



# Welcome Back to Fundamentals of Multimedia (MR412) Fall, 2012

ZHU Yongxin, Winson  
[zhuyongxin@sjtu.edu.cn](mailto:zhuyongxin@sjtu.edu.cn)





上海交通大學

Shanghai Jiao Tong University

# Chapter 5

## Fundamental Concepts in Video

- [5.1 Types of Video Signals](#)
  - [5.2 Analog Video](#)
  - [5.3 Digital Video](#)
  - [5.4 Further Exploration](#)
-



# 5.1 Types of Video Signals

- **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 - 1 Signal

- **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.
  - c) The chrominance and luminance components can be separated at the receiver end and then the two color components can be further recovered.
  - d) 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.





上海交通大學

Shanghai Jiao Tong University

# Composite Video - 1 Signal

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



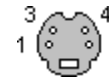


# S-Video - 2 Signals

- **S-Video:** as a compromise, (Separated video, or Supervideo, e.g., in S-VHS) uses two wires, one for luminance and another for a composite chrominance signal
- As a result, there is less crosstalk between the color information and the crucial gray-scale information.



4 pin S-Video female



4 pin S-Video male

Pin	Name	Definition
1	GND	Y GND
2	GND	C GND
3	Y	Luminance
4	C	color

Source: yesky.com





# S-Video - 2 Signals

- The reason for placing luminance into its own part of the signal is that black-and-white information is most crucial for visual perception.



- In fact, 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.



7针 S-Video MINI母头



7针 S-Video MINI公头




Pin	Name	Description
1	GND	亮度地
2	GND	色彩地
3	Y	亮度信号
4	C	色信号
5	-	-
6	V	复合视频信号
7	VGND	复合地

Pin	Name	Description
1	GND	Ground (Y)
2	GND	Ground (C)
3	Y	Intensity (Luminance)
4	Pr	Pr 红色差
5	-	-
6	V	Composite Video
7	VGND	Composite Ground
8	Pb	Pb 蓝色差
9	GND	Ground(Pb)

Source: [www.infoavchina.com](http://www.infoavchina.com)



# Quality vs. Transmission Distance?

-  RGB: best quality, around 3-15m
-  S-video: close to the optimal quality, around 15-30m, possibly over 60m
-  Composite video: acceptable quality, more than 30m, possibly over 500m





## 5.2 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.
- In TV, and in some monitors and multimedia standards as well, 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.
  - b) 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.



# "interlaced" scanning

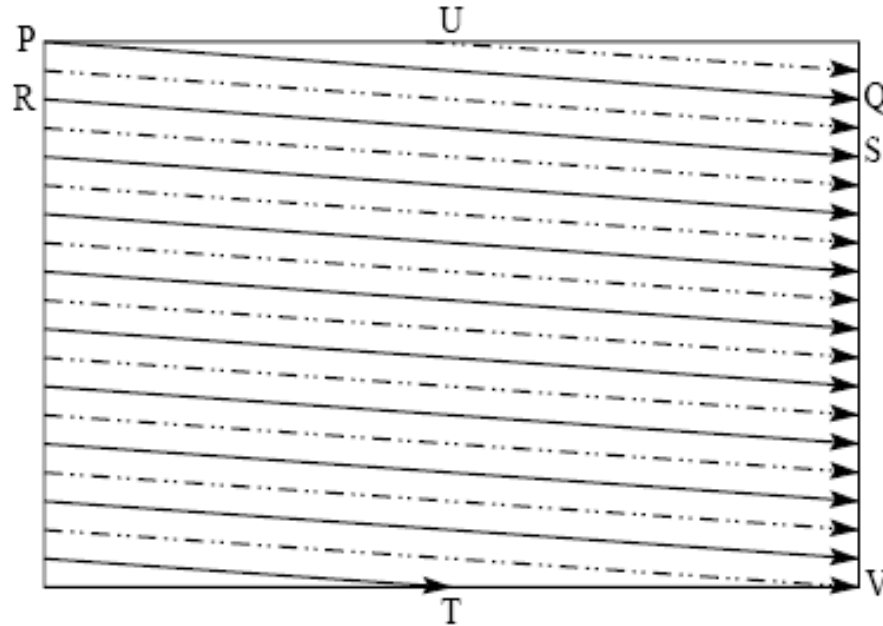


Fig. 5.1: Interlaced raster scan

- 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**.



# "interlaced" scanning (cont'd)

- 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.



# "interlaced" scanning (cont'd)



(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



# De-interlace

- Since it is sometimes necessary to change the frame rate, resize, or even produce stills from an interlaced source video, various schemes are used to "de-interlace" it.
  - a) The simplest de-interlacing method consists of discarding one field and duplicating the scan lines of the other field. The information in one field is lost completely using this simple technique.
  - b) Other more complicated methods that retain information from both fields are also possible.
- Analog video use a small voltage offset from zero to indicate "black", and another value such as zero to indicate the start of a line. For example, we could use a "blacker-than-black" zero signal to indicate the beginning of a line.



# NTSC scan line

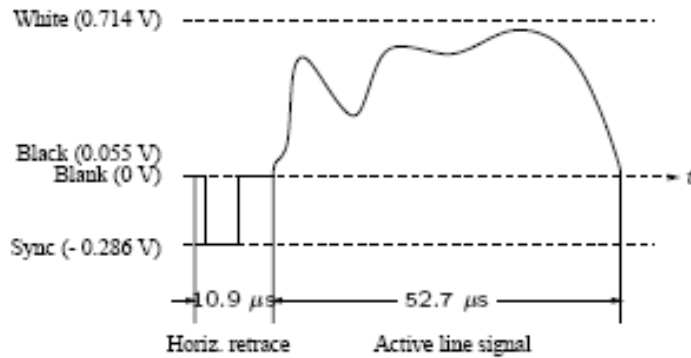


Fig. 5.3 Electronic signal for one NTSC scan line.



# 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) Thus the horizontal sweep frequency is  $525 \times 29.97 \approx 15,734$  lines/sec, so that each line is swept out in  $1/(15.734 \times 10^3)$  usec  $\approx 63.6$  usec.
  - c) Since the horizontal retrace takes 10.9  $\mu$ sec, this leaves 52.7  $\mu$ sec for the active line signal during which image data is displayed (see Fig.5.3).



# NTSC Video raster

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

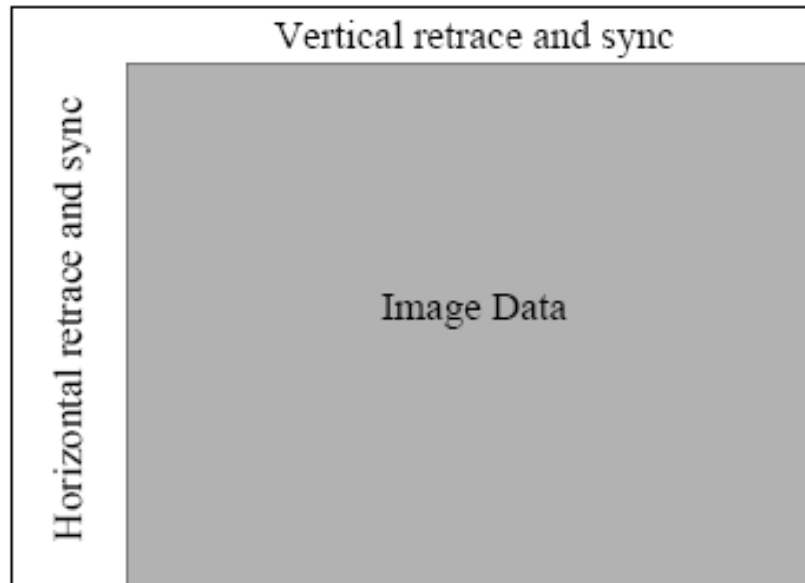


Fig. 5.4: Video raster, including retrace and sync data.





# NTSC Video raster (cont'd)

- a) 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.
- b) Similarly, almost 1/6 of the raster at the left side is blanked for horizontal retrace and sync. The non-blanking pixels are called *active pixels*.
- c) Since the horizontal retrace takes 10.9  $\mu\text{sec}$ , this leaves 52.7  $\mu\text{sec}$  for the active line signal during which image data is displayed (see Fig.5.3).
- d) It is known that pixels often fall in-between the scan lines. Therefore, even with non-interlaced scan, NTSC TV is only capable of showing about 340 (visually distinct) lines, i.e., about 70% of the 485 specified active lines. With interlaced scan, this could be as low as 50%.



# NTSC Video (cont'd)

- 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 mm	300
Hi-8 mm	425



# Color Model and Modulation of NTSC

- NTSC uses the YIQ color model, and the technique of **quadrature modulation** is employed to combine (the spectrally overlapped part of)  $I$  (in-phase) and  $Q$  (quadrature) signals into a single chroma signal  $C$ :

$$C = I \cos(F_{sc}t) + Q \sin(F_{sc}t) \quad (5.1)$$

- This modulated chroma signal is also known as the **color subcarrier**, whose magnitude is  $\sqrt{I^2 + Q^2}$ , and phase is  $\tan^{-1}(Q/I)$ . The frequency of  $C$  is  $F_{sc} \approx 3.58$  MHz.
- The NTSC composite signal is a further composition of the luminance signal  $Y$  and the chroma signal as dened below:

$$\text{composite} = Y + C = Y + I \cos(F_{sc}t) + Q \sin(F_{sc}t) \quad (5.2)$$



# NTSC spectrum

- Fig. 5.5: NTSC assigns a bandwidth of 4.2 MHz to Y, and only 1.6 MHz to I and 0.6 MHz to Q due to humans' insensitivity to color details (high frequency color changes).

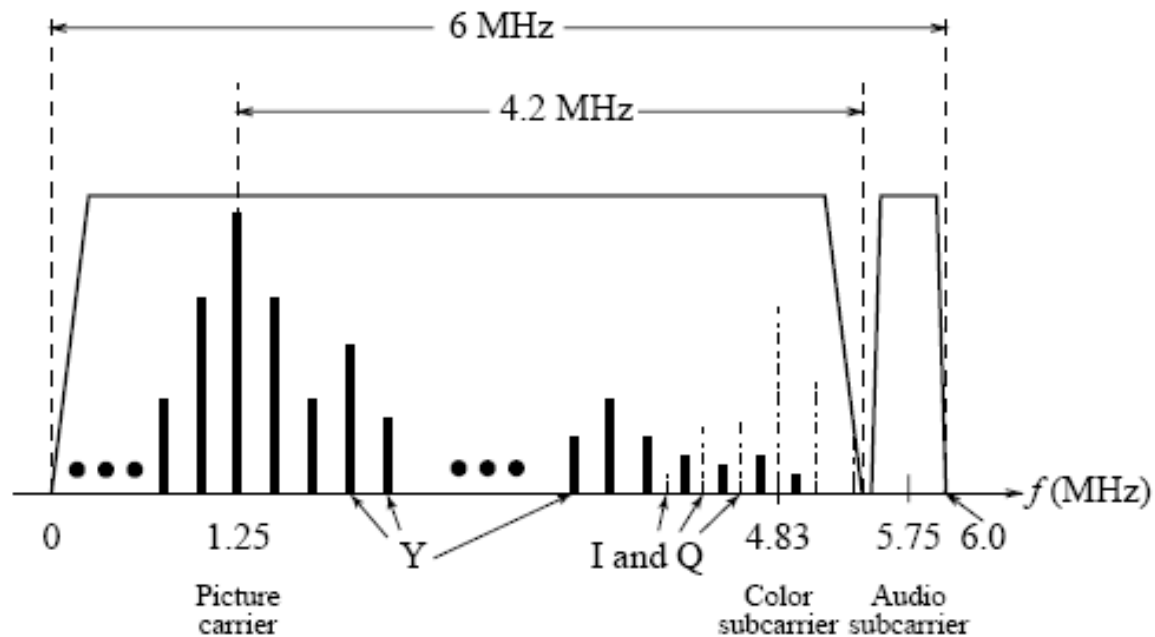


Fig. 5.5: Interleaving Y and C signals in the NTSC spectrum.



# Decoding NTSC Signals

- The first step in decoding the composite signal at the receiver side is the separation of  $Y$  and  $C$ .
- After the separation of  $Y$  using a low-pass filter, the chroma signal  $C$  can be demodulated to extract the components  $I$  and  $Q$  separately. To extract  $I$ :

- 1. Multiply the signal  $C$  by  $2 \cos(F_{sc}t)$ , i.e.,

$$\begin{aligned} C \cdot 2\cos(F_{sc}t) &= I \cdot 2\cos^2(F_{sc}t) + Q \cdot 2\sin(F_{sc}t) \cos(F_{sc}t) \\ &= I \cdot (1 + \cos(2F_{sc}t)) + Q \cdot 2\sin(F_{sc}t) \cos(F_{sc}t) \\ &= I + I \cdot \cos(2F_{sc}t) + Q \sin(2F_{sc}t). \end{aligned}$$

- 2. Apply a low-pass filter to obtain  $I$  and discard the two higher frequency ( $2F_{sc}$ ) terms.



# Decoding NTSC Signals

- Similarly,  $Q$  can be extracted by first multiplying  $C$  by  $2 \sin(F_{sc}t)$  and then low-pass filtering:

$$\begin{aligned} C \cdot 2\sin(F_{sc}t) &= I \cdot 2\sin(F_{sc}t)\cos(F_{sc}t) + Q \cdot 2\sin^2(F_{sc}t) \\ &= I \cdot \sin(2F_{sc}t) + Q \cdot (1 - \cos(2F_{sc}t)) \\ &= Q + I \cdot \sin(2F_{sc}t) - Q \cos(2F_{sc}t). \end{aligned}$$



# Decoding NTSC Signals (cont'd)

- The NTSC bandwidth of 6 MHz is tight. Its audio subcarrier frequency is 4.5 MHz. The Picture carrier is at 1.25 MHz, which places the center of the audio band at  $1.25+4.5 = 5.75$  MHz in the channel (Fig. 5.5). But notice that the color is placed at  $1.25+3.58 = 4.83$  MHz.
- So the audio is a bit too close to the color subcarrier - a cause for potential interference between the audio and color signals. It was largely due to this reason that the NTSC color TV actually slowed down its frame rate to  $30 \times 1,000 / 1,001 \approx 29.97$  fps.
- As a result, the adopted NTSC color subcarrier frequency is slightly lowered to

$$f_{sc} = 30 \times 1,000 / 1,001 \times 525 \times 227.5 \approx 3.579545 \text{ MHz};$$

where 227.5 is the number of color samples per scan line in NTSC broadcast TV.



# 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".
  - (c) This facilitates the use of a (line rate) comb filter at the receiver - the signals in consecutive lines are averaged so as to cancel the chroma signals (that always carry opposite signs) for separating Y and C and obtaining high quality Y signals.





# 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 at 4.25 MHz and 4.41 MHz respectively.
  - (b) They are sent in alternate lines, i.e., only one of the U or V signals will be sent on each scan line.



# Comparison

- 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



## 5.3 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.



# 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:
  - (a) 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.



# Chroma Subsampling (cont'd)

- (b) 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)(Cb4, Y4), and so on (or averaging is used).
- (c) The scheme "4:1:1" subsamples *horizontally* by a factor of 4.
- (d) 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.
- Scheme 4:2:0 along with other schemes is commonly used in JPEG and MPEG (see later chapters in Part 2).



# Chroma subsampling

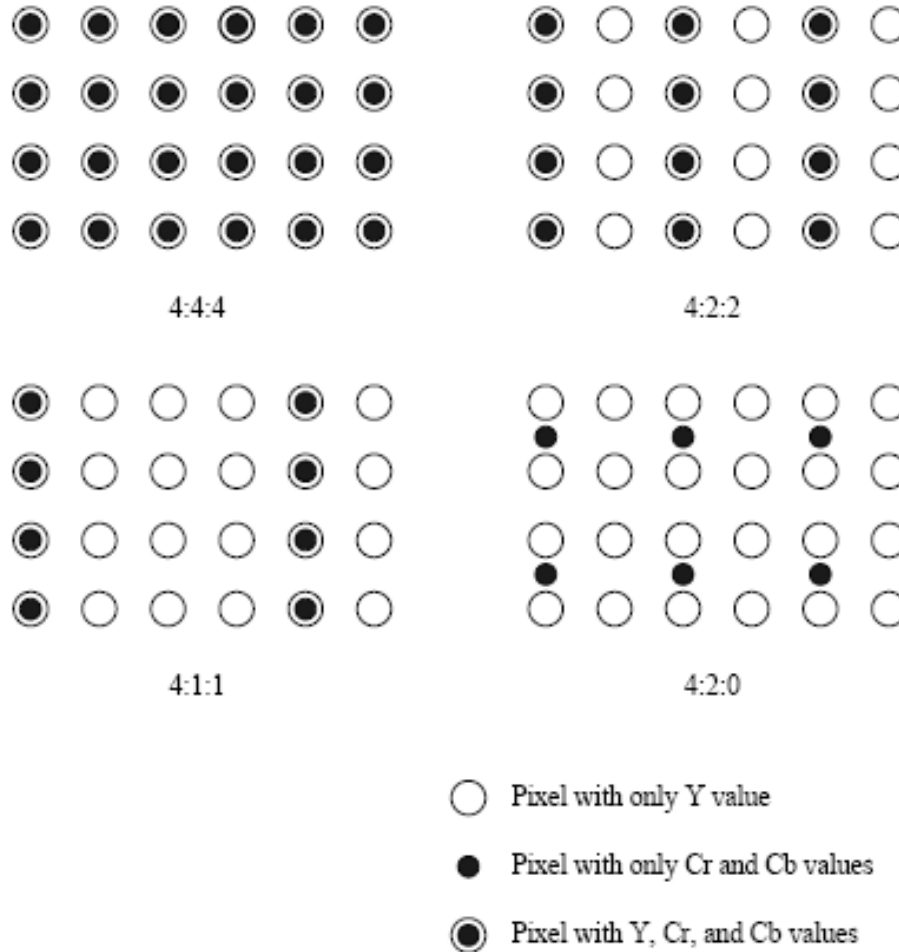


Fig. 5.6: Chroma subsampling.



# CCIR 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).
- $525 * 858 * 30 * 2 * 8 = 216\text{Mbps}$



- 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.
  - (d) 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.





# Digital video specifications

Table 5.3: Digital video specifications

	CCIR 601 525/60 NTSC	CCIR 601 625/50 PAL/SECAM	CIF	QCIF
Luminance resolution	720 × 480	720 × 576	352 × 288	176 × 144
Chrominance resolution	360 × 480	360 × 576	176 × 144	88 × 72
Color Subsampling	4:2:2	4:2:2	4:2:0	4:2:0
Aspect Ratio	4:3	4:3	4:3	4:3
Fields/sec	60	50	30	30
Interlaced	Yes	Yes	No	No



# HDTV (High Definition TV)

- 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) MUSE (MUltiple sub-Nyquist Sampling Encoding) was an improved NHK HDTV with hybrid analog/digital technologies that was put in use in the 1990s. 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.



# history of HDTV evolution

- A brief history of HDTV evolution:
  - (a) In 1987, the FCC decided that HDTV standards must be compatible with the existing NTSC standard and be conned to the existing VHF (Very High Frequency) and UHF (Ultra High Frequency) bands.
  - (b) 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.
  - (c) 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, Sarno and others.
  - (d) This eventually led to the formation of the ATSC (Advanced Television Systems Committee) - responsible for the standard for TV broadcasting of HDTV.
  - (e) In 1995 the U.S. FCC Advisory Committee on Advanced Television Service recommended that the ATSC Digital Television Standard be adopted.



# ASTC Formats

- 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	Picture Rate
1,920	1,080	16:9	60I 30P 24P
1,280	720	16:9	60P 30P 24P
704	480	16:9 & 4:3	60I 60P 30P 24P
640	480	4:3	60I 60P 30P 24P



# TV vs HDTV

- 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.



# Digital TV Broadcasting

- The FCC has planned to replace all analog broadcast services with digital TV broadcasting by the year 2006. The services provided will include:
  - **SDTV (Standard Definition TV)**: the current NTSC TV or higher.
  - **EDTV (Enhanced Definition TV)**: 480 active lines or higher, i.e., the third and fourth rows in Table 5.4.
  - **HDTV (High Definition TV)**: 720 active lines or higher.



## 5.4 Further Exploration

<http://www.cs.sfu.ca/mmmbook/furtherv2/node5.html>

- Links given for this Chapter on the text website include:
  - Tutorials on NTSC television
  - The official ATSC home page
  - The latest news on the digital TV front
  - Introduction to HDTV
  - The official FCC (Federal Communications Commission) home page



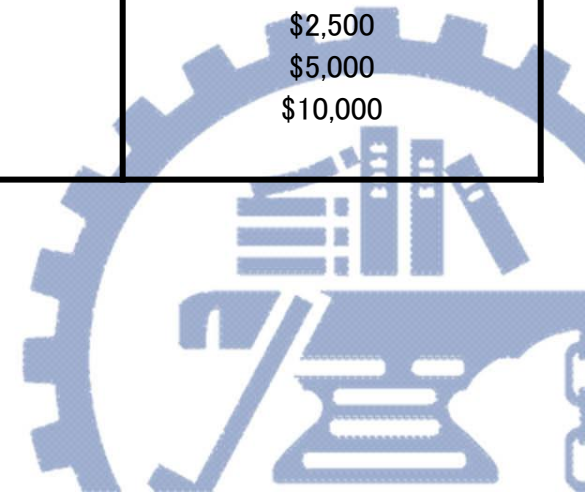
- 1991: MPEG-1
  - No IP charge
- 1994: MPEG-2
  - 1997, MPEG LA Patent Pool start, one stop shop
  - Before 2002, 4 US\$/device
  - After 2002, 2.5 US\$/device
- 1996: MPEG LA Inc. no relationship with MPEG
- 1999: MPEG-4 (Part 2)
  - Complex, Charge to device, content (per title), usage (per year)
  - AOL-Times Warner against
- 2003: H.264/MPEG-4 AVC (Part 10)
  - 2003.11.17 licensing term announced
  - EBU against ([2003第96号声明](#))
  - 2004.05.20 licensing term fixed



## MPEG4 AVC licensing term for content and usage



Categories	MPEGLA	Via	Total
<b>Title-by-title</b>			
Less than 12 minutes	\$0.00	\$0.005	\$0.005
12 – 30 minutes	2% or \$0.02	\$0.005	2% or \$0.025
31 – 89 minutes	2% or \$0.02	\$0.015	2% or \$0.035
More than 90 minutes	2% or \$0.02	\$0.025	2% or \$0.045
<b>Subscriber/year</b>			
0 – 10,000	0		0
10,000 – 250,000	\$25,000		\$25,000
250,000 – 500,000	\$50,000	-	\$50,000
500,000 – 1,000,000	\$75,000		\$75,000
More than 1,000,000	\$100,000		\$100,000
<b>Free to Air Broadcasting</b>			
<b>Viewers/year</b>			
100,000 – 500,000	\$2,500		\$2,500
500,000 – 1,000,000	\$5,000	-	\$5,000
More than 1,000,000	\$10,000		\$10,000





# MPEG4 AVC licensing term for device

Unit/Year	MPEGLA	Via	Total
0 – 50,000	\$0.00	\$0.00	\$0.00
50,000 – 100,000	\$0.00	\$0.25	\$0.25
100,000 – 5,000,000	\$0.20	\$0.25	\$0.45
5,000,000 – 20,000,000	\$0.10	\$0.25 (Reach the CAP at 10M or 16M)	\$0.35
More than 20,000,000	\$0.10 (Reach the CAP at 30M, 37.5M or 45M)	–	\$0.10

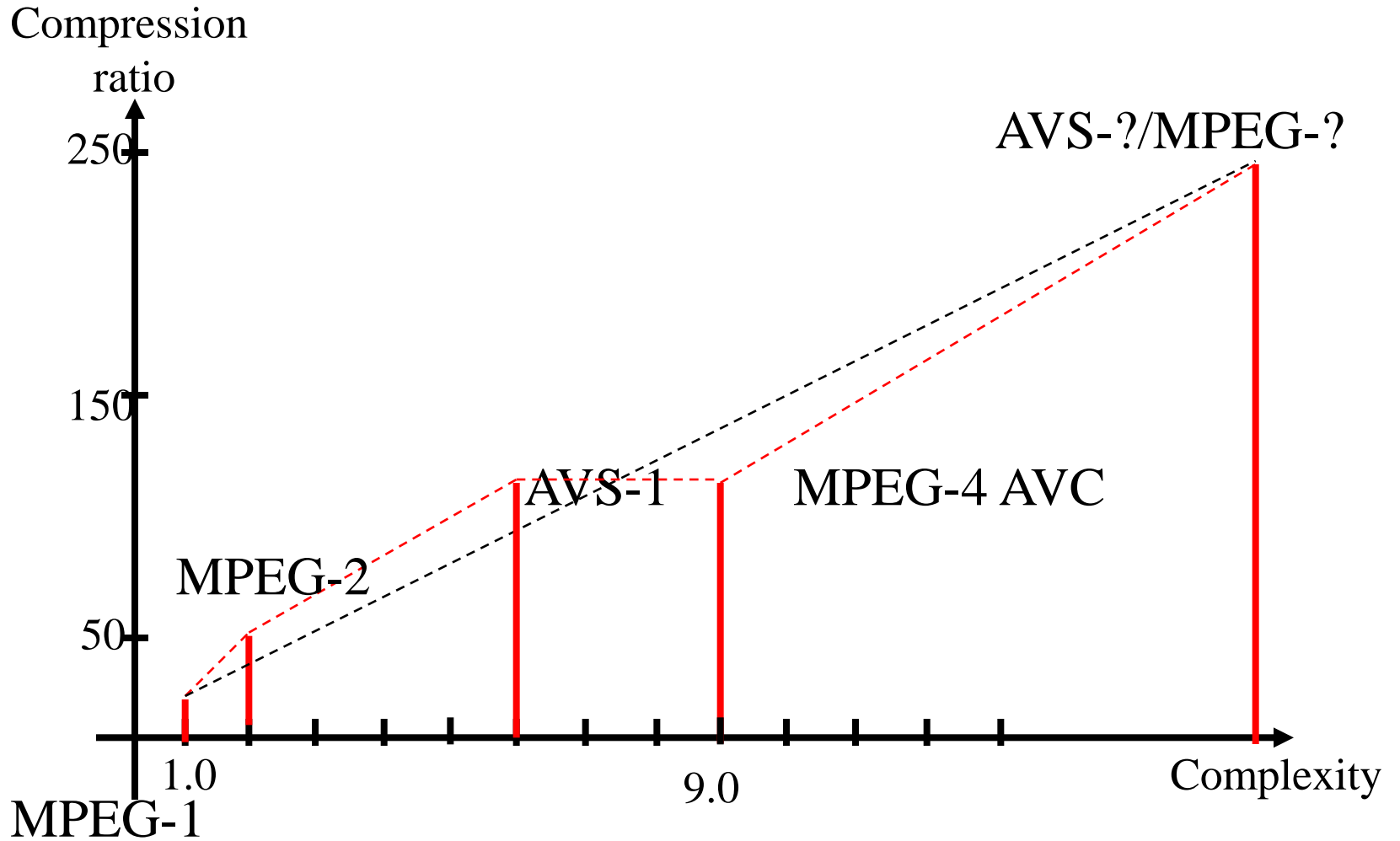


# Differences between AVS, H.264, and MPEG-2

tools	AVS	H.264	MPEG-2
<b>1/4 pixel MC</b>	1/2 pixels 4-tap 1/4 pixels 4-tap	1/2 pixels 6-tap 1/4 pixels 2-tap	1/2 pixels 2-tap
<b>Transform and quantization</b>	8x8 integer transform, encoding site normalization only	4x4 integer transform, both encoding and decoding sites need to normalize	8x8 float DCT
<b>Entropy coding</b>	Adaptive 2D VLC	CAVLC CABAC	VLC
<b>Loop filter</b>	8x8 based Less boundaries Less BS-levels (0..2), Less pixels filtered (p0, p1, q0, q1)	4x4 based More boundaries More BS-levels (0..4), More pixels filtered (p0..p3, q0..q3)	N/A



# Cost efficiency





# Cost efficiency analysis

Tools	Estimated cost increase	
	AVS	H.264
Multiple reference	1	2
Variable block-size MC	1	2
Quarter pixel	3	3
Entropy coding	0.5	1
Deblock filter	0.5	1
Total	6	9