Q.2 a. Define a multimedia system. Describe about the different components of Multimedia. (2+3)

Answer:

Multimedia ---- An Application which uses a collection of multiple media sources e.g. text, graphics, images, sound, animation and video.

Multimedia is the field concerned with the computer controlled integration of text, graphics, drawings, still and moving images (Video), animation, audio, and any other media where every type of information can be represented, stored, transmitted and processed digitally

Basic components in multimedia

Text

A text is a coherent set of symbols that transmits some kind of informative message. Text Inclusion of textual information in multimedia is the basic step towards development of multimedia software. Text can be of any type, may be a word, a single line, or a paragraph. The textual data for multimedia can be developed using any text editor. However to give special effects, one needs graphics software which supports this kind of job. The text can have different type, size, color and style.

Images & graphic

A digital image is a representation of a two-dimensional image using ones and zeros (binary). Depending on whether or not the image resolution is fixed, it may be of vector or raster type. Without qualifications, the term "digital image" usually refers to raster images also called bitmap images.

Another interesting element in multimedia is graphics. As a matter of fact, taking into consideration the human nature, a subject is more explained with some sort of pictorial/graphical representation

Audio

Audio is sound within the acoustic range available to humans. An audio frequency (AF) is an electrical alternating current within the 20 to 20,000 hertz (cycles per second) range that can be used to produce acoustic sound. Sound is a sequence of naturally analog signals that are converted to digital signals by the audio card, using a microchip called an analog-to-digital converter (ADC). When sound is played, the digital signals are sent to the speakers where they are converted back to analog signals that generate varied sound.

Animation

A simulation of movement created by displaying a series of pictures, or frames. Cartoons on television is one example of animation. Animation on computers is one of the chief ingredients of multimedia presentations. There are many software applications that enable you to create animations that you can display on a computer monitor.

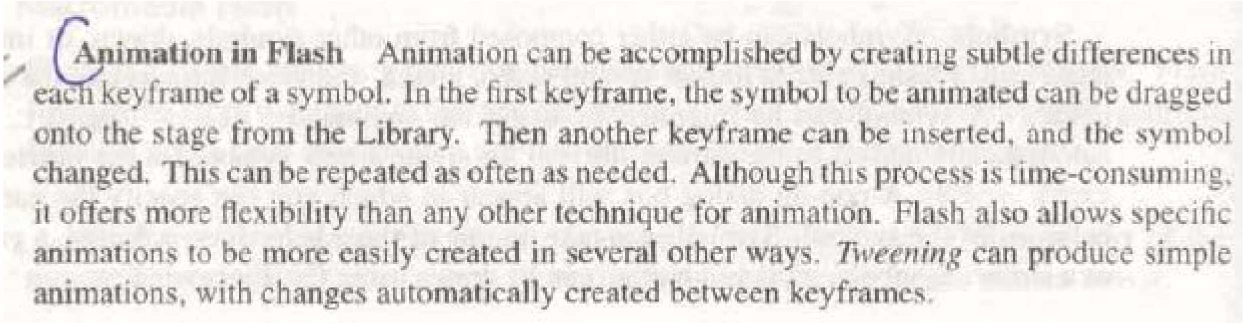
Video

Beside animation there is one more media element, which is known as video. With latest technology it is possible to include video impact on clips of any type into any multimedia creation, be it corporate presentation, fashion design, entertainment games, etc.

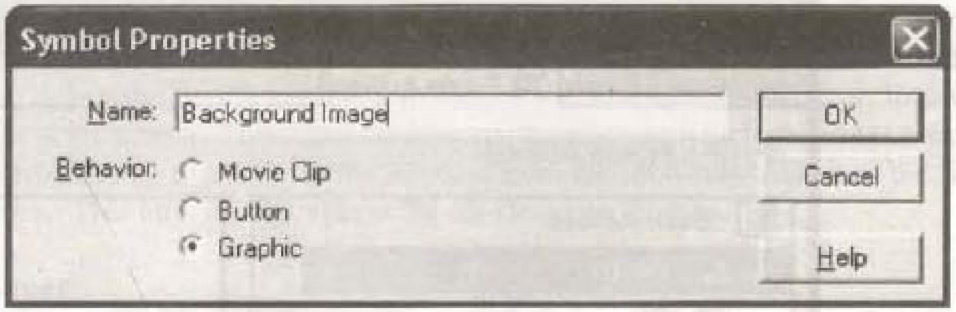
The video clips may contain some dialogues or sound effects and moving pictures. These video clips can be combined with the audio, text and graphics for multimedia presentation. Incorporation of video in a multimedia package is more important and complicated than other media elements. One can procure video clips from various sources such as existing video films or even can go for an outdoor video shooting.

b. Discuss the method of accomplishing Animation in Flash. (5)

Answer:



Animation in Flash Animation can be accomplished by creating subtle differences in each keyframe of a symbol. In the first keyframe, the symbol to be animated can be dragged onto the stage from the Library. Then another keyframe can be inserted, and the symbol changed. This can be repeated as often as needed. Although this process is time-consuming, it offers more flexibility than any other technique for animation. Flash also allows specific animations to be more easily created in several other ways. *Tweening* can produce simple animations, with changes automatically created between keyframes.



Symbol Properties

Name: Background Image

Behavior: ○ Movie Clip ○ Button ● Graphic

FIGURE 2.25: Create symbol dialog.

Tweening There are two types of tweening: *shape* and *movement* tweening. Shape tweening allows you to create a shape that continuously changes to a different shape over time. Movement tweening allows you to place a symbol in different places on the Stage in different keyframes. Flash automatically fills in the keyframes along a path between the start and finish. To carry out movement tweening, select the symbol to be tweened, choose `Insert > Create Motion Tween`, and select the end frame. Then use the command `Insert > Frame` and move the symbol to the desired position. More advanced tweening allows control of the path as well as of acceleration. Movement and shape tweenings can be combined for additional effect.

Mask animation involves the manipulation of a layer mask — a layer that selectively hides portions of another layer. For example, to create an explosion effect, you could use a mask to cover all but the center of the explosion. Shape tweening could then expand the mask, so that eventually the whole explosion is seen to take place. Figure 2.26 shows a scene before and after a tweening effect is added.

Action Scripts Action scripts allow you to trigger events such as moving to a different keyframe or requiring the movie to stop. Action scripts can be attached to a keyframe or symbols in a keyframe. Right-clicking on the symbol and pressing `Actions` in the list can modify the actions of a symbol. Similarly, by right-clicking on the keyframe and pressing `Actions` in the pop-up, you can apply actions to a keyframe. A Frame Actions window will come up, with a list of available actions on the left and the current actions being applied symbol on the right. Action scripts are broken into six categories: *Basic*

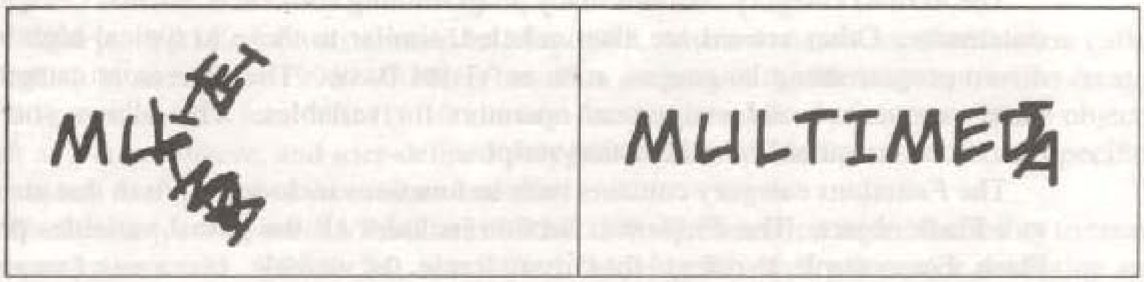


FIGURE 2.26: Before and after tweening letters.

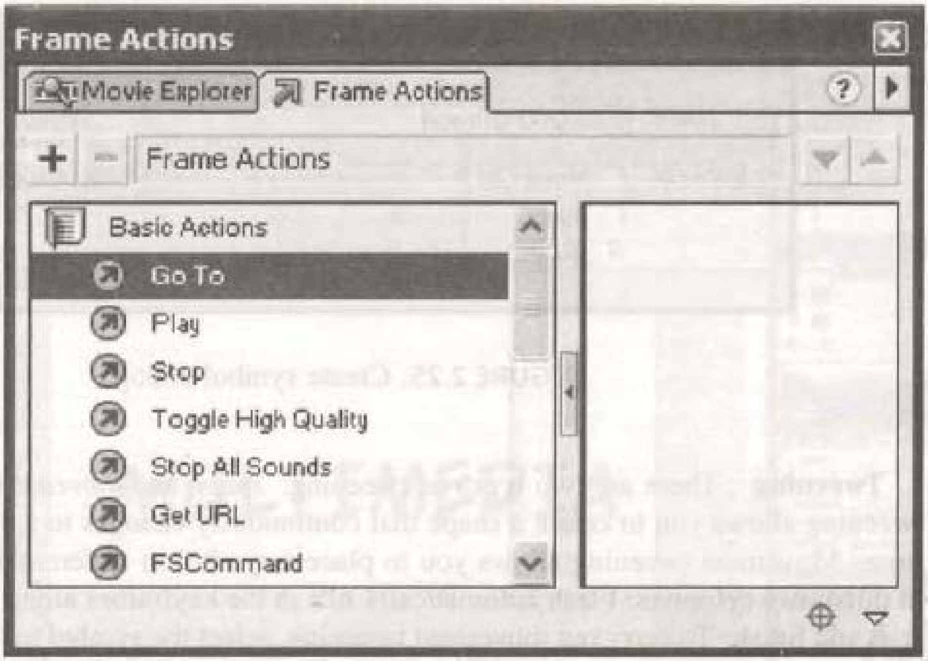


FIGURE 2.27: Action scripts window.

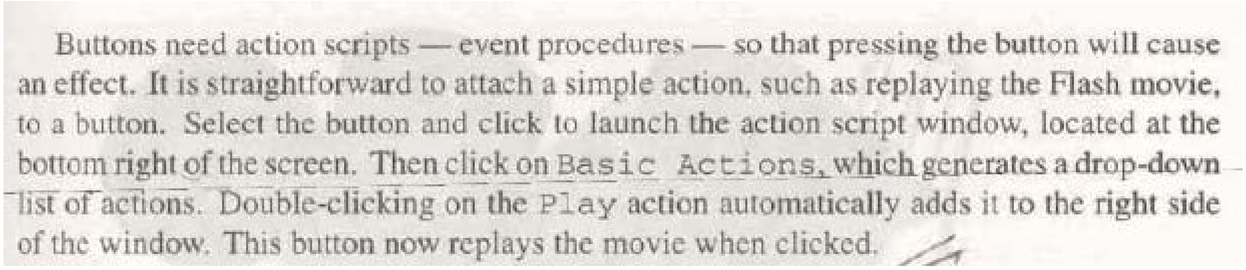
Actions, *Actions*, *Operators*, *Functions*, *Properties*, and *Objects*. Figure 2.27 shows the Frame Actions window.

Basic Actions allow you to attach many simple actions to the movie. Some common actions are

- `Goto`. Moves the movie to the keyframe specified and can optionally stop. The stop action is commonly used to stop interactive movies when the user is given an option.
- `Play`. Resumes the movie if the movie is stopped.
- `Stop`. Stops the movie if it is playing.
- `Tell Target`. Sends messages to different symbols and keyframes in Flash. It is commonly used to start or stop an action on a different symbol or keyframe.

The *Actions* category contains many programming constructs, such as `Loops` and `Goto` statements. Other actions are also included, similar to those in typical high-level, event-driven programming languages, such as Visual Basic. The *Operators* category includes many comparison and assignment operators for variables. This allows you to perform operations on variables in the action script.

The *Functions* category contains built-in functions included in Flash that are not specific to a Flash object. The *Properties* section includes all the global variables predefined in Flash. For example, to refer to the current frame, the variable `_currentframe` is defined. The *Objects* section lists all objects, such as movie clips or strings and their associated functions.



Buttons need action scripts — event procedures — so that pressing the button will cause an effect. It is straightforward to attach a simple action, such as replaying the Flash movie, to a button. Select the button and click to launch the action script window, located at the bottom right of the screen. Then click on `Basic Actions`, which generates a drop-down list of actions. Double-clicking on the `Play` action automatically adds it to the right side of the window. This button now replays the movie when clicked.

c. Define VRML. Write short notes on VRML 1.0 and VRML 2.0. (3+3)

Answer:

The Virtual Reality Modeling Language (VRML) allows us to describe 3D objects, and combine them into interactive scenes and worlds. The virtual worlds - which can integrate 3D graphics, multimedia, and interactivity - can be accessed through the WWW (http). The remote users can explore the content interactively in much more sophisticated ways than clicking/scrolling. VRML is not a programming language like JAVA, nor is it a "Markup Language" like HTML. It is a modelling language, which means we use it to describe 3D scenes. It's more complex than HTML, but less complex (except for the scripting capability) than a programming language. VRML is a (text)file-format that integrates 3D graphics and multimedia: a simple language for describing 3D shapes and interactive environments. We can create(write) a VRML file using either any text editor or "wordbuilder" authoring software. To view a VRML file we need either a standalone VRML browser or a Netscape plug-in

VRML 1.0 allowed to create static 3D worlds assembled from static objects, which could be hyperlinked to other worlds, as well as to HTML documents. Visitors of the worlds were able to "fly" or "walk" around the static objects, and the only way of interaction was possible by "clicking" on a hyperlinked object, which worked like a hyperlink on a www-page: dropped to the target of the link.

In **VRML 2.0** objects can be animated, and they can respond to both time-based and user-initiated events. VRML 2.0 also allows us to incorporate multimedia objects(for example sound and movies) in our scenes.

Q.3 a. Describe the color models YUV, YIQ and YCbCr used to describe the colors in video. (3+3+3)

Answer: YUV Color Model

First, it codes a luminance signal (for gamma-corrected signals) equal to *Y* ‘ The luma *Y* ‘ is similar, but not exactly the same as, the CIE luminance value Y, gamma-corrected.

As well as magnitude or brightness we need a colorfulness scale, and to this end *chrominance* refers to the difference between a color and a reference white at the same luminance. It can be represented by the color *differences U, V*:

$U = B' - Y'$

$V = R' - Y'$

$$\begin{bmatrix} Y' \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.299 & -0.587 & 0.886 \\ 0.701 & -0.587 & -0.114 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}$$

YIQ Color Model

YIQ (actually, *Y ‘ I Q*) is used in NTSC color TV broadcasting. Again, gray pixels generate zero *(I, Q)* chrominance signal. The original meanings of these names came from combinations of analog signals, *I* for "in-phase chrominance" and *Q* for "quadrature chrominance" signal—these names can now be safely ignored.

It is thought that, although *U* and *V* are more simply defined, they do not capture the most-to-least hierarchy of human vision sensitivity. Although *U* and *V* nicely define the color differences, they do not best correspond to actual human perceptual color sensitivities. In NTSC, *I* and *Q* are used instead.

$$\begin{aligned} I &= 0.492111(R' - Y')\cos 33^\circ - 0.877283(B' - Y')\sin 33^\circ \\ Q &= 0.492111(R' - Y')\sin 33^\circ + 0.877283(B' - Y')\cos 33^\circ \end{aligned}$$

This leads to the following matrix transform:

$$\begin{bmatrix} Y' \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.595879 & -0.274133 & -0.321746 \\ 0.211205 & -0.523083 & 0.311878 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}$$

YCbCr Color Model

The international standard for component (3-signal, studio quality) *digital* video uses another color space,*YCbCr* , often simply written YCbCr. The YCbCr transform is closely related to the YUV transform. YUV is changed by scaling such that *Cb* is *U*, but with a coefficient of 0.5 multiplying *B'*. In some software systems, *Cb* and *Cr* are also shifted such that values are between 0 and 1. This makes the equations as follows:

$$\begin{aligned} C_b &= \big((B' - Y')/1.772\big) + 0.5 \\ C_r &= \big((R' - Y')/1.402\big) + 0.5 \end{aligned}$$

$$\begin{bmatrix} Y' \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.168736 & -0.331264 & 0.5 \\ 0.5 & -0.418688 & -0.081312 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} + \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix}$$

b. Write the advantages of digital representation of video. (3)

Answer:

Digital Video

The advantages of digital representation for video are many. It permits

- Storing video on digital devices or in memory, ready to be processed (noise removal, cut and paste, and so on) and integrated into various multimedia applications.
- Direct access, which makes nonlinear video editing simple.
- Repeated recording without degradation of image quality.
- Ease of encryption and better tolerance to channel noise.

c. Write short note on NTSC video standard. (4)

Answer:

The NTSC TV standard is mostly used in North America and Japan. It uses a familiar 4:3 *aspect ratio* (i.e., the ratio of picture width to height) and 525 scan lines per frame at 30 fps. The NTSC television standard defines a composite video signal with a refresh rate of 60 half-frames (interlaced) per second. Each frame contains 525 lines with up to 16 million different colors.

N*ational* ***T****elevision* ***S****ystem* ***C****ommittee*(NTSC) is responsible for setting television and video standards in the United States (in Europe and the rest of the world, the dominant television standards are PAL and SECAM). The NTSC standard for television defines a composite video signal with a refresh rate of 60 half-frames(interlaced) per second. Each frame contains 525 lines and can contain 16 million different colors.

The NTSC standard is incompatible with most computer video standards, which generally use *RGB* video signals. However, you can insert special video adapters into your computer that convert NTSC signals into computer video signals and vice versa.

Q.4 a. What do you understand by Huffman coding? What is the principle in generating the Huffman code? (3+5)

Answer:

Huffman coding is a statistical technique which attempts to reduce the amount of bits required to represent a string of symbols. The Huffman code for an alphabet (set of symbols) may be generated by constructing a binary tree with nodes containing the symbols to be encoded and their probabilities of occurrence. Huffman coding is based on the frequency of occurrence of a data item (pixel in images). The principle is to use a lower number of bits to encode the data that occurs more frequently. Codes are stored in a *Code Book* which may be constructed for each image or a set of images. In all cases the code book plus encoded data must be transmitted to enable decoding.

Algorithm :

1. Initialization: put all symbols on the list sorted according to their frequency counts.

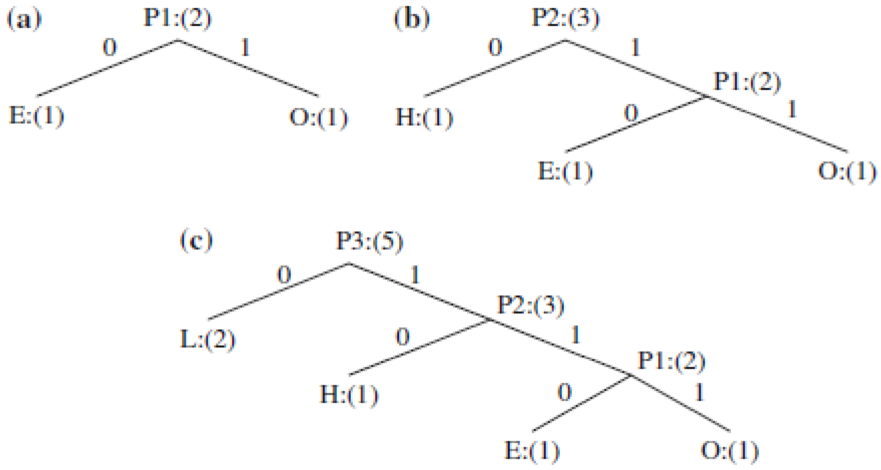
2. Repeat until the list has only one symbol left.

(a) From the list, pick two symbols with the lowest frequency counts. Form a Huffman subtree that has these two symbols as child nodes and create a parent node for them.

(b) Assign the sum of the children’s frequency counts to the parent and insert it into the list, such that the order is maintained.

(c) Delete the children from the list.

3. Assign a codeword for each leaf based on the path from the root.



Huffman algorithm are described in the following *bottom-up* manner. Let us use the example word, HELLO. A binary coding tree will be used as above, in which the left branches are coded 0 and right branches 1.
For instance, the code 0 assigned to L ,10 for H or 110 for E or 111 for O,
110 for E or 111 for O.

b. Differentiate between DPCM and ADPCM. (2+2)

Answer:

DPCM: Differential Pulse Code Modulation is exactly the same as Predictive Coding, Predictive coding except that it incorporates a quantizer step. Quantization is as in PCM and can be uniform or nonuniform. Stores a multibit difference value. A bipolar D/A converter is used for playback to convert the successive difference values to an analog waveform.
ADPCM: Stores a difference value that has been mathematically adjusted according to the slope of the input waveform. Bipolar D/A converter is used to convert the stored digital code to analog for playback. Example to be based on the above stated difference.

c. What is MIDI? Discuss the basic MIDI message structure. (2+2)

Answer:

MIDI messages can be classified into two types, as in Figure 6.12 — channel messages and system messages — and further classified as shown. Each type of message will be examined below.

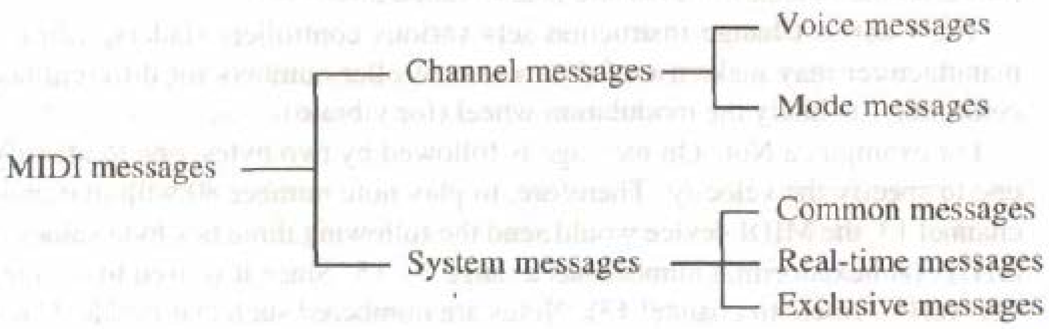


FIGURE 6.12: MIDI message taxonomy.

Q.5 a. What is the significance of JPEG standard? Describe any two modes that JPEG standard support. (4+4)

Answer:

JPEG is designed for compressing either full-color or gray-scale images of natural, real-world scenes. JPEG is a lossy compression algorithm. When we create a JPEG or convert an image from another format to a JPEG, we are asked to specify the quality of image we want. Since the highest quality results in the largest file, we can make a trade-off between image quality and file size. The lower the quality, the greater the compression, and the greater the degree of information loss. JPEGs are best suited for continuous tone images like photographs or natural artwork; not so well on sharp-edged or flat-color art like lettering, simple cartoons, or line drawings. JPEG compression introduces noise into solid-color areas, which can distort and even blur flat-color graphics. All Web browsers most support JPEGs, and a rapidly growing number support progressive JPEGs.

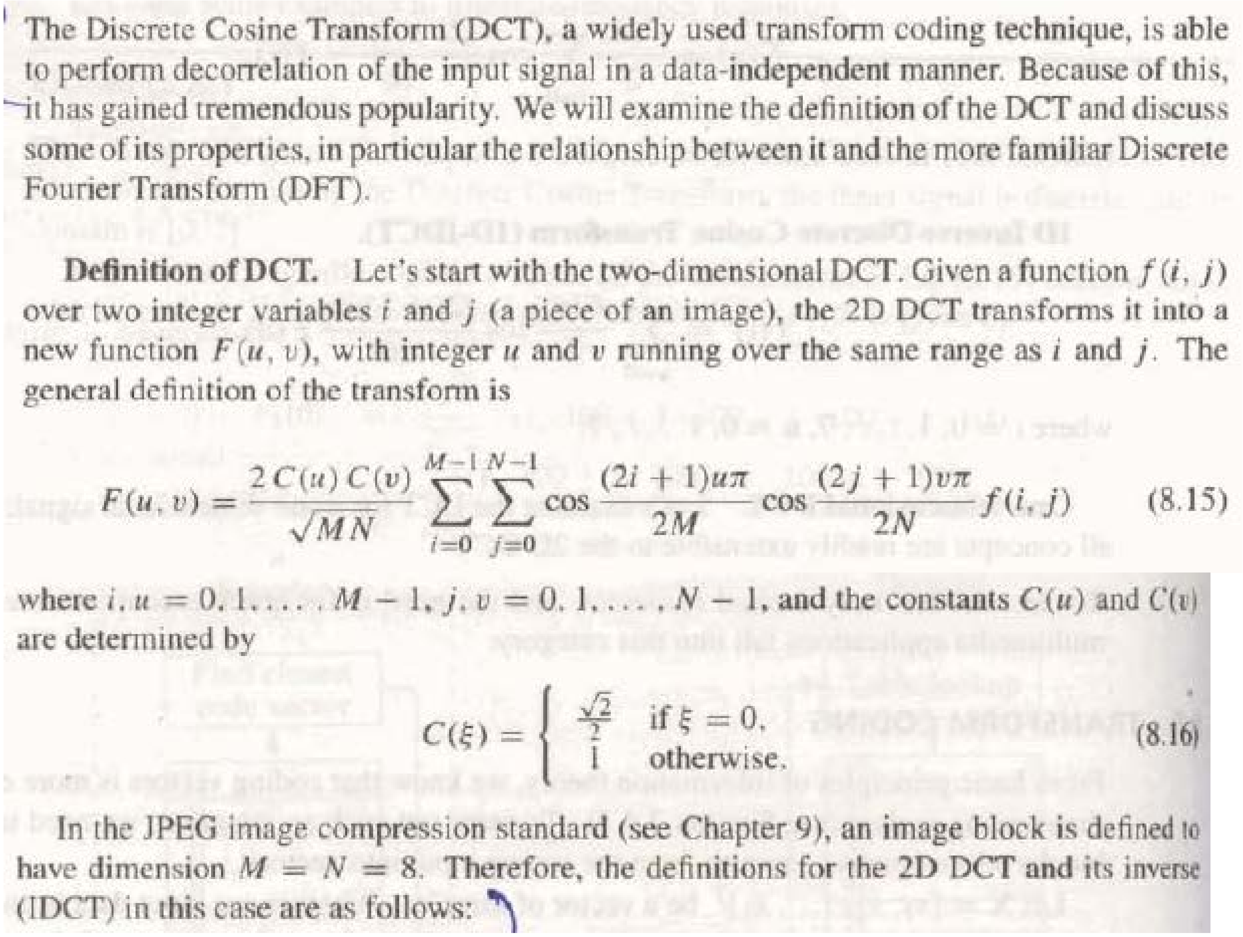
JPEG Modes: The JPEG standard supports numerous modes (variations). Some of the commonly used ones are:

Sequential Mode. This is the default JPEG mode. Each gray-level image or color image component is encoded in a single left-to-right, top-to-bottom scan. We implicitly assumed this mode in the discussions so far. The ‘Motion JPEG’ video codec uses Baseline Sequential JPEG, applied to each image frame in the video.

Progressive Mode. Progressive JPEG delivers low-quality versions of the image quickly, followed by higher-quality passes, and has become widely supported in web browsers. Such multiple scans of images are of course most useful when the speed of the communication line is low. In Progressive Mode, the first few scans carry only a few bits and deliver a rough picture of what is to follow. After each additional scan, more data is received, and image quality is gradually enhanced. The advantage is that the user-end has a choice whether to continue receiving image data after the first scan(s). Progressive JPEG can be realized in one of the following two ways. The main steps (DCT, quantization, etc.) are identical to those in Sequential Mode.

b. Define Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT). List out the different characteristics of DCT. (4+4)

Answer:



The Discrete Cosine Transform (DCT), a widely used transform coding technique, is able to perform decorrelation of the input signal in a data-independent manner. Because of this, it has gained tremendous popularity. We will examine the definition of the DCT and discuss some of its properties, in particular the relationship between it and the more familiar Discrete Fourier Transform (DFT).

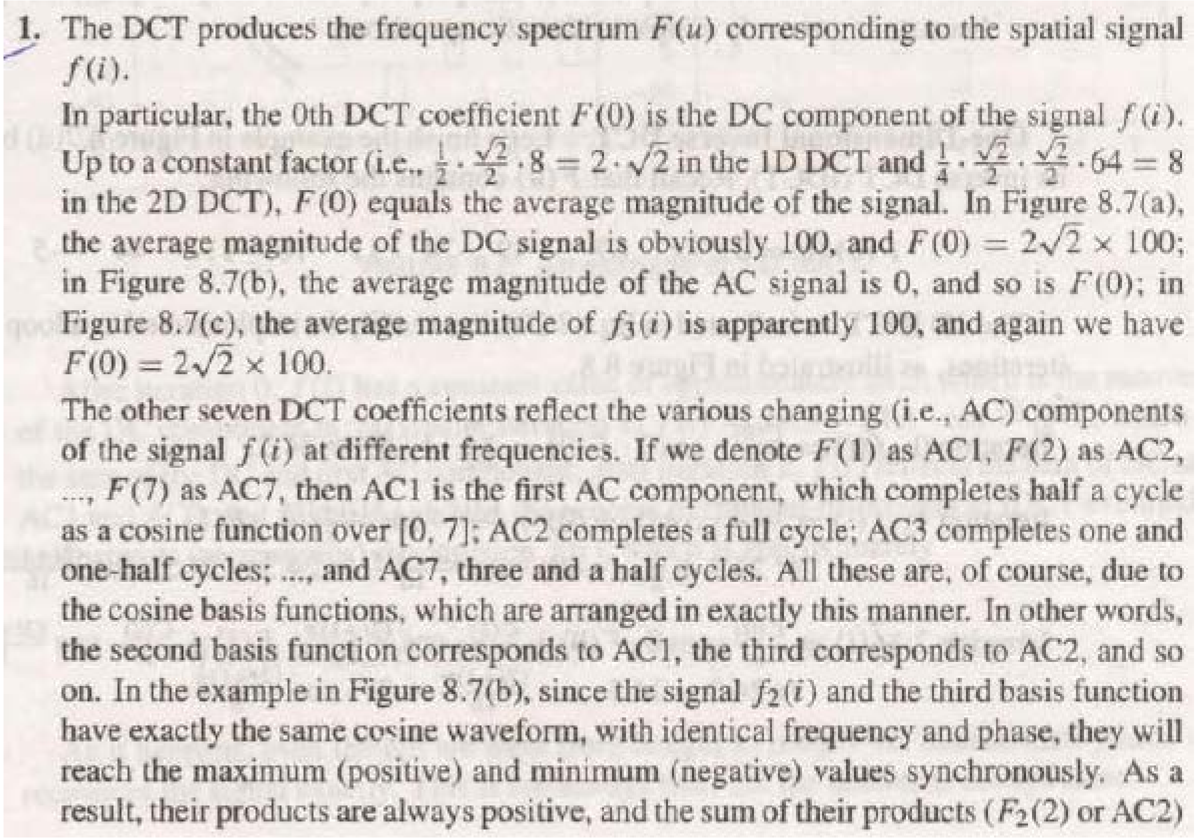
Definition of DCT. Let’s start with the two-dimensional DCT. Given a function $f(i, j)$ over two integer variables i and j (a piece of an image), the 2D DCT transforms it into a new function $F(u, v)$, with integer u and v running over the same range as i and j. The general definition of the transform is

$$F(u,v) = \frac{2\,C(u)\,C(v)}{\sqrt{MN}} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \cos\frac{(2i+1)u\pi}{2M} \cos\frac{(2j+1)v\pi}{2N} f(i,j) \qquad (8.15)$$

where $i, u = 0, 1, \ldots, M-1$, $j, v = 0, 1, \ldots, N-1$, and the constants $C(u)$ and $C(v)$ are determined by

$$C(\xi) = \begin{cases} \frac{\sqrt{2}}{2} & \text{if } \xi = 0, \\ 1 & \text{otherwise.} \end{cases} \qquad (8.16)$$

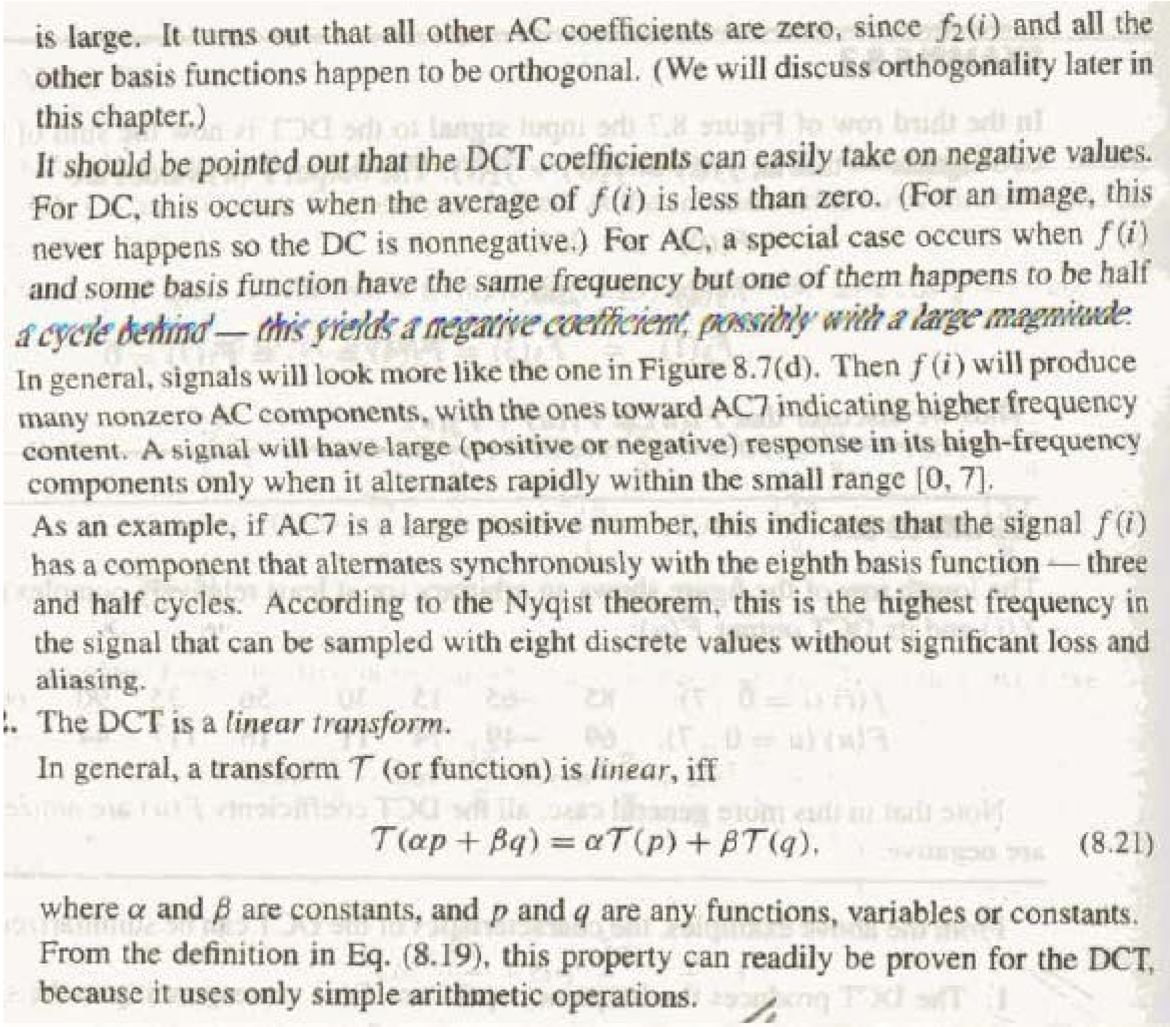
In the JPEG image compression standard (see Chapter 9), an image block is defined to have dimension $M = N = 8$. Therefore, the definitions for the 2D DCT and its inverse (IDCT) in this case are as follows:



1. The DCT produces the frequency spectrum $F(u)$ corresponding to the spatial signal $f(i)$.

In particular, the 0th DCT coefficient $F(0)$ is the DC component of the signal $f(i)$. Up to a constant factor (i.e., $\frac{1}{2} \cdot \frac{\sqrt{2}}{2} \cdot 8 = 2 \cdot \sqrt{2}$ in the 1D DCT and $\frac{1}{4} \cdot \frac{\sqrt{2}}{2} \cdot \frac{\sqrt{2}}{2} \cdot 64 = 8$ in the 2D DCT), $F(0)$ equals the average magnitude of the signal. In Figure 8.7(a), the average magnitude of the DC signal is obviously 100, and $F(0) = 2\sqrt{2} \times 100$; in Figure 8.7(b), the average magnitude of the AC signal is 0, and so is $F(0)$; in Figure 8.7(c), the average magnitude of $f_3(i)$ is apparently 100, and again we have $F(0) = 2\sqrt{2} \times 100$.

The other seven DCT coefficients reflect the various changing (i.e., AC) components of the signal $f(i)$ at different frequencies. If we denote $F(1)$ as AC1, $F(2)$ as AC2, ..., $F(7)$ as AC7, then AC1 is the first AC component, which completes half a cycle as a cosine function over [0, 7]; AC2 completes a full cycle; AC3 completes one and one-half cycles; ..., and AC7, three and a half cycles. All these are, of course, due to the cosine basis functions, which are arranged in exactly this manner. In other words, the second basis function corresponds to AC1, the third corresponds to AC2, and so on. In the example in Figure 8.7(b), since the signal $f_2(i)$ and the third basis function have exactly the same cosine waveform, with identical frequency and phase, they will reach the maximum (positive) and minimum (negative) values synchronously. As a result, their products are always positive, and the sum of their products ($F_2(2)$ or AC2)



is large. It turns out that all other AC coefficients are zero, since $f_2(i)$ and all the other basis functions happen to be orthogonal. (We will discuss orthogonality later in this chapter.)

It should be pointed out that the DCT coefficients can easily take on negative values. For DC, this occurs when the average of $f(i)$ is less than zero. (For an image, this never happens so the DC is nonnegative.) For AC, a special case occurs when $f(i)$ and some basis function have the same frequency but one of them happens to be half a cycle behind — this yields a negative coefficient, possibly with a large magnitude.

In general, signals will look more like the one in Figure 8.7(d). Then $f(i)$ will produce many nonzero AC components, with the ones toward AC7 indicating higher frequency content. A signal will have large (positive or negative) response in its high-frequency components only when it alternates rapidly within the small range [0, 7].

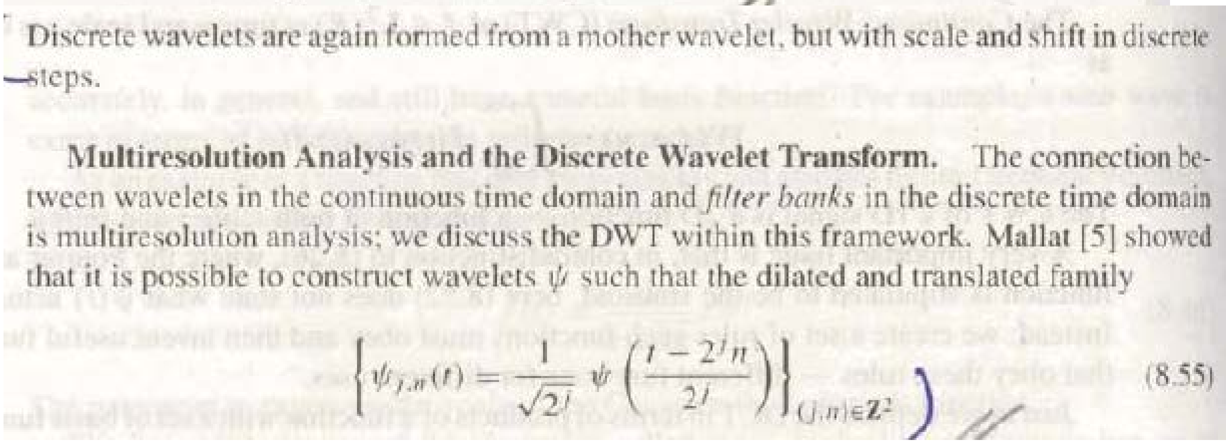
As an example, if AC7 is a large positive number, this indicates that the signal $f(i)$ has a component that alternates synchronously with the eighth basis function — three and half cycles. According to the Nyqist theorem, this is the highest frequency in the signal that can be sampled with eight discrete values without significant loss and aliasing.

2. The DCT is a *linear transform*.

In general, a transform $\mathcal{T}$ (or function) is *linear*, iff

$$\mathcal{T}(\alpha p + \beta q) = \alpha \mathcal{T}(p) + \beta \mathcal{T}(q), \qquad (8.21)$$

where α and β are constants, and p and q are any functions, variables or constants. From the definition in Eq. (8.19), this property can readily be proven for the DCT, because it uses only simple arithmetic operations.



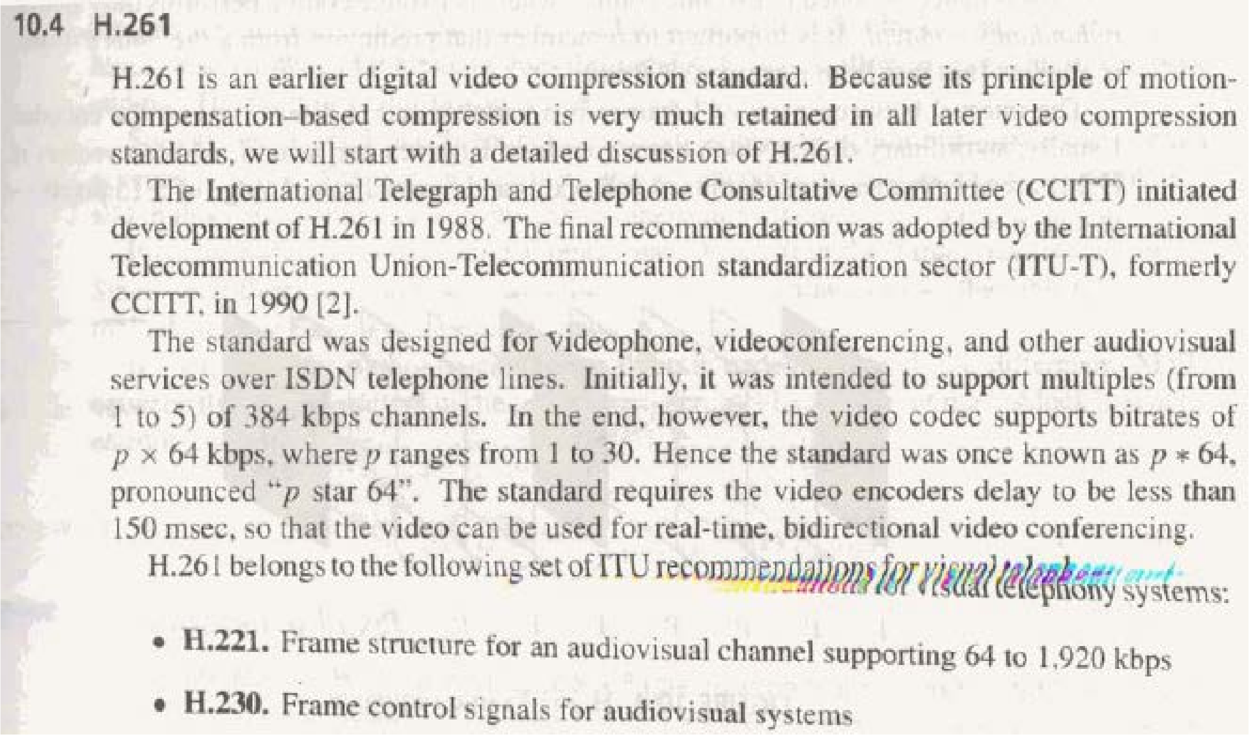
Discrete wavelets are again formed from a mother wavelet, but with scale and shift in discrete steps.

Multiresolution Analysis and the Discrete Wavelet Transform. The connection between wavelets in the continuous time domain and *filter banks* in the discrete time domain is multiresolution analysis; we discuss the DWT within this framework. Mallat [5] showed that it is possible to construct wavelets ψ such that the dilated and translated family

$$\left\{ \psi_{j,n}(t) = \frac{1}{\sqrt{2^j}}\, \psi\left(\frac{t - 2^j n}{2^j}\right) \right\}_{(j,n)\in Z^2} \qquad (8.55)$$

Q.6 a. Explain the characteristic of data stream used by H.261 and H.263. (4+4)

Answer:



10.4 H.261

H.261 is an earlier digital video compression standard. Because its principle of motion-compensation–based compression is very much retained in all later video compression standards, we will start with a detailed discussion of H.261.

The International Telegraph and Telephone Consultative Committee (CCITT) initiated development of H.261 in 1988. The final recommendation was adopted by the International Telecommunication Union-Telecommunication standardization sector (ITU-T), formerly CCITT, in 1990 [2].

The standard was designed for videophone, videoconferencing, and other audiovisual services over ISDN telephone lines. Initially, it was intended to support multiples (from 1 to 5) of 384 kbps channels. In the end, however, the video codec supports bitrates of $p \times 64$ kbps, where p ranges from 1 to 30. Hence the standard was once known as $p * 64$, pronounced "p star 64". The standard requires the video encoders delay to be less than 150 msec, so that the video can be used for real-time, bidirectional video conferencing.

H.261 belongs to the following set of ITU recommendations for visual telephony systems:

- **H.221.** Frame structure for an audiovisual channel supporting 64 to 1,920 kbps
- **H.230.** Frame control signals for audiovisual systems

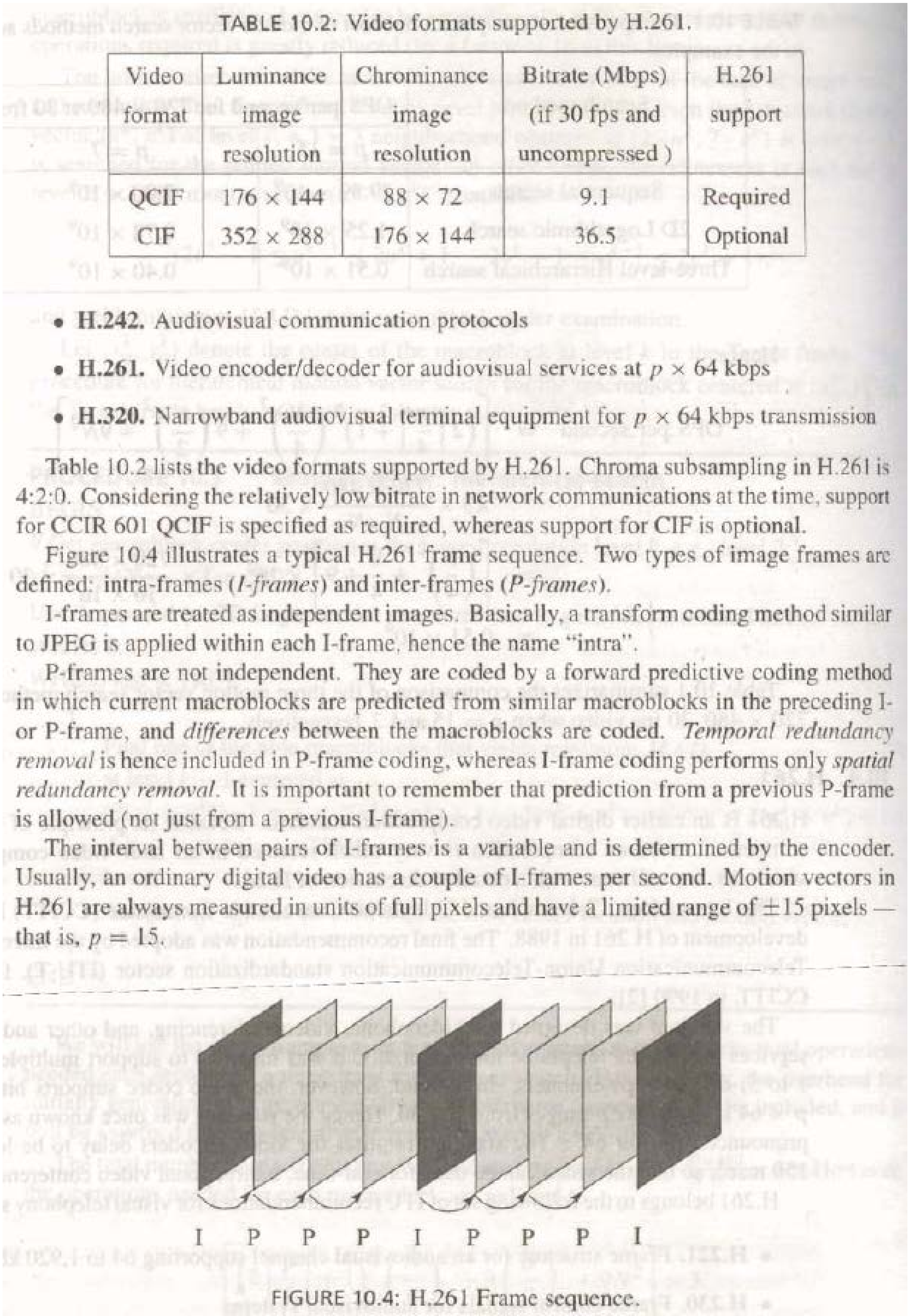


TABLE 10.2: Video formats supported by H.261.

Video format	Luminance image resolution	Chrominance image resolution	Bitrate (Mbps) (if 30 fps and uncompressed)	H.261 support
QCIF	176 × 144	88 × 72	9.1	Required
CIF	352 × 288	176 × 144	36.5	Optional

- **H.242.** Audiovisual communication protocols
- **H.261.** Video encoder/decoder for audiovisual services at $p \times 64$ kbps
- **H.320.** Narrowband audiovisual terminal equipment for $p \times 64$ kbps transmission

Table 10.2 lists the video formats supported by H.261. Chroma subsampling in H.261 is 4:2:0. Considering the relatively low bitrate in network communications at the time, support for CCIR 601 QCIF is specified as required, whereas support for CIF is optional.

Figure 10.4 illustrates a typical H.261 frame sequence. Two types of image frames are defined: intra-frames (*I-frames*) and inter-frames (*P-frames*).

I-frames are treated as independent images. Basically, a transform coding method similar to JPEG is applied within each I-frame, hence the name "intra".

P-frames are not independent. They are coded by a forward predictive coding method in which current macroblocks are predicted from similar macroblocks in the preceding I- or P-frame, and *differences* between the macroblocks are coded. *Temporal redundancy removal* is hence included in P-frame coding, whereas I-frame coding performs only *spatial redundancy removal*. It is important to remember that prediction from a previous P-frame is allowed (not just from a previous I-frame).

The interval between pairs of I-frames is a variable and is determined by the encoder. Usually, an ordinary digital video has a couple of I-frames per second. Motion vectors in H.261 are always measured in units of full pixels and have a limited range of ±15 pixels — that is, $p = 15$.

FIGURE 10.4: H.261 Frame sequence.

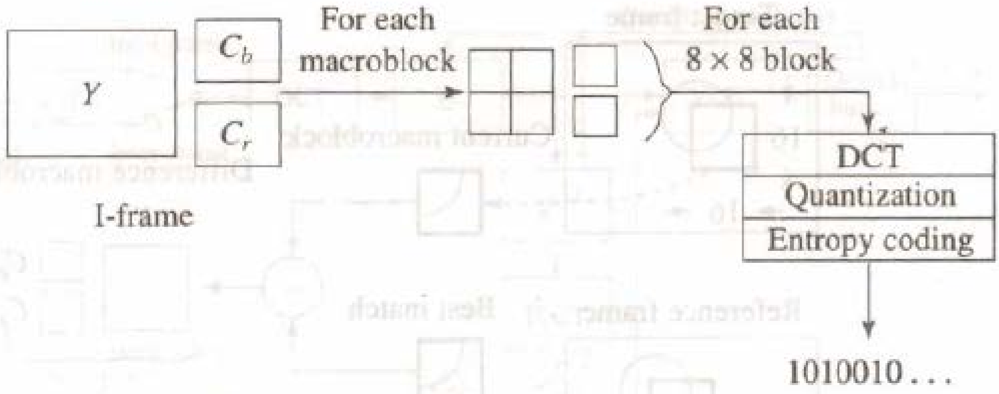
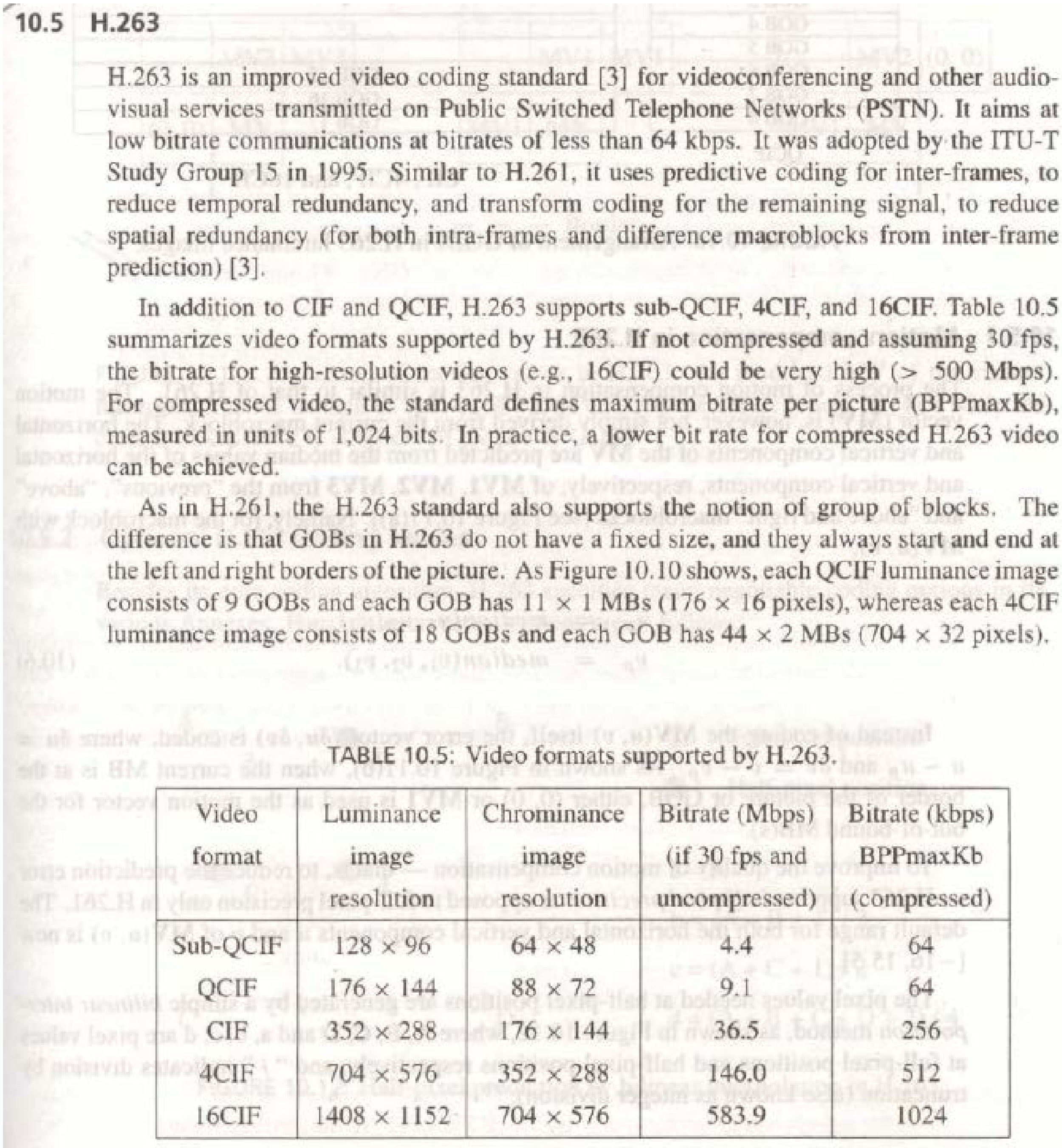


FIGURE 10.5: I-frame coding.



10.5 H.263

H.263 is an improved video coding standard [3] for videoconferencing and other audio-visual services transmitted on Public Switched Telephone Networks (PSTN). It aims at low bitrate communications at bitrates of less than 64 kbps. It was adopted by the ITU-T Study Group 15 in 1995. Similar to H.261, it uses predictive coding for inter-frames, to reduce temporal redundancy, and transform coding for the remaining signal, to reduce spatial redundancy (for both intra-frames and difference macroblocks from inter-frame prediction) [3].

In addition to CIF and QCIF, H.263 supports sub-QCIF, 4CIF, and 16CIF. Table 10.5 summarizes video formats supported by H.263. If not compressed and assuming 30 fps, the bitrate for high-resolution videos (e.g., 16CIF) could be very high (> 500 Mbps). For compressed video, the standard defines maximum bitrate per picture (BPPmaxKb), measured in units of 1,024 bits. In practice, a lower bit rate for compressed H.263 video can be achieved.

As in H.261, the H.263 standard also supports the notion of group of blocks. The difference is that GOBs in H.263 do not have a fixed size, and they always start and end at the left and right borders of the picture. As Figure 10.10 shows, each QCIF luminance image consists of 9 GOBs and each GOB has 11 × 1 MBs (176 × 16 pixels), whereas each 4CIF luminance image consists of 18 GOBs and each GOB has 44 × 2 MBs (704 × 32 pixels).

TABLE 10.5: Video formats supported by H.263.

Video format	Luminance image resolution	Chrominance image resolution	Bitrate (Mbps) (if 30 fps and uncompressed)	Bitrate (kbps) BPPmaxKb (compressed)
Sub-QCIF	128 × 96	64 × 48	4.4	64
QCIF	176 × 144	88 × 72	9.1	64
CIF	352 × 288	176 × 144	36.5	256
4CIF	704 × 576	352 × 288	146.0	512
16CIF	1408 × 1152	704 × 576	583.9	1024

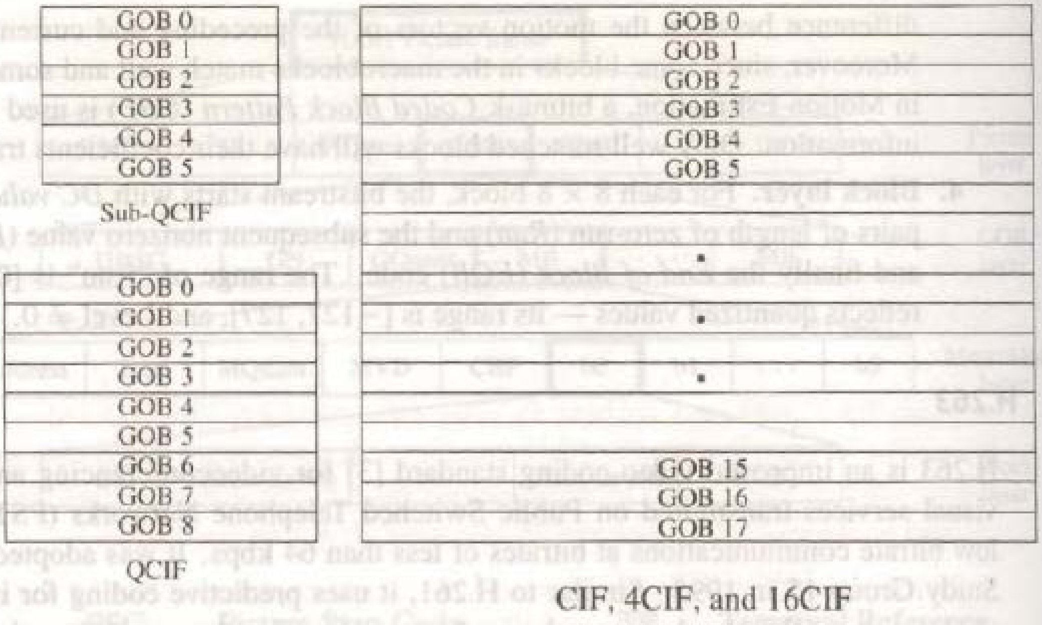


FIGURE 10.10: Arrangement of GOBs in H.263 luminance images.

b. Explain various parts of the MPEG-1 standard. Describe the MPEG-1 video standard mentioning the roles of I-,P- and B- frames. (4+4)

Answer:

The MPEG-1 standard, also referred to as ISO/IEC 11172 , has five parts:
11172-1 Systems,
11172-2 Video,
11172-3 Audio,
11172-4 Conformance, and
11172-5 Software.

Briefly, Systems takes care of, among many things, dividing output into packets of bitstreams, multiplexing, and synchronization of the video and audio streams. Conformance (or compliance) specifies the design of tests for verifying whether a bitstream or decoder complies with the standard. Software includes a complete software implementation of the MPEG-1 standard decoder and a sample software implementation of an encoder.

In the field of video compression a video frame is compressed using different algorithms These different algorithms for video frames are called **picture types** or **frame types**. The major picture types used in the different video algorithms are **I** and **P** . They are different in the following characteristics:

An **I-frame** is an 'Intra-coded picture', in effect a fully specified picture, like a conventional static image file: that is it is treated as independent image . P-frames hold only part of the image information, so they need less space to store than an I-frame, and thus improve video compression rates. **I**-frames are the least compressible but don't require other video frames to decode. I-frames coding performs only spatial redundancy removal

A **P-frame** ('Predicted picture') are not independent holds only the changes in the image from the previous frame. They are coded by forward predictive coding method. For example, in a scene where a car moves across a stationary background, only the car's movements need to be encoded. The encoder does not need to store the unchanging background pixels in the P-frame, thus saving space. **P**-frames can use data from previous frames to decompress and are more compressible than I-frames. Temporal redundancy removal is included in P-frame coding.

B-frame*s* and their accompanying bidirectional motion compensation. In addition to the forward prediction, a backward prediction is also performed, in which the matching macroblock is obtained from a future I- or P-frame in the video sequence. A **B-frame** ('Bi-predictive picture') saves even more space by using differences between the current frame and both the preceding and following frames to specify its content.

Q.7 a. Define MPEG-21 and its various key elements. (3+5)

Answer:

MPEG-21: As we stepped into the new century (and millennium), multimedia had seen its ubiquitous use in almost all areas, An ever-increasing number of content creators and content consumers emerge daily in society. However, there is no uniform way to define, identify, describe, manage and protect multimedia frame work to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities.

Its seven key elements are:

(i) Digital item declaration, to establish a uniform and flexible abstraction and interoperable schema for declaring digital items.

(ii) Digital item identification and description, to establish a frame work for standardized identification and description of digital items, regardless of their origin, type or granularity.

(iii) Content management and usage, to provide an interface and protocol that facilitate management and use of the content.

(iv) Intellectuals' property management and protection (IPMP), to enable contents to be reliably managed and protected.

(v) Terminals and networks, to provide interoperable and transparent access to content with quality of service (CEOS) across a wide range of networks and terminals.

(vi) Content representation, to represent content in a adequate way to pursuing the objective of MPEG-21, namely "content any time anywhere"

(vii) Event reporting, to establish metrics and interfaces for reporting events, so as to understand performance and alternatives.

b. Distinguish between channel vocoder and formant vocoder by briefly describing each one of them. (4+4)

Answer:

Vocoders are specifically voice coders. Vocoders are concerned with modeling speech, so that the salient features are captured in as few bits as possible. They use either a model of the speech waveform in time (*Linear Predictive Coding* (LPC) vocoding), or else break down the signal into frequency components and model these (channel vocoders and formant vocoders).

Channel Vocoder

A *channel vocoder* first applies a filter bank to separate out the different frequency components, The filter bank derives relative power levels for each frequency range. A subband coder would not rectify the signal and would use wider frequency bands.

A channel vocoder also analyzes the signal to determine the general pitch of the speech—low (bass), or high (tenor)—and also the *excitation* of the speech. Speech excitation is mainly concerned with whether a sound is *voiced* or *unvoiced*.

Formant Vocoder

It turns out that not all frequencies present in speech are equally represented. Instead, only certain frequencies show up strongly, and others are weak. This is a direct consequence of how speech sounds are formed, by resonance in only a few chambers of the mouth, throat, and nose. The important frequency peaks are called *formants* . The peak locations however change in time,as speech continues. For example, two different vowel sounds would activate different sets of formants—this reflects the different vocal-tract configurations necessary to form each vowel. Usually, a small segment of speech is analyzed, say 10–40 ms, and formants are found. A *Formant Vocoder* works by encoding only the most important frequencies.

Q.8 a. When should RTP be used and when should RTSP be used? Is there any advantage in combining the protocols? (2+2)

Answer:

Real-Time Transport Protocol (RTP), is designed for the transport of real-time data, such as audio and video streams. As we have seen, networked multimedia applications have diverse characteristics and demands; there are also tight interactions between the network and the media. Hence, RTP's design follows two key principles, namely *application layer framing*, i.e., framing for media data should be performed properly by the application layer, and *integrated layer processing*, i.e., integrating multiple layers into one to allow efficient cooperation.

The Real Time Streaming Protocol (RTSP) is a network control protocol designed for use in entertainment and communications systems to control streaming media servers. The protocol is used for establishing and controlling media sessions between end points. Clients of media servers issue VCR-like commands, such as play and pause, to facilitate real-time control of playback of media files from the server. The transmission of streaming data itself is not a task of the RTSP protocol. Most RTSP servers use the Real-time Transport Protocol (RTP) for media stream delivery, however some vendors implement proprietary transport protocols. The RTSP server from RealNetworks, for example, also features RealNetworks' proprietary RDT stream transport.

b. State any four parameters on which Quality of service for multimedia depends. (4)

Answer:

- Quality of service parameters:
- Supply time for initial connection
- Fault rate
- Fault repair time
- Unsuccessful call ratio
- Call set-up time
- Response times for operator services
- Response time for directory enquiry services

c. Explain MP3 coding technique with a block diagram. (4+4)

Answer:

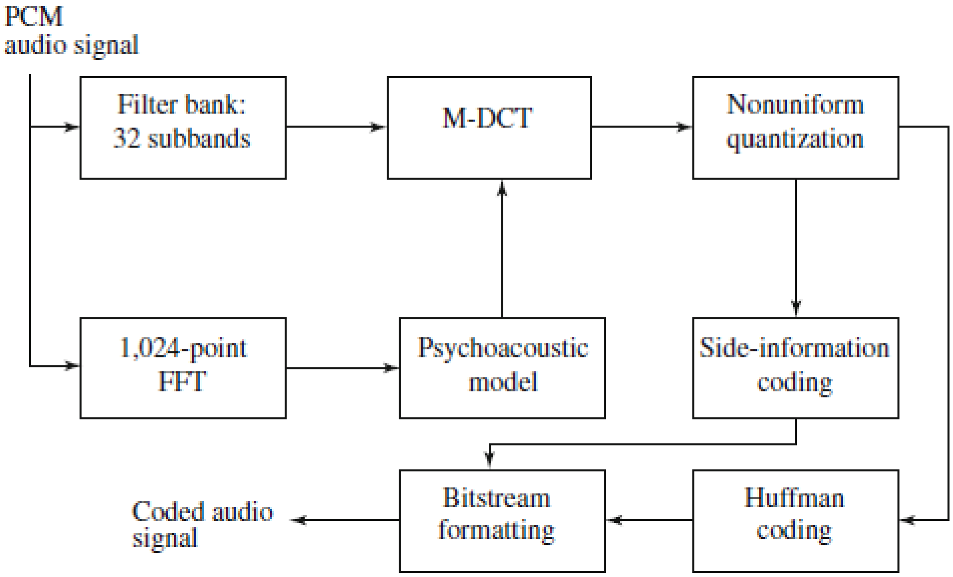
The overall algorithm is broken up into 4 main parts.

Part 1 divides the audio signal into smaller pieces, these are called frames. An MDCT filter is then performed on the output.

Part 2 passes the sample into a 1024-point FFT, and then the psychoacoustic model is applied. Another MDCT filter is performed on the output.

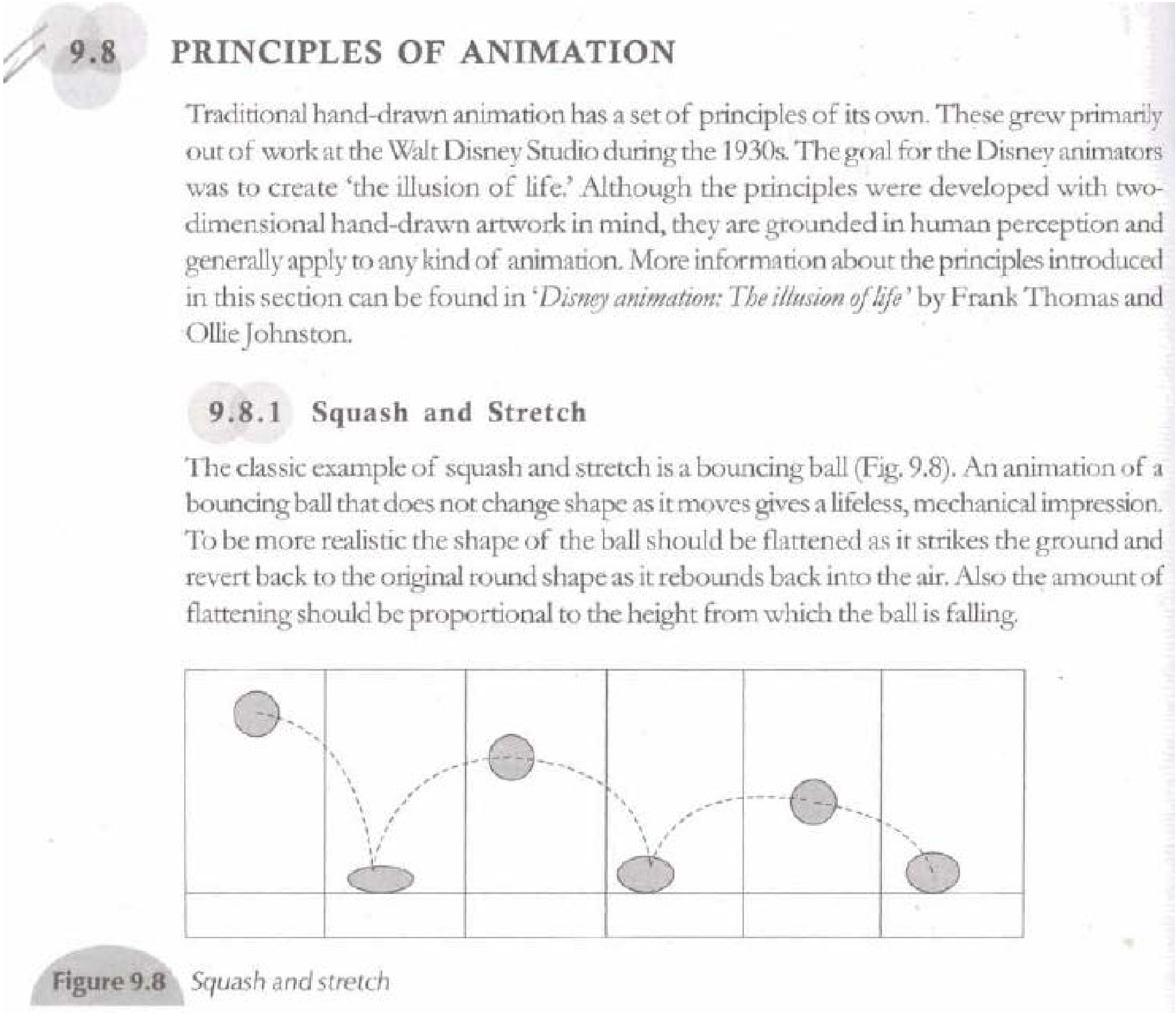
Part 3 quantifies and encodes each sample. This is also known as noise allocation. The noise allocation adjusts itself in order to meet the bit rate and sound masking requirements.

Part 4 formats the bitstream, called an audio frame. An audio frame is made up of 4 parts, The Header, Error Check, Audio Data, and Ancillary Data.



Q.9 a. What are the various techniques of animation in multimedia? Explain principles of animation. (4+4)

Answer:



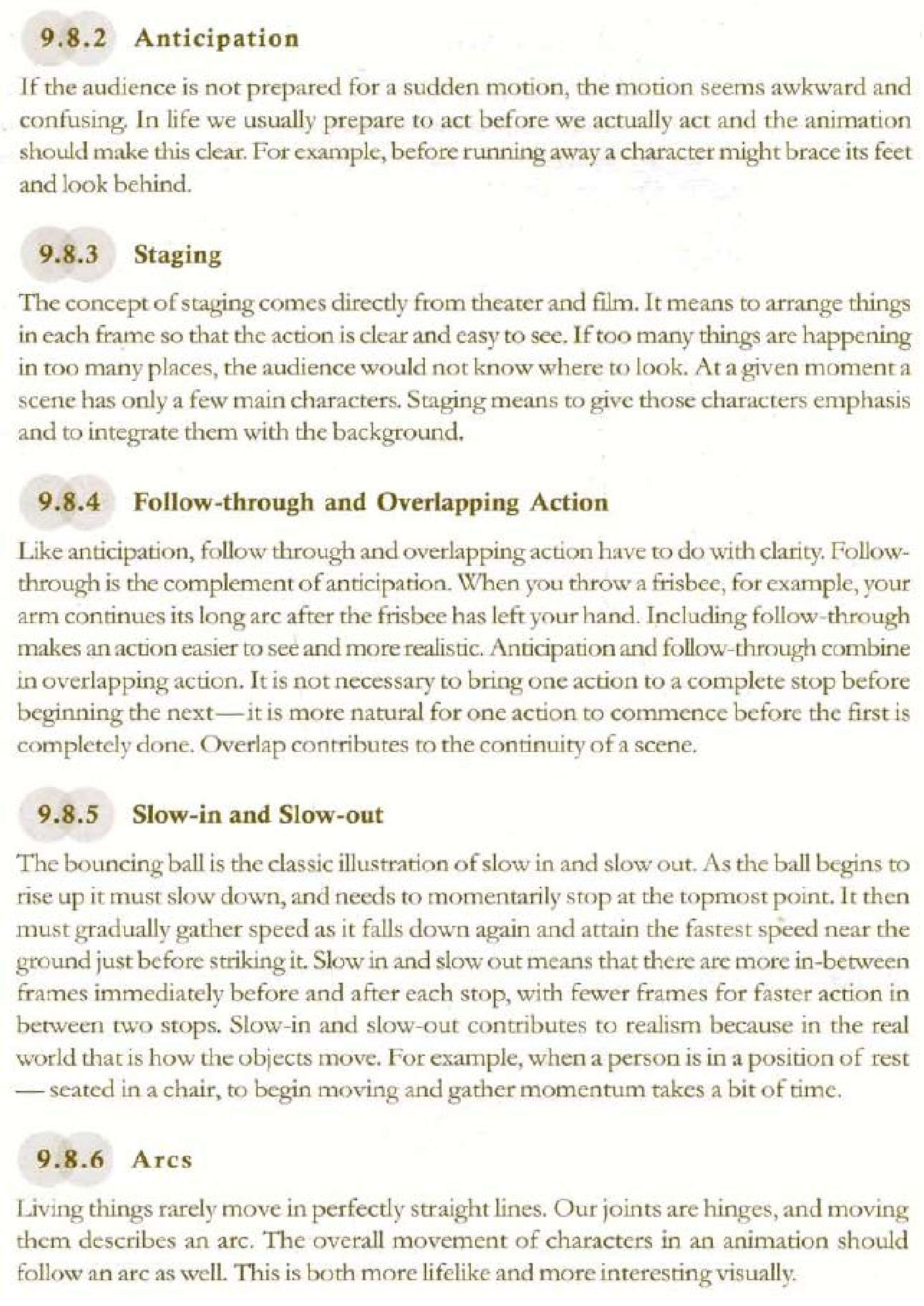
9.8 PRINCIPLES OF ANIMATION

Traditional hand-drawn animation has a set of principles of its own. These grew primarily out of work at the Walt Disney Studio during the 1930s. The goal for the Disney animators was to create 'the illusion of life.' Although the principles were developed with two-dimensional hand-drawn artwork in mind, they are grounded in human perception and generally apply to any kind of animation. More information about the principles introduced in this section can be found in '*Disney animation: The illusion of life*' by Frank Thomas and Ollie Johnston.

9.8.1 Squash and Stretch

The classic example of squash and stretch is a bouncing ball (Fig. 9.8). An animation of a bouncing ball that does not change shape as it moves gives a lifeless, mechanical impression. To be more realistic the shape of the ball should be flattened as it strikes the ground and revert back to the original round shape as it rebounds back into the air. Also the amount of flattening should be proportional to the height from which the ball is falling.

Figure 9.8 *Squash and stretch*



9.8.2 Anticipation

If the audience is not prepared for a sudden motion, the motion seems awkward and confusing. In life we usually prepare to act before we actually act and the animation should make this clear. For example, before running away a character might brace its feet and look behind.

9.8.3 Staging

The concept of staging comes directly from theater and film. It means to arrange things in each frame so that the action is clear and easy to see. If too many things are happening in too many places, the audience would not know where to look. At a given moment a scene has only a few main characters. Staging means to give those characters emphasis and to integrate them with the background.

9.8.4 Follow-through and Overlapping Action

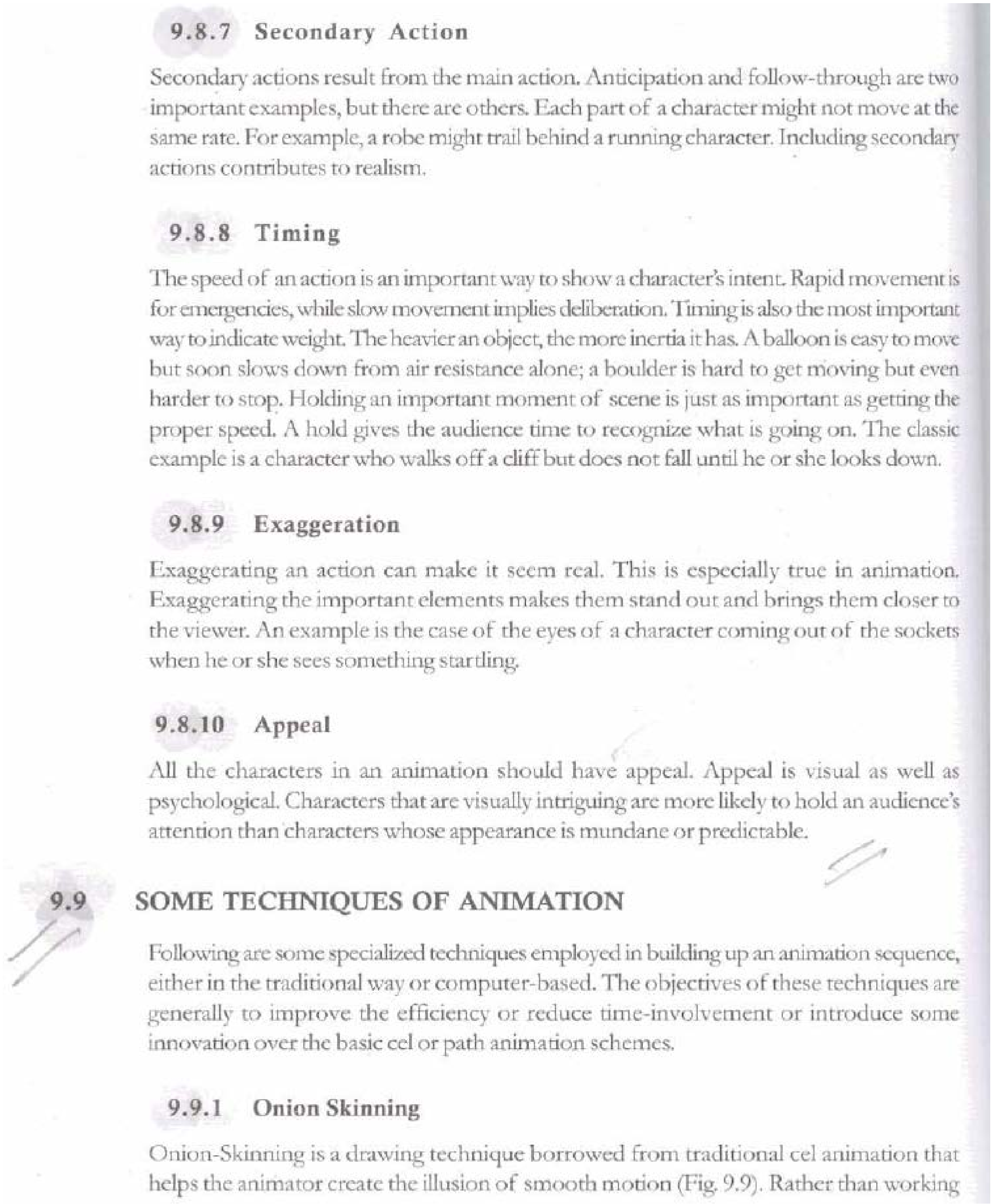
Like anticipation, follow through and overlapping action have to do with clarity. Follow-through is the complement of anticipation. When you throw a frisbee, for example, your arm continues its long arc after the frisbee has left your hand. Including follow-through makes an action easier to see and more realistic. Anticipation and follow-through combine in overlapping action. It is not necessary to bring one action to a complete stop before beginning the next—it is more natural for one action to commence before the first is completely done. Overlap contributes to the continuity of a scene.

9.8.5 Slow-in and Slow-out

The bouncing ball is the classic illustration of slow in and slow out. As the ball begins to rise up it must slow down, and needs to momentarily stop at the topmost point. It then must gradually gather speed as it falls down again and attain the fastest speed near the ground just before striking it. Slow in and slow out means that there are more in-between frames immediately before and after each stop, with fewer frames for faster action in between two stops. Slow-in and slow-out contributes to realism because in the real world that is how the objects move. For example, when a person is in a position of rest — seated in a chair, to begin moving and gather momentum takes a bit of time.

9.8.6 Arcs

Living things rarely move in perfectly straight lines. Our joints are hinges, and moving them describes an arc. The overall movement of characters in an animation should follow an arc as well. This is both more lifelike and more interesting visually.



9.8.7 Secondary Action

Secondary actions result from the main action. Anticipation and follow-through are two important examples, but there are others. Each part of a character might not move at the same rate. For example, a robe might trail behind a running character. Including secondary actions contributes to realism.

9.8.8 Timing

The speed of an action is an important way to show a character's intent. Rapid movement is for emergencies, while slow movement implies deliberation. Timing is also the most important way to indicate weight. The heavier an object, the more inertia it has. A balloon is easy to move but soon slows down from air resistance alone; a boulder is hard to get moving but even harder to stop. Holding an important moment of scene is just as important as getting the proper speed. A hold gives the audience time to recognize what is going on. The classic example is a character who walks off a cliff but does not fall until he or she looks down.

9.8.9 Exaggeration

Exaggerating an action can make it seem real. This is especially true in animation. Exaggerating the important elements makes them stand out and brings them closer to the viewer. An example is the case of the eyes of a character coming out of the sockets when he or she sees something startling.

9.8.10 Appeal

All the characters in an animation should have appeal. Appeal is visual as well as psychological. Characters that are visually intriguing are more likely to hold an audience's attention than characters whose appearance is mundane or predictable.

9.9 SOME TECHNIQUES OF ANIMATION

Following are some specialized techniques employed in building up an animation sequence, either in the traditional way or computer-based. The objectives of these techniques are generally to improve the efficiency or reduce time-involvement or introduce some innovation over the basic cel or path animation schemes.

9.9.1 Onion Skinning

Onion-Skinning is a drawing technique borrowed from traditional cel animation that helps the animator create the illusion of smooth motion (Fig. 9.9). Rather than working on each frame in isolation, animators lay these transparent cels one on top of the other. This enables them to see previous and following frames while they are drawing the current frame. Onion skinning is an easy way to complete sequence of frames at a glance and to see how each frame flows into the frames before and after.

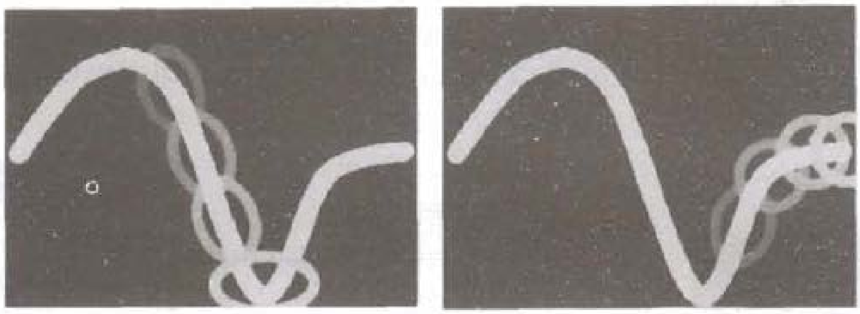


Figure 9.9 *Onion-skinning*

9.9.2 Motion Cycling

Human and animal motion such as walking, running and flying, is mainly a repetitive action that is best represented by a cycle. A walk cycle requires from 8 to 12 frames (Fig. 9.10). The sequence usually falls into two halves. The first half begins at an extreme (frames 1–2): the feet are at their farthest apart, with the back toe and front heel touching the ground. In the remainder of the first half (frames 3 – 4) the legs trade position. So do the arms – but when the left leg is forward, the right arm is forward, and vice versa. The second half of the cycle (frames 5 – 8) is simply a variation of the first half, but with arms and legs reversed. The finished walking character can be used as a moving cel, i.e. a sprite.

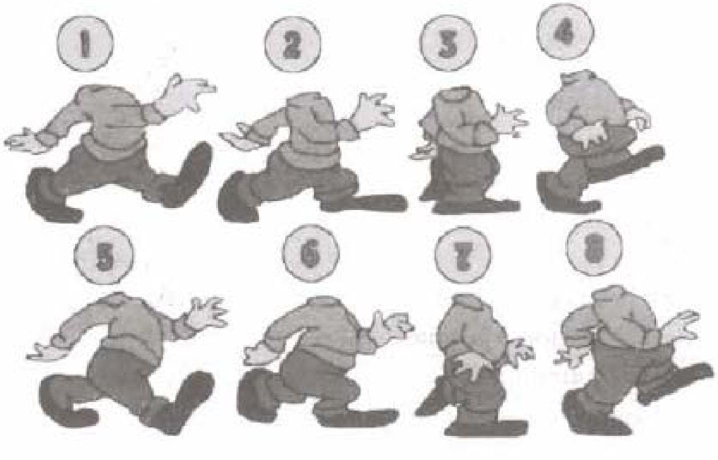


Figure 9.10 *Motion-cycling*

9.9.3 Masking

Artists often use cutouts of a red plastic called rubylith to protect parts of their work from the application of paint. A mask in a computer program is in a sense a model of the plastic masks — it protects parts of a frame from effects of other editing tools. This technique can be used to make an animated object move "behind" the protected area. In Fig. 9.11 the frame of the TV is masked, so that the scrolling text do not appear in front of the frame, but is only visible within the TV screen.

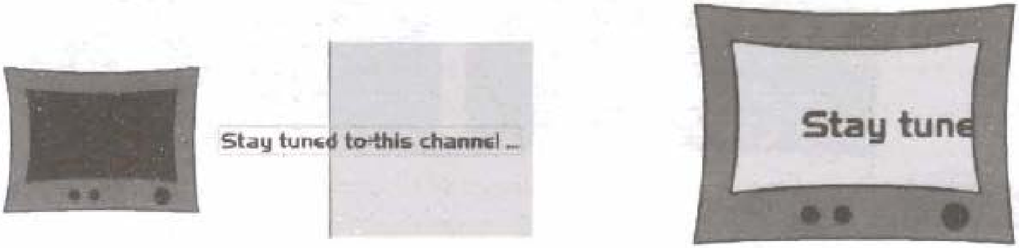


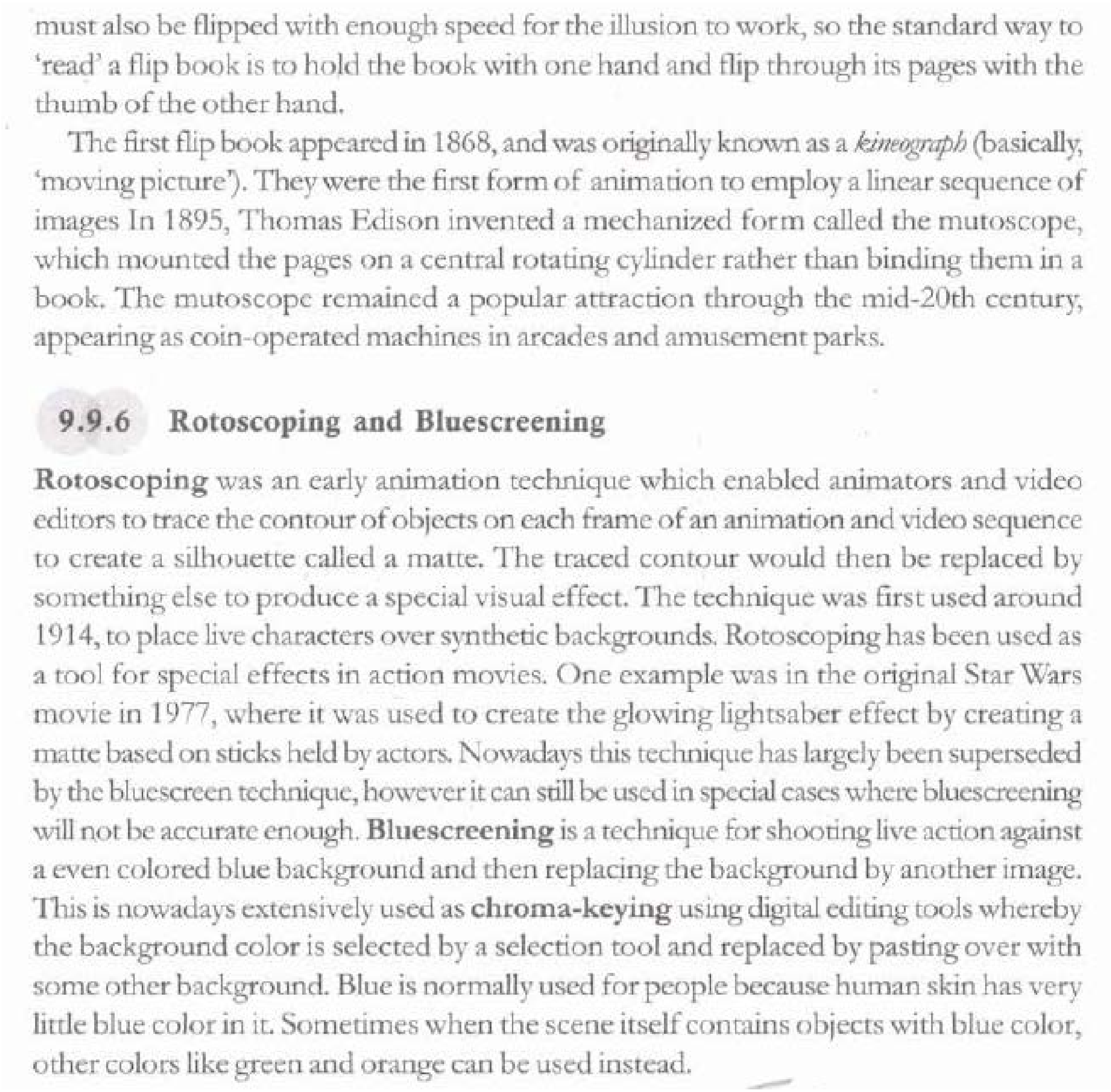
Figure 9.11 *Masking*

9.9.4 Adding Sound

Sound is an important enhancement to moving images. Background music can evoke emotions. Sounds that arise from the actions being viewed can clarify what is happening and create an effect of realism. Narration or voice-over can provide information that is missing from visual images. When sound matches the visuals — for example, a door opening or a person speaking, it is called synchronous sound. When sound is independent of the visuals, such as background music, it is called asynchronous sound.

9.9.5 Flip-book Animation

A **flip book** is a book with a series of pictures varying gradually from one page to the next, so that when the pages are turned rapidly, the pictures appear to animate, simulating motion or some other change. Flip books are often illustrated books for children, but may also be geared towards adults and employ a series of photographs rather than drawings. Flip books are not always separate books, but may appear as an added feature in ordinary books or magazines, often in the page corners. Flip books are essentially a primitive form of animation. Like motion pictures, they rely on persistence of vision to create the illusion that continuous motion is being seen rather than a series of discontinuous images being exchanged in succession. Because of this, flip book images will only appear on one side of the book's pages, so that rather than 'reading' left to right, a viewer simply stares at the same location of the pictures in the flip book as the pages turn. The book



must also be flipped with enough speed for the illusion to work, so the standard way to 'read' a flip book is to hold the book with one hand and flip through its pages with the thumb of the other hand.

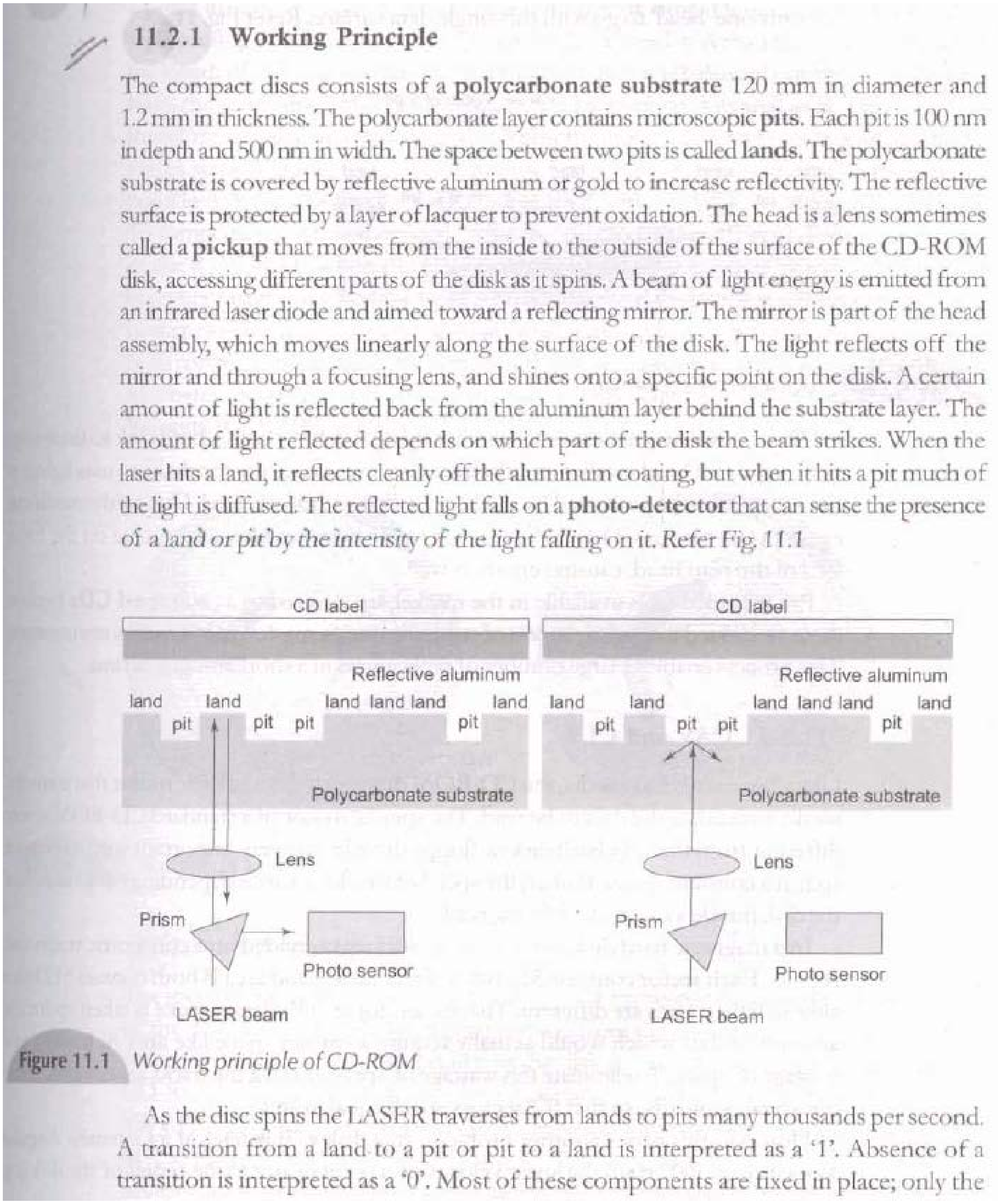
The first flip book appeared in 1868, and was originally known as a *kineograph* (basically, 'moving picture'). They were the first form of animation to employ a linear sequence of images In 1895, Thomas Edison invented a mechanized form called the mutoscope, which mounted the pages on a central rotating cylinder rather than binding them in a book. The mutoscope remained a popular attraction through the mid-20th century, appearing as coin-operated machines in arcades and amusement parks.

9.9.6 Rotoscoping and Bluescreening

Rotoscoping was an early animation technique which enabled animators and video editors to trace the contour of objects on each frame of an animation and video sequence to create a silhouette called a matte. The traced contour would then be replaced by something else to produce a special visual effect. The technique was first used around 1914, to place live characters over synthetic backgrounds. Rotoscoping has been used as a tool for special effects in action movies. One example was in the original Star Wars movie in 1977, where it was used to create the glowing lightsaber effect by creating a matte based on sticks held by actors. Nowadays this technique has largely been superseded by the bluescreen technique, however it can still be used in special cases where bluescreening will not be accurate enough. **Bluescreening** is a technique for shooting live action against a even colored blue background and then replacing the background by another image. This is nowadays extensively used as **chroma-keying** using digital editing tools whereby the background color is selected by a selection tool and replaced by pasting over with some other background. Blue is normally used for people because human skin has very little blue color in it. Sometimes when the scene itself contains objects with blue color, other colors like green and orange can be used instead.

b. Describe the working principle of encoding digital data on a CD Surface. Differentiate between CD-R and CD-RW. (4+4)

Answer:

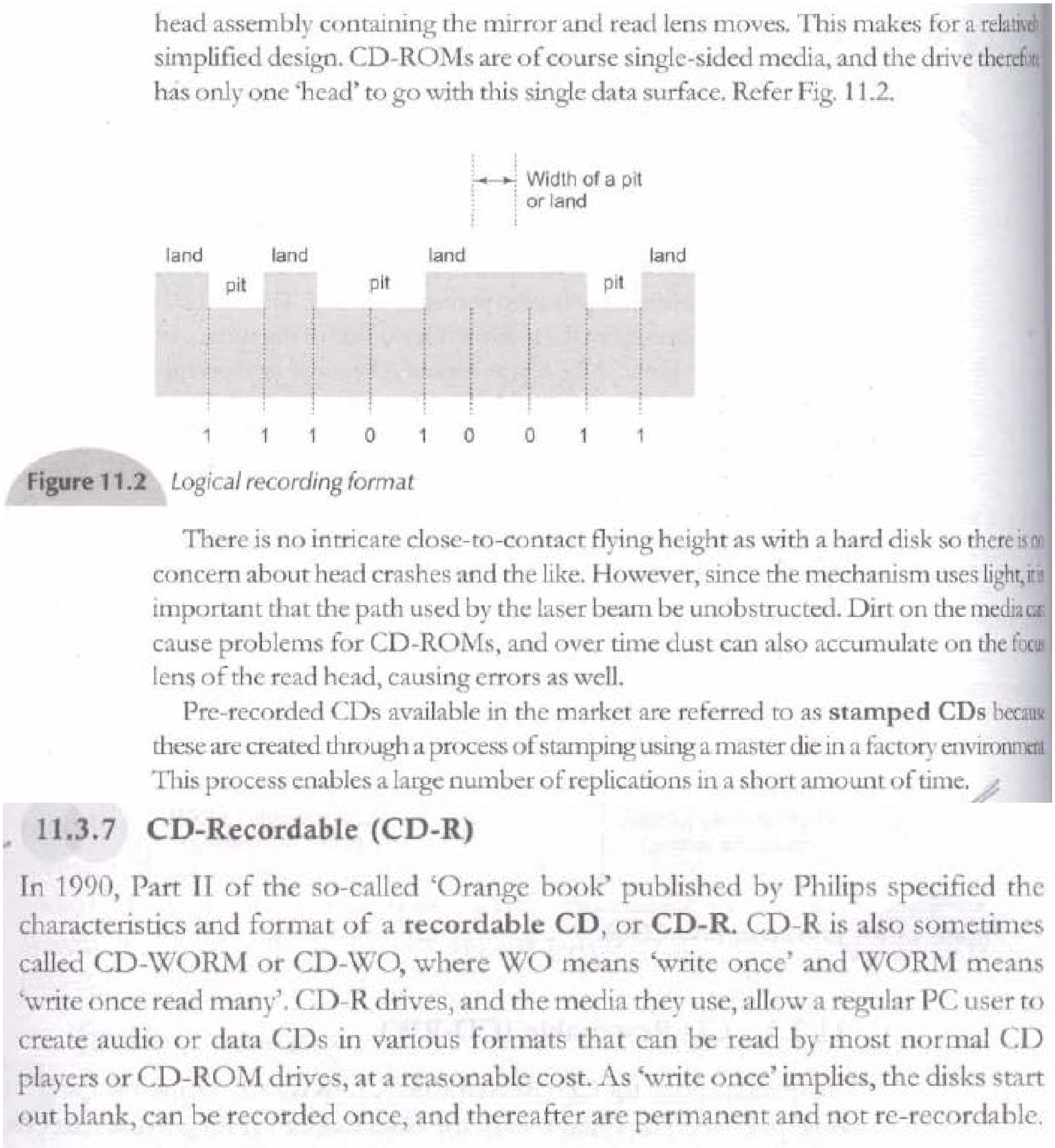


11.2.1 Working Principle

The compact discs consists of a **polycarbonate substrate** 120 mm in diameter and 1.2 mm in thickness. The polycarbonate layer contains microscopic **pits.** Each pit is 100 nm in depth and 500 nm in width. The space between two pits is called **lands.** The polycarbonate substrate is covered by reflective aluminum or gold to increase reflectivity. The reflective surface is protected by a layer of lacquer to prevent oxidation. The head is a lens sometimes called a **pickup** that moves from the inside to the outside of the surface of the CD-ROM disk, accessing different parts of the disk as it spins. A beam of light energy is emitted from an infrared laser diode and aimed toward a reflecting mirror. The mirror is part of the head assembly, which moves linearly along the surface of the disk. The light reflects off the mirror and through a focusing lens, and shines onto a specific point on the disk. A certain amount of light is reflected back from the aluminum layer behind the substrate layer. The amount of light reflected depends on which part of the disk the beam strikes. When the laser hits a land, it reflects cleanly off the aluminum coating, but when it hits a pit much of the light is diffused. The reflected light falls on a **photo-detector** that can sense the presence of a *land or pit by the intensity of the light falling* on it. Refer Fig. 11.1

Figure 11.1 *Working principle of CD-ROM*

As the disc spins the LASER traverses from lands to pits many thousands per second. A transition from a land to a pit or pit to a land is interpreted as a '1'. Absence of a transition is interpreted as a '0'. Most of these components are fixed in place; only the



head assembly containing the mirror and read lens moves. This makes for a relatively simplified design. CD-ROMs are of course single-sided media, and the drive therefore has only one 'head' to go with this single data surface. Refer Fig. 11.2.

Width of a pit or land
land
pit
land
pit
land
pit
land
1 1 1 0 1 0 0 1 1

Figure 11.2 *Logical recording format*

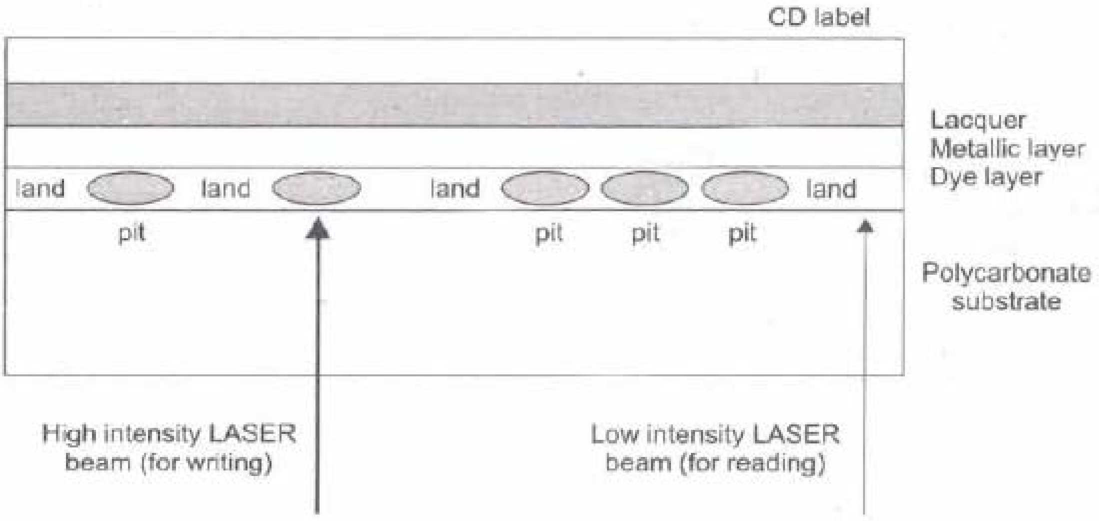
There is no intricate close-to-contact flying height as with a hard disk so there is no concern about head crashes and the like. However, since the mechanism uses light, it is important that the path used by the laser beam be unobstructed. Dirt on the media can cause problems for CD-ROMs, and over time dust can also accumulate on the focus lens of the read head, causing errors as well.

Pre-recorded CDs available in the market are referred to as **stamped CDs** because these are created through a process of stamping using a master die in a factory environment. This process enables a large number of replications in a short amount of time.

11.3.7 CD-Recordable (CD-R)

In 1990, Part II of the so-called 'Orange book' published by Philips specified the characteristics and format of a **recordable CD**, or **CD-R.** CD-R is also sometimes called CD-WORM or CD-WO, where WO means 'write once' and WORM means 'write once read many'. CD-R drives, and the media they use, allow a regular PC user to create audio or data CDs in various formats that can be read by most normal CD players or CD-ROM drives, at a reasonable cost. As 'write once' implies, the disks start out blank, can be recorded once, and thereafter are permanent and not re-recordable.

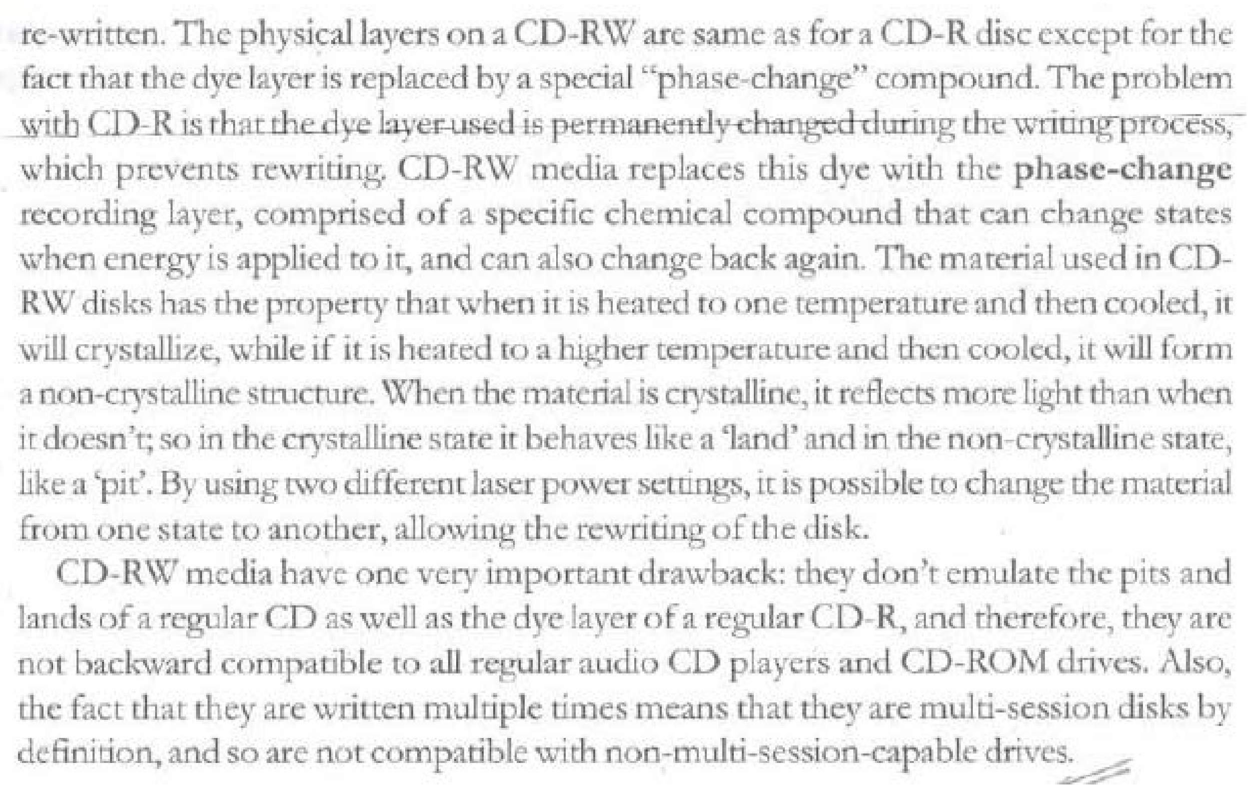
Part III of the 'Orange book' defines rewriteable CDs, which are erasable, unlike CD-R CD-Recordable (CD-R) media starts with a polycarbonate substrate, just like regular CDs do. On top of the polycarbonate, a special photosensitive dye layer is deposited; on top of that a metal reflective layer is applied (such as a gold or silver alloy) and then finally, a protective lacquer layer. It is these different layers that give CD-R media their different visual appearance from regular CDs. The key to the media is the dye layer (and the special laser used in the drives.) It is chosen so that it has the property that when light from a specific type and intensity of laser is applied to it, it heats up rapidly and changes its chemical composition. As a result of this change in chemical composition, the area 'burned' reflects less light than the areas that do not have the laser applied. This system is designed to mimic the way light reflects cleanly off a 'land' (from the metallic layer behind) on a regular CD, but is scattered by a 'pit', so an entire disk is created from burned and non-burned areas, just like how a regular CD is created from pits and lands. The result is that the created CD media will play in regular CD players as if it were a regular CD, in most cases. Since the media is being physically altered by a process of heat and chemistry, the change is permanent and irreversible. Once any part of the CD has been written, the data is there forever. Some drives allow you to record some information in one sitting, and then more information later on, if the disk is not yet full. This is called **multi-session** recording, and requires a CD player capable of recognizing multi-session disks in order to use the burned disk. See Fig. 11.8.



Recording on a CD-R

11.3.8 CD-Rewritable (CD-RW)

The specifications for **CD-Rewritable** (CD-RW) are codified as Part III of the 'Orange book' published by Philips. with the objective of creating a disc that can be erased and



re-written. The physical layers on a CD-RW are same as for a CD-R disc except for the fact that the dye layer is replaced by a special "phase-change" compound. The problem with CD-R is that the dye layer used is permanently changed during the writing process, which prevents rewriting. CD-RW media replaces this dye with the **phase-change** recording layer, comprised of a specific chemical compound that can change states when energy is applied to it, and can also change back again. The material used in CD-RW disks has the property that when it is heated to one temperature and then cooled, it will crystallize, while if it is heated to a higher temperature and then cooled, it will form a non-crystalline structure. When the material is crystalline, it reflects more light than when it doesn't; so in the crystalline state it behaves like a 'land' and in the non-crystalline state, like a 'pit'. By using two different laser power settings, it is possible to change the material from one state to another, allowing the rewriting of the disk.

CD-RW media have one very important drawback: they don't emulate the pits and lands of a regular CD as well as the dye layer of a regular CD-R, and therefore, they are not backward compatible to all regular audio CD players and CD-ROM drives. Also, the fact that they are written multiple times means that they are multi-session disks by definition, and so are not compatible with non-multi-session-capable drives.

TEXT BOOK

I. Fundamentals of Multimedia, Ze-Nian Li and Mark S. Drew, Pentice Hall, Edition – 2007

II. Principles of Multimedia, Ranjan Parekh, Tata McGraw-Hill, Edition 2006