# Chapter 16
# Multimedia

## 16.1 What is Multimedia?:

The use of different tools addressing different human senses is called multimedia. Specifically, this refers to use of all of signal representations including text-images-audio-video-movies and 3D imagery.

The multimedia terminal could be a PC, TV or a large projection system. It could be passive if the information flow is in one direction or interactive meaning that a dialog is set up between man and machine or between two users via machine. The machine in this case could be a nearby PC or could be a server located thousands of miles away with a PC acting as a means of communication through a network. In person to system communication, the PC accesses server files. With the advent of mobiles, PC's, TV's and mobiles form an integrated set to access a massive network called internet web.

There are different uses for multimedia. Interpersonal communication involves computer telephony – voice mail - teleconferencing - voice over IP - email and desktop videoconference. In computer telephony, the user can take part in telephone calls through the PC using a microphone, telephone card and computer telephony integration software. Voice mail is used in the event that the called party is on the move in which case a spoken message can be left for him in his voice mail box located in a voice mail server which he could access at a later time.

Teleconferencing involves multiple interconnected telephones/PCs among different simultaneous participants. The internet can now support telephone calls through PC-PC link which is called voice over IP (internet protocol). In addition, the internet web can of course support email. This could be text only or text on image or speech and video. Altogether, such email is called multimedia (MM) mail. The mailbox is actually located at the server and is accessed only upon request. The web can also support video telephony and desktop conferencing, where two or more parties may be involved in a virtual simultaneous meeting. There are other forms of interactive applications over the internet.

The total information is stored in different servers which provide a vast library of documents. Each document is a linked set of pages. Linkages between pages are called hyperlinks. They appear as pointers, flashers or references to other pages of the same document or other documents within the web. The optional linkage points within the document are defined by the designer of the document and are known as anchors. In a text document, such links are called has a unique address: a uniform resource locator (URL), which identifies both the location of the server on the internet where the first (front or home) page of the document is stored, and also the file reference on that server .All hyperlinks on this and other pages have similar URLs associated with them. The physical location of the page is transparent to the user and can be located anywhere on the web. The format used for writing documents is Hypertext Mark up of Language (HTML). It is also used for writing client software to explore contents of web servers This is called Browser. It is the means of communication with the server. The user may click at any anchor point to activate a linkage. The information stored at that point is thus retrieved or the user is referred to another anchor point.

There are many applications for multimedia in addition to personal communication. Multimedia is used in education, where each student can learn at his pace. It is also used in culture, tourism, electric journalism information gathering, teleshopping, telebanking, reservations, research and electronic government, where different paper work could be done through the internet. Multimedia is also exclusively used in entertainment. Video on demand (VOD) or movie on demand (MOD) is being available on a number of sites on the internet or on TV for paid viewing. This is done using a video server which has a huge collection of digital movies displayed upon request. Interactive TV is becoming a reality using a set top box (STB) or email, mobile SMS (short message signal).

All types of multimedia must be stored and processed within a computer, i.e. they must be in a digital form. In case of text, character strings are translated to bit strings with a codeword for each character. Computer generated graphical images are made up of a mix of lines, circles and squares, each represented in a digital form. A line, for example, is represented by start and end pairs of coordinates. Microphones and video cameras must be connected to A/D converters, while for speakers and display devices, D/A converters must be used. Thus, different media types may be integrated together to give a bit stream, which can be transmitted over a single all digital network. The bandwidth of the communication network limits the bandwidth of the information to be transmitted.

### 16.2 Text Representation:

Text may be divided into 3 categories: unformatted (plain) text, formatted (rich) text, and hyper text. In plain text, strings of fixed size characters forming limited character sets are used, each gives a 7 bit code. This is used in videotex and teletext, and appears as broadcast information services in TV. Formatted text involves strings of characters of different styles, sizes, and shapes. It involves also tables, graphics and image insertions. It uses format control character sequences. Hypertext employs HTML which is used in the creation of web pages. browser enabling and specification of hyperlinks. It also enables the user to browse interactively and use sound and video clips. Specification of a hyperlink is made by specifying the URL of the required page and the name of the link.

### 16.3 Image Representation:

Images may be divided into 3 categories: graphics, digitized document and digitized picture. All those types are displayed and printed as a 2D matrix of individual pixels in computer memory or file for display or printing. Graphics refer to computer generated images. To create graphics, tools are used to draw freehand objects, lines, arcs, squares, etc. by using a combination of cursor symbols, a pencil (or paint brush) and a mouse. Objects may be edited, i.e., change shape – color – size, or by introducing predrawn images, either previously created or from a gallery of images (clipart). Textural information can then be added from a previously stored library. Filling and adding light and shadows are used to give a 3D – like appearance The transition from a model to final graphic is called rendering. A computer display screen can be considered as being made up of a two dimensional matrix of individual picture elements (pixels). VGA (video graphic array) is a common type of display. It consists of a matrix of 640 horizontal pixel, x 480 vertical pixels. With 8 bits/pixel, we have 256 different colors.

In creating graphics including freehand objects, we note that each object is made up of a series of lines that are connected to each other. What may appear as a curved line is a series of very short lines, each made up of a string of pixels which in the limit have the resolution of a pair of adjacent pixels on the screen.

Each object has a number of attributes associated with it. These include: shape – size – color – shadow. Editing an object involves changing one or more of the attributes. An open object has the start of the first line and the end of the last line not connected, while a closed object has no start or end. Usually, objects can be drawn on the screen by specifying the name of the object and its attributes including color fill, shadow effect and pixel locations. Graphics consist of a set of commands each with attributes that are necessary to draw the object. The representation of a complete graphic is in a way similar to writing a program in a high level language. Representing a computer graphic with a high level language is similar to a source code of a high level program. Alternatively, we may use low level (machine) code to describe the actual pixel/image of the graphic. This is similar to the byte string corresponding to the machine code of a program. This is called bit map format. A graphic can be transferred over a network in either form. But high level language form is more compact, and requires less memory to store the image and requires less bandwidth for this transmission. Fractals are also used to describe an object mathematically and send the mathematical representation in order to use high level language, the destination must be able to interpret the various commands, I,e. the end user must have the same high level language compiler. So the bit map form is often used instead. To help with it, some standardization forms, such as GIF (graphical interchange format), TIFF (tagged image file format) and SRGP (simple raster graphics package) are often used to convert high level language form into the pixel image form which the end user can easily identify.

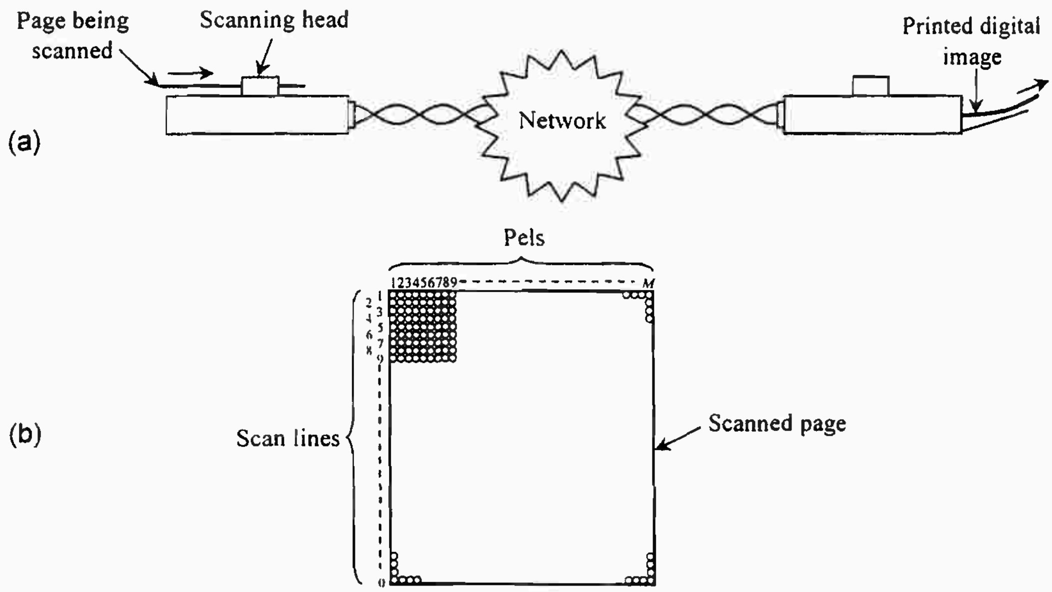
In a digitized document, such as produced by a scanner or fax where we have 3.85 or 7.7 lines/mm, each line is scanned and the output of the scanner is digitized to a resolution of about 8 pixels/mm. Each pixel is either 0 (while) or 1 (black). A typical page yields a stream of 2 million bits. A fax printer can reproduce the original image by printing out the received stream of bits. The use of a single binary digits / pixel is best suited to scanning bitonal (black and white), images such as printed document comprising mainly textual information. (Fig. 16.1).

For a digitized picture, scanners may provide continuous tone monochromatic scan. Good quality black and white pictures may be obtained by using 8 bits/ pixel, leading to 256 grey level / element. For color scan, we must recall first that there are two color mixing systems (additive) and subtraction (Fig. 16.2). For digitized color picture, we use the subtractive color system.

**16.4 The Monitor:**

In a TV or computer monitor, the image changes from frame to frame. The persistence of the light changes but the phosphor must decay quickly. Hence it is necessary to refresh the screen. The frame refresh rate must be high enough to ensure that the eye does not notice the refreshing procedure. A low refresh rate leads to flicker caused by the fading of the image on the eye retina. To avoid this, the refresh rate must be 50 – 60 Hz. Most displays work in an analog mode. In digital TV and digitized pictures stored within the memory of a computer, the color signals are in a digital form and comprise a string of pixels at fixed number per scan line. In order to display a stored image, the pixels that make up each line are read from memory in time synchronism with the scanning process and converted to analog using a DAC.

Since the area of the computer memory that holds the string of pixels making up the image must be accessed continuously as each line is scanned, a separate block of memory known as video RAM is used to store the pixel image. In this way, the graphics program needs to be written into the video RAM, whenever either selected pixels or the total image changes.




Pel = picture element  Resolution = $M \times N$

Pel = resolution = 8 per mm

Line resolution = 3.85 / 7.7 per mm = 100 / 200 per inch

**Fig. (16.1) Fax**

*a) schematic  b) digitization format*

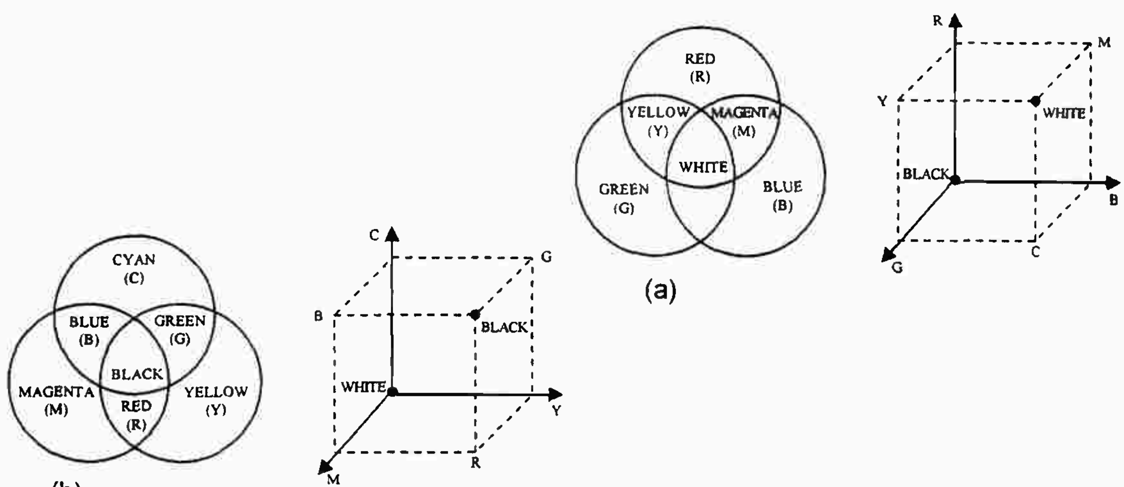



**Fig. (16.2) Color models**

*a) additive  b) subtractive*

The graphics program creates a high level version of the image interactively, and the display controller part of the program interprets sequences of display commands and converts them into displayed objects. A video controller is a hardware subsystem that reads pixel values stored in the video RAM in time synchronism with the scanning process. It converts each set of pixel values to an equivalent set of R, G, B analog signals (Fig. 16.3).

The number of bits / pixel is known as pixel depth. It determines the range of different colors that can be produced. For 4 bits / color, we have 4096 different colors and for 8 bits / color, we have $2^{24}$ (16 millions). In practice, however the eye cannot discriminate between such a wide range of colors. Hence, a select subset is sufficient. Such a set are stored in a table, and each pixel value is used as an address to a location within the table which contains the corresponding three color values. This is called color look up table (CLUT). If each pixel is 8 bits (out of 24 bit entries), then the table will provide a subset of 256 $(2^8)$ from a palette of 16 million. In this way, The amount of memory required to store an image can be reduced significantly.

In TV, we use interlaced scan, while for a PC monitor we use non-interlaced scan. Due to the proximity of the viewer in the case of PC the voltage and the velocity are high enough, and hence, the time spent by the e-beam per pixel is small. The amount of radiation in this case must be low. In the case of a TV monitor, we need higher illumination, so the time spent by the beam on a pixel must be long. This makes non interlaced scan possible. In the case a high definition TV (a wide screen TV), the refresh rate is increased to $100Hz$ to make the image brighter and flicker free for both interlaced and non interlaced display.

The aspect ratio of the monitor is defined as the ratio of the width to the height. It is equal to 4/3 in normal monitor and 16/9 for wide screen. The number of visible lines / frame (vertical resolution) is 480 (NTSC) or 576 (Europe) and 1152 (HDTV). For an aspect ratio of 4/3, and in order to avoid distortion of a square of $1\times1$ pixel, we must have $Y/480 = X/640$ where $Y$ is the vertical distance and $X$ the horizontal distance (Fig. 16. 4). We note that $X/Y = 4/3$. We call $XY$ the spatial resolution. Table (16.1) gives spatial resolution and memory requirements for 4/3 aspect ratio. The resolution of HDTV for 4/3 aspect ratio is $1440\times1152$ and for 16/9 wide screen $1920\times1152$.

**Ex. 16.1**

Derive the time to transmit the following digitized images at both 64 kbps and 1.5mbps using $640\times480\times8$ VGA

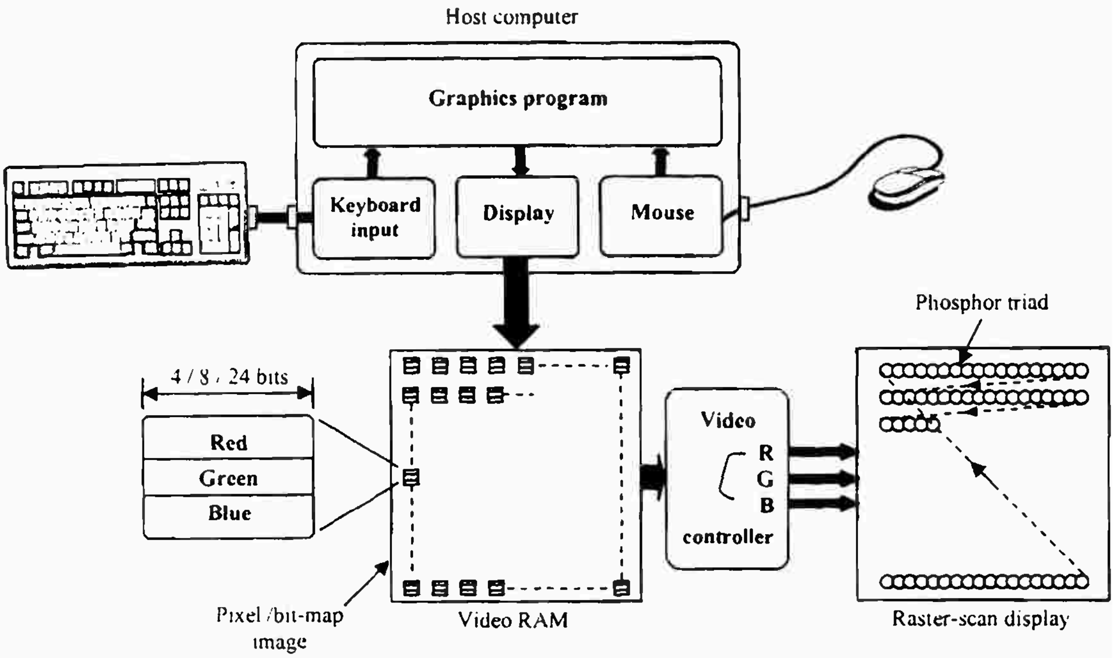



**Fig. (16.3) Raster scan display system**

**Solution**

VGA image $= 640 \times 480 \times 8 = 2.4576 M\,bits$

Time to transmit each image at 64kbps $= \dfrac{2.4576 \times 10^6}{64 \times 10^3} = 38.42$ s

Time to transmit each image at 1.5Mbps $= \dfrac{2.4576 \times 10^6}{1.5 \times 10^6} = 1.6384$ s

**Table (16.1) Example display resolutions and memory requirements.**

| Standard | Resolution | Number of colors | Memory required per frame (pytes) |
|---|---|---|---|
| VGA | $640 \times 480 \times 8$ | 256 | 307.2 kB |
| XGA | $640 \times 480 \times 16$<br>$1024 \times 768 \times 8$ | 64k<br>256 | 614.4 kB<br>786.432 kB |
| SVGA | $800 \times 600 \times 16$<br>$1024 \times 768 \times 8$<br>$1024 \times 768 \times 24$ | 64k<br>256<br>16M | 960 kB<br>786.432 kB<br>2359.296 kB |

### 16.5 Digital Camera and Scanner:

In both cases of a digital camera and a scanner, the captured image is transferred directly to the computer as it is produced. Alternatively, in the case of a digital camera, a set of digitized images can be stored within the camera itself and then downloaded into the computer. The image is captured in either case using a CCD, which is ID in the scanner and 2D in the camera. When the camera shutter is activated, each photo site stores the level of intensity of light that falls on its surface in the form of an equivalent charge. The charge is read out and converted into a digital value, using ADC. In the scanner, the photo sites comprise one row. The stored charges are read out and digitized before the next scan occurs (Fig. 16.5). For color, we may use filters for $R, G, B$ with multiple exposure, 3 coatings in the array or use a beam splitter and three CCD arrays. If the camera is connected to the computer, then the bit map can be loaded directly to the frame buffer ready to be displayed (Fig. 16.6). If the images are to be stored within the camera an IC memory is used either as a card to be inserted in the PC into the PCMCIA slot a fixed card within the camera through a cable link. Once within the computer, software can be used to insert the digital image into a document and may be sent by email.

Photoediting software may be used to manipulate a stored image. Tagged file format (TIFF / EP) is used in electronic photography to allow different types of data to be stored in the image file, such as data and time for each image.

### 16.6 Audio:

Audio signals may be divided to: speech and music. Speech bandwidth extends from 50Hz-10kHz. For speech, the sampling frequency is 20 kHz and we have 12 bits / sample. While for music, the sampling rate is 40 kHz, and we have 16 bits / sample. Prior to sampling and conversion by ADC, the signal is usually passed through a compressor to compress the amplitude of the input signal, then linear quantization is applied. This nonlinear processing-called companding - is justified since the ear is more sensitive to noise on quiet signals than it is on loud signals. Hence, the effect of quantization noise is reduced with narrower intervals for small amplitude signals. At the destination, each received codeword is first fed to DAC. The analog output from DAC is then passed on to the expander (Fig. 16.7).

There is a standard called CD-DA associated with discs used in CD players and CD ROMs used in music and multimedia information. The sampling rate is 44.1 k samples /s (23 $\mu s$ intervals), 16 bits / sample (65536 quantization level), bit rate per channel is 44 x $10^{13}$ x 16 = 705.6 kbps with 2 x 705.6 = 1.411 Mbps for stereo.

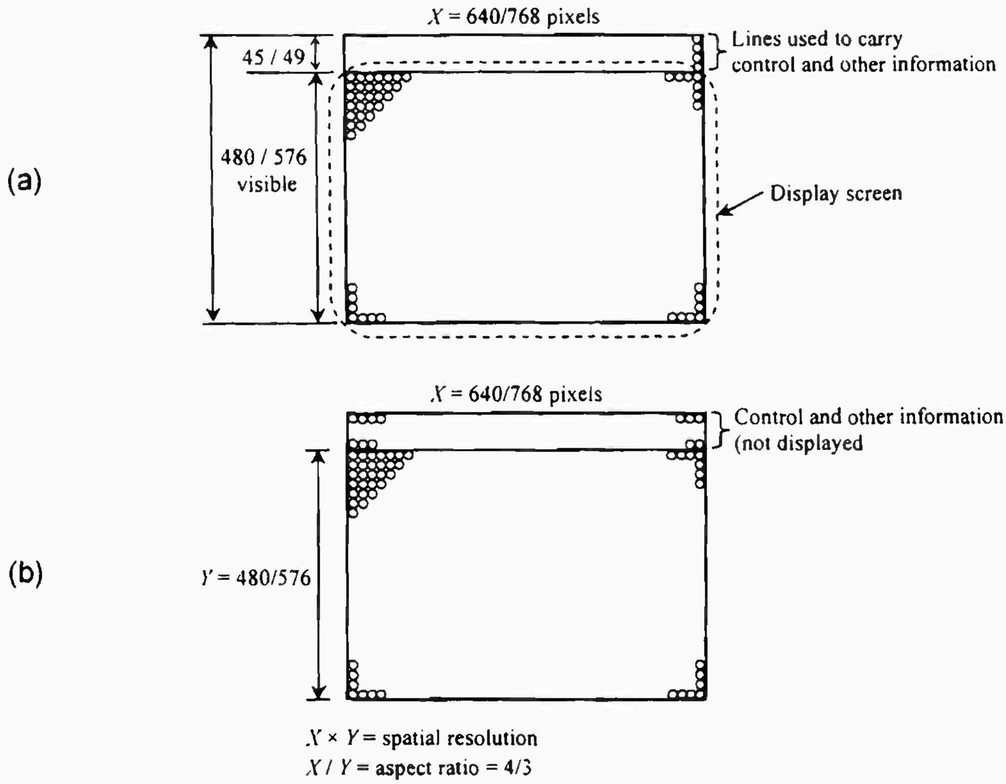



**Fig. (16.4) Screen resolutions**

*a) visible lines per frame* *b) digitization spatial resolution*

**Ex. 16.2**

Derive the bit rate and memory required to store 10 minute passage of stereophonic music.

**Solution**

Bit rate per channel = 40k x 16 = 640 kbps

Bit rate for stereo = 2 x 640k = 1280 kps

Memory size $= \dfrac{bit\ rate\ bps \times time\ s}{8bits}$

$= \dfrac{1280 \times 10^3 \times 600}{8} = 96MB$

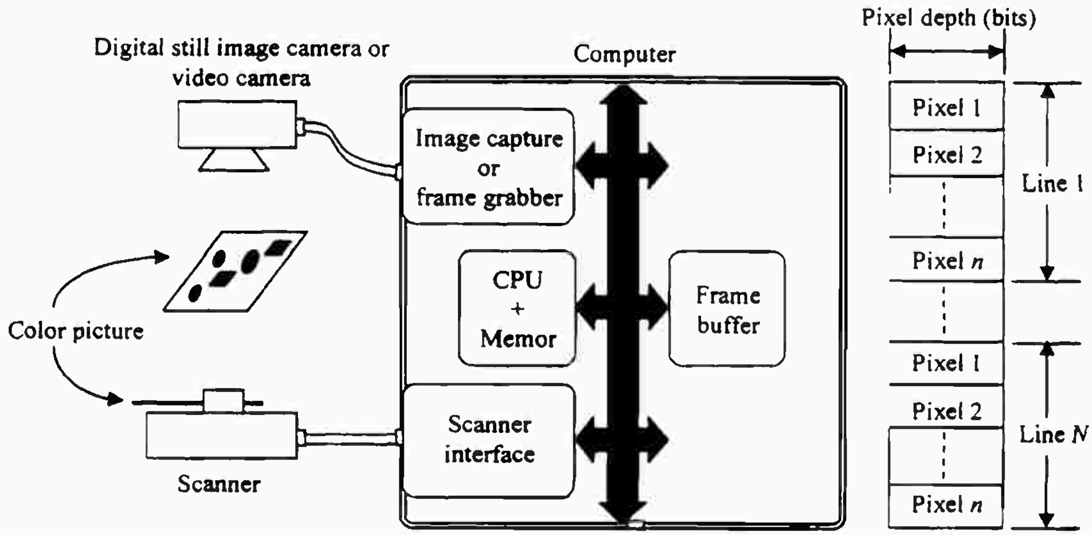



**Fig. (16.5) Image Capture**

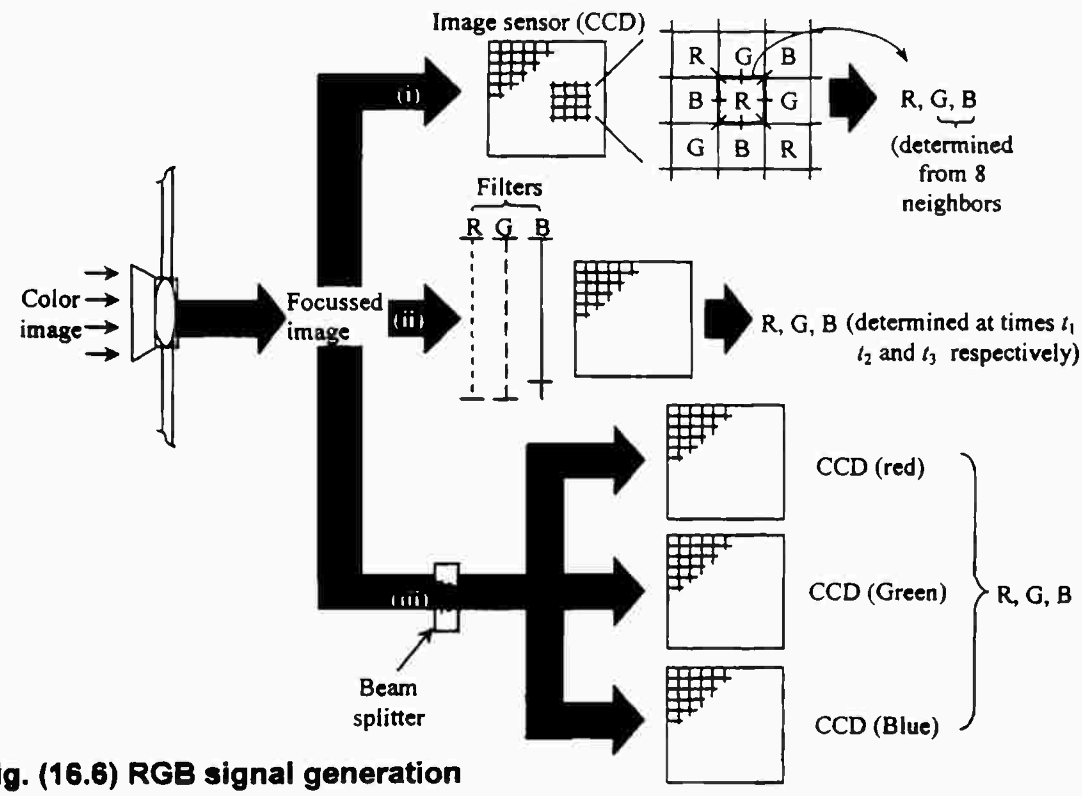



**Fig. (16.6) RGB signal generation**

**Ex. 16.3**

Derive the storage capacity of a CD ROM to store 10 minute multimedia (CD quality), and the time to transmit it using bit rate 64kbps. What is the time for 30s interaction?

**Solution**

CD–DA digitization produces bit rate 1.411Mbps.

Storage capacity for 10 minutes $= \dfrac{1.411 \times 600}{8} = 105.825 MB = 8.466 \times 10^{8} bits$

Transmission time $= \dfrac{8.466 \times 10^{8}}{64 \times 10^{3}} = 1.3228 \times 10^{4} s = 3.67 hours$

For 30 s interaction

Transmission time $= \dfrac{42.33 \times 10^{6}}{64 \times 10^{3} \times 60} = 11$ minutes

This is too high for interaction.

**16.7 Synthesized Audio:**

We have seen from the previous two examples that the amount of memory required to store a digitized audio even for small passages is quite high. For this reason, synthesized audio is often used in multimedia applications. It saves 2 – 3 orders of magnitude of memory storage and it is much easier to edit and mix passages together. An audio synthesizer uses a set of sound generators, a keyboard (piano) and a PC. The PC takes input commands from the keyboard and outputs them to DAC, then to the sound generators, then into the speakers. For every key pressed, a message of a certain code is produced and read by the PC program. The message includes which key and the pressure applied. A control panel contains a range of switches and slides to indicate to the computer program possible sound effects associated with the key pressing (Fig. 16.8). A secondary storage interface allows the sequence of messages including those associated with the control panel to be saved. There are also programs to allow the user to edit a previously stored passage or mix passages together.

The sequencer program provides synchronization as it feeds the merged passage to the sound generators

A MIDI (music instrument digital interface) connects several inputs such as electric guitar in addition to the keyboard and control panel. Different types of devices have a MIDI code associated with them. Thus, a passage of audio produced by a synthesizer consists of a very compact sequence of messages, each comprising a string of bytes which can be either played out by the sequencer program directly or saved in a file. In multimedia, this file is recalled at a link in the text.

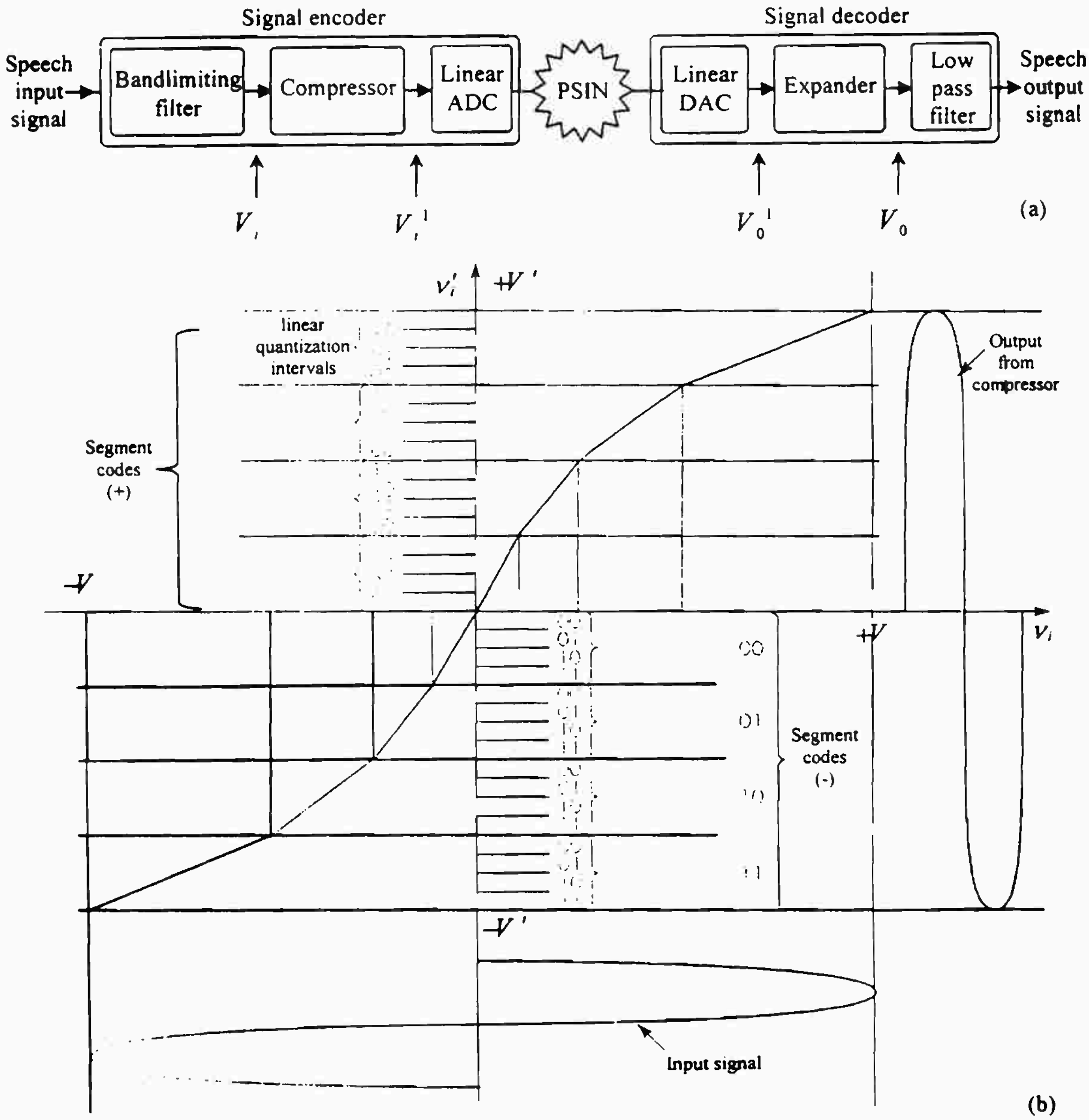



**Fig. (16.7) Audio encoding and companding**
*a) block diagram* *b) characteristic*

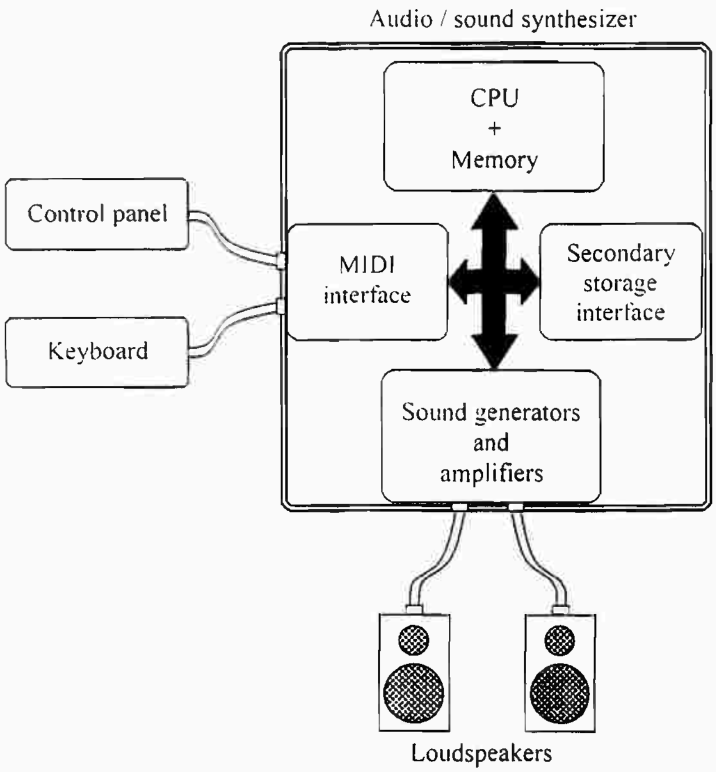



**Fig. (16.8) Audio / sound synthesizer**

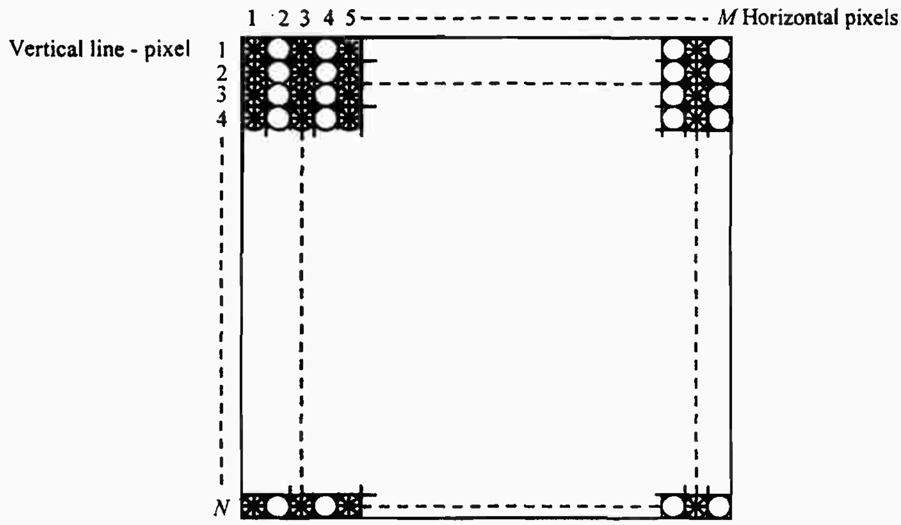
Since the music is in the form of MIDI messages, it is necessary to have a sound card to interpret the sequence of messages and generate the appropriate sound. The sound generators either use FM synthesis or prestored samples of sound produced by real instruments. This technique is called wavelet synthesis.

**16.8 Video:**

Three video signals are needed, R, G, B. But luminance (amount of white is given by).

$$Y = 0.299R + 0.587G + 0.144B$$

Where $Y$ is the amplitude of luminance and $R$ , $G$ and $B$ are the magnitudes of the three color components. Hence, we need to transmit $Y$ , $C_b$ and $C_r$ where $C_b = B - Y$ and $C_r = R - Y$ , Thus, $C_b$ and $C_r$ contain the chrominance signal.




$O = Y$, $+ = C_b$, $\times = C_r$ sample position

525-line system: $M = 720$, $N = 480$, 60 Hz refresh rate (interlaced)
$Y = 720 \times 480$, $C_b = C_r = 360 \times 480$

625-line system: $M = 720$, $N = 576$, 50 Hz refresh rate (interlaced)
$Y = 720 \times 576$, $C_b = C_r = 360 \times 576$

**Fig. (16.9) Sample positions in 4 : 4 : 2 format**

In digital systems, video signals need to be in a digital form, since it then becomes possible to store them in the memory of a computer and to edit and integrate them with other media types. In analog TV broadcasts, the three component signals are combined while in digital TV, it is usual to digitize the three component signals separately prior to their transmission. This enables editing and processing, since the three component signals are treated separately in digital TV.

However, if we require the same resolution in terms of sampling rate and bits per sample for the three components this will lead to extended bandwidth. But since the eye is less sensitive to color than to luminance, we may tolerate a reduced resolution for chrominance signal. Hence, we can achieve reduced bandwidth by using luminance and two color difference signals instead of the RGB signals directly.

The three component analog video signals have bandwidth up to $6MHz$ for luminance and less than half this for the two chrominance signals. Thus, the minimum sampling rate is 12 $MHz$ for luminance and $6MHz$ for chrominance.

## 16.9 The Basic Digital Format 4 : 2 : 2:

Digitization parameters for video signals in component form is based on $Y, C_b C_r$ signals in one of two formats 4:2:2: or 4:2:0, while 4:4:4, stands for digitization of the RGB signals. The 4:2:2 format means 4Y sampling with $2C_b$ samples and $2C_r$ samples with quantization involving 8bits. The sampling frequency is 13.5 $MHz$ for luminance and 6.75 $MHz$ for chrominance regardless of the standard of the input signal (525 lines / 60 $Hz$ or 625 lines / 50 $Hz$ ). The 13.5 $MHz$ was chosen since it is the nearest frequency to 12 $MHz$ which results in a whole number of samples per line for both 525 and 625 (line) systems.

**Ex. 16.4 (case study)**

Derive the sampling frequency in 4:2:2 format and then show the sampling positions.

**Solution**

In a 525 line system, the total sweep time in 63.56 $\mu s$ . But during this time, each beam is turned off for retrace for 11.56 $\mu s$ which yields an active line sweep time of 52 $\mu s$ . Similarly, in a 625 line system, the total sweep time is 64 $\mu s$ with a blanking time of 12 $\mu s$ , which also yields an active line sweep time of 52 $\mu s$ . Hence, in both cases a sampling rate of 13.5 $MHz$ yields.

$$52 \times 10^{-6} \times 13.5 \times 10^{6} = 702 \text{ samples / line.}$$

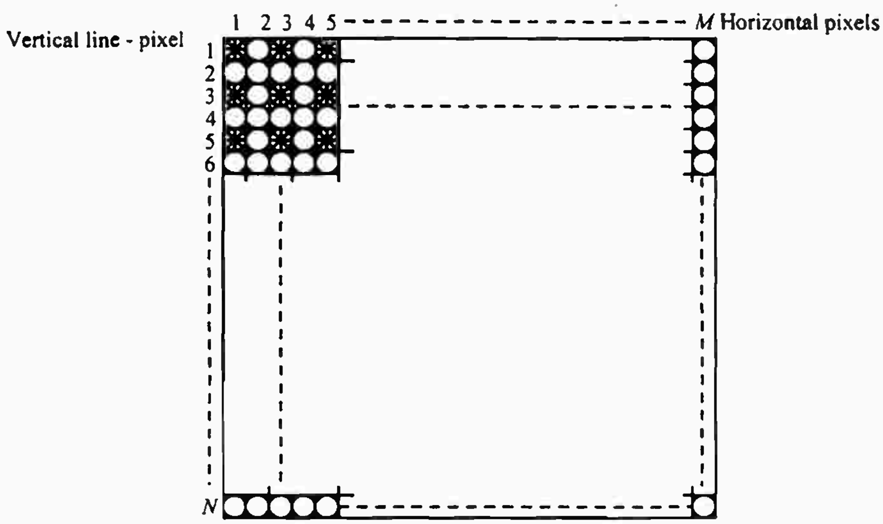
In practice, the number of samples /line is increased to 720 by taking a slightly longer active line time, which results in a small number of black samples at the beginning and end of each line. The number of bits per sample is 8; hence, we have 256 quantization levels. The vertical resolution for all three signals is the same. We have 480 active lines in 525, line system and 276 lines in 625 line system. In TV, interlaced scanning is used at a frame refresh rate of 60 $Hz$ for 525 line system, or 50 $Hz$ for 625 line system. Since each line is sampled at a constant rate (13.5 and 6.75 $MHz$ ) with a fixed number of samples per line (720 and 360), the samples for each line are in a fixed position which repeats from frame to frame, hence called orthogonal sampling. The sampling positions for each of the three signals relative to a rectangular grid are shown (Fig. 16.9).

In 525 line system, $Y = 720 \times 480 \quad C_b = C_r = 360 \times 480$

In 625 line system, $Y = 720 \times 576 \quad C_b = C_r = 360 \times 576$

**Ex. 16.5**

For the 4:2:2 format, derive the bit rate and the memory requirement to store each frame and the total memory to store 1.5 hour movie for 525 line system.




$O=Y$, $+=C_b$, $\times=C_r$ sample position

525-line system: $M = 720$, $N = 480$, 60 Hz refresh rate (interlaced)
$Y = 720 \times 480$, $C_b = C_r = 360 \times 480$
625-line system: $M = 720$, $N = 576$, 50 Hz refresh rate (interlaced)
$Y = 720 \times 576$, $C_b = C_r = 360 \times 576$

**Fig. (16.10) Sample positions in 4 : 2 : 0**

**Solution**

In 525 line system, we have 720 samples per line and 480 visible lines.

$$Y = 720 \times 480$$

$$C_b = C_r = 360 \times 480$$

Line sampling is at 13.5 $MH_z$ for Y and 6.75 $MHz$ for both $C_b$ and $C_r$ with 8 bit / sample.

Bit rate $= 13.5 \times 10^6 \times 8 + 2(6.75 \times 10^6 \times 8) = 216M\ bps$

Hence, memory required per line $= 720 \times 8 + 2(360 \times 8) = 11520 bits$ = 1440 bytes

Memory per frame (480 lines) $= 480 \times 11520 = 5.5296\ Mb = 691.2 kB$

Memory to store 1.5 hour (assuming 60 frames per second)

$$= 691.2 \times 60 \times 1.5 \times 3600 kB$$

$$= 223.9488 GB$$

Both bit rate and memory requirements are very large, and for this reason various lower resolution formats have been defined.

### 16.10 Other Formats:

Another format is 4:2:0 where chrominance samples are present in alternate lines. In this case, we have the same luminance resolution as 4:2:2, but half the chrominance resolution (Fig. 16.10). Interlaced scan is used. Also, to avoid flicker effects receivers of large screen televisions store the incoming digitized signals of each field in a memory buffer. A refresh rate of double the normal rate $(100/120/Hz)$ is then used with the stored set of signals used for the second field. for 525 line system,

$$Y = 720 \times 480$$

$$C_b = C_r = 360 \times 240$$

For 625 line system,

$$Y = 720 \times 576$$

$$C_b = C_r = 360 \times 240$$

There are a number of digitization formats for high definition television (HDTV). The resolution which relates to the 4/3 aspect ratio can be $440 \times 1152$ pixels. For 16/9 with screen tubes, the resolution is $1920 \times 1152$ pixels. In both cases, the number of visible lines / frame is 1080, which produces a square pixel lattice structure. Both use 4:2:2 format for studio applications or 4: 2: 0 format for broadcast applications. The corresponding refresh rate may be 50 / 60 $Hz$ for the 4:2:2 format or 25 / 30$Hz$ for the 4:2:0 format or higher in modern sets.

In source intermediate format (SIF), half the spatial resolution is used in both horizontal and vertical directions as that used in the 4:2:0 format. This is known as subsampling. The refresh rate is $30Hz$ for 525 line system and $25Hz$ for 625 line system. This is called temporal resolution, SIF also is intended for storage applications. Progressive (non – interlaced) scanning is used. In the digitization format is known as 4:1:1, the positions of the three sampling instants per frame are shown (Fig. 16.11). At the receiver the missing samples are estimated by interpolating between each pair of values sent. The common intermediate format (CIF) is defined for use in video conferencing application. It uses a combination of the spatial resolution used for the SIF in the 625 line system and the temporal resolution used in the 525 line system. This yields a spatial resolution of.

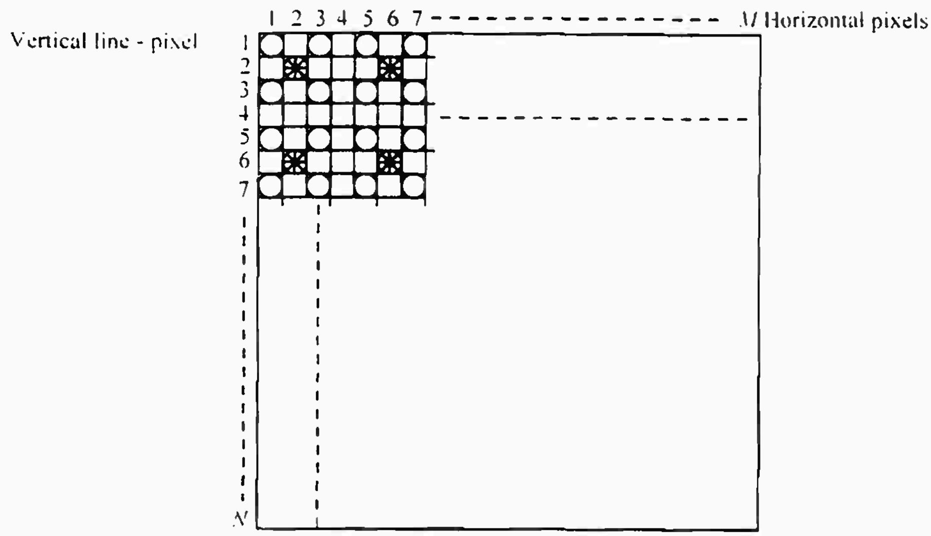
$$Y = 360 \times 288$$

$$C_b = C_r = 180 \times 144$$

$$Y = 360 \times 288$$

$$C_b = C_r = 180 \times 144$$

With a temporal resolution of $30Hz$ using progressive (non – interlaced) scanning, the positions of the sampling instants per frame are the same as for SIF, and hence, the digitization format is 4: 1: 1.




$O = Y$, $\cdot = C_b$, $\times = C_r$ sample position

525-line system: $M = 720$, $N = 480$ with 30 Hz refresh rate (non-interlaced)
$Y = 360 \times 240$, $C_b = C_r = 180 \times 120$

625-line system: $M = 720$, $N = 576$, with 25 Hz refresh rate (non-interlaced)
$Y = 360 \times 288$, $C_b = C_r = 180 \times 144$

CIF: $M = 720$, $N = 576$, with 30 Hz refresh rate (non-interlaced)
$Y = 360 \times 288$, $C_b = C_r = 180 \times 144$

**Fig. (16.11) Sample position for SIF and CIF**

## 16.11 PC Video Format:

All digitization formats described above are intended for use with standard TV receivers. Often, a window is used on a PC monitor that involves live video (video in a window). Most picture tubes operate in an analog mode, i.e., the amplitude of each of the three color signals is continuously varying as each line is scanned. In the case of digital TV and digitized pictures stored within the memory of a computer, the color signals are in digital form, and comprise a string of pixels with a fixed number of pixels per scan line. Hence, in order to display a stored image, the pixels that make up each line are read from memory in time synchronism with the scanning process and converted into a continuously varying analog form by means of digital to analog converter (DAC).

The aspect ratio of current TV tubes and for PC monitors is 4/3, while for wide screen TV tubes it is 16/9. In order to avoid distortion on a screen which has 4/3 aspect ratio when displaying a sequence of $N \times N$ pixels, it is necessary to have 640 pixels (480x4/3) per line with an NTSC monitor and 768(576x4/3) pixels per line for a European monitor. This leads to a lattice structure producing proper square pixels. The memory required to store a single image varies between $307.2 kB$ for VGA (video graphics array) screen with 8 bits per pixel through $2.36 MB$ for SVGA (Super VGA) with 24 bits per pixel. To avoid distortion on a PC screen when displaying a square of $N$ x $N$, we have seen that we need to use a horizontal resolution of 640 (480 x 4/3) pixels per line with a 625 line PC monitor. Hence, for applications that involves live video with other information on a PC screen, the line sampling rate is modified to obtain the required horizontal resolution. For a 525 line monitor, the sampling rate is reduced from 13.5 MHz to 12.2727 MHz, while for a 625 line monitor, the line sampling rate is decreased from 13.5 MHz 14.75 MHz. (why? Prob. 16.9). In the case of a digital TV broadcast, a conversion is necessary before the video is displayed. All PC monitors use progressive (non – interlaced) scanning. PC video digitization formats are shown (Table 16.2).

**Table (16.2) PC video digitization formats**

| *Digitization format* | *System* | *Spatial resolution* | *Temporal resolution* |
|---|---|---|---|
| 4:2:0 | 525-line | $Y = 640 \times 480$<br>$C_b = C_r = 320 \times 240$ | 60 $Hz$ |
| | 625-line | $Y = 768 \times 576$<br>$C_b = C_r = 388 \times 288$ | 50 $Hz$ |
| SIF | 525-line | $Y = 320 \times 240$<br>$C_b = C_r = 160 \times 240$ | 30 $Hz$ |
| | 625-line | $Y = 384 \times 288$<br>$C_b = C_r = 192 \times 144$ | 25 $Hz$ |
| CIF | | $Y = 384 \times 288$<br>$C_b = C_r = 192 \times 144$ | 30 $Hz$ |

## Problems

1. Derive the bit rate and memory requirements to store each frame assuming 4:2:2 format. Also, find the total memory required to store a 1.5 hour movie in a 625 line system. Compare the results with those in Ex. 16.5. What do you conclude?

2. Repeat prob. (1) for 4:2:0 format for both 525 and 625 line systems.

3. For HDTV of $1440 \times 1152$ resolution, calculate the bit rate and compare it with that of standard resolution $720 \times 480$. Compare the result with large screen resolution of $1420 \times 1152$ .

4. Calculate the bit rate for SIF, CIF.

5. Calculate the memory per frame per VGA, SVGA.

6. Repeat the above problem for a PC monitor $1280 \times 1024 \times 24$ for and refresh rate of 75 frames / s.

7. Repeat the above problem for 100 $Hz$ refresh rate. What do you conclude?

8. Calculate the time to transmit the digitized images at both 64 kbps and 1.5 $Mbps$ for 1024 x 768 x 24 SVGA.

9. Calculate the sampling rate for PC monitor and show why it is different from that of TV monitor.

10. Show how the aspect ratio guarantees distortionless display for HDTV for both 4/3 and 16/9 ratios.

11. Verify PC video digitization formats in relation to sampling rates for both 525 and 625 line systems.

12. Show how TV standards must be modified upon displaying live video on a PC monitor.

## References


1. "Multimedia Communications", F. Halsall Addison. Wesley, Pearson Education, London, 2001.

2. "Video Engineering", A. Luther and A. Inglis, 3rd ed., McGraw Hill, N. Y. 1999.

3. Digital TV Fundamentals", M. Robin and M. Poulin, 2nd ed., McGraw Hill, N. Y., 2000.

4. "Color TV", G. Hutson, P. Shepherd and J. Brice, McGraw Hill, N. Y., 1999.

5. "Video Demystified", K. Jach, LLH, Technology Eagle Rock 2001.

6. "Introduction to Communication Electronics" M. S. Said, Shorouk, Cairo, 2004.