional analog video recorders such as the B- and C-type recorders and the Betacam and M-machines (see, e.g., Ref. 6).

Comparison of VHS and 8 mm Systems

Table 21.3.1 depicts the key parameters of VHS and 8 mm systems. The numbers of the Betamax system have been added for comparison. Table 21.3.1 shows that Betamax and VHS are comparable and are nearly equal systems. The picture quality of Betamax was slightly better, because the drum was larger and more tape was consumed, allowing a broader recording frequency band. For home recorders, VHS finally won the battle because more prerecorded movies were available in the market, and it offered a longer playing time. The 8 mm system was introduced later and offers a higher picture quality than VHS, merely as a result of using newer, thinner tapes with an improved magnetic layer. Consequently, the system size could be reduced (drum, cassette), combined

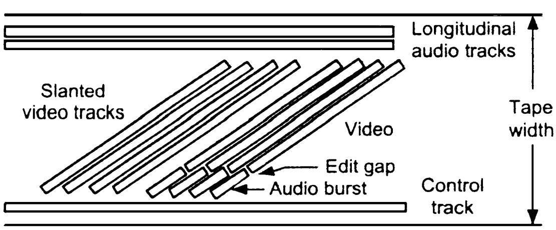

FIGURE 21.3.4 Tape format as used in consumer recorders.

TABLE 21.3.1 Key Parameters of Betamax, VHS, and 8 mm Helical-Scan Cassette Recorders

	Betamax	VHS	8 mm
Tape width	12.65 mm (0.5 in.)	12.65 mm (0.5 in.)	8 mm
Tape thickness	20/16.5 μm	20/16.5 μm	13 (10 Hi-8) μm
Tape material		CoγFe oxide	Metal particle
Head drum diameter	75 mm	62 mm	40 mm
Tape wrap angle	180°	180°	180°
Head-tape speed	7.0 m/s	5.8 m/s	3.76 m/s
Linear tape speed	4.0/2.0/1.33 cm/s	3.34/1.67/1.12 cm/s	1.43 cm/s
Video track width	58.6/29.2/19.5 μm	58.6/29.2/19.5 μm	20.5 μm
Audio track width	1.05 mm	1.0 mm	0.65 mm
Track length		97 mm	62.6 mm
Track angle		6°	4° 53′
Azimuth angle		+/−6°	+/−10°
Video modulation	FM composite	FM composite	FM composite
Color modulation	color under	color under	color under
FM carrier, sync tip	3.5 MHz	3.4 MHz	4.2 MHz
P.White in spectrum	4.8 MHz	4.4 MHz	5.4 MHz
		7 MHz (SVHS)	7.7 MHz (Hi-8)
Chrominance carrier	688 kHz	629 kHz	743 kHz
Cassette size	156 × 96 × 25 mm	162 × 104 × 25 mm	95 × 62.5 × 15 mm
Playing time	60/120/180/240 min	120/240/360/480 min	90 min

with a smaller track width. The carrier frequencies were raised to a higher level. Because of its smaller size, the 8 mm camcorder products became an almost immediate success and they have gradually pushed VHS camcorders out of the consumer market.

Nowadays, the VHS and VHS-C camcorders are being used less often, and camcorders are typically based on either the analog 8 mm format or the digital DV system discussed later. The success of the DV system has also resulted in the reintroduction of the 8 mm format, but now based on digital recording using the video compression technique that was developed for the DV system. This product was introduced at the end of the 1990s by Sony and is called “Digital 8.” Camcorders are basically video tape recorders in small form with a lens, optical pickup device based on a CCD sensor (resolution 250,000 pixels or higher), a viewfinder using a color liquid crystal display (LCD), and a microphone for capturing audio signals.

Color Modulation in Analog Recorders

The video signal recording on VHS and 8 mm is such that the color is modulated with a specific technique, called “color under.” This will be explained now, starting from the limitation of the magnetic recording channel properties. The television signal contains about 5 MHz bandwidth. The magnetic recording channel is bandlimited, that is, it cannot record either zero frequency or very high frequencies. Analog professional and consumer recorders therefore modulate the signal to a bandpass signal. The modulation is based on frequency modulation (FM).

Professional analog recorders such as the type C machines apply direct recording, in which the frequency carrier is modulated by the composite video signal. The composite video signal contains the baseband luminance (Y) signal with a 5-MHz spectrum onto which the chrominance or color (C) signal is added with amplitude modulation using a specific carrier frequency. For PAL this is 4.43 MHz and for NTSC it is 3.58 MHz. With direct recording, the sync tip (lowest voltage) of the video signal has a carrier frequency of 7.06 MHz, the blanking signal (zero video) is modulated with 7.8 MHz and the peak white signal (full video) with 9.3 and 10 MHz, for PAL/Secam and NTSC, respectively. These numbers refer to the analog type C recorder. This type

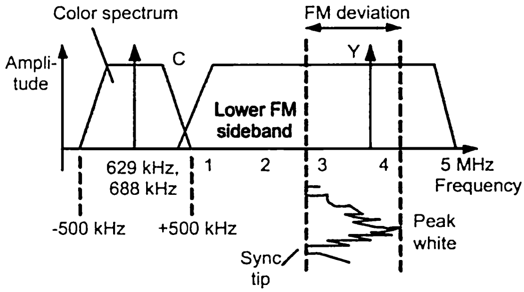

FIGURE 21.3.5 Spectrum of a full color video signal after applying the "color under" system.

of modulation requires a total bandwidth of 15 MHz, which is too expensive for consumer systems. The advantage of direct recording is the excellent quality.

Consumer analog recorders have a different system for adding the color signal to the modulated spectrum. In order to save bandwidth, the luminance signal is modulated with a lower carrier frequency. The color signal is then modulated with a carrier frequency residing at the low-frequency part of the luminance spectrum, so that the color signal is located close to zero frequency and below the luminance signal. This approach is therefore called the "color under" system. The principle is visualized in Fig. 21.3.5 and Fig. 21.3.6. In analog consumer recorders such as VHS, the modulation spectrum looks typically like Fig. 21.3.5. Figure 21.3.5 shows the resulting spectrum with the modulation carrier frequencies and Fig. 21.3.6 depicts an example of an encoder block diagram required for modulation. The required bandwidth for the modulated video signal is well below 10 MHz. However, owing to the bandwidth restrictions, the color bandwidth is in practice limited to about 300 kHz for the VHS system.

Digital Video Recording and the Growth of Bit-Rate Reduction

The advent of digital signal processing techniques and its subsequent rapid development has resulted in a reconsideration of the current concepts for video recording. Digital processing and recording not only outperforms analog recording with respect to time stability, robust picture reconstruction, and (flawless) reproduction after multiple copying, but also allows for advanced image manipulation techniques that were impossible with the conventional analog technology. In the past decade, we have witnessed a transition from analog recording to digital recording systems. As always, at the introduction stage of a new technology where costs are still critical, the first systems are targeted for professional (broadcast) applications. The first full-digital video recorder brought to the market was the D-1,[7] which was introduced around 1985. It was a YUV-component video recorder, capable of recording a CCIR-601 video signal. Soon thereafter, a digital recorder, referred to as D-2, was agreed upon, for composite NTSC[8] and PAL video signals. From 1995 onwards, the gradual introduction of a digital format for consumer applications took place, called the DV system. The DV system established itself as a compact camcorder format, initially for semiprofessional and ENG applications but recently also for the consumer. Principal manufacturers are Sony, JVC, and Matsushita. From a technological viewpoint, a small-sized digital video recorder became feasible and the IEEE-1394 digital interface to computers for off-line digital editing has proven its value.

In addition to the developments in consumer magnetic recording, video recording has further broadened its scope of application by new video (image) storage equipment based on optical disc recording. One of the first systems was the compact disc interactive (CD-i) system of Philips[9] from about 1990, of which the video-coding technique was standardized for multimedia applications MPEG-1,[10,11] electronic photography using magnetic and optical storage equipment, or even solid-state memory. The CD-i was soon succeeded by the

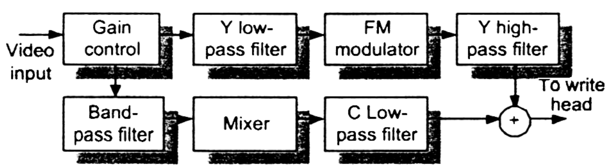

FIGURE 21.3.6 Modulation block diagram for the "color under" system.

video-CD (VCD) format, offering straightforward MPEG-1 compressed video[10] on 1.2 Mb/s on a compact disc. The system is still being used in Asia. Both CD-i and VCD are discussed later in this chapter. In the past years, the optical recording density was improved to a quad-fold of the audio CD, and the new video compression standard MPEG-2[12,13] was used to store video in strongly compressed form. The disk format was standardized as the Digital Versatile Disc (DVD). The high density of the DVD and the compression factor of 25 of MPEG-2 for broadcast video enables recording of a full movie on a single disk. This format has resulted in a number of recordable DVD formats[14] of which DVD − RW and DVD + RW are most known. Other DVD formats are DVD-ROM and DVD-RAM. Details of the above-cited systems are disclosed in the subsequent sections.

The importance of video bit-rate reduction techniques (data compression) cannot be underestimated. The implementation of this technology in various forms combined with the implementation in cost-efficient silicon chips has fueled the realization of many new products in consumer electronics as described above. Video compression has been widely studied, which is proven by the excellent introductory textbooks on digital image processing and associated video coding,[15–17] which are available for the interested reader. More details about the algorithms video data compression for recording can be found in reviews, see e.g. Ref. 18. From this literature, it can be deduced that transform coding[15] has proven its merits for efficient coding of video signals. During the 90s, advanced video compression has become an emerging and enabling technology for new systems in digital transmission and storage. An example of successful video compression supporting digital video broadcasting in the MPEG standard,[10–13] which will be discussed in a succeeding section.

From Composite to YUV-Component Video

Except for the video cassette recorders such as VHS and 8 mm systems, which store the video signal in composite form, all modern recording systems are based on YUV- or YCrCb-component recording. Composite recording means that the video signal is basically one signal with the color signal modulated into it around the color subcarrier frequency. With component recording, the input video signal, which is mostly generated by a camera in three color components—red, green, and blue (RGB), is also recorded in three separate signals, mostly YCrCb. The three RGB components are converted to YCrCb components at the input of the recorder (e.g., with camcorders) or the signal is already ordered in YCrCb format (with, e.g., DVD playback). The conversion is a 3 × 3 matrix multiplication with a RGB sample triplet at the input and the new YCrCb triplet at the output. The Y-signal refers to the luminance portion (black and white) of the full-color signal, the UV or CrCb to the color-difference signals, i.e., B-Y and R-Y. The CrCb components are shifted and centered around the value 128 in an 8-bit format between 0 and 255, in accordance with the Y component, whereas U and V signals are centered around zero. Since the color-difference signals contain less energy than the Y-component, they are usually sampled with a lower frequency than the Y sample frequency, e.g., with half or a quarter of the Y sample frequency.

TABLE 21.3.2 Key Parameters of Typical YCrCb Sampling Standards Used in Digital Video Products.

Sampling standard	4:1:1	4:2:0	4:2:2
Sample frequency Y	13.5 MHz	13.5 MHz	13.5 MHz
Y samples/line active	720 (704)	720 (704)	720
Y lines active (50/60)	576/480	576/480	576/480
Cr/Cb sample frequency	3.375 MHz	6.75 MHz	6.75 MHz
C samples/line active	180 (176)	360 (352)	360
C lines active	576/480	288/240	576/480
Active bit rate (8-bit)	124 Mb/s	124 Mb/s	166 Mb/s

Note: When interlacing is used, the number of lines per picture splits into two groups (fields) of odd and even lines.

Table 21.3.2 summarizes the most important properties of common sampling standards. The 4:2:2 sampling standard is mostly found in professional recorders, such as the D-1. The DV format applies 4:2:0 in Europe and 4:1:1 in the United States and Japan. The DVD standards are all based on 4:2:0 sampling in both 50 and 60 Hz countries. Older products such as VCD use 4:2:0, albeit at halved horizontal and vertical resolution of the input signal, also known as SIF or CIF resolution (standard or common intermediate format). The halved resolution at the input is required to obtain the low bit rate on CD. It should be noted that the conversion from RGB to YUV or YCrCb signals reduces the quality of the full-color signal, because the difference signals are sampled at reduced frequencies. The advantage is a smaller input bit rate (see Table 21.3.2) at the expense of some loss of sharpness in color transients.

PROFESSIONAL DIGITAL VIDEO RECORDING

The Lossless D-1, D-2, and D-3 Systems

The gradual introduction of digital signal processing in the studio and an increased use of computer graphics in video signals resulted in the introduction of various digital video tape recorders for professional use. The advantages of digital signal storage in a studio environment are, apart from conventional issues such as increased robustness, that multiple editing generations (copy of copy) are allowed. The deterioration in quality is much less than that with analog recorders. In the early 1990s, both composite (D-2, D-3) and component digital studio recorders (D-1) were used. D-1 is the oldest one (introduced in the mid-1980s), but due to its robustness for editing it is still being used actively. All systems are cassette recorders, based on helical-scan recording and use 8 bits for video sampling.[6,19] The recording mechanics are constructed in such a way that three different types of cassettes can be inserted without adaptors, leading to various playing times. Another common property is that they do not use video compression in order to offer maximum quality for studio applications. For further comparison, see Table 21.3.3.

When uncompressed video is used, the bit rates after sampling with sufficiently high frequency are too high (e.g., about 100 Mb/s at minimum) for using a single recording head. Therefore, a head *pair* is used for recording two tracks simultaneously. Professional recorders sometimes use a plurality of heads to offer high picture quality during special modes such as "still picture," slow motion, and fast forward or backward search. The picture quality of digital VTRs in such modes is further improved by *shuffling* the pixels in a special way prior to recording on tape.[7] The shuffling balances and smoothes the effect on the picture quality of crossing various tracks that inevitably lead to data discontinuities.[18] D-1 can search up to 64 times the normal speed and

TABLE 21.3.3 Key Parameters of Digital Professional Video Recorders (f_{sc} = Color Subcarrier Frequency, 3.58 MHz for NTSC and 4.43 MHz for PAL)

	D-1	D-2	D-3
Video recording type	Component	Composite	Composite
Tape width	19 mm (3/4 in.)	19 mm (3/4 in.)	12.65 mm (0.5 in.)
Tape magnetic layer	Fe oxide	Metal	Metal particle
Tape speed	28.6 cm/s	13.2 cm/s	8.4 cm/s
Track width	40 μm	39 μm	20 μm
Bit length	0.45 μm	0.43 μm	0.38 μm
Recording density	55 kb/mm^2	59 kb/mm^2	130 kb/mm^2
Video sample freq. (Y/C)	13.5/6.75 MHz	$4f_{sc}$	$4f_{sc}$
Video sampling	8 bits	8 bits	8 bits
Audio sampling	20 bits	20 bits	20 bits
Audio recording	4 tracks	4 tracks	4 tracks
Recording bit rate	216 Mb/s	105 Mb/s	105 Mb/s
Playing time	76/34/11 min	208/94/32 min	240/125 min

still show a complete—though noisy—picture. The noise is caused by a picture that is constituted from data blocks of pixels originating from various pictures with different time origin. The data shuffling redistributes the pixels over the image leading to an effect perceived as pixel “noise.” The noise increases with the search speed.

The recording density continuously improves because of the introduction of better magnetic tapes, using special magnetic materials for higher density. The D-3 that was introduced in 1990 by Panasonic uses a higher recording density than the D-1 and D-2 machines. It is based on the small size of Betacam of Sony and a format called M-II of Panasonic, but now designed for digital recording. Another difference is that the D-3 cassettes contain half-inch tapes, whereas D-1 and D-2 use $^3/_4$-in. tapes. The high recording density reduces the robustness against channel errors, such as scratches and dirt, but this is combated by *error correction codes*. The aforementioned recorders apply Reed-Solomon (R-S) block codes for error protection. A group of pixels is mapped into a data block onto which a string of parity check symbols is appended. These codes can reduce a fluctuating Byte Error Rate (BER) in the order of 10^{-4} 10^{-5} to as low as 10^{-10} 10^{-12}. This extra coding step is also an aid in achieving sufficient robustness for multiple editing iterations to, e.g., 100 times.

D-5 Recorder

The D-5 of Panasonic is an extension of the D-3 machine and is backward compatible to D-3. The unique feature of D-5 is the recording of 16:9 aspect ratio television signals, using three video components of 10-bit resolution. For wide screen, an 18-MHz sampling frequency is used. Additionally, an adapter can be used with 4:1 video compression for recording HDTV signals of 1.2 Gb/s.

Professional Recorders with Compression

The DV format that is being discussed later in the chapter has resulted in three different formats:

- **DVCAM** (Sony) is very similar to DV, using 25 Mb/s video with DV compression, 8-bit sampling on 4:1:1, and the 6.35-mm DV tape with a metal evaporated layer. However, 15 μm tracks are used, giving higher robustness. It is DV backwards compatible. Audio is done with 2-track recording, 16-bit sampling, and 738 kb/s subcode can be added.
- **DVCPRO** (Panasonic) is also compatible with DV and even DVCAM. The major differences are the use of two longitudinal tracks for improved editing (cue and control). The track pitch is 18 μm for robustness. This system was later extended to 50 Mb/s recording, called DVCPRO50, offering 4:2:2 broadcast quality recording with less compression (slightly more than 3:1) than DV (5:1).
- **Digital-S** (JVC) uses 0.5-in. tape based on S-VHS recording mechanics and a twin DV compression system for high-quality 4:2:2 broadcast recording. This system also records with a compression of 3:1:1. The 50 Mb/s machines have 4-track audio and the double subcode capacity of DVCAM, i.e., 1.47 Mb/s.

The Betacam SX recorder has also 8-bit video with compression using a 0.5-in. tape. The compression is higher—about 10:1—than DV, since it applies the MPEG 4:2:2 Studio Profile. This was chosen for compatibility reasons with emerging MPEG-based DVB standards. The video data rate is 18 Mb/s.

Finally, a number of digital high-definition recorders were introduced with the first experimental transmissions of digitally compressed HDTV in the United States. Examples are HDCAM, HDD-1000, and HDD-2700 machines. Typical resolutions are 1440 through 1920 samples by 720 progressive, 1035 or 1080 interlaced lines, using 10-bit video sampling. Audio is performed with 20-bit sampling and recorded with 4 or 8 tracks.

D-VHS

D-VHS (or Data-VHS) is the latest development of the VHS video format developed by JVC that enables digital recording of high-definition broadcasts and multichannel broadcasts. Three different D-VHS modes have been standardized: the STD (standard) mode, the HS (high speed) mode which is compatible with high-definition (HD) digital images and multichannel broadcasts, and the LS (low speed) mode which enables a maximum recording time of 49 h (Fig. 21.3.7 and Table 21.3.4). The technical format specifications of D-VHS were standardized in April 1996. D-VHS is compatible with all types of digital broadcasts in the world including new digital high-definition broadcasts.

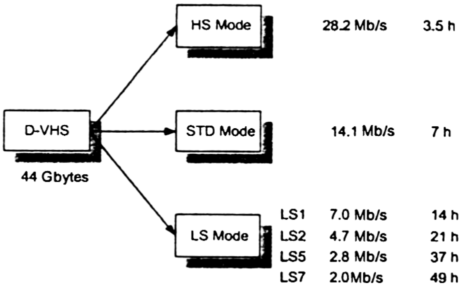

FIGURE 21.3.7 Various D-VHS formats.

Main Characteristics of D-VHS

D-VHS[39] enables bit-stream recording of digital broadcasts, in addition to analog recording of conventional broadcasts. D-VHS takes advantage of the characteristics of tape as a medium to provide extremely large

TABLE 21.3.4 Basic Specification of the Three D-VHS Modes

		HS Mode	STD	LS Mode			
				LS2	LS3	LS5	
Cassette		D-VHS cassette					
Head Configuration		Dual × 2	Single × 2	Single × 1 / Dual × 1	Single × 2		
Mechanics		Based on conventional VHS system					
Recording time							
Using DF-420		3.5 h	7 h	14 h	21 h	35 h	49 h
Using DF-300		2.5 h	5 h	10 h	15 h	25 h	35 h
Track Composition	Tape speed (mm/s)	33.35	16.67	8.33	5.55	3.33	2.38
	Head azimuth	+/–30°					
	Drum rotation	1,800 rpm					
	Tracking system	CTL track system					
	Main data input rate	28.2 Mbit/s	14.1 Mbit/s	7 Mbit/s	4.7 Mbit/s	2.8 Mbit/s	2 Mbit/s
Recording Specifications	Sub-data input rate	0.292 Mbit/s	0.146 Mbit/s	73.0 kbit/s	48.7 kbit/s	29.3 kbit/s	20.9 kbit/s
	Recording rate	38.28 Mbit/s	19.14 Mbit/s				
	Track structure	1 sector					
	Sync block length	112 Bytes					
	Inner correction code	RS code					
	Outer correction code	RS code					
	Code word shuffling	6 track					
	Modulation system	SI-NRZI					
	Interface	IEEE-1394 digital interface					

capacity digital data storage. D-VHS maintains compatibility with conventional VHS by allowing recording and playback of the enormous amount of VHS software available as well as stored around the world. D-VHS can readily evolve and spread as a household device owing to the possibility of using current VHS technologies, parts, and manufacturing facilities.

The data that are recorded from digital video broadcasts are actually an MPEG-compressed video and audio program, as standardized by, e.g., the DVB standard. This standard is based on MPEG-2 compression and so-called MPEG-2 transport stream (TS) packets of 188 bytes. The DVB standard has adopted most of the MPEG-2 standard as the basis for transmission of digital programs. The MPEG standard is explained later in this chapter. The recording of digital data on a basically analog machine is not without problems because the required media such as magnetic heads and tape for digital data differ from heads and tape that are optimal for analog signals. Fortunately, the size of the VHS headwheel (62 mm) is large enough to contain additional heads and transformers for constituting a compatible recording system.

DV RECORDING SYSTEM

Introduction

The design of the DV format has shown that for digital TV recording (and HDTV), small-sized recording mechanics giving sufficient playing time can only be achieved if the recording bit rate is reduced substantially. The initial bandwidth expansion due to the digitization of the video signal, combined with binary recording instead of the analog multilevel recording, must be compensated to a large extent by bit-rate reduction.[43] The relation between recording mechanics and video compression has played an important role in setting the DV standard so that both will be addressed.

For consumer recording,[19] the trend is toward small recording mechanics, especially for portable applications, and digital recording of compressed digital video signals. These aspects have been combined into the DV system, which was first announced in 1993[20] and which has been further developed by a substantial group of companies. The main focus of the standard was to introduce attractive small camcorder products in the mid-90s. Experimental digital video recording systems were initiated at the end of the 80s.[21–25] Since September 1995, a number of companies have introduced DV camcorders into the market. Based on the DV technology employing also new tapes,[26–30] the professional systems DVCPRO and DVCAM have emerged, which are increasingly popular.

Despite the acceptance of the MPEG and JPEG standards in numerous applications, the development of the DV video compression standard was largely guided and justified by the targeted product and cassette size, power consumption, and the consumer price level. Irrespective of the previous issues, a number of recording system constraints have to be satisfied in any case and are rather typical, namely:

- Editing, preferably on picture basis
- Robust for repeated (de-)compression, resulting from analog video dubbing
- Multitrack format, using a set of tracks for one frame (see Fig. 21.3.9)
- Forward and backward search on tape, from small data bursts pictures have to be recovered (see Fig. 21.3.9)
- Very high robustness
- High picture quality

DV System Architecture

Figure 21.3.8 from Ref. 31 portrays the block diagram of the recording system. Signal processing is divided into two main classes:

- Source coding employing video compression and source data formatting
- Channel coding using error-correcting codes and modulation codes for spectrum shaping of the bit stream

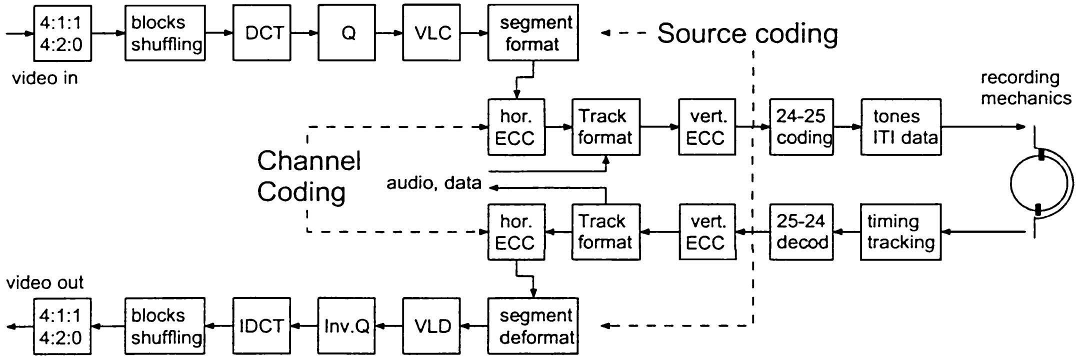

FIGURE 21.3.8 Block diagram of complete processing of consumer DV recording system.

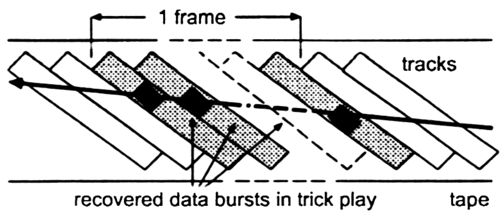

FIGURE 21.3.9 Multitrack format and trick play.

The video compression is based on the discrete cosine transform (DCT), and subsequent quantization and variable-length coding (VLC) of the transformed video data. Since the operations are performed on blocks of 8 × 8 samples, the video input signal is stored for line-to-block conversion. Shuffling of data blocks is performed for improving the picture quality under normal and trick play. The sampling of the video signal is derived from the CCIR-601 standard (4:2:2) with 13.5 MHz sampling frequency for the luminance (Y) signal. The color-difference signals (Cr and Cb) are subsampled either horizontally with a factor 2 (4:1:1) for 60 Hz or vertically with a factor 2 (4:2:0) for 50 Hz systems.

For error-correction coding (ECC), the DV recording format employs the obligatory Reed-Solomon (R-S) product codes, of which the error-correcting properties are defined within a track. After compression, the data are organized in small packets, called sync blocks. Each sync block gets some parity bytes (HECC) assigned to it for random byte errors and for error detection of burst errors. Subsequently, the complete video data block (or audio) is extended with extra sync blocks containing vertical parity bytes (VECC). The performance of these codes is very good for removing random errors and general burst errors. As a second step in channel coding,[40] a high-efficiency DC-free 24-25 channel code was adopted, described in Refs. 32, 33, to constitute pilot tones for tracking, which are embedded in the data.

The recording mechanism is based on a small drum of 21.7 mm, which rotates at 9000 rpm. The basic recording mode (2 × 1 head) has a capacity of 41.85 Mb/s (equivalent to 25 Mb/s video), but the system can be upgraded mechanically and electronically to 83.7 Mbit/s (50 Mbit/s net video rate) for HDTV recording.[31] Possibilities to go to 12.5 Mbit/s video rate or lower, required for recording of MPEG signals (MP@ML SDTV), have been indicated earlier.[32] Table 21.3.5 portrays the key parameters of the experimental recording system.

TABLE 21.3.5 Key Parameters of Digital Consumer DV Recorder

Scanner	2 heads
Scanner diameter	21.7 mm
Scanner speed	9000 rpm
Tape width	6.35 mm
Tape thickness	7 μm
Track pitch	10 μm
Recording rate	41.85 Mb/s
Tracking	Embedded tracking tones
Channel modulation	24-25 code
Error correction code	R-S product code
Video bit rate	25 Mb/s
Video coding	8 × 8 intraframe DCT
Trick mode implementation	Video macroblock shuffling

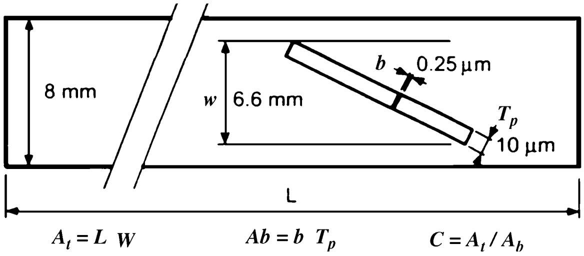

FIGURE 21.3.10 Important 8-mm tape parameters.

DV Cassette Size and Capacity

Cassette Size Related to the 8 mm System. The well-known analog 8 mm cassette was the reference and the starting point for establishing the cassette parameters in relation to the playing time. The largest tape length in the 8 mm system is about 125 m. Some relevant parameters of the 8 mm tape format are depicted in Fig. 21.3.10. The effective width w of the tape area for analog video and PCM audio (221 ± wrap) equals 6.6 mm, which is somewhat lower than the tape width of 8 mm. The edges cannot be used for the helical tracks because of the limited head-tape contact. The product of L and w results in 0.83 m^2 maximum usable tape area A_t. The cassette capacity C in bits is the ratio of A_t to the bit area A_b: It was estimated from experiments on ME tape that a track width of 10 μm with a bit length of 0.25 μm would result in sufficient SNR and robustness for the DV system. The gross 8 mm cassette capacity is therefore 330 Gb.

To derive the net capacity for compressed video, the estimated practical contributions for overhead are tabulated in the left column of Table 21.3.6. The estimations are educated guesses of the expected overhead, assuming a sync block size of around 1000 bits and a not usable track margin of ±5 – 6. For clarification, the real values of the DV system have been added in the right columns of Table 21.3.6. The overhead for channel modulation and digital audio is now briefly clarified. A well-known block code such as 8-10 modulation needs an overhead of 20 percent. Some form of channel modulation is a must, primarily for tracking and robustness reasons, but the overhead should be limited. The sophisticated 24-25 modulation from Ref. 33 with only 4 percent overhead has proven to be an attractive candidate. The overhead for digital audio is in principle more an absolute value in bit rate and less a relative overhead. However, we have assumed that the

TABLE 21.3.6 Contributions to Recording Rate

Topic	Estim. (%)	DV (kbit)	DV fraction (%)
Channel modulation	4	5.4	3
Tracking information	0	1.8	1
Synchronization + identification	4	6.5	5
Two-dim. error correction	15	20.3	15
Digital audio	6	5.2	4
Auxiliary data	2	2.2	2
Subcode information	2	1.2	1
Margins at track begin/end	3	(1.2)	
Edit gaps and pre/post-amble	3	9.1	7
Video	61	83.2	62
Total one track	100	134.9	100

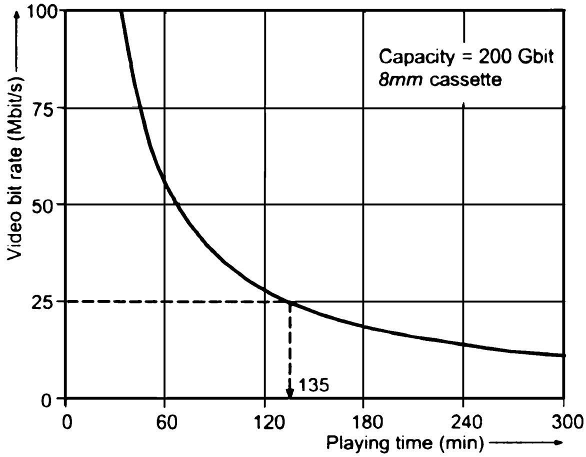

FIGURE 21.3.11 Video bit rate and playing time.

bits needed for digital audio are a reasonable 10 percent fraction of the bits for compressed video, hence 6 percent of the total.

From the previously discussed overhead inventory it is clear that, even with an economical solution for channel modulation, only around 60 percent of the gross capacity is available for compressed video. The net cassette capacity for the 8 mm system is therefore 200 Gb. From this number a simple relation is obtained between the compressed video bit rate and the playing time, which is graphically presented in Fig. 21.3.11. This curve explains why the video bit rate of 25 Mb/s was chosen.

Requirements of the DV System. Prior to discussing the principal DV system parameters, we list the basic aims and assumptions for defining the recording system:

- Pocketable camcorders, which are substantially more compact than the analog 8 mm system. Playing time is less important; 60 min is regarded as sufficient. Picture quality and editability are of utmost importance.
- Home-use recording of HDTV signals with a minimum playing time of 135 min needed to record almost all movies in one piece. At the time of standardization, HDTV systems such as MUSE in Japan, HDMAC in Europe, and ATV in the United States were seen as very important for the near future.
- Home recording of SDTV signals, with the same compression as for the camcorder. Ratio of HDTV to SDTV bit rate of 2:1, resulting in a minimum playing time of 270 min for SDTV.
- Preferably, a one-cassette system has to be adopted.

It will be clear that the last point can be conflicting with the first two, owing to the significantly higher bit rate and playing time demands of the HDTV system, when compared to the camcorder application. This will become more apparent from a further evaluation of Fig. 21.3.11. At a playing time of 135 min for HDTV, the available video bit rate for an 8 mm cassette equals 25 Mb/s, and consequently 12.5 Mb/s for SDTV. It is elucidated later that this bit rate is considered too low for an economical, well editable, and high-quality compression scheme for the camcorder. In fact, the double of this bit rate was regarded as realistic. This inevitably leads to a two-cassette system with a larger cassette for home use and a compact cassette with less playing time for the camcorder. Currently, the camcorder cassette has proven to be the main cassette in practical use.

Cassette Sizes for DV System. The net cassette capacity for home use was planned to be 400 Gb, requiring a substantially larger cassette compared to the analog 8 mm system. The required capacity of the small

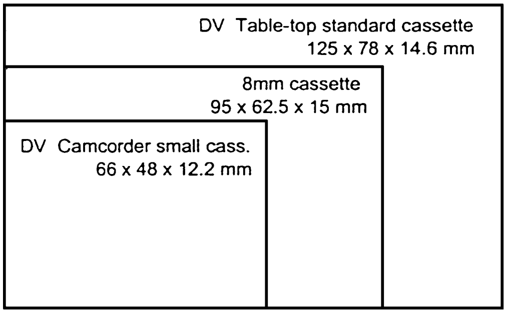

FIGURE 21.3.12 Preferably, a one-cassette system has to be adopted.

camcorder cassette is roughly one quarter of the standard cassette, owing to the playing time ratio, and has therefore half its size. This is substantially smaller than the 8 mm cassette. The relative sizes of the final DV cassettes are depicted in Fig. 21.3.12.

Thus far the thickness of the cassettes was assumed identical to the 8 mm cassette. This is suitable for the larger standard cassette, but a small camcorder cassette with this thickness is not very elegant. Additionally, the compactness of the camcorder is influenced by this parameter, which was therefore reduced as far as possible to the well-known quarter inch (6.35 mm) owing to the camcorder application. The final thicknesses for the standard and small cassettes are 14.6 and 12.2 mm, respectively.

Drum Diameter and Speed

For a camcorder the focus is on an economical drum for the camcorder, which is the well-known drum concept having two heads of different azimuth on opposite sides and using a wrap angle of 180°. This not only leads to the lowest drum complexity, but also minimizes the required read/write electronics and therefore power dissipation, because only single recording channel electronics are needed. Alternative drum configurations do exist and their relations to other applications will be elucidated briefly.

The relation between the drum and video signals asks for a single worldwide recording system with identical rates and drum speeds everywhere. Furthermore, an integer number of tracks/frame was realized to support editing, combined with a small drum size for compactness. The results are given in Table 21.3.7. A comparison between Table 21.3.7 and the wish list at the beginning of this section shows that 10/12 tracks/frame is a well-balanced choice, taking into account a few additional considerations. First, an even number of tracks/frame is preferred for both systems, owing to azimuth recording. Second, higher drum speeds generally yield a higher power dissipation of the drum motor, but too small drum diameters pose problems with mounting internal electronics. Fine-tuning of the tape format resulted in 41.85 Mb/s channel rate, 34.024 mm track length, 21.7 mm drum size, and 9.1668° track angle for the DV system. The effective track length is 32.890 mm, which corresponds to an effective wrap of 174°.

Alternative drum concepts are important for possible extensions toward HDTV, using a doubled recording rate. This is at best achieved with a drum containing two head pairs that are mounted diametrically. The second double head requires continuous two-channel recording, allowing for an HD mode with two times the bit rate of SD. Full compatibility to the SD and even longplay (LP) modes is realized by using only one head pair. If the second double head is mounted with a slightly modified height position, it can also be used for read after write in generic data recording applications (see Ref. 31 for further details).

DV Video Blocks and Sync Block Format

In the DV system, video data is mapped onto the tracks, where each track consists of data packets, called sync blocks. Video data in the form of images is divided into macroblocks, i.e., a 16 × 16 pixel area from an image.

TABLE 21.3.7 Various Drum Configurations

Tracks/ frame (30/25)	Track rate (tr/s)	Drum speed (rpm)	Track length (mm)	Drum diam. (mm)	Track angle (deg.)
5/6	150	4500	67	42	5
10/12	300	9000	33	21	10
15/18	450	13500	22	14	15
20/24	600	18000	17	11	20

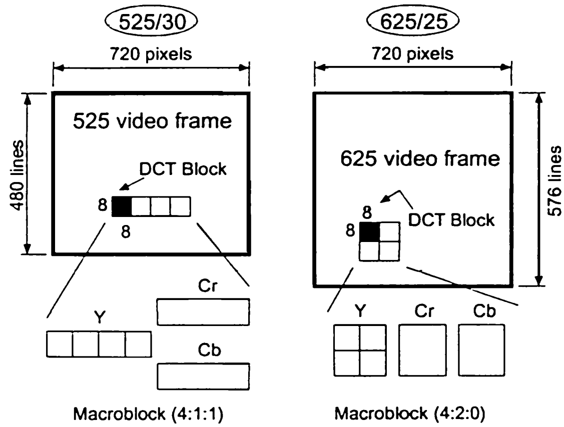

FIGURE 21.3.13 Macroblocks (MBs) taken from 525/30 and 625/25 frames and corresponding MB structure.

This section addresses the mapping of those macroblocks (MBs) on the sync blocks in the tracks, under the assumptions of 10/12 tracks/frame and 25 Mbit/s for compressed video. After sampling of the analog signal, the effective area of the video frame consists of 720 pixels horizontally by 480 or 576 lines vertically for the 525/30 and 625/25 systems, respectively. This is depicted in Fig. 21.3.13. The pixels are grouped in DCT blocks of 8 × 8 pixels, resulting in 5400 or 6480 luminance DCT blocks/frame. With 10/12 tracks/frame, this leads to an identical number of 540 luminance DCT blocks/track for both systems. Although the subsampling of the chrominance signals is different for 525/30 and 625/25, the ratio of luminance to chrominance is 4:2 in both cases. Macroblocks are formed, consisting of four luminance and two chrominance DCT blocks, resulting in 135 MBs per track.

There are restrictions on sync block (SB) mapping. For reasons of convenience in hardware implementation, robustness, and trick-play performance, a fixed mapping of a limited number of MBs (a unit) onto an integer number of SBs has been adopted. Moreover, a multiple of such mapping units should exactly fit into one track for obvious reasons. As a result, the number of MBs per unit is a divider of 135. Possibilities are 1, 3, 5, 9, 15, . . . Similarly, the same units, in the sequel called segments, will be used as independent coding units for the video compression. The video segment size was fixed to 5 MBs, because it was found from simulations that no essential improvement in picture quality could be obtained for larger segments (see Fig. 21.3.20) in the subsection on DV video compression). This results in 27 video segments per track.

The remaining key parameter in the mapping is the number of SBs per segment. It is clear that this number will vary the total amount of video SBs in a track, which, as a consequence of the previous result, should be a multiple of 27.

For determining the last parameter of the format, the highest possible trick-play speed was a decisive factor (thereby adopting a constant bit-rate concept). Assuming a reasonable positioning of the head for reading (mapped into a "quality factor" $Q = 1/2$) and relative tape speed $n > 1$, the recovered data burst length D during fast reading can be approximated by $D = T_L/n$. The track length is denoted T_L. Naturally, as the speed increases, the amount of data read becomes smaller. Figure 21.3.14 portrays the data recovered as a function from the search speed (relative to the reference playback speed). Because the position of the track crossings with respect to the SBs is arbitrary at high speeds, the data burst must be at least two SBs in order to retrieve one correctable SB from tape. The maximum trick-play speed n_{max} is then given by

$$n_{max} = \frac{T_L}{D_{min}} = \frac{T_L}{2} \quad (1)$$

with T_L expressed in SBs being the total length of a track for 180° wrap. The total overhead in a 180° track is assumed to be 40 percent, of which 15 percent is SB-based. The remaining 25 percent of the overhead can be expressed as SBs on top of the video SBs. Therefore, the total number of SBs in a track (T_L) is 4/3 times the

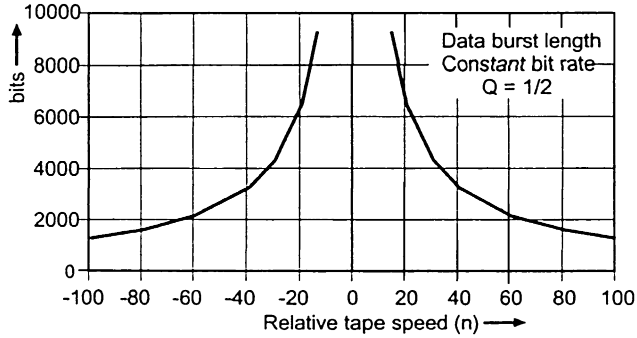

FIGURE 21.3.14 Data burst length vs. search speed.

number of video SBs. Results for T_L and n_{max} are given in Table 21.3.8. In practice, the value of n_{max} will be somewhat lower.

Table 21.3.8 shows why the 5-5 mapping has been chosen in the DV standard. First, this mapping leads to the highest possible trick-play speed. Second, note that with this mapping, in contrast to the other possible mappings, one MB is stored in one SB, thereby enabling a fixed allocation of the most important data. For trick play, this enables the decoding on SB basis, which was assumed implicitly for the calculation of n_{max}. Third, as a bonus, the 5-5 mapping results in the shortest SB length. This property, combined with the fixed data allocation, proves to be very beneficial for the robustness of the system. With the chosen 5-5 mapping, the size of an SB can be determined. With 83 kbit and 135 MBs in a track for video, a 77-byte data area is required to store one macroblock. The addition of a 2-byte Sync pattern, a 3-byte ID, and 8-byte parity for the inner (horizontal) ECC results in a total SB length of 90 bytes. For completeness, it is mentioned that this number must be a multiple of 3 bytes, because of the 24-25 channel coding. The SB structure of the DV system is given in Fig. 21.3.15.

DV Video Compression

It has been explained that video compression is required to establish a low bit rate and sufficient playing time. For this reason, an intraframe coding system was adopted. Intraframe coding means that each picture is compressed independent of other pictures. However, to support the high-speed search the system goes one step further and concentrates on coding segments of a picture as independent units. This will be elaborated in this section. We consider a feedforward-buffered bit-rate reduction scheme, based on DCT coding, in which the pictorial data are analyzed prior to coding. The aim is to define a compression system with fixed-length coding of a relatively small group of MBs because this is beneficial for recording applications such as trick play and robustness. Given the system constraints from previous sections, the target system is based on fame-based DCT coding with VLC. Such a system operates well using compression factors 5 to 8. This results globally in a bit rate after compression of 20 to 25 Mb/s. Numerous subjective quality experiments during development

TABLE 21.3.8 Various Mappings of MBs to SBs

MB's per segment	SB's per segment	Video SB's per track	Total SB's per track	n_{max}
5	3	81	108	54
5	4	108	144	72
5	5	135	180	90

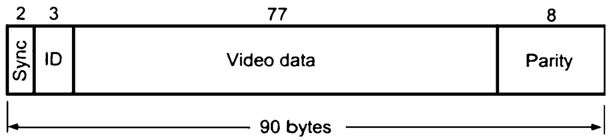

FIGURE 21.3.15 Format of a single sync block (SB).

of the DV standard provided evidence that about 25 Mb/s yielded the desired picture quality, which is well above analog consumer recorders such as VHS and 8 mm.

DV Feedforward Coding Concept. In most transform coding systems using VLC techniques, the variable-rate output is buffered and monitored by feedback quantizer control to obtain—on the average—a constant bit rate, although the bit rate is locally varying. With feedback compression systems having a locally varying output bit rate, the relation between the recovered data bits and the location of the data in the picture is lost. The major advantage of the feedforward coding system is that relatively small groups of DCT blocks, henceforth termed segments, are coded independently and, in contrast with a feedback system, as a fixed entity. This property makes the segments independently accessible on the tape, while the fixed code length ensures a unique relation between the segment location on tape and its location in the reconstructed picture. The latter property, in combination with the 1 macroblock-per-SB data allocation (see previous section on mapping), is exploited for optimizing the picture quality during trick modes (see later).

In the feedforward coding system (see Fig. 21.3.16), video is first organized into blocks and subsequently in groups of blocks, called segments. Each segment is then compressed with DCT coding techniques into a fixed code length (bit cost), despite the application of VLCs. Fixed-length coding of a small group of DCT blocks (several tenths only) can only be realized if the transformed data are analyzed prior to coding, requiring temporal data storage. During the storage of a segment, several coding strategies are carried out simultaneously, from which only one is chosen for final quantization and coding. This "analysis of the limited future" explains the term feedforward buffering. Feedforward coding has two important advantages: a fixed relation between the data on tape and the reconstructed image, and a high-intrinsical robustness as the channel error propagation is principally limited within a video segment.

DV Motion-Adaptive DCT. The DCT has become the most popular transform in picture compression, since it has proven to be the most efficient transform for energy compaction, whereas the implementation has limited complexity. The definition of the DCT used in the DV system is

$$F(u,v) = C(u)C(v)\sum_{i=0}^{N-1}\sum_{i=0}^{N-1} f(i,j)\cos\frac{(2i+1)u\pi}{2N}\cos\frac{(2j+1)v\pi}{2N} \tag{2}$$

where a block of samples $f(i, j)$ has size $N \times N$. The two constants $C(u)$ and $C(v)$ are defined by $C(w) = \frac{1}{2}$, for $w = 0$ and $C(0) = \frac{1}{2}\sqrt{2}$. The DV standard applies a block size of 8 × 8 samples because it provides the best compromise between compression efficiency and complexity and robustness.

One of the first main parameters for coding efficiency is to choose between intrafield and intraframe coding. In the latter system, the odd and even fields are first combined into a complete image frame prior to block

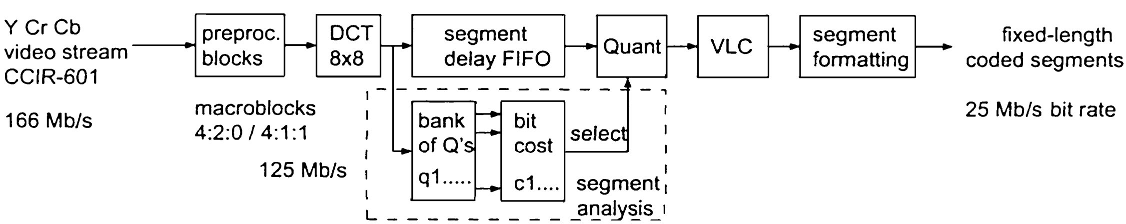

FIGURE 21.3.16 Architecture of feedforward video compression system.

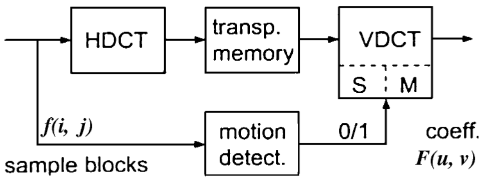

FIGURE 21.3.17 Architecture of a motion-adaptive DCT transformer.

coding. It has been found that intraframe coding is about 20 to 30 percent more efficient than intrafield coding, or, for the available 25 Mb/s bit rate, it offers a considerable better quality. For this reason, intraframe coding was adopted in the standard. However, it was found in earlier investigations that local motion in sample blocks leads to complicated data structures after DCT transformation, which usually are particularly difficult to code. The solution for this problem is to split the vertical transform into two field-based transforms of length $N/2$. Hence, first an N-point horizontal transform (HDCT) is performed yielding intermediate data $F_h(I, v)$, and subsequently, two vertical ($N/2$)-point transforms (VDCT), specified by

$$F(u,v) = C(u)C(v) \sum_{i=0}^{N/2-1} \sum_{j=0}^{N-1} g_s(i,j) \cos \frac{(2i+1)u\pi}{N} \cos \frac{(2j+1)v\pi}{2N}$$
$$F(u+4,v) = C(u)C(v) \sum_{i=0}^{N/2-1} \sum_{j=0}^{N-1} g_d(i,j) \cos \frac{(2i+1)u\pi}{N} \cos \frac{(2j+1)v\pi}{2N} \tag{3}$$

$$g_s(i, j) = [f(2i, j) + f(2i + 1, j)]$$
$$g_d(i, j) = [f(2i, j) - f(2i + 1, j)] \tag{4}$$

Note that in the first field-based output coefficient block, the sum of the two fields is taken as an input, while in the second coefficient block the difference of the two fields is considered. Hence, this corresponds to a two-point transform in the temporal domain.

The required DCT processor architecture is depicted in Fig. 21.3.17.

DV Adaptive Quantization

The primary elements for quantization of the coefficients $F(u, v)$ are frequency-dependent weighting, adaptivity to local image statistics, and global bit-rate control. These elements are individually addressed briefly. The weighting is based on the decaying sensitivity of the transfer function H of the human visual system (HVS). As an example, as HVS model from the literature has been plotted in Fig. 21.3.18, in which the transfer function of the HVS has been normalized.

When using that model the image is divided in 8×8 blocks, in Fig. 21.3.18 we find that the HVS decreases exponentially, is multiplicative in nature, and hence $F_W(u, v) = W(u, v)F(u, v)$. The weighting function can be simplified using special multiplying factors (see Table 21.3.10). For simplicity, the matrix of factors $W(u, v)$

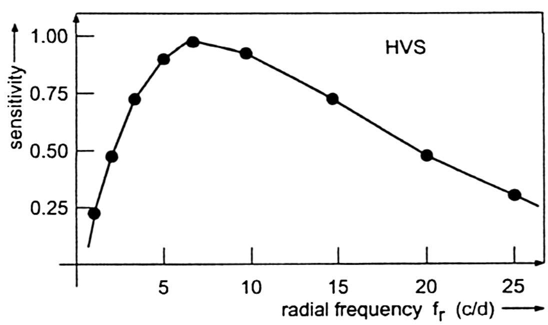

FIGURE 21.3.18 HVS function of radial frequency f_r.

TABLE 21.3.9 Area Numbers of Weighting for Static 8 × 8 Block (left) and Moving 2 × (4 × 8) Block

Static 8 × 8 block: W(u, v), u ↓, v →

X	0	0	1	1	1	2	2
0	0	1	1	1	2	2	2
0	1	1	1	2	2	2	3
1	1	1	2	2	2	3	3
1	1	2	2	2	3	3	3
1	2	2	2	3	3	3	3
2	2	2	3	3	3	3	3
2	2	3	3	3	3	3	3

Moving 2 × (4 × 8) block: W(u, v), u ↓, v →

sum	X	0	1	1	1	2	2	3
	0	1	1	2	2	2	3	3
	1	1	2	2	2	3	3	3
	1	2	2	2	3	3	3	3
difference	0	0	1	1	2	2	2	3
	0	1	1	2	2	2	3	3
	1	1	2	2	2	3	3	3
	1	2	2	3	3	3	3	3

are not different for each (u, v) combination. Instead, groups of (u, v) combinations apply the same weighting factor, according to Table 21.3.9.

The second element of quantization is the adaptivity to local image statistics. The adaptivity is to be worked out by measuring the contents of the DCT coefficient block. One of the most well-known metrics for local activity is the "ac energy" of the block, $\Sigma_{u,v} F(u, v)^2$. However, simpler metrics, such as the maximum of all $F(u; v)$ within a block, perform satisfactorily as well. The DV system allows that any metric in the encoder is acceptable: the decoder simply follows the two decision bits reserved for the quantizer classification. This freedom also allows different quantization of luminance (Y) and color (Cr, Cb) blocks. Generally, more activity or information content results in more coarse quantization. The third element in the quantizer is a final block quantization by a linear division with a step size S. The advantage of this approach is its simplicity and it leads to uniform quantization. The variable S defines the accuracy of the global block quantization. When taking into account the elements previously discussed, the overall quantization is specified by $F_Q(u, v) = W(u, v)F(u, v)/S$. For camcording, low implementation cost is of utmost importance. A particular simple system is obtained by taking $W(u, v) = S = 2^{-p}$ with p being an integer that is controlled by all three elements discussed in this subsection. The final quantization table is shown in Table 21.3.10. The weighting area numbers in Table 21.3.10 refer to Table 21.3.9.

TABLE 21.3.10 Table of Step Sizes Using Area Indication for Weighting and Strategy Number for Global Uniform Quantization

Q class 0 (Q strategy ↓)	Q class 1	Q class 2	Q class 3*	W area number 0	W area number 1	W area number 2	W area number 3
15							
14							
13							
12	15						
11	14						
10	13		15	1	1	1	1
9	12	15	14				
8	11	14	13	1	1	1	2
7	10	13	12	1	1	2	2
6	9	12	11				
5	8	11	10	1	2	2	4
4	7	10	9				
3	6	9	8	2	2	4	4
2	5	8	7				
1	4	7	6	2	4	4	8
0	3	6	5				
	2	5	4	4	4	8	8
	1	4	3				
	0	3	2	4	8	8	16
		2	1				
		1	0	8	8	16	16
		0					

*If class 3 occurs, all step sizes are multiplied by 2.

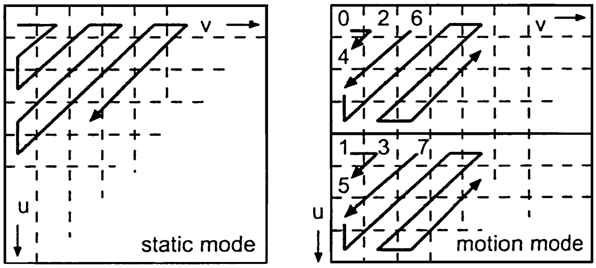

FIGURE 21.3.19 Diagonal zigzag scanning in 8×8 and $2 \times (4 \times 8)$ mode for clustering of zeros.

DV Variable-Length Coding

A bit-assignment using one coding table only was chosen for simplicity. The coding algorithm is fundamentally the same as in MPEG compression and is based on runlength counting of zeros only and the use of an end-of-block (EOB) codeword.

The principle of the algorithm is that first the block of quantized coefficients is scanned using diagonal zigzag scanning in order to create a one-dimensional stream of numbers. The scanning is adapted to motion. The purpose of the scanning is to cluster zero coefficients (see Fig. 21.3.19), so that they can be coded efficiently. Second, from the start of the string, zeros are counted until a nonzero coefficient is noticed. The magnitude of this nonzero coefficient is combined with the preceding length of zeros into a single event (run, amplitude), which is jointly coded with a single codeword. The sign of the coefficient is appended at the end of the codeword. An excerpt of the encoding table showing the variable wordlengths and codewords is given in Table 21.3.11. The coding table is optimized to prevent large codewords and low implementation cost.

Total DV Video Performance

The control and overall performance of the compression system is now indicated. The system optimizes the picture quality by testing a number of quantizer settings in parallel. The optimal quantization strategy m_{opt} is the quantizer setting (strategy) that yields a bit rate just below or equal to the desired bit rate. The choice of m_{opt} can vary between 0 and $M - 1$ when M different quantization strategies are used. The picture quality of the compression system for various segment sizes K and number of quantization strategies M was measured. The resulting picture quality is expressed in SNR (dB), which refers to the mean squared error (MSE) with the original picture compared to squared peak white (255^2). The results of the measurements are shown in Fig. 21.3.20.

The optimal choice of the segment size K and the number of strategies M can be derived from Fig. 21.3.20. Evidently, the recording system designer opts for the smallest possible value of K (small segments), because it yields a high robustness and it enables higher search speeds. However, Fig. 21.3.20 shows that if the size becomes too small, i.e., $K < 30$, the picture quality deteriorates rapidly. For $K = 30$ to 60 DCT blocks, the resulting SNR remains nearly constant. Therefore, a segment size of $K = 30$ was adopted as being the best compromise. The 30 DCT blocks are not arbitrarily chosen from the picture, but clustered in groups, called MBs. An MB (see Fig. 21.3.13) is a group consisting of 2×2 DCT blocks of 8×8 Y samples each and the two corresponding 8×8 color blocks Cr and Cb. In order to improve the picture quality of the compression system, MBs are selected from different areas of the picture, so that the influence of local image statistics is smoothed. The result is a stable and high picture quality for a large class of images. However, after compression, the MBs are redistributed in order to improve the visual performance during search (see later in this DV section). A second conclusion is that $M = 16$ gives a substantial improvement in picture quality, compared to $M = 8$. The quality improvement can be fully explained by a more efficient use of the available bit rate, which becomes particularly important for small segment sizes.

TABLE 21.3.11 Table of Wordlengths (a) and Codewords (b) of DV System

Run ↓ \ Amplitude (abs.) →	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	11	2	3	4	4	5	5	6	6	7	7	7	8	8	8	8	8
1	11	4	5	7	7	8	8	8	9	10	10	10	11	11	11	12	12
2	12	5	7	8	9	9	10	12	12	12	12	12					
3	12	6	8	9	10	10	11	12									
4	12	6	8	9	11	12											
5	12	7	9	10													
6	13	7	9	11													
7	13	8	12	12													
8	13	8	…	…													

coeff. sign not incl.
EOB = 4

(a)

Event	Codeword
(0, 1)	00s
(0, 2)	010s
EOB	0110s
(1, 1)	0111s
(0, 3)	1000s
(0, 4)	1001s
(2, 1)	10100s
(1, 2)	10101s
(0, 5)	10110s
(0, 6)	10111s
(3, 1)	110000s

Event	Codeword
(4, 1)	110001s
(0, 7)	110010s
(0, 8)	110011s
(5, 1)	1101000s
(6, 1)	1101001s
(2, 2)	1101010s
(1, 3)	1101011s
(1, 4)	1101100s
(0, 9)	1101101s
(0, 10)	1101110s
(3, 11)	1101110s

(b)

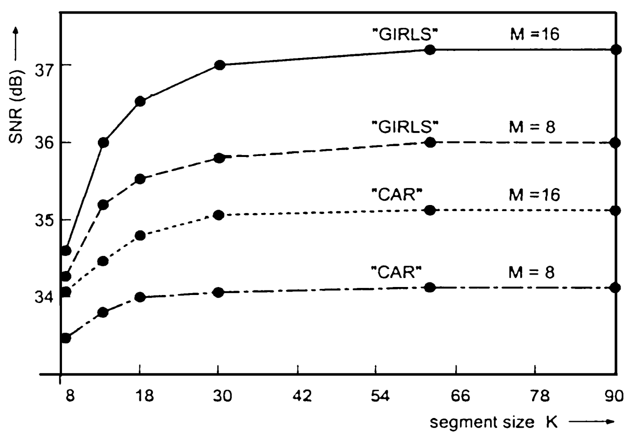

FIGURE 21.3.20 SNR of DV compression system for various segment sizes K and number of strategies M (CCIR-601 images, 4:2:0 sampling, 25 Mb/s).

The subjective picture quality of the system is known as excellent and regarded as very close to levels for professional use. For regular input pictures, the SNR approaches 40 dB, and the resulting subjective image quality of the system comes rather close to the quality of existing professional recording systems. For complex and detailed imagery, the SNR is a few dBs lower.

Macroblock-Based SB Format

The DV format uses a special compressed data format inside a sync block (SB) to improve robustness for high-speed search, where a part of the error-correction coding (ECC) cannot be applied, because only data portions of a track are recovered. This considerably multiplies the chance of having errors in the signal. Second, at higher tape speeds, the head-to-tape contact is reduced and less stable, which also leads to a lower robustness. The special data format enables the compression decoder to cope with residual errors in the video data. In order to construct a robust format, it is absolutely essential to limit the propagation of errors that emerge from erroneous variable-length decoding. The propagation of errors is limited in three ways, which are briefly discussed.

First, *fixed-length coding of segments* is realized by the choice of a feedforward coding scheme. Every segment is compressed into a fixed bit cost, so that the decoder should periodically reset itself at a segment border. A proper numbering of SBs allows to identify the start of a new segment without using the compressed data. Error propagation from segment to segment is therefore impossible. Second, *identification of individual macroblocks* is enabled. A segment consists of five full-color MBs as indicated in the previous section. For robustness, a single MB is put into a single SB. Note that the MB is sometimes smaller than a SB and sometimes larger. Furthermore, every SB has a fixed unique location on tape. As a result, each MB—at least the low-frequency information—can be addressed. Third, *identification of individual DCT blocks* is possible. Within an SB (thus one MB) six DCT blocks are located. Each compressed DCT block is of variable length. Similarly, to the MBs, by putting the low-frequency data of each DCT block on a fixed position, each DCT block can be addressed and partially decoded, and error propagation is limited to high-frequency components of DCT blocks only.

The internal SB format is depicted in Fig. 21.3.21. A group of five SBs forms a fixed-length segment, preventing error propagation. As a bonus, the fixed-length segment compression allows replacement of individual segments for post editing or error concealment in the compressed domain, without decoding the full picture. The individual segments can also easily be transmitted over a digital interface (IEEE-1394) to other equipment. Figure 21.3.21 also shows the fixed positions of the start of each DCT block.

DV Data Shuffling for Trick Play

The DV standard applies special data shuffling techniques to optimize the picture quality both for normal playback and for high-speed search on tape. It was elucidated in preceding subsections that a coding unit, called video segment, consists of five MBs only. When considering MBs for shuffling, the following general statement applies. For smoothing statistics, a regular distribution over the picture area of the MBs for one segment results in the highest average picture quality for all practical video material. This subsection describes MB shuffling in more detail, taking the 625/25 system as an example.

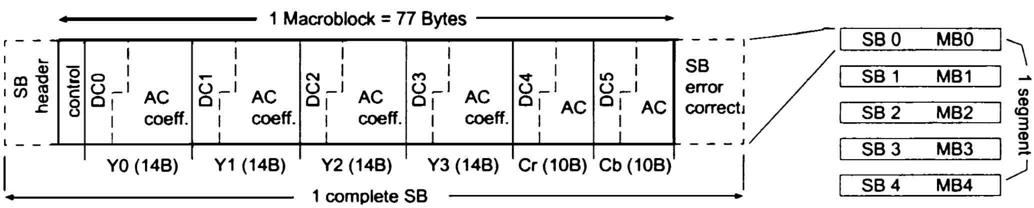

FIGURE 21.3.21 Inner SB data format showing the fixed predetermined positions of low-frequency DCT coefficients of each block and the construction of segments using these sync blocks.

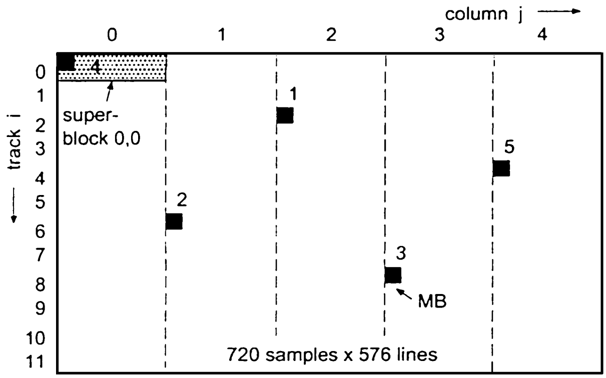

FIGURE 21.3.22 Selection of MBs for segment construction and assignment of picture areas on tracks.

Data Ordering for Optimal Picture Quality at Normal Speed. As depicted in Fig. 21.3.13, a picture consists of 480/576 lines of 720 pixels wide. With 10/12 tracks per frame, it is clear that the data of 48 lines or 135 MBs is stored in one track. In the case of the 625/25 video system, the picture of 36 by 45 MBs is divided into 12 horizontal rows of three MBs high and 45 MBs wide, where each row corresponds to one track. On the other hand, one segment consists of five MBs, which should originate from distributed parts of the picture, preferably with the highest distance. A maximum horizontal distance is achieved when they are distributed regularly, leading to a horizontal pitch of nine MBs. Consequently, the picture is divided into five columns of nine MBs wide. The row-column structure is depicted in Fig. 21.3.22. Despite the difference in video lines, the 525/30 picture is divided similarly in 10 rows and five columns. In the vertical direction, a regular distribution with maximum distance is also the optimal for a good picture quality. Taking into account the 10/12 rows for both systems and a universal algorithm, a distance of two rows is the best option.

A unit of 3 by 9 MBs is called a superblock. A picture consists of superblocks $S_{i,j}$ with *i, j* being the row and column number, respectively. The numbering of the MBs within the superblock can be found in Fig. 21.3.23. The construction of a superblock for the 525/30 system is somewhat different because of modified MB dimensions (see Fig. 21.3.13). Since the distances are known now, a suitable algorithm has to be defined such that the five MBs forming one segment are spread out over the picture area. The following algorithm has been adopted:

$$V_{i,k} = \sum_{p=0}^{4} MB[(i+2p) \bmod n,\ 2p \bmod 5, k] \tag{5}$$

with $0 \le i < n - 1$, $0 \le k \le 26$, and $n = 10$ or 12. Furthermore, $M\,[i, j, k]$ denotes the macroblock number k in superblock $S_{i,j}$. The MBs forming segment $V_{0,0}$ are indicated in Fig. 21.3.22.

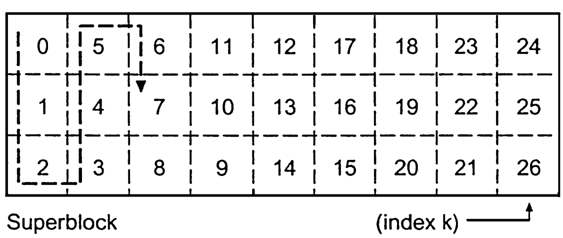

FIGURE 21.3.23 Ordering of macroblocks in repetitive clusters called superblocks.

Macroblock Mapping for Trick Play. Editing and picture reconstruction in normal play are performed on a picture basis, so that the mapping can be optimized for trick play. Consequently, the data of one picture is recorded in an alternative order within the boundaries of the set of tracks assigned to the complete picture. The following aspects are used in the mapping: *Coherency and speed adaptivity.* Coherency means that the best subjective trick-play picture quality is obtained if the largest possible coherent picture parts are refreshed in one piece.18 The picture quality should be adaptive to the speed, meaning that the highest quality is achieved for the lowest search speeds. At the introduction of the MBs, it was shown that the recovered data burst length from tape increases for lower speeds (see Fig. 21.3.14). Consequently, neighboring pictorial information should be recorded side by side in a track.

The previous insight contradicts with the MB shuffling for efficient video encoding at normal speed requiring that the pictorial information should be spread out over the picture. As indicated in Fig. 21.3.21, the majority of the information for one MB is stored in one SB. This is an attractive property for trick play. It enables decoding on a SB rather than on a segment basis, resulting in a slightly reduced but still adequate picture quality for trick play. This feature allows for a block mapping on MB basis instead of segments, thereby solving the paradox between MB shuffling and actual mapping on tape.

The selection of the optimal mapping for a large set of trick-play speeds is a difficult issue, especially if the possibilities for alternative scanners have to be considered simultaneously. The highest flexibility in all situations is achieved by a more or less direct projection of the picture on the tape. This is realized by recording the superblocks in one row of Fig. 21.3.22 one after the other in a track, with the row number equal to the track number. MBs within a superblock are stored in the order indicated in Fig. 21.3.23. A schematic representation of the final mapping is given in Fig. 21.3.24. On top of this mapping, trick-play speeds are chosen carefully at special fractional noninteger values, leading to sufficient refresh of data over time. More details can be found in Ref. 18. In general, it can be concluded that the search speed should be such that all other picture parts are refreshed between the updates of a specific part.

Conclusion on DV System

It has been clarified that the adopted cassette size and tape thickness have determined the required compression factor and the resulting bit rate of 25 Mb/s after compression. At this bit rate, sufficient playing time can be obtained, both for an ultrasmall camcorder cassette and the larger desktop cassette. For desktop video recording, the trend is clearly to disk-based recording, which is discussed later in this chapter. The compression system has been optimized for low-cost intraframe coding because of the use in portable equipment. The feedforward coding architecture is different from MPEG-based recording in DVD, but it offers high robustness for high-speed search and error concealment. The compression in fixed-length segments offers besides robustness also easy packet transmission over digital links. The system yields a high picture quality, even in the case of motion, owing to a special motion adaptivity.

The independent compression and coding of segments, based on five macroblocks only, allows for high search speeds during trick play and provides a very robust tape format. In normal play, the powerful R-S product ECC ensures an error-free and flawless reproduction of the recorded images. A special macroblock format mapped

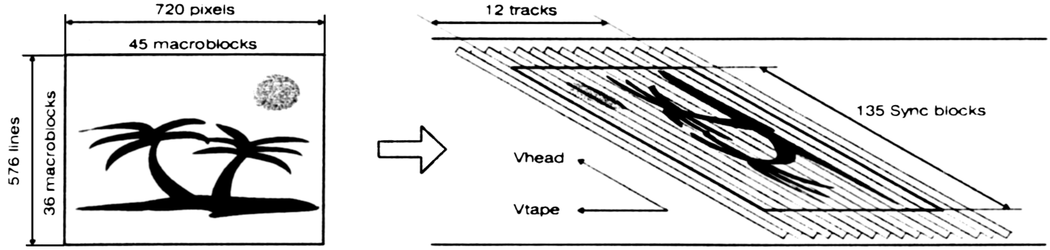

FIGURE 21.3.24 Reorganization of macroblocks on tape for trick play so that the picture is stored in coherent form on tape.

onto single channel sync blocks enables data recovery at very high search speeds. Even at high search speeds, the format limits error propagation severely and shows only errors in high-frequency information. Moreover, the special data shuffling scheme enables a relatively high perceptual performance during trick-play modes.

CD-I AND VIDEO CD

Introduction to Optical Recording

Optical recording can support much higher packing densities than magnetic recording. A typical optical recorder has a 1.6 μm track spacing, and a minimum recorded wavelength of 1.2 μm. A single side of a 30-cm optical disc can record as much as 60 min of video. Several systems are in use since the early 90s.

The first system is designed for mass replication. An expensive master produces numerous very cheap copies, in a similar manner to phonograph discs. However, the process is more refined because of the smaller dimensions involved. This system is highlighted subsequently in more detail. The second system uses an irreversible change produced by laser heating selected spots in the medium. The change may be melting, bubble formation, a dye change, or a chemical change. The systems in which writing is irreversible are termed *write once* or WORM (write once, read many times) systems. A typical example is the Panasonic 20-cm optical recorder that stores nearly 14 min of an NTSC signal. The disc rotates at 1800 rpm and the system can either record a single spiral track, or 24,000 concentric tracks.

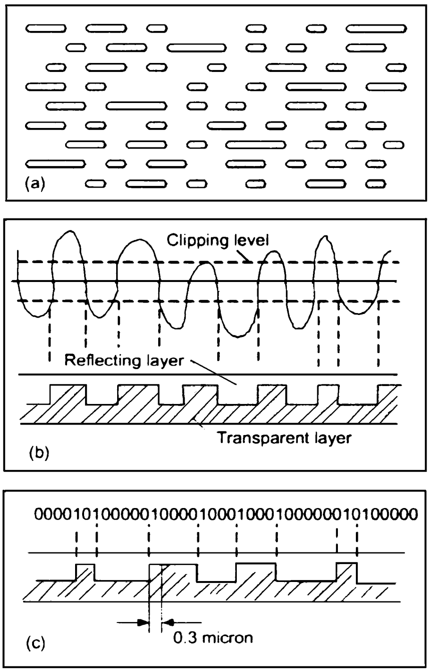

FIGURE 21.3.25 (*a*) The pits in an optical disk of Laservision or CD with tracks of about 0.5 μm width, (*b*) the analog waveform in Laservision is clipped and the transitions defining pits and walls on the disc, (*c*) the conversion of a binary string into pits and walls in reflective layer on the compact disc where each transition corresponds with a "1" in the binary signal. Between each binary 1, every flat distance of 0.3 μm represents a binary 0. The reflective layer is the backside of the optical disc.

The third method uses laser heating to define the small recorded area, but the recording is magnetic and therefore erasable. The system is similar to WORM systems except that the disc is coated with a thin magnetic layer. Like any magnetic recording, the signal can be erased or it can be overwritten. Such a system using analog recording can store over 30 min of NTSC television per side of a 30-cm disc. If digital recording is used, 50 s on each side of the 13-cm disc can be recorded.

We describe herein briefly a type of system that is suited for mass replication. There have been several approaches to the home optical disc system, including one pioneered by Philips, named Laservision. The Philips video disc consists of a transparent plastic material with a diameter of 30 cm and a thickness of 2.5 mm for a double-sided recording. The audio disc has grooves, whose walls move a transducer to reproduce the audio signals. The video disc must meet a requirement of much higher information density. It has no grooves and has tracks with a much finer structure. The track spacing is about one-sixtieth of that of an audio disc. The rotational speed is synchronous with the vertical frame rate to enable recording of one complete frame per revolution (1800 rpm for NTSC, 1500 rpm for PAL and SECAM). A longer-playing version of the Philips system maintains a constant track-to-pickup velocity, allowing approximately 1 h per side of recording.

Let us now briefly look into the basic optical readout technology. In both Laservision and the CD-based system, the signal is recorded in the form of a long spiral track, consisting of pits of about 0.5 μm width (see Fig. 21.3.25 (*a*)). The pits have been pressed into a plastic type of substrate with a special stamp. The stamp originates from a parent disc that was generated by a real laser-burned disc. The recorded layer is protected by a transparent coating. Since the coating is much thicker than the recorded layer, scratches and dirt on the surface of the disc are outside the focal plane of the optical pickup that reads out the data.

An essential difference between Laservision and all CD successors is that Laservision registers analog and CD holds digital information. Figure 21.3.25(*b*) depicts how the analog signal is recorded on a Laservision disc. The video signal is frequency modulated (FM) and an analog sound signal is superimposed on the FM-modulated video signal. The resulting signal is clipped for not exceeding maximum positive and negative signal values. The transitions of the signal coincide with the pits and walls on the disc. A predecessor of the video-CD was a system called *CD-Video* of Philips that recorded a digital audio signal in special modulation area on top of the FM-video signal.

Figure 21.3.25(*c*) portrays the recording of digital data on the CD. The length of each pit or wall is 0.3 μm or a multitude of this length with a certain maximum. With a special coding algorithm, called *eight-to-fourteen modulation* (EFM), the number of zeros between two ones is kept between a pre-defined minimum and maximum value. Such a code is often referred to as a *run-length-limited* (RLL) code. The bounding of zeros helps in robustness by avoiding DC energy and it helps with synchronization. As can be noticed, each transition corresponds with the bit value 1 and the intermediate bits get the bit value 0. These bits are called *channel bits* and they originate via a coding procedure from the data bits. The data bits for digital audio (CD-DA) are resulting from 16-bit audio samples that represent a stream of binary numbers, called *pulse code modulation* (PCM). An illustration of the focus motor and detection optics is illustrated in Fig. 21.3.26. To track the line of very narrow pits by a closed-circuit servo like the one depicted in the illustration, two auxiliary light beams are formed, which are slightly displaced from the centerline of the track, one on each side of the correct position. Two photodiodes are introduced into the optical paths on either side of the quadrant detectors, as portrayed in Fig. 21.3.26. The error signal generated by the difference in output of these diodes, after appropriate filtering, is used to deflect a galvanometer mirror and thus move the beam laterally to correct the error. Similar techniques can be used for optical tracking, where the signal envelope of recovered waveform after opto-electrical transformation by receiving photodiodes can be kept as high as possible.

Figure 21.3.27 depicts an outline of the signal path of the compact disc. Assuming correct servo control and focus of the laser beam, the electrical signal coming from photodiodes is amplitude controlled by a gain amplifier.

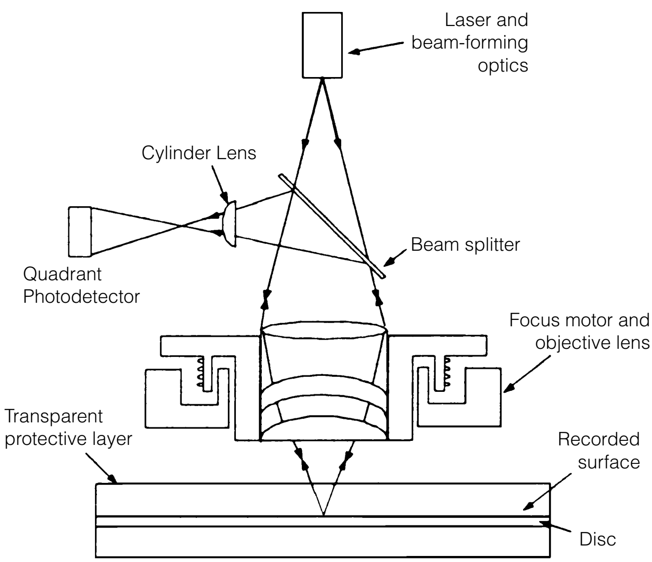

FIGURE 21.3.26 Example construction for focusing the laser beam on the optical disc.

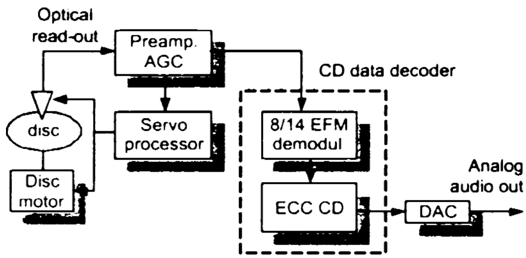

FIGURE 21.3.27 Block diagram of signal path in compact disc.

The output in the form of a digital string of channel bits is buffered and demodulated by the EFM demodulator. This involves a mapping of groups of 14 channel bits and three merging bits into consecutive bytes. Subsequently, these bytes are buffered in the error-correction decoder (ECC). This ECC decoder recomputes data packets and corrects possible errors. The decoder consists of a Reed-Solomon block decoder combined with a so-called *cyclic interleaved redundancy check* (CIRC). The latter step involves data interleaving to spread out errors and cut the length of long burst errors into shorter pieces that can be handled by the Reed-Solomon decoder. After the complete ECC decoding, the audio samples come available as PCM data and are available for digital-to-analog conversion (DAC).

History and CD-i Concept

The CD format was invented in 1980 and was introduced in Europe and Japan in 1982 and in the United States in 1983. This originally introduced format was CD-audio (CD-A, CD-DA, or "Red Book" CD, as specified in the International Electrotechnical Commission IEC-908 standard available from the American National Standards Institute). Additional CD formats were subsequently introduced: CD-ROM (1984), CD-i (1986), CD-WO [CD-R] (1988), Video-CD (1994), and CD-RW (1996).

The primary focus in this section is on the compact disc interactive (CD-i). Even if CD-i is not very much used anymore, its importance to the evolution of the digital multimedia recording systems aimed at applications such as home entertainment, education, and training has been fundamental. It should be noted that the strong development of the CD-based formats was the fundament for the recent success of the DVD,[5] which is discussed later in this chapter. Furthermore, the CD-i format resembles very much the video-CD format that is currently very popular in Asia. CD-i combines audio, video, text, graphics, and animation, while providing features such as interactivity.

At the time of its introduction, CD-i provided several advantages over PC-based interactive systems, some of which still hold today:

- *Cost* compared with that of building a PC with the same audiovisual performance and functionality.
- *Compatibility*, since each CD-i disc is compatible with every CD-i player. There are no "system requirements," such as a type of display adapter, sound card, version of the operating system, screen resolution, CD-ROM drive speeds, drivers, hardware conflicts, and so forth, like in the PC-based scenario.
- *Ease-of-use.* A CD-i player and software are very easy to use and do not require software setup, adjusting the hardware or other complex installation procedures. Moreover, CD-i can be connected to a variety of devices, such as TV sets and stereosystems. As an additional advantage, the user interfaces resemble those of CE devices, making it a far more comfortable system for many people to use over a PC.
- *Worldwide standard*. CD-i is a worldwide standard, crossing the borders of various manufacturers and TV systems. Every disc is compatible with every player, regardless of its manufacturer or the TV system (PAL, NTSC, or SECAM) that is being used.

CD-i System Specification

The CD-i players are based on a 68000 CPU running at least at 15 MHz, with at least 1 Mbyte of RAM, a single speed CD-drive, dedicated audio and video decoding chips, at least 8 kbyte of nonvolatile storage memory, and a dedicated operating system called CD-RTOS, which stands for *compact disc real-time operating system*. CD-RTOS is based on version 2.4 of Microware's OS-9/68K operating system that is very similar to the Unix operating system, and supports multitasking and multiuser operation. The operating system as well as other player-specific software such as the player's start-up shell are hard coded in a ROM of at least 512 kbyte.

CD-i Sector Format

CD-i is able to retrieve in real time the audiovisual data stored on the disc and send this information to the appropriate decoder ICs, without putting a heavy load on the overall system performance. Hence, a CD-i player does not need much RAM or processing power, since all audio and video decoding is performed in real time without storing large amounts in RAM for later decoding. To enable the simultaneous retrieval of both audio and video information, data are interleaved on the CD-i disc.

Since a CD-i disc is read at a constant continuing speed, the designer needs to be aware of the audio or video bit-stream quality. For instance, when a lower audio quality is used, fewer sectors will be occupied than with a higher quality. Alternatively, it is also possible to read only the sectors belonging to one audio channel at a time, and then move back to the beginning of the disc and read the sectors of another audio channel. Since a CD-i disc lasts for 74 min and the lowest audio quality only uses one out of every 16 sectors, the audio can be digitally recorded for (16 × 74 min) over 19 h.

Because of this real-time reading of sectors, every CD-i player reads data at the same speed, sometimes referred to as normal speed or single speed. It would be unnecessary to make a CD-i player running at a higher speed CD-drive, since data are to be read in real-time according to the specifications (thus single speed) and audio, video, and animation would be out of sync when being read at a higher speed. Special attention has been paid to the development of encoding techniques that enable high-quality audio and video within the single data speed and hence resulting in a longer playing time, instead of using a high-speed drive and by such reducing the playing time.

For CD-ROM, the mode 1 sector format is defined that allows for 2048 bytes of user data in every sector, with an accompanying 280 bytes of error correction information in each sector. When data are read at 75 sectors per second (the normal CD speed), this results in a data rate of 150 kbytes per second. For CD-i, it is not always necessary to have error correction in each sector. For example, audio and video need a much lower degree of correction than data. Instead, the 280 bytes used for error correction in mode 1 could be added to the 2048 of user bytes, resulting in 2324 bytes of user data per sector. This larger sector size results in an improved data rate of about 170 kbyte per second, which is referred to as mode 2. Within mode 2, two forms were defined: form 1 does incorporate the original error correction and is used for data and form 2 which lacks the error correction is used for multimedia. Mode 2 added an additional header to the header of mode 1, which holds information about the type of data that are contained in a sector (audio, video, data, and so forth), the way it is coded (for example, which audio level is used), and an indication of the used sector form. This header is interpreted by the CD-i system for each sector, which is then processed by the appropriate decoders. Both forms of mode 2 sectors can be interleaved, so that program data and audio and video can be read instantaneously from the disc.

Note that when all sectors are form 1, the disc holds 648 Mbyte. When all sectors are form 2, the capacity is 744 Mbyte. CD-i's disc capacity can hence be between 648 and 744 Mbyte. Although a CD-i disc consists of only mode 2 sectors, a CD-i system must be able to read mode 1 sectors on CD-ROM discs, and of course the audio sectors that are defined for CD audio.

Physical Dimensions and Specifications

A CD disc is 120 mm in diameter (60 mm radius), with a hole 15 mm diameter (7.5 mm radius) and a thickness of 1.2 mm. Starting at the hole edge at 7.5 mm radius, there is a clamping area extending from 7.5 to 23 mm radius (this area is partly clear and partly metalized, and may include a visible inscription stamped by the manufacturer), then a 2 mm wide lead-in area extending from radius 23 to 25 mm (containing information used to control the player), then the 33 or 33.5 mm wide data area (program area) extending from radius 25 to a maximum of c. 58 mm, a lead-out area (which contains digital silence or zero data) of width 0.5 to 1 mm from radius starting maximally at c. 58 mm, and finally at c. 1 mm unused area extending to the outer edge.

CD-i Disc Structure

A CD-i disc is divided into tracks. A CD-i disc contains at least one CD-i track, and may also optionally contain additional CD-Audio tracks that may also be played on a standard CD-Audio player. The first 166 sectors of the CD-i track are message sectors, followed by the disc label. Subsequently, an additional 2250 message sectors follow that contain a spoken message in CD-Audio format, which informs users who put the disc in a regular CD-Audio player about the possible damage to equipment or speakers when the disc is not taken out immediately. Usually, a modern CD-Audio player will recognize the CD-i track as a data track and will not play it, so you won't hear the message. The disc label contains some specified fields that offer a lot of information about the disc, such as the title and creator, but also the name of the CD-i application file that needs to be run at start-up. Furthermore, the disc label contains the file structure volume descriptor, which is loaded into RAM at start-up. This allows the system to find a certain file on a CD-i disc in only one stroke. After these message sectors and disc label, the actual CD-i data start.

CD-i Audio

A minimal configuration (denominated also "Base Case") of a CD-i player should be able to decode standard PCM audio as specified for CD-Audio, as well as a dedicated audio coding scheme called *adaptive delta pulse code modulation* (ADPCM). The difference with PCM is that audio is not stored individually per time segment, but that only the difference (delta) to the previous sample is recorded. Owing to the existing correlation between adjacent samples, a significant decrease in the used storage space on the disc can be achieved, and hence in the bit stream being read from the disc. When normal PCM CD-Audio would be used (which occupies all successive sectors), this would not leave room for video or animations to be read without interrupting the audio playback.

CD-i provides three levels of ADPCM audio, all of which can be used either in mono or stereo, as shown in Table 21.3.12. Note that level A provides Hifi quality, Level B gives FM radio quality, while Level C is for voice quality. Level C mono can be used for up to 16 voice channels, e.g., in different languages. The sector structure facilitates switching between languages on the fly, as the sectors are interleaved on the disc.

Thus, when ADPCM Level C is used, only 1 out of every 16 sectors needs to be used for audio, leaving all other sectors for other data such as video or animation. It is also possible to record different audio channels at once, allowing for the seamless switching between, e.g., various languages. The disc may also be read from the beginning while decoding a different audio channel, allowing for increased audio-playing times, as indicated in Table 21.3.12.

A CD-i player equipped with a digital video cartridge is also able to decode MPEG-1 Layer I and II audio. MPEG is far more efficient in coding audio, resulting in an even longer playing time while providing a highly increased audio quality when compared to ADPCM. This is because of the fact that MPEG audio is based on precision adaptive subband coding (PASC),[35] which uses perceptual coding to only store those audio signals that are audible, while filtering out other signals. Note that CD-i offers a very flexible way of using MPEG audio (for example, at various bit rates and quality levels), but cannot decode MPEG-1 Level III, or MP3 files.

CD-i Video

The video image of a CD-i player consist of four "planes," which are overlaid on top of each other. The first plane is used by a cursor and its size is limited to 16 × 16 pixels. The second and third planes are shown

TABLE 21.3.12 Different ADPCM Formats

Format CD-audio PCM	Frequency (kHz) 44.1	Numbers of bits per sample 16	Used sectors all sectors	Recording time up to 74 min
ADPCM Level A stereo	37.8	8	1 in 2 sectors	up to 2.4 h
ADPCM Level A mono	37.8	8	1 in 4 sectors	up to 4.8 h
ADPCM Level B stereo	37.8	4	1 in 4 sectors	up to 4.8 h
ADPCM Level B mono	37.8	4	1 in 8 sectors	up to 9.6 h
ADPCM Level C stereo	18.9	4	1 in 8 sectors	up to 9.6 h
ADPCM Level C mono	18.9	4	1 in 16 sectors	up to 19.2 h

underneath the cursor and are used for full screen images. The fourth plane is used for a single-colored background or for MPEG full motion video (or to display video from an external source on some players). Parts of an image on one of the middle two planes can be transparent, so that the underlying plane becomes visible. This can be used, for example, to show subtitles or menus on an image. Both planes can also be used for blending and fading effects.

There are various encoding techniques for video that can be used in CD-i, which are indicated below.

DYUV. DYUV or Delta YUV is used for the encoding of high-quality photographs and other natural images. It is based on the fact that the human eye is more sensible to differences in brightness than to differences in color. Therefore, it stores one color for a set of pixels, and a brightness value for each pixel. The result is an image of slightly more than 100 kbyte. Owing to the complexity of a DYUV image, the storage on the disc must take place in advance, and it cannot be created nor modified in the player. DYUV is used mostly in CD-i titles because of its high quality and efficient storage.

RGB555. RGB555 is a compression format that allows only 5 bits per *R*, *G*, and *B* value, resulting in a picture with a maximum of over 32,000 colors. Since RGB555 uses both planes to display the image, it cannot be used in combination with other graphics. An RGB555 image is roughly 200 kbyte in size. The image can be altered by the player at run time. RGB555 is actually never used in regular CD-i titles because of its inefficiency and limitations in usage.

CLUT. CLUT, or color look-up table, is a compression method aimed at coding simple graphics. The colors used in a certain picture are stored in a CLUT-table, which reduces the size of the image dramatically, because color values refer to the appropriate CLUT-entry instead of indicating, for example, a 24-bit color value. In CD-i, a CLUT image can have an 8-bit (256 colors), 7-bit (128 colors), 4-bit (16 colors), or 3-bit (8 colors) resolution.

Run Length Encoding (*RLE*). RLE is a variation of the CLUT compression method that besides storing the CLUT-color table in an image, further reduces the image size by storing certain "run lengths" of repeating horizontal pixels with the same color. The results are usually pictures between 10 and 30 kbyte in size. This makes RLE ideal for compressing animations that contain large continuous areas with similar colors.

QHY. QHY, or quantized high Y, is an encoding technique that combines DYUV and RLE, resulting in a very sharp high-quality natural image, that is displayed in CD-i's high-resolution mode. A QHY image is usually about 130 kbyte in size. Since it consists of a DYUV component, it cannot be modified by the player. QHY is, for example, used to display the images of a photo-CD in high resolution on a CD-i player.

CD-i can display both main planes in either normal, double, or high resolution, which are 384×280, 768 times a DYUV image is always standard resolutions. It is possible for the images on each of the planes to be displayed 280 and 768×560 pixels, respectively. Some encoding techniques are limited to a single resolution, for example at once, even if they are in different resolutions. For example, a double-resolution CLUT4 menu bar can be overlayed on a standard resolution DYUV image. CD-i highest resolution (768×560 pixels), used for QHY images, is the highest resolution that can be made visible on a normal TV set.

To enable audiovisual data to be coded at CD-i bit-rates, MPEG-1 compression has been employed. This standard allows CD-i to display 384×280 progressive video sequences, at a video quality roughly comparable to standard VHS.

Full Motion CD-i Players

The quality of the "Base Case" CD-i players can be further extended to "Full Motion" audiovisual quality.[9] For this, two coding formats are defined for video sectors: "MPEG video" and "MPEG still picture" for storage of MPEG-coded progressive video and interlaced still picture data. One new coding format is defined for audio sectors: "MPEG audio" for storage of coded MPEG audio data. All these video, still picture, and audio sectors are a form 2 sector with a usable datafield of 2324 bytes, while the Full Motion application program is stored in a normal data sector.

To code audio, CD-i applies MPEG layer I and layer II. Audio can be coded in stereo and in mono, at varying bit rates. For instance, for speech and other audio not requiring a high quality, the lowest bit rate for mono of 32 kbit/s can be used. For high-quality audio, a bit rate of 192 kbit/s in stereo is necessary to achieve similar quality to the compact disc. Audio can be coded in stereo, joint stereo (i.e., intensity stereo), dual channel, or single channel. All bit rates allowed by MPEG for these layers are supported. The applied audio sampling frequency is 44.1 kHz. The bit rates can vary from 32 to 448 kbit/s for layer I and from 32 to 384 kbit/s for layer II. For more information, see Ref. 9.

The Full Motion system in CD-i[41] supports the video parameters defined in the video part of the MPEG standard, but some parameters have a larger range in the CD-i system. For instance, while in the MPEG-1 specification the bit rate should not exceed 1.856 Mb/s; in the Full Motion system a maximum bit rate of about 5 Mb/s is allowed. Also, in the Full Motion system, video coding with a variable bit rate is allowed, leading to a higher visual quality. The maximum picture width is 768 pixels and the maximum height is 576 lines. The supported picture rates are 23.976, 24, 25, 29.97, and 30 Hz. The maximum size of the VBV buffer is 40 kbyte. More information on the specific video parameters employed for CD-i Full Motion video can be found in Ref. 9. Digital video is displayed on the background plane and can be overlayed with images coded in CLUT or RLE format.

The system part of the MPEG standard applies a multiplex structure consisting of packs and packets. In each MPEG video sector, still picture sector, and audio sector, one pack is stored. Each pack consists of a pack header and one or more packets containing a packet header followed by data from one audiovisual stream. More information can be found in Ref. 9.

The authoring system provided by the CD-i[41] allows to optimize the audiovisual compression parameters to the requirements of the application. On a CD-i disc multiple MPEG audio and video streams can be recorded in parallel, e.g., for applications requiring audio in different languages. Moreover, during the authoring process, trade-offs can be made between the number of streams, the bit rate, the audio or picture quality of each stream, and, in the case of video, the picture size.

Compatibility to Other Formats

A CD-i player only plays discs that incorporate a dedicated application program that was designed for CD-i's operating system and hardware components. For some disc types, such as video-CDs, photo-CDs, and CD-BGM, this CD-i application is a mandatory part of the disc's specification. Next to this, the CD-i standard requires the player to be able to play back standard CD-Audio discs.

Differences between Video-CD and CD-i. A video-CD is a compact disc with up to 75 min of VHS quality video with accompanying sound in CD quality. Audio and video are coded according to the MPEG-1 standard and the disc layout (see Fig. 21.3.28) is based on the CD-i Bridge (see Fig. 21.3.29) specification to allow for the playback on a variety of plackback devices such as CD-i players and dedicated video-CD players.

Video-CD became very popular mainly in Asia, while elsewhere, it is mainly used as a prototype tool. Although video-CD compatibility is not required for DVD-video players, it is very likely that video-CD playback

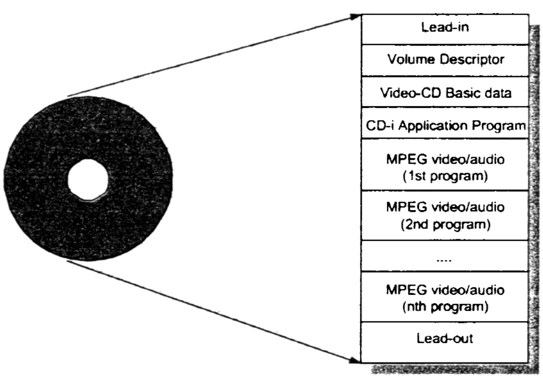

FIGURE 21.3.28 Video CD—disc data allocation.

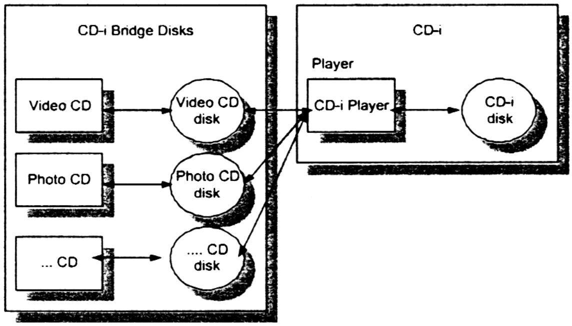

FIGURE 21.3.29 CD-i bridge.

functionality is included since every DVD-video player must be able to decode MPEG-1 as well. Another difference is that the resolution of the MPEG video on a CD-i movie disc is slightly higher than the defined resolution of a video-CD disc (384 × 280 for the CD-i "Base Case" players instead of 352 × 240 for video-CD). This also prevents extracting the video from a CD-i disc in order to burn a video-CD. To do this, reencoding of the video according to the White Book (video-CD) specification is necessary, leading to a decreased picture quality.

BRIEF DESCRIPTION OF MPEG VIDEO CODING STANDARD

The Meaning of I, P, and B Pictures

MPEG video coding is described here because it applies both to VCD (and CD-i) and the DVD system that will be described next. Many good overviews of the MPEG video coding standards can be found in a variety of books[13] and articles.[11] The MPEG video compression algorithm relies on two basic techniques: block-based motion compensation for the reduction of the temporal redundancy and DCT-based compression for reducing spatial correlations. The DCT-based compression has been described in detail in the section on DV recording. The intraframe DCT coder of DV comes close to the MPEG video processing with respect to DCT, quantization, and VLC coding. Therefore, the focus is in this section on exploiting the temporal redundancy.

In MPEG, three picture types are considered: intra pictures (I), predicted pictures (P), and bidirectional prediction pictures (B). Intra pictures are intraframe coded, i.e., temporal correlation is not exploited for the compression of these frames. I pictures provide access points for random access but achieve only moderate compression, because as their name indicates, compression is limited within the same picture. Predicted pictures, or P pictures, are coded with reference to a past I or P picture and will in general be used as a reference for a future frame. Note that only I and P pictures can be used as a reference for temporal prediction. Bidirectional pictures provide the highest level of compression by temporally predicting the current frame with respect to both a previous and current reference frame. B pictures can achieve higher compression since they can handle effectively uncovered areas, since an area just uncovered cannot be predicted from the past reference, but can be properly predicted from the future reference. They also have the additional benefit that they decouple prediction and coding (no error propagation is incurred from the prediction errors of B pictures).

In all cases, when a picture is coded with respect to a reference, motion compensation is applied to improve the prediction and the resulting coding efficiency. The relationship with respect to the three picture types is illustrated in Fig. 21.3.30. The I pictures and the B and P pictures predicted based on it form a group-of pictures (GOP). The GOP forms an information layer in MPEG video and is from the data point of view buildup as shown at the lower side of Fig. 21.3.30. First, the I picture is coded at the start and kept in a memory in the encoder. These memories are indicated at the bottom in Fig. 21.3.31. Second, the next

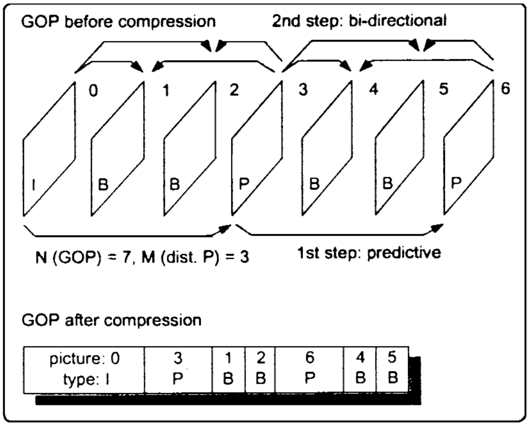

FIGURE 21.3.30 A group-of-pictures (GOP) divided in I, P, and B pictures.

P picture is selected and coded with reference to the I picture. The pictorial result after local decoding is also stored in a second memory. Third, the two B pictures in between are processed in sequential order. Each B picture is coded depending on the past I picture and the near P picture in future. This explains the term *bidirectional*. When the B pictures in between have been processed, the process repeats itself and the next P picture is first coded and stored as reference, and so on. Note that MPEG video coding as a result changes the transmission order of pictures over the channel (see bottom of Fig. 21.3.30 for the modified order).

MPEG Coding and Motion Compensation

Coding and processing with reference to a past and future picture involves motion compensation (MC). This means that the motion of objects is taken into account and predicted and only the difference between

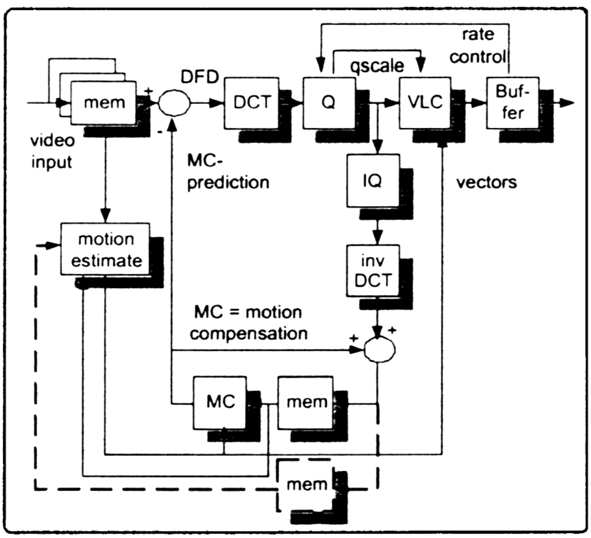

FIGURE 21.3.31 MPEG video encoder block diagram.

the temporal prediction and the actual video data is coded (see Fig. 21.3.31). In order to predict the motion of video data, it is first measured in an initial step, called *motion estimation* (ME). Motion compensation and estimation in MPEG are based on block-based processing. Hence, the blocks of video samples in actual and reference picture are compared. The typical block size for comparison in MPEG is a macroblock of 16×16 pixels.

There is a trade-off between the coding gain due to motion compensation and the cost of coding the necessary motion information (prediction type, motion vectors relative to the reference pictures, and so forth). Hence, in MPEG, the choice of 16×16 macroblocks for the motion compensation unit is the result of such a trade-off. The motion information is coded differentially with respect to the motion information of the previous adjacent macroblock. The differential motion information is likely to be small, except at object boundaries, and it is coded using a variable-length code to provide greater efficiency.

To reduce the spatial correlation in I, P, and B pictures, DCT-based coding is employed. After the computation of the transform coefficients, they are quantized using a visually weighted quantization matrix. Subsequently, after quantization the coefficients are zigzag scanned and grouped into (run, amplitude) pairs. To further improve the coding efficiency, a Huffman-like table for the DCT coefficients is used to code (run, amplitude) pairs. Only those pairs with a high probability of occurrence are coded with a VLC, while less likely pairs are coded with an escape symbol followed by fixed-length codes, in order to avoid extremely long codewords and reduce the cost of implementation. All these steps are indicated in Fig. 21.3.31.

The quality of video compressed with the MPEG-1 algorithm at rates about 1.2 Mb/s is comparable to VHS quality recording. The quality of MPEG-2 video is higher, but it is also operated at a higher bit rate between 3 to 5 Mb/s. This is discussed further in the next section about DVD.

THE DVD SYSTEM

Introduction and History

The DVD system was introduced in 1996[14] and has experienced a strong growth of interest and sales in the consumer area. The key factors of its success are the very high quality of the reproduced video and audio signals and the fact that the storage is based on a high-capacity optical disc medium. The latter factor builds clearly on the success of the compact disc technology, albeit with strongly improved densities to boost the total storage capacity of the disc.

The success of DVD results also from the perfect fit with the trends that are taking place in the consumer and computer industry. In the 90s, computer applications have been expanding continuously toward high-quality audio and video recording and playback functions. For example, video compression of MPEG-1[10] and MPEG-2[12] have been adopted gradually and the CD audio playback function evolved to CD-ROM and later CD-R/CD-RW recording subsystems in the multimedia computers. Nowadays, DVD-ROM has become an integral part of the modern multimedia computer, thereby solving (temporarily) the ever increasing request for more storage capacity on a removable medium. DVD-video players share most of the same physical and mechanical components with DVD-ROM drives in the computer. In the coming years, DVD recording (R/RW), which is already introduced, will occupy a significant part of consumer and general computing applications.

A DVD holds 4.4 to 15.9 GB (gigabytes), which is the same as 4.7 to 17 billion bytes. The capacity improvement of the typical DVD of 4.7 GB compared to CD (of 0.65 GB) results from different factors, which are shown in Table 21.3.13. The expansion to larger capacities results from extra bit layers, which will be elucidated in the next subsection. The DVD capacity was tuned to satisfy one of the main applications: storing a full-length and full-quality movie on a single disc. Over 95 percent of the Hollywood movies are shorter than 2 h 15 min, so that 135 min was taken as a guideline for the design of a digital video disc system. Uncompressed movies, e.g., with 4:2:2 10-bit sampling, run at 270 Mb/s and consume more than 32 Mbyte/s, requiring more than 250 gigabytes storage capacity. With resampling on a somewhat lower resolution to 124 Mb/s and MPEG-2 video coding offering a compression factor of 30, the average bit rate is reduced to 3.5 Mb/s for the video signal. Additionally, on the average, three tracks of audio consuming 384 kb/s each are added to come to 4.4 Mb/s average bit rate.

TABLE 21.3.13 Key Parameters of CD and DVD Standard Where Improvement Factors in Both Density and Channel Coding Lead to an Improvement of a Factor of 7 in Capacity

	CD	DVD	Factor
Pit length	0.83 μm	0.40 μm	2.08
Track pitch	1.6 μm	0.74 μm	2.16
Data area surface	8605 mm^2	8759 mm^2	1.02
Channel modulation ratio	8/17 EFM	8/16 EFM+	1.06
Error correction ratio	1064/3124 (34%)	308/2366 (13%)	1.32
Sector overhead ratio	278/3390 (8.2%)	62/2418 (2.6%)	1.06

Unlike the DV format, almost everything in the DVD standard is variable. The previous discussion on playing time was entirely based on average numbers. Let us consider some variations on the application. If the movie producer skips two audio tracks, the playing time increases to a near 160 min. The difference can, e.g., be used to increase the picture quality. The maximum bit rate can be as high as 9.8 Mb/s for the video and the total bit rate including audio is limited to 10.08 Mb/s. This maximum rate can be sustained and would lead to a playing time of 62 min. On the other side of the spectrum, if a movie is recorded in MPEG-1 at about 1.2 Mb/s, the playing time becomes more than 9 h. This discussion applies to DVD-video. When considering DVD-ROM, the flexibility in usage and the recorded data is infinite. If a new video coding technique is found in the future, the disc may be used to store more than 3 h of high-quality film.

Bit Layers in DVD

The increase in storage capacity from 4.7 to 15.9 gigabytes results from a new technology that was applied in the DVD system, where information bits can be stored in more than one layer. The laser that reads the disc can be focused at two different levels. Hence it can read the top layer or it can look through the top layer and read the layer beneath. This process is enabled by a special coating on the top layer, which is semireflective, thereby allowing the laser wavelengths to pass through. The reading process of a disc starts at the inside edge and gradually moves toward the outer edge, following a spiral curve. The length of this curve is about 11.8 km. If the laser reaches the end of the first layer, it quickly refocuses on the second layer and starts reading backwards. The switch to the second layer takes place in the order of 100 ms, which is fast enough to match this electronically with data buffering in order to prevent hiccups in presentation of the video or audio data. The reading process of bits from the disc is faster than the decoding of it to compensate for time losses from refocusing. Possibilities for various forms of reading and layering are shown in Fig. 21.3.32. First, DVDs can be single-sided or double-sided. Second, each side can have one or two information layers, indicated with the tiny square waves in Figs. 21.3.32(*a*) through 21.3.32 (*e*). The information layers are in the middle of the disc for protection and enabling the double-sided options. A double-sided disc is constructed by binding two

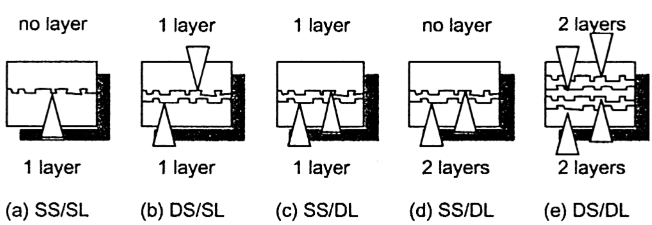

FIGURE 21.3.32 Various configurations of bit layers and sides: SS = single side, DS = double side, SL = single layer, DL = double layer.

TABLE 21.3.14 DVD Capacities Resulting from the Various Combinations of Layers, Sides, and Disc Sizes. The Triangles Represent the Focused Laser Beam. CD-ROM is Added for Reference (Assuming Four Times the Speed)

Type	Diam.	Sides	Layers	Gigabytes	Pl.-time/(h)
DVD-5	12 cm	SS	SL	4.70	2.25
DVD-9	12 cm	SS	DL	8.54	4.0
DVD-10	12 cm	DS	SL	9.40	4.50
DVD-18	12 cm	DS	DL	17.08	8.0
	8 cm	SS	SL	1.46	0.75
	8 cm	SS	DL	2.66	1.25
	8 cm	DS	SL	2.92	1.5
	8 cm	DS	DL	5.32	2.5
CD-ROM	12 cm	SS	SL	0.68	0.25

stamped substrates back to back. Single-sided discs have a blank substrate on one side. There are even more possibilities, since DVDs can also vary in size, i.e., by using 8 cm or 12 cm diameter for the disc. The most important options are shown in Table 21.3.14. Physically, each substrate has a thickness of 0.6 mm. The thickness of a DVD is thus 1.2 mm so that it equals the thickness of a CD. The glued construction of the DVD makes them more rigid than a CD, thereby enabling more accurate tracking by the laser and thus a higher density.

A special topic is backward compatibility to the CD-ROM format. Although this is not required, the key manufacturers see this point of vital importance, because of large market presence of CD formats. One of the problems is that with CD, the data are within a stamped metal layer deep inside the disc (virtually at the backside), whereas in DVD it is in the middle. The focusing is therefore different. This leads to a player with three focusing levels. Another problem is that CD discs do not well reflect the 635 to 650 nm wavelength transmitted by a DVD player. This sometimes leads to extra lasers or optical devices. A last aspect is the compatibility with data formats (VCD, CD-i, photo CD, and so forth) requiring extra electronics inside data decoding circuitry. In practice, most DVD players have backward compatibility with VCD and sometimes with one or more CD-R/RW formats.

The DVD Disc

The disc is read from the bottom, starting with data layer 0. The second layer, called layer 1, is further from the readout surface. The layers are actually very closed, spaced about 55 μm from each other. The backside coating of the deepest layer is made from metal (such as aluminium) and is thus fully reflective, while it has a semitransparent coating on top. Layer 1 has a semireflective coating to enable reading of layer 0. The first layer is about 20 percent reflective and the second layer 70 percent. The substrate is transparent. The laser light has a wavelength of 650 or 635 nm (red laser). The lens has a numerical aperture of 0.60 and the optical spot diameter is 0.58 to 0.60 μm.

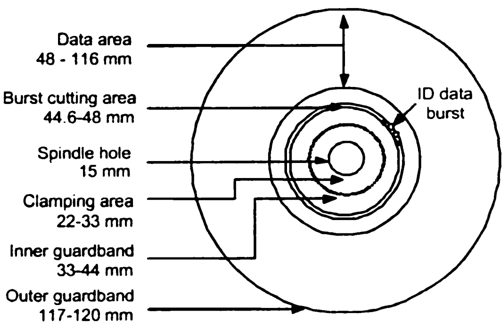

FIGURE 21.3.33 Physical 12 cm DVD disc area parameters.

Figure 21.3.33 portrays the key areas of the disc. The burst cutting area is for unique identification of a disc via a 188-byte code and can be used for automatic machines with multiple disc storage capacity (e.g., a juke box). Table 21.3.15 shows relevant parameters for recovering bits from a DVD-ROM or DVD-video. What strikes the eye here is the robustness: a scratch of 6 mm can be covered by the ECCs, which clearly outperforms the CD channel coding.

TABLE 21.3.15 Key Physical Bit Reading Parameters for DVD-ROM

Pit length SL	0.40–1.87 μm
Pit length DL	0.44–2.05 μm
Track pitch	0.74 μm
Average data bit length	0.267 (SL)/0.293 μm (DL)
Average channel bit length	0.133 (SL)/0.147 μm (DL)
Velocity	570–1630 rpm (574–1528 rpm data)
Scanning speed	3.49 m/s (SL), 3.84 m/s (DL)

DVD Data Format

The logical data format of DVD is indicated in Table 21.3.16. A key parameter is the easy to use sector size of user data of 2048 bytes that nicely fits with computer applications. The modulation code is called EFMPlus, a more efficient successor of the EFM code that was applied for the CD. This code enables a high density of bits on the disc, by limiting the number of zeros between two transitions. In this case, the code bounds the number of zeros between 2 and 8, between each group of ones. The modulation also helps keep DC energy low and provides synchronization in the channel. The ECC is a powerful R-S product code having considerably less overhead than in the CD. By using row-column processing of the rectangular data sector block of 192 by 172 bytes, it is well suited for removing burst errors resulting from scratches and dirt and so on. Note that the gross number of user data here is 11.08 Mb/s, instead of the 10 Mb/s earlier. The extra 1 Mb/s is available for extra insertion of user data such as subpictures and the like. The following figures explain the data sector construction in more detail.

The construction of a DVD data sector is portrayed by Fig. 21.3.34. The sector consists of 12 rows of 172 bytes each. The first sector starts with identification bytes (sector number with error detection on it) and copy protection bytes and the last sector ends with four EDC bytes for a payload ECC. Figure 21.3.35 depicts the conversion from data sectors into ECC blocks. All data sectors together then form a block of 192 rows of 172 bytes each. The rows of the ECC block are interleaved to spread out possible burst errors, thereby enhancing the removal of those errors. Looking vertically in Fig. 21.3.35 each of the 172 columns of the ECC block gets an outer parity check data of 16 bytes assigned to it. This extra data forms the outer parity block. Similarly, for each of the 208 rows (192 + 16 bytes), a 10-byte inner parity check is computed and appended. The power of the R-S product code is such that a (random) error rate of 2 percent is reduced to less than one error out of 10^{15} bits (million-billion). The product code works on relatively small blocks of data, resulting in a maximum correctable burst error of approximately 2800 bytes, which corresponds to about 6 mm scratch length protection.

To come to recording sectors, each group of 12 rows of the complete ECC block gets one row of parity added to it, leading to a spread of parity codes as well. Thus, a recording sector consists of 13 (12 + 1) rows

TABLE 21.3.16 Data Format Parameters of DVD

Data sector (user data)	2048 bytes
Logical data sector size	2064 bytes (2048 + 12 header + 4 EDC)
Recording sector size	2366 bytes (2064 + 302 ECC bytes)
Unmodulated physical sector	2418 bytes (2366 + 52 sync bytes)
Modulated physical sector	4836 bytes (2418 × 2 bytes)
Modulation code	8–16 (EFMPlus)
Error correction code	R-S product code (208,192,17) × (182,172,11)
ECC block size	16 sectors (32,768 bytes user data)
ECC overhead	15% (13% of recording sector)
Format overhead	16% (37,856 bytes for 32,768 user B)
User/Channel data rate	11.08/26.16 Mb/s

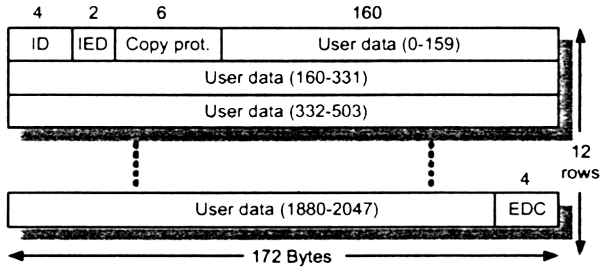

FIGURE 21.3.34 Construction of a DVD data sector.

of 182 (172 + 10) bytes. A recording sector is split in the middle and 2 bytes are inserted for synchronization and 2 sync bytes are added at the beginning of the first half. This leads to two parts of 2 + 91 bytes (total 186 bytes) and 13 rows, which form 2418 bytes together. This block is modulated to 4836 bytes by the EFMPLus code.

DVD-Video

The DVD-video system is an application of the DVD-ROM standard and aims at playback of high-quality movies together with multichannel audio. Presently, this is one of the mostly accepted standards of the DVD products. DVD-video consists of one stream of MPEG-2 video coded at variable rate, up to eight channels of Dolby digital multichannel audio or MPEG-2-coded multichannel audio or linear PCM audio, and up to 32 streams of subpicture graphics with navigation menus, stills pictures, and control for jumping forward and backward through the video program. With respect to video resolution, the format intends to serve and output both analog standard resolution TV as well as high-quality digital TV and possibly in future digital HDTV. Many multimedia computers can playback DVD-video as well.

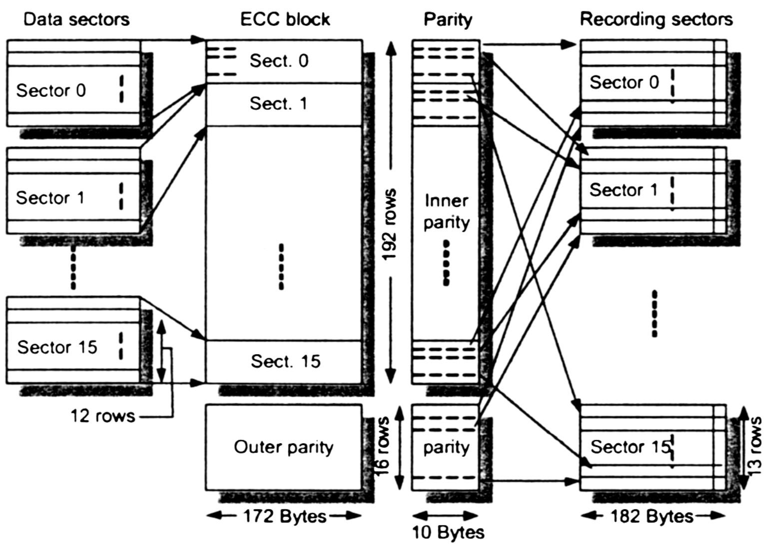

FIGURE 21.3.35 Data sectors mapped into DVD ECC blocks that are subsequently converted into recording sectors.

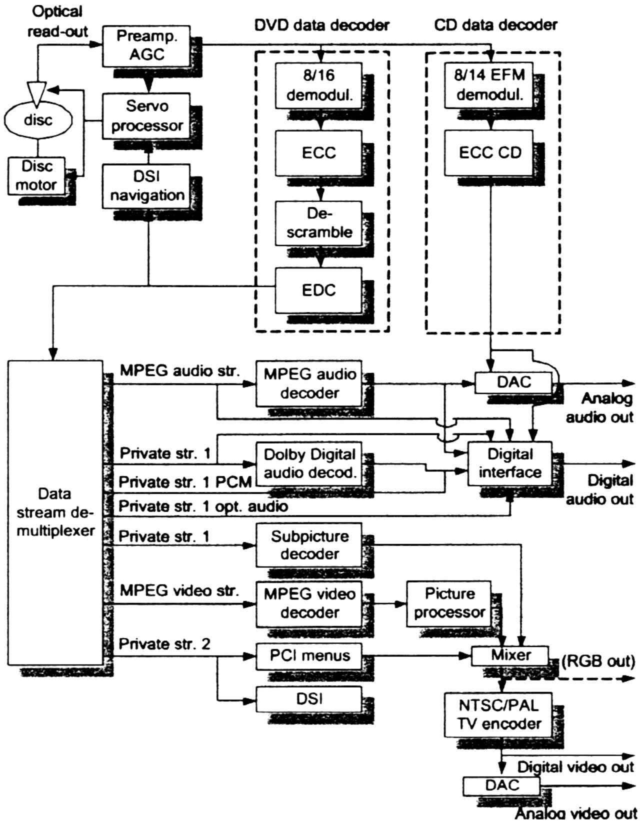

FIGURE 21.3.36 Block diagram of DVD player with CD compatibility.

The system diagram of DVD-video is depicted in Fig. 21.3.36. The diagram depicts audio data reading on disc with either CD audio data decoding or via the DVD demultiplexing of audio streams. The video results from demultiplexing the DVD data into the various data streams. Audio is decoded with MPEG or Dolby digital decoders. The amount of postprocessing is growing continuously and may involve noise reduction, sharpness improvement, and even 100 Hz conversion (Europe). The NTSC/PAL encoder ensures seamless connection to existing TV receivers.

The data flow and buffering operates as follows. The channel bits from disc are read in a constant 26.16 Mb/s rate. The EFMPlus 8-16 demodulator halves this to 13.08 Mb/s and this is reduced after ECC to a constant user data stream of 11.08 Mb/s. Data search information (DSI) is copied from this stream prior to writing this in a track buffer. A so-called MPEG program stream of variable rate is recovered at 10.08 Mb/s, which contains five different packetized elementary streams (PES): video, audio, subpicture, presentation control information (PCI), and the data search information (DSI). The latter two streams are system overhead data. The 10.08 Mb/s data rate is the maximum rate of the video and audio stream together. Video may be coded as MPEG-1 or MPEG-2 streams and has the form of a PES stream; audio, if coded as MPEG, is also a PES stream. The PCI and DSI streams are mapped as MPEG private streams. Another private stream contains the subpicture data and alternative audio data. Since audio consumes data rate, the video is limited to either 9.8 Mb/s for MPEG-2 and 1.85 Mb/s for MPEG-1. The pulse code modulation (PCM) audio stream is maximally 6.14 Mb/s, which can consist of eight channels audio sampled at 16 to 48 kHz. Note that each AV decoder has its own buffers for keeping data at the input. The buffer sizes for video, audio, subpicture, and PCI are 232, 8, 52, and 3 kbyte, respectively.

DVD Audio and Data

Audio streams in DVD are based on three different possibilities as follows:

- *PCM audio*. This is a straightforward stream of digital samples of audio without compression, referred to as LPCM. The minimum bit rate is 768 kb/s and for multiple channels, it can be maximally 6.144 Mb/s. The typical bit-rate is 1.536 Mb/s for 48 kHz sampling frequency.
- *Dolby digital.* This is based on the Dolby digital coding scheme, having a minimum, typical, and maximum bit rate of 64, 384, and 448 kb/s, respectively. The coding scheme is also known as AC-3. There can be one, two, three, four, or five channels with an optional subwoofer channel (all together indicated as Dolby 5.1). All Dolby digital decoders are required to downmix 5.1 channels to two channels stereo PCM and analog output.
- *MPEG audio*. MPEG-1 audio layer II is based on coding the audio spectrum in individual subbands. It is limited to 384 kb/s bit rate, the typical rate is 192 kb/s. MPEG-2 audio can extend this from 64 kb/s to maximally 912 kb/s. The typical rate is 384 kb/s. At maximum, seven channels can be used (7.1 with subwoofer). The advanced audio coding (AAC) mode is optimized for perception and is part of MPEG-2 audio coding. The primary audio track of DVD is always MPEG-1 compatible.

The subpicture data of 10 kb/s may grow to 3.36 Mb/s maximally, whereas the DTS data can have a rate up to 1.536 Mb/s. Typically, the rates for this type of data are rather limited in capacity. Various coding types and parameters can be used in the above options. Coding algorithms can be besides the MPEG equal to LPCM, AC-3, DTS, and SDDS. The last two modes involve optional multichannel formats (e.g., 5.1) using compression on PCM-based channels. Audio sampling can be performed at 16, 20, or 24-bit resolution. Sampling rate is 48 or 96 kHz and the dynamic range compensation (DRC) of the audio signals can be switched on and off. Along with certain applications, a code can be used, such as surround (multichannel HQ audio), karaoke (for singing with supporting subtitling), or an unspecified code.

The data are organized in files. Each side of the disk contains one logical volume, using one partition and one set of files. Each file is a contiguous area of data bytes to support easy transfer and decoding. The complete list of constraints applies to the normal DVD-video application that are used at playback. However, the format allows to record special data files that do not satisfy the above constraints. This is allowed if they are formatted into the DVD-video "Data" stream, behind the regular data that are required for playback. Then a specially designed player can access the extra data. This option allows many special applications in the future.

DVD Conclusions

DVD is an evolutionary product that successfully builds further on the widely accepted and distributed optical recording technology of the CD. The principal advantage is the excellent video and audio quality recorded on a single disc. Table 21.3.17 provides the main system parameters of DVD compared with DV and D-VHS. The table clearly shows that VHS is lagging behind in picture quality and that the digital formats of DV and DVD are comparable with each other.

The compression factor of DVD is high but the quality is approaching the DV quality in spatial resolution. The MPEG compression relies heavily on the motion compensation, which is not used in DV. Thus the temporal quality of DV clearly outperforms that of DVD and this explains why there are semiprofessional DV systems (e.g., DVCPRO) for portable studio applications.

The coming years will show further penetration of recordable DVD formats in the market (DVD-RW, DVD + RW, DVD-R[36] with the above-mentioned resolution and quality. Meanwhile, even higher capacity optical discs are being developed to pave the way for HDTV and to stay comparable with computer hard disks.

TABLE 21.3.17 DVD Compared to Other Existing Recording Systems

	DV	D-VHS	DVD
Video	Component digital	Composite analog or digital bit stream	Component digital
Playing time	0.5–1 h (mini) 4.5 h (standard)	2–6 h (VHS) 4–40 h (D-VHS)	2.25 h (1 layer) 4.5 h (2 layers)
Data capacity	5–50 Gbytes	32–40 Gbytes (300–400 m tape)	4.7–17 Gbytes
Compression	Intraframe DV factor 5	External (MPEG-2)	MPEG-1/MPEG-2 factor 30
Data rate	25.146 Mb/s video 1.536 Mb/s audio	28.2 Mb/s HD, 14.1 Mb/s STD, 2–7 Mb/s LP	Up to 9.8 Mb/s combined VBR or CBR V + A
Error correction	R-S product code + track interleave	R-S (inner/outer)	R-S product code
Channel modulation	24-25 code	NA	EFMPlus (8–16 block)
TV video systems	525/60/2:1 29.97 Hz 625/50/2:1 25 Hz	NA	525/60, 1:1 24 Hz/ 2:129.97 Hz 625/50/2:1 25 Hz 1:1/24 Hz
V 525 resolution	720 × 480 (346 kpixels)	320 × 480 (154 kpixels) VHS	720 × 480 (346 kpixels)
V 625 resolution	720 × 576 (415 kpixels)	320 × 580 (187 kpixels) VHS	720 × 480 (415 kpixels)
Audio	2 tracks of 2 channels (32 kHz, 12-bit nonlinear PCM) or 1 track of channels (32/44.1/48 kHz 16-bit PCM)	Analog VHS stereo HiFi or digital bit stream (MPEG/Dolby like)	8 tracks of up to 8 channels each (LPCM/ Dolby digital/MPEG-2)
Audio SNR	72 dB (12-bit), 96 dB (16-bit)	40 dB(mono), 90 dB HiFi	96–144 dB
Trick play, searching	Up to 70–90 times, with complete picture	Up to 10–20 times, with distorted picture (VHS)	Jumping to I pictures several speeds, stills

NETWORKED AND COMPUTER-BASED RECORDING SYSTEMS

Personal Video Recorder (PVR)

The final part of the chapter is devoted to new emerging technologies and products in recording of which most are related to further applying the computer technology into consumer systems. The personal video recorder (PVR) is a TV recorder that records on a computer hard disk as opposed to tape. The PVR is also called personal digital recorder, digital video recorder, digital network recorder, smart TV, video recording computer, time-shifted television, hard disk recorder, personal television receiver, television portal, or on-demand TV. PVR has evolved as a direct consequence of digital recording and low-cost hard disk storage. The advantages of hard-disk-based recording are:

- Allowing random access to content
- Metadata can be stored along with the content (allowing easier searching, selection and management of content, segmentation of content, richer interaction, and so forth
- Easier transfer and redistribution of content

Some of the features provided by the PVR justify its popularity in first product trials.

Simultaneous record and replay, allowing "trick" modes such as pausing and rewinding live TV. This allows pausing of a live TV program when the phone rings, later viewing of a program that has already started but is not yet finished, "fast-forward" through parts of programs when watching a time-shifted program, and so forth.

Sophisticated interactive multimedia information services, regular program and news updates that are instantly available to consumers. For instance, downloads of electronic program guides (EPGs) with the TV signal or via the phone network simplify program selection, browsing (by time, by channel, and so forth), and recording.

Automatic indexing of recorded programs allowing easy retrieval and manipulation of content.

"Intelligent" agents that can automatically record or recommend TV programs that they think the viewer might want to watch. For instance, the EPG can also include metadata describing the programs (e.g., the actors, the genre) or a brief text summary, that can be used by the PVR to manage the recording process. Consequently, the PVR can use this metadata to automatically record or recommend similar types of programs the users have recorded in the past or preferences they have entered as part of the set-up procedure.

Nonlinear viewing of programs that were recorded in a linear manner. The viewing is performed by, e.g., skipping certain segments.

Several brands of PVRs exist—TiVo, Replay TV, Ultimate TV, Axcent, and so on. Subsequently, we will explain some of the technology and issues behind the PVR, by giving some implementation examples.

Content-Format I/O and Storage. The various PVR brands use different formats for the I/O and for the storage format. This choice is very important since it influences the overall PVR architecture, the necessity for content protection as well as the resulting quality of the stored content. For instance, TiVo's I/O interfaces are analog, even for the digital services, thereby eliminating issues such as conditional access and scrambling. For storage, the analog signals are converted to digital and then compressed using MPEG-2 coding. This implies that in the case digitally broadcasted content is stored using TiVo, the audiovisual quality can potentially be decreased by the cascaded encoding/decoding process. To solve this problem, other PVR configurations store the original broadcast digital signal in scrambled form on its hard disk, giving the potential for better audiovisual quality and the possibility for the conditional access operator to control the replay of stored programs. On the positive side, however, because TiVo compresses the content, easy trade-offs between quality and storage capacity can be made. For example, if a TiVo has 40 Gigabytes of hard disk capacity, this represents about 40 h storage at "basic" quality or about 12 h at "best" quality. It should be noted, however, that the "best" quality is limited by the incoming audiovisual content quality.

Event- Versus Time-Driven Programming. To better understand the differences between these types of programming, let us consider two different implementations. For the TiVo PVR, data are downloaded daily via the telephone network, and depending on the availability of the data, TiVo can have program details for up to two weeks in advance. The TiVo data are entirely time-driven, such that the program is stored based on the a priori available time information. Alternatively, other PVRs are event-driven, i.e., detect the program junctions. For instance, Axcent PVR uses a 64 kb/s channel on the satellite transponder to which the PVR tunes by default to keep recording when a program runs late.

Software Downloading to Existing PVRs for Enabling New Services. Different services can be enabled using the PVR functionality. For instance, "trailer selection" services are expected to appear enabling the viewer to simply press a button when a trailer for a program is presented, and the PVR will automatically record the program whenever this is broadcast. This can be done using the program metadata (see next paragraph). Furthermore, other e-commerce services can also be enabled via similar mechanisms, as described below.

Metadata. Since the introduction of the PVR, it was argued that the scheduling of the programs at attractive hours will lose its power to attract viewers, because the PVR can present desired video programs that were already stored on its hard disk on-demand. If scheduling will indeed lose its power to attract the viewers, its place will probably be taken by the metadata and services that support PVR functions. Consequently, the battle for the attention of the viewers will be won by the provider that can describe programs and services most fully and attractively using metadata. Moreover, with a PVR it is possible for the viewer to skip all advertisements. Nevertheless, the content still has to be paid for, such that new mechanisms are necessary to encourage the viewers to watch commercials. To enable this, ubiquitous metadata can be employed to target advertisements at specific interest groups. Furthermore, PVR's combination of storage, processing power, and supporting metadata offers potential for new kinds of programs (e.g., educational), where the viewer interactively navigates around the stored program.

Note that standards are necessary for describing multimedia content, i.e., metadata, which is required by these services. MPEG-7 provides such a standard[37] (see the "MPEG-7" section). Furthermore, while many PVRs are already existing on the market, proprietary solutions restrict the user to a single service or content provider and lock together the broadcaster, service provider, and the PVR manufacturer. Consequently, to enable PVR proliferation, standards are necessary for metadata to choose programs, for finding the programs in a multichannel, web-connected broadcasting environment, for management of the rights that must be paid for in order to support the services, and so forth. These standards are generated in a new worldwide body, the TV-Anytime forum (see "TV-Anytime" section).

Brief MPEG-7 Overview

As mentioned in the PVR section, digital audiovisual recorded material is increasingly available to users in a variety of formats. For instance, persistent large-volume storage that allows nonlinear access to audiovisual content, such as hard-disk storage in powerful PC platforms and personal video recorders (PVR), is becoming available in many consumer devices. Consequently, there is a need for rapid navigation and browsing capabilities to enable users to efficiently discover and consume the contents of the stored audiovisual programs. Users will also benefit from having nonlinear access to different views of a particular program, adapted to the user's personal preferences, interests, or usage conditions such as the amount of time the user wants to spend in consuming the content or the resources available to the user's terminal. Such adaptability will enhance the value provided by the multimedia content. The MPEG-7 standard,[38] formally named *Multimedia Content Description Interface*, provides a rich set of standardized tools to describe such multimedia content (see Fig. 21.3.37). A good description of the MPEG-7 standard can be found in Refs. 37 and 38. In this section, we will summarize only those MPEG-7 components that can be useful in finding, retrieving, accessing, filtering, and managing digitally recorded audiovisual content.

The main elements of the MPEG-7 standard[37,38] are presented below.

- *Descriptor* (*D*) is a representation of a feature. A descriptor defines the syntax and the semantics of the feature representation.
- *Description Scheme* (*DS*) specifies the structure and semantics of the relationships between its components, which may be both Ds and DSs.

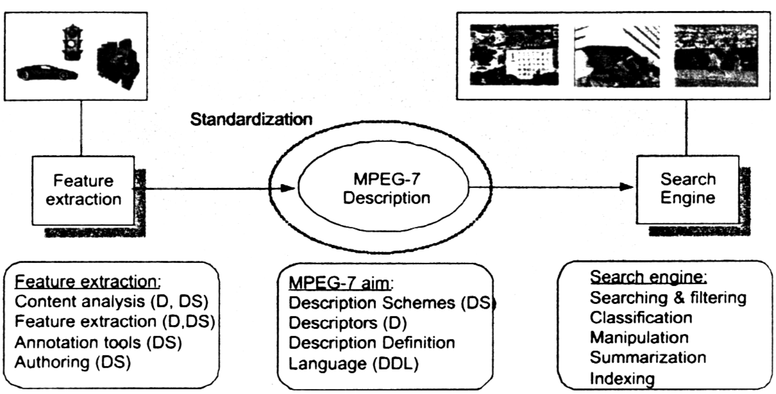

FIGURE 21.3.37 MPEG-7 standardization role.

- *Description Definition Language* (*DDL*) is a language to specify description schemes. It also allows the extension and modification of existing DS. MPEG-7 decided to adopt XML Schema Language as the MPEG-7 DDL. However, DDL requires some specific extensions to XML Schema Language to satisfy all MPEG-7 requirements.
- *Binary representation* provides one or more ways (e.g., textual, binary) to encode descriptions. A coded description is a description that has been encoded to fulfill relevant requirements such as compression efficiency, error resilience, random access, and so forth.

MPEG-7 offers a comprehensive set of audiovisual description tools that can be used for effective and efficient access (search, filtering, and browsing) to multimedia content. For instance, MPEG-7 description tools allows the creation of descriptions of content that may include:[37]

- Information describing the creation and production processes of the content (director, title, short feature movie)
- Information related to the usage of the content (copyright pointers, usage history, broadcast schedule)
- Information of the storage features of the content (storage format, encoding)
- Structural information on spatial, temporal, or spatio-temporal components of the content (scene cuts, segmentation in regions, region motion tracking)
- Information about low-level features in the content (colors, textures, sound timbres, melody description)
- Conceptual information of the reality captured by the content (objects and events, interactions among objects)
- Information about how to browse the content in an efficient way (summaries, variations, spatial and frequency subbands, and so forth)
- Information about collections of objects
- Information about the interaction of the user with the content (user preferences, usage history)

The process of generating metadata for content description that can later be used for retrieving, accessing, filtering, and managing digitally recorded audiovisual content is portrayed in Fig. 21.3.38. Note that the MPEG-7 metadata can be either physically co-located on the same storage system with the associated audiovisual content or could also be stored elsewhere. In the latter case, mechanisms that link multimedia and its MPEG-7 descriptions are needed.

Example: MPEG-7 Usage in PVR Applications

For a better understanding of the MPEG-7 importance to the field of digital recording, let us consider its benefits in a PVR application. In this application, content descriptions could be generated by a service provider,

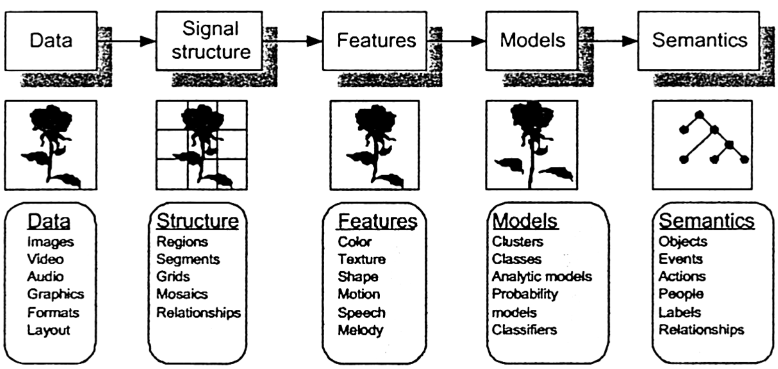

FIGURE 21.3.38 Generating metadata for content description.

separate from the original content provider or broadcaster. Certain high-level content descriptors, such as the program name and channel name, can be downloaded by the PVR in advance to provide an on-screen electronic program guide (EPG). This EPG enables the user to efficiently navigate at the program level and record programs easily. Moreover, summary descriptions can be made available in advance as well (e.g., in the case of movies), or downloaded at the end of a program (e.g., in the case of sports events). Furthermore, low-level descriptions that describes the various features of the content can also be provided. These descriptions can include information for video transcoding (i.e., transcoding hints), that can be used by the PVRs or transcoding proxies whenever transcoding from the "source" (service provider) content quality to a lower quality that can be used for local storage is desired.

Additionally, other descriptions may be generated by the user, e.g., by marking highly entertaining segments for later review. Such a feature can be simply provided by the PVR by copying the XML-fragment associated with the selected segment, including its name and locators and storing this element separately, along with some high-level elements that will allow its easy identification at a later time. Such fragments can be exchanged with friends or relatives. MPEG-7 is also developing description schemes that can capture the user's preferences with respect to a specific content, and store them on the PVR under the user's control. These schemes support personalized navigation and browsing, by allowing the user to indicate the preferred type of view or browsing, and automatic filtering of content based on the user's preferences.

TV Anytime

The *TV Anytime Forum*[42] is an international consortium of companies dedicated to producing standards for PVRs. TV Anytime aims at developing a generic framework that incorporates standards, tools, and technologies for an integrated system providing a multitude of services such as movies on demand, broadcast recording, broadcast searching and filtering, retrieving associated information from web pages, home banking, e-commerce, home shopping, and remote education. To enable this vision, TV Anytime will define specifications

- That will enable applications to exploit local persistent storage in consumer electronics platforms
- That are network independent with regard to the means for content delivery to consumer electronics equipment, including various delivery mechanisms (e.g., ATSC, DVB, DBS, and others) and the Internet and enhanced TV
- For interoperable and integrated systems, from content creators/providers, through service providers, to the consumers
- That provide the necessary security structures to protect the interests of all parties involved

Two important components in the TV Anytime framework are the digital storage and recording and the metadata, because they allow consumers to access the content they want, whenever they want it and how they

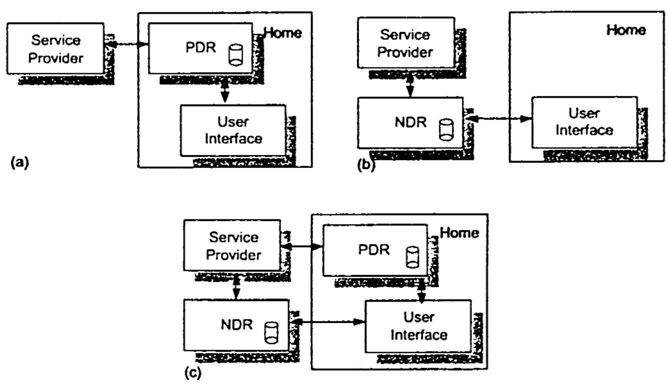

FIGURE 21.3.39 Various digital storage models: (*a*) in-home consumer storage—personal digital recorder (PDR), (*b*) remote consumer storage—network digital recorder (NDR), (*c*) PDR + NDR combination.

want it (i.e., presented and tailored according to the user preferences/requests). Metadata can be easily stored along with the content to enable searching, selection, and management of the content in a much easier fashion as compared to the current analog VCRs, and also allow a richer interaction with the stored content. Other benefits of local digital storage are that the TV viewers can order a program to be recorded using a single button during a trailer, making available intelligent agents that based on the stored user preferences can record TV programs that they think a viewer might want to watch and consuming a program in a nonlinear rather than linear manner (e.g., a news program). Note also that the digital recording process can be performed within the TV Anytime framework locally, remotely, or in a combined manner (see Fig. 21.3.39).

Hence, digital recording of content on local storage and digital broadcasting together with the framework provided by TV Anytime for content referencing and location resolution, metadata and rights management and protection, are providing significant benefits as compared with alternative forms of content delivery such as analog broadcasting, Internet, and broadband networks.

REFERENCES

1. Y. Shiraishi, "History of Home Videotape Recorder Development," *SMPTE J.*, pp. 1257–1263, December 1985.
2. K. Sadashige, "Transition to Digital Recording: An Emerging Trend Influencing All Analog Signal Recording Applications," *SMPTE J.*, pp. 1073–1078, November 1987.
3. H. Sugaya, and K. Yokoyama, Chapter 2 from C. D. Mee and E. D. Daniel, *Magnetic Recording*, Vol. 3, McGraw-Hill, 1988.
4. S. B. Luitjens, "Magnetic Recording Trends: Media Developments and Future (video) Recording Systems," *IEEE Trans. Magnetics*, Vol. 26, No. 1, pp. 6–11, January 1990.
5. M. Umemoto, Y. Eto, and T. Fukinuki, "Digital Video Recording," *Proc. IEEE*, Vol. 83, No. 7, pp. 1044–1054, July 1995.
6. J. C. Whitaker, and K. Blair Benson, Standard Handbook of Video and Television Engineering, 3rd ed., Chapter 8, McGraw-Hill, 2000.
7. S. Gregory, *Introduction to the 4:2:2 Digital Video Tape Recorder*, Pentech Press, 1988.
8. R. Brush, "Design Considerations for the D-2 NTSC Composite DVTR," *SMPTE J.*, pp. 182–193, March 1988.
9. J. van der Meer, "The Full Motion System for CD-I," *IEEE Trans. Consum. Electron.*, Vol. 38, No. 4, pp. 910–920, November 1992.
10. *Coding of Moving Pictures and Associated Audio*, Committee Draft of Standard ISO 11172, ISO MPEG 90/176, December 1990.
11. D. LeGall, "MPEG: A Video Compression Standard for Multimedia Applications," *Commun. ACM*, Vol. 34, No. 4, pp. 46–58, April 1991.
12. *Coding of Moving Pictures and Associated Audio*, International Standard, ISO/IEC 13818, November 1994.
13. J. L. Mitchell, W. B. Pennebaker, C. Fogg, and D. LeGall, *MPEG Video Compression Standard*, Chapman and Hall, 1996.
14. J. Taylor, *DVD Demystified: The Guide Book for DVD-ROM and DVD-Video*, McGraw-Hill, ISBN 0-07-064841-7, 1998.
15. R. J. Clarke, *Transform Coding of Images*, Academic Press, 1985.
16. N. S. Jayant, and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice Hall, 1984.
17. M. Rabbani, and P. W. Jones, *Digital Image Compression Techniques*, SPIE Tutorial Texts, Vol. TT 7, SPIE Opt. Engineering Press, Bellingham, 1991.
18. P. H. N. de With, *Data Compression Techniques for Digital Video Recording*, Ph.D. Thesis, University of Technology Delft, June 1992.
19. J. Watkinson, The Art of Digital Video, Focal Press, Chapter 6, pp. 226–274, 1990.
20. *Outline of Basic Specifications for Consumer-Use Digital VCR*, Matsushita, Philips, Sony, Thomson, July 1993.
21. S. M. C. Borgers, W. A. L. Heijnemans, E. de Niet, and P. H. N. de With, "An Experimental Digital VCR with 40 mm Drum and DCT-Based Bit-Rate Reduction," *IEEE Trans. Consum. Electron.*, Vol. CE-34, No. 3, pp. 597–605, August 1988.

22. N. Doi, H. Hanyu, M. Izumita, and S. Mita, "Adaptive DCT Coding for Home Digital VTR," *IEEE Proc. Global Telecomm. Conf., Hollywood* (*USA*), Vol. 2, pp. 1073–1079, November 1988.
23. T. Kondo, N. Shirota, K. Kanota, Y. Fujimori, J. Yonemitsu, and M. Nagai, "Adaptive Dynamic Range Coding Scheme for Future Consumer Digital VTR," *IERE Proc. 7th Int. Conf. Video, Audio & Data Recording*, York (U.K.), Publ. No. 79, pp. 219–226, March 1988.
24. C. Yamamitsu, A. Ide, A. Iketani, T. Juri, S. Kadono, C. Matsumi, K. Matsushita, and H. Mizuki, "An Experimental Study for Home-Use Digital VTR," *IEEE Trans. Consum. Electron*., Vol. CE-35, No. 3, pp. 450–457, August 1989.
25. H.-J. Platte, W. Keesen, and D. Uhde, "Matrix Scan Recording, a New Alternative to Helical Scan Recording on Videotape," *IEEE Trans. Consum. Electron.*, Vol. CE-34, No. 3, pp. 606–611, August 1988.
26. K. Kanota, H. Inoue, A. Uetake, M. Kawaguchi, K. Chiba, and Y. Kubota, "A High Density Recording Technology for Digital VCR's," *IEEE Trans. Consum. Electron.*, Vol. 36, No. 3, pp. 540–547, August 1990.
27. M. Kobayashi, H. Ohta, and A. Murata, "Optimization of Azimuth Angle for Some Kinds of Media on Digital VCR's," *IEEE Trans. Magnetics*, Vol. 27, No. 6, pp. 4526–4531, November 1991.
28. Y. Eto, "Signal Processing for Future Home-Use Digital VTR's," *IEEE J. Sel. Areas Commun*., Vol. 10, No. 1, pp. 73–79, January 1992.
29. M. Shimotashiro, M. Tokunaga, K. Hashimoto, S. Ogata, and Y. Kurosawa, "A Study of the Recording and Reproducing System of Digital VCR Using Metal Evaporated Tape," *IEEE Trans. Consum. Electron.*, Vol. 41, No. 3, pp. 679–686, August 1995.
30. C. Yamamitsu, A. Iketani, J. Ohta, and N. Echigo, "An Experimental Digital VCR for Consumer Use," *IEEE Trans. Magnetics*, Vol. 31, No. 2, pp. 1037–1043, March 1995.
31. P. H. N. de With, and A. M. A. Rijckaert, "Design Considerations of the Video Compression System of the New DV Camcorder Standard," *IEEE Trans. Consum. Electron*., Vol 43, No. 4, pp. 1160–1179, November 1997.
32. R. W. J. J. Saeijs, P. H. N. de With, A. M. A. Rijckaert, and C. Wong, "An Experimental Digital Consumer Recorder for MPEG-coded Video Signals," *IEEE Trans. Consum. Electron.*, Vol. 41, No. 3, pp. 651–661, August 1995.
33. J. A. H. Kahlman, and K. A. S. Immink, "Channel Code with Embedded Pilot Tracking Tones for DVCR," *IEEE Trans. Consum. Electron*, Vol. 41, No. 1, pp. 180–185, February 1995.
34. P. H. N. de With, A. M. A. Rijckaert, H.-W. Keessen, J. Kaaden, and C. Opelt, "An Experimental Digital Consumer HDTV Recorder Using MC-DCT Video Compression," *IEEE Trans. Consum. Electron*., Vol. 39, No. 4, pp. 711–722, November 1993.
35. G. C. P. Lokhoff, "Precision Adaptive Subband Coding (PASC) for the Digital Compact Cassette (DCC)," *IEEE Trans. Consum. Electron.*, Vol. 38, No. 4, pp. 784–789, November 1992.
36. S. G. Stan and H. Spruit, "DVD + R—A write-once optical recording system for video and data applications," *Int. Conf. Consum. Electron.*, Los Angeles, 2002, Digest of Techn. Papers, pp. 256–257, June 2002.
37. S. F. Chang, T. Sikora, and A. Puri, "Overview of the MPEG-7 standard," *IEEE Trans. Circuits Syst*. Video Technol., Vol. 11, No. 6, pp. 688–695, June 2001.
38. B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*, Wiley, April 2002.
39. D-VHS: A First Look at a New Format, http://www.thedigitalbits.com/articles/dvhs.
40. K. A. S. Immink, *Codes for Mass Data Storage Systems*, Shannon Found. Publishers, ISBN 90-74249-23-X, Venlo, The Netherlands, November 1999.
41. F. Sijstermans, and J. van der Meer, "CD-i Full-Motion Video on a Parallel Computer," *Commun. ACM*. Vol. 34, No. 4, pp. 82–91, April 1991.
42. http://www.tv-anytime.org
43. P. H. N. de With, M. Breeuwer, and P. A. M. van Grinsven, "Data Compression Systems for Home-Use Digital Video Recording," IEEE J. Sel. Areas Commun., Vol. 10, No. 1, pp. 97–121, January 1992.