Chapter 5
Fundamental Concepts in Video

5.1 Analog Video

- An analog signal $f(t)$ samples a time-varying image.
- So-called "progressive" scanning traces through a complete picture (a frame) row-wise for each time interval.
- Another system, called "interlaced" scanning is used:

a) The odd-numbered lines are traced first, and then the even-numbered lines are traced.

  - This results in "odd" and "even" fields — two fields make up one frame.
  - In fact, the odd lines (starting from 1) end up at the middle of a line at the end of the odd field, and the even scan starts at a half-way point.

5.1 Analog Video

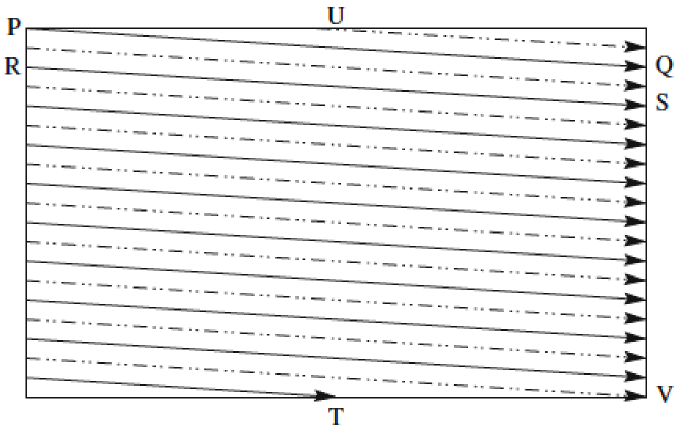
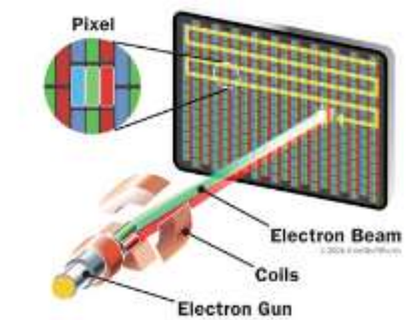


Fig. 5.1: Interlaced raster scan



c) Figure 5.1 shows the scheme used. First the solid (odd) lines are traced, P to Q, then R to S, etc., ending at T; then the even field starts at U and ends at V.

d) The jump from Q to R, etc. in Figure 5.1 is called the **horizontal retrace**, during which the electronic beam in the CRT is blanked. The jump from T to U or V to P is called the **vertical retrace**.

5.1 Analog Video

- Because of interlacing, the odd and even lines are displaced in time from each other
  - Generally not noticeable except when very fast action is taking place on screen, when blurring may occur.
- For example, in the video in Fig. 5.2, the moving helicopter is blurred more than is the still background.

5.1 Analog Video

(a)

(b)

(c)

(d)

Fig. 5.2: Interlaced scan produces two fields for each frame. (a) The video frame, (b) Field 1, (c) Field 2, (d) Difference of Fields.

5.1.1 NTSC Video

- NTSC (National Television System Committee) TV standard is mostly used in North America and Japan. It uses the familiar 4:3 **aspect ratio** (i.e., the ratio of picture width to its height) and uses 525 scan lines per frame at 30 frames per second (fps).

  a) NTSC follows the interlaced scanning system, and each frame is divided into two fields, with 262.5 lines/field.

  b) Vertical retrace takes place during 20 lines reserved for control information at the beginning of each field. Hence, the number of active *video lines* per frame is only 485.

- Fig. 5.4 shows the effect of "vertical retrace & sync" and "horizontal retrace & sync" on the NTSC video raster.

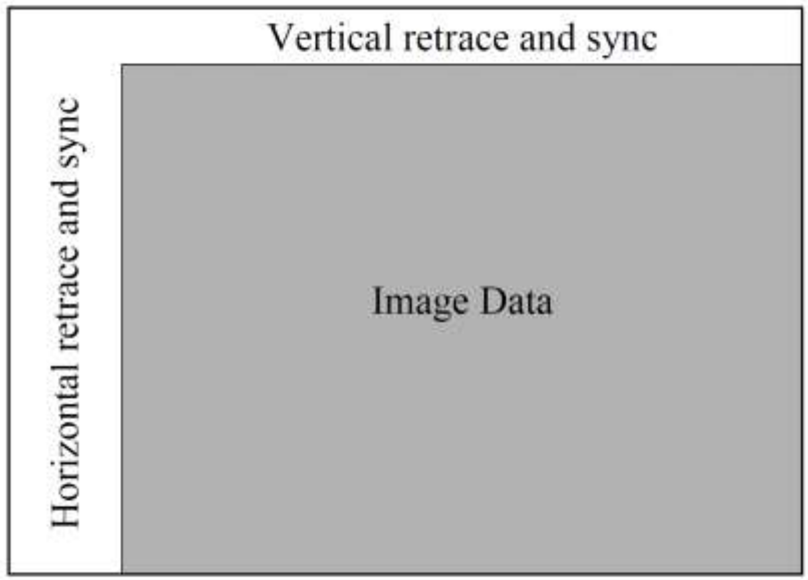


Fig. 5.4: Video raster, including retrace and sync data

- NTSC video is an analog signal with no fixed horizontal resolution. Therefore one must decide how many times to sample the signal for display: each sample corresponds to one pixel output.
- A "pixel clock" is used to divide each horizontal line of video into samples. The higher the frequency of the pixel clock, the more samples per line there are.
- Different video formats provide different numbers of samples per line, as listed in Table 5.1.

Table 5.1: Samples per line for various video formats

Format	Samples per line
VHS	240
S-VHS	400-425
Betamax	500
Standard 8 m	300
Hi-8 mm	425

5.1.2 PAL Video

- **PAL (Phase Alternating Line)** is a TV standard widely used in Western Europe, China, India, and many other parts of the world.
- PAL uses 625 scan lines per frame, at 25 frames/second, with a 4:3 aspect ratio and interlaced fields.

(a) PAL uses the YUV color model. It uses an 8 MHz channel and allocates a bandwidth of 5.5 MHz to Y, and 1.8 MHz each to U and V. The color subcarrier frequency is $f_{sc} \approx 4.43$ MHz.

(b) In order to improve picture quality, chroma signals have alternate signs (e.g., +U and -U) in successive scan lines, hence the name “Phase Alternating Line”.

5.1.3 SECAM Video

- **SECAM** stands for *Système Electronique Couleur Avec Mémoire*, the third major broadcast TV standard.
- SECAM also uses 625 scan lines per frame, at 25 frames per second, with a 4:3 aspect ratio and interlaced fields.
- SECAM and PAL are very similar. They differ slightly in their color coding scheme:
  (a) In SECAM, U and V signals are modulated using separate color subcarriers.
  (b) They are sent in alternate lines, i.e., only one of the U or V signals will be sent on each scan line.

- Table 5.2 gives a comparison of the three major analog broadcast TV systems.

Table 5.2: Comparison of Analog Broadcast TV Systems

TV System	Frame Rate (fps)	# of Scan Lines	Total Channel Width (MHz)	Bandwidth Allocation (MHz)		
				Y	I or U	Q or V
NTSC	29.97	525	6.0	4.2	1.6	0.6
PAL	25	625	8.0	5.5	1.8	1.8
SECAM	25	625	8.0	6.0	2.0	2.0

Exercise: Interlacing Simulation

Task: Given two images: Field A (odd lines) and Field B (even lines), combine them to create a full progressive frame manually or in MATLAB.

Solution (Conceptual):

- Overlay the lines from both fields in alternating rows.

Exercise: Interlacing Simulation

- **Why This Happens**

This simulates how **interlaced video** works:

- In analog video (like NTSC or PAL), each frame is split into two fields to **reduce bandwidth**:
  - Field A is shown first (odd lines)
  - Field B is shown next (even lines)
- To display it on modern screens (which use **progressive scan**), we must **reconstruct the full frame**.

🔧 **Conceptual Solution**

To rebuild a **full image**, we need to:

- Place **Field A's rows in the odd positions**: 1st, 3rd, 5th...
- Place **Field B's rows in the even positions**: 2nd, 4th, 6th...

This gives us a complete image: line 1 from A, line 2 from B, line 3 from A, line 4 from B, and so on.

Solution

MATLAB Code Explanation

```
progressive_img = zeros(size(fieldA));
```
% Create an empty image of same size
% Creates a placeholder (matrix of zeros) to store the final image.

```
progressive_img(1:2:end, :) = fieldA(1:2:end, :);
```

% This line says: Fill the odd rows of progressive_img (1, 3, 5, ...)
% With the odd rows from Field A.

```
progressive_img(2:2:end, :) = fieldB(2:2:end, :);
```
% This line fills the even rows of progressive_img (2, 4, 6, ...)
% With the even rows from Field B.

5.2 Digital Video

- The advantages of digital representation for video are many. For example:

    (a) Video can be stored on digital devices or in memory, ready to be processed (noise removal, cut and paste, etc.), and integrated to various multimedia applications;

    (b) Direct access is possible, which makes nonlinear video editing achievable as a simple, rather than a complex, task;

    (c) Repeated recording does not degrade image quality;

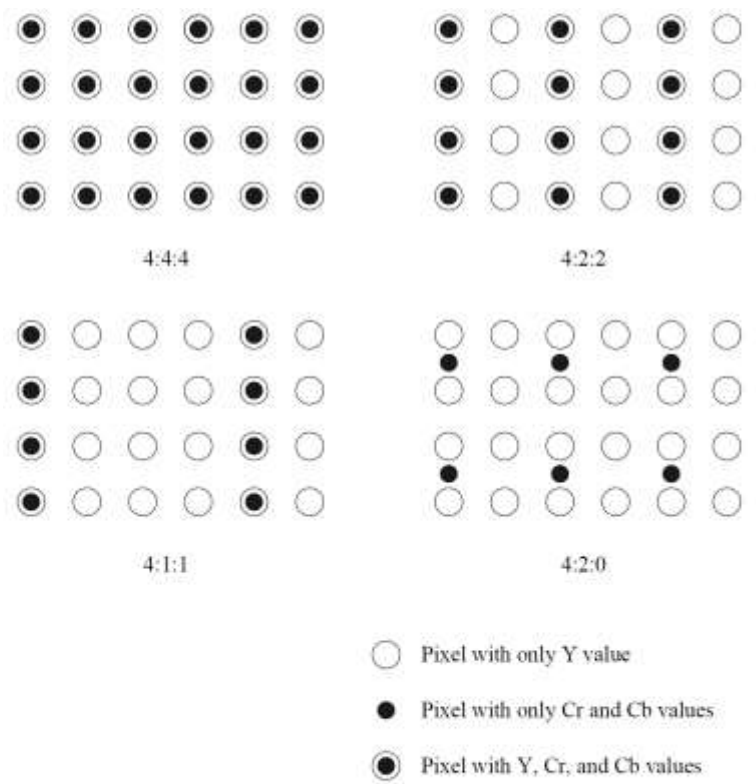
    (d) Ease of encryption and better tolerance to channel noise.

5.2.1 Chroma Subsampling

- Since humans see color with much less spatial resolution than they see black and white, it makes sense to "decimate" the chrominance signal.
- Interesting (but not necessarily informative!) names have arisen to label the different schemes used.
- To begin with, numbers are given stating how many pixel values, per four original pixels, are actually sent (kept).

5.2.1 Chroma Subsampling

- The chroma subsampling scheme "4:4:4" indicates that no chroma subsampling is used: each pixel's Y, Cb and Cr values are transmitted, 4 for each of Y, Cb, Cr.
- The scheme "4:2:2" indicates horizontal subsampling of the Cb, Cr signals by a factor of 2.
  - That is, of four pixels horizontally labelled as 0 to 3, all four Ys are sent, and every two Cb's and two Cr's are sent, as (Cb0, Y0)(Cr0, Y1)(Cb2, Y2)(Cr2, Y3), and so on (or averaging is used).
- The scheme "4:1:1" subsamples *horizontally* by a factor of 4.
- The scheme "4:2:0" subsamples in both the *horizontal* and *vertical* dimensions by a factor of 2.
  - Theoretically, an average chroma pixel is positioned between the rows and columns as shown Fig.5.6.
  - is commonly used in JPEG and MPEG (see later chapters in Part 2).



5.2.2 CCIR and ITU-R Standards for Digital Video

- **CCIR** is the Consultative Committee for International Radio, and one of the most important standards it has produced is CCIR-601, for component digital video.
  - This standard has since become standard ITU-R-601, an international standard for professional video applications
    - adopted by certain digital video formats including the popular DV video.

- Table 5.3 shows some of the digital video specifications, all with an aspect ratio of 4:3. The CCIR 601 standard uses an interlaced scan, so each field has only half as much vertical resolution (e.g., 240 lines in NTSC).

- CIF stands for Common Intermediate Format specified by the CCITT.

  (a) The idea of CIF is to specify a format for lower bitrate.

  (b) CIF is about the same as VHS quality. It uses a progressive (non-interlaced) scan.

  (c) QCIF stands for “Quarter-CIF”. All the CIF/QCIF resolutions are evenly divisible by 8, and all except 88 are divisible by 16; this provides convenience for block-based video coding in H.261 and H.263, discussed later in Chapter 10.

a) Note, CIF is a compromise of NTSC and PAL in that it adopts the 'NTSC frame rate and half of the number of active lines as in PAL.

Table 5.3: ITU-R digital video specifications

	CCIR 601 525/60 NTSC	CCIR 601 625/50 PAL/SECAM	CIF	QCIF
Luminance resolution	720 x 480	720 x 576	352 x 288	176 x 144
Chrominance resolution	360 x 480	360 x 576	176 x 144	88 x 72
Colour Subsampling	4:2:2	4:2:2	4:2:0	4:2:0
Fields/sec	60	50	30	30
Interlaced	Yes	Yes	No	No

5.2.3 High Definition TV (HDTV)

- The main thrust of **HDTV** (High Definition TV) is not to increase the "definition" in each unit area, but rather to increase the visual field especially in its width.
  a) The first generation of HDTV was based on an analog technology developed by Sony and NHK in Japan in the late 1970s.
  b) It has 1,125 scan lines, interlaced (60 fields per second), and 16:9 aspect ratio.
  c) Since uncompressed HDTV will easily demand more than 20 MHz bandwidth, which will not fit in the current 6 MHz or 8 MHz channels, various compression techniques are being investigated.
  d) It is also anticipated that high quality HDTV signals will be transmitted using more than one channel even after compression.

- A brief history of HDTV evolution:

  - In 1987, the FCC decided that HDTV standards must be compatible with the existing NTSC standard and be confined to the existing VHF (Very High Frequency) and UHF (Ultra High Frequency) bands.

  - In 1990, the FCC announced a very different initiative, i.e., its preference for a full-resolution HDTV, and it was decided that HDTV would be simultaneously broadcast with the existing NTSC TV and eventually replace it.

  - Witnessing a boom of proposals for digital HDTV, the FCC made a key decision to go all-digital in 1993. A "grand alliance" was formed that included four main proposals, by General Instruments, MIT, Zenith, and AT&T, and by Thomson, Philips, Sarnoff and others.

  - This eventually led to the formation of the ATSC (Advanced Television Systems Committee) — responsible for the standard for TV broadcasting of HDTV.

  - In 1995 the U.S. FCC Advisory Committee on Advanced Television Service recommended that the ATSC Digital Television Standard be adopted.

- The standard supports video scanning formats shown in Table 5.4. In the table, "I" mean interlaced scan and "P" means progressive (non-interlaced) scan.

Table 5.4: Advanced Digital TV formats supported by ATSC

# of Active Pixels per line	# of Active Lines	Aspect Ratio	Frame Rate
1,920	1,080	16:9	60P 60I 30P 24P
1,280	720	16:9	60P 30P 24P
704	480	16:9 or 4:3	60P 60I 30P 24P
640	480	4:3	60P 60I 30P 24P

- For video, MPEG-2 is chosen as the compression standard. For audio, AC-3 is the standard. It supports the so-called 5.1 channel Dolby surround sound, i.e., five surround channels plus a subwoofer channel.

- The salient difference between conventional TV and HDTV:

  a) HDTV has a much wider aspect ratio of 16:9 instead of 4:3.

  b) HDTV moves toward progressive (non-interlaced) scan. The rationale is that interlacing introduces serrated edges to moving objects and flickers along horizontal edges.

5.2.4 Ultra High Definition TV (UHDTV)

- UHDTV is a new generation of HDTV. The standards initiated in 2012 support 4K UHDTV: 2160P (3,840×2,160, progressive scan) and 8K UHDTV: 4320P (7,680×4,320, progressive scan).
- The aspect ratio is 16:9. The bit-depth is 10 or 12 bits per sample, and the chroma subsampling can be 4:2:0, 4:2:2, or 4:4:4.
- The supported frame rate has been gradually increased to 120 fps.
- The UHDTV will provide superior picture quality, comparable to IMAX movies, but it will require a much higher bandwidth and bitrate.

- 16K UHDTV has been demonstrated in 2018, targeting applications such as Virtual Reality with true immersion. Its resolution is 15,360 × 8,640 for a total of 132.7 megapixels.

Table 5.5: A Summary of UHDTV

Type of UHDTV	Resolution	Bit Depth	Aspect Ratio	Frame Rate
4K UHD (2160P)	3840 x 2160	10 or 12 bits	16:9	Up to 120P
8K UHD (4320P)	7680 x 4320	10 or 12 bits	16:9	Up to 120P
16K UHD (8640P)	15360 x 8640	10 or 12 bits	16:9	Up to 240P

5.3 Video Display Interfaces

5.3.1 Analog Display Interfaces

Analog video signals are often transmitted in one of three different interfaces: *Component video*, *Composite video*, and *S-video*. Figure 5.7 shows the typical connectors for them.

Fig. 5.7: Connectors for typical analog display interfaces. From left to right: Component video, Composite video, S-video, and VGA.

Component video

- **Component video**: Higher-end video systems make use of three separate video signals for the red, green, and blue image planes. Each color channel is sent as a separate video signal.
  a) Most computer systems use Component Video, with separate signals for R, G, and B signals.
  b) For any color separation scheme, Component Video gives the best color reproduction since there is no "crosstalk" between the three channels.
  c) This is not the case for S-Video or Composite Video, discussed next. Component video, however, requires more bandwidth and good synchronization of the three components.

Composite Video

- **Composite video**: color ("chrominance") and intensity ("luminance") signals are mixed into a *single* carrier wave.

  a) Chrominance is a composition of two color components (I and Q, or U and V).

  b) In NTSC TV, e.g., I and Q are combined into a chroma signal, and a color subcarrier is then employed to put the chroma signal at the high-frequency end of the signal shared with the luminance signal.

a) The chrominance and luminance components can be separated at the receiver end and then the two color components can be further recovered.

b) When connecting to TVs or VCRs, Composite Video uses only one wire and video color signals are mixed, not sent separately. The audio and *sync* signals are additions to this one signal.

- Since color and intensity are wrapped into the same signal, some interference between the luminance and chrominance signals is inevitable.

S-Video

- **S-Video**: as a compromise, (separated video, or Super-video, e.g., in S-VHS) uses two wires, one for luminance and another for a composite chrominance signal.
- Less crosstalk between the color information and the crucial gray-scale information.
- The reason for placing luminance into its own part of the signal is that black-and-white information is most crucial for visual perception.
  - Humans are able to differentiate spatial resolution in grayscale images with a much higher acuity than for the color part of color images.
  - As a result, we can send less accurate color information than must be sent for intensity information — we can only see fairly large blobs of color, so it makes sense to send less color detail.

5.3.2 Digital Display Interfaces

Digital interfaces emerged in 1980s (e.g., Color Graphics Adapter (CGA)), and evolved rapidly. Today, the most widely used digital video interfaces include Digital Visual Interface (DVI), High-Definition Multimedia Interface (HDMI), and DisplayPort.

Fig. 5.8: Connectors of different digital display interfaces. From left to right: DVI, HDMI, DisplayPort.

High-Definition Multimedia Interface (HDMI)

- HDMI is a newer digital audio/video interface developed to be backward compatible with DVI.

   1. HDMI doesn't carry analog signal and hence is not compatible with VGA.
   2. HDMI supports both RGB and YCbCr 4:4:4 or 4:2:2. [DVI is limited to the RGB color range (0-255).]
   3. HDMI supports digital audio, in addition to digital video.

- The maximum pixel clock rate for HDMI 1.0 is 165 MHz, which is sufficient to support 1080P (1,920 × 1,200) at 60 Hz.
- HDMI 2.0 was released in 2013, which supports 4K resolution at 60 frames per second.

Exercise 2: Chroma Subsampling Calculation

Task: For a 1920x1080 frame, calculate the total number of samples for Y, Cb, and Cr in each format:

4:4:4

4:2:2

4:2:0

Link chroma subsampling to file size and image quality in compression.

Exercise 2: Chroma Subsampling Calculation

- **4:4:4**:
  - Y = 1920×1080 = 2,073,600
  - Cb = 1920×1080
  - Cr = 1920×1080
  - **Total = 6,220,800 samples**
- **4:2:2**:
  - Y = 1920×1080 = 2,073,600
  - Cb = 960×1080 = 1,036,800
  - Cr = 960×1080
  - **Total = 4,147,200 samples**
- **4:2:0**:
  - Y = 1920×1080 = 2,073,600
  - Cb = 960×540 = 518,400
  - Cr = 960×540
  - **Total = 3,110,400 samples**