Part 3

Practical matters

This page intentionally left blank

Component video colour coding for SD

29

Various scale factors are applied to the basic colour difference components $B'-Y'$ and $R'-Y'$ for different applications. In the previous chapter, I introduced luma and colour difference coding; in this chapter, I will detail the following coding systems:

- $B'-Y'$, $R'-Y'$ components form the numerical basis for all the other component sets; otherwise, they are not directly used.
- $P_B P_R$ components are used for component analog video (including analog interfaces in devices such as DVD players and set-top boxes).
- $C_B C_R$ components as defined in BT.601 and BT.709 are used for component digital video, including studio video, DV, MPEG, and H.264.
- "Full-swing" (or "full-range") $C_B C_R$ components are used in JPEG/JFIF.
- UV components are used for composite NTSC or PAL, as described in *UV components,* in Chapter 5 of *Composite NTSC and PAL: Legacy Video Systems*.
- IQ components were historically used for composite NTSC until about 1970, as described in Chapter 4, *NTSC Y'IQ system*, of *Composite NTSC and PAL: Legacy Video Systems*.

Video uses the symbols U and V to represent certain colour difference components. The CIE defines the pairs $[u, v]$, $[u', v']$, and $[u^*, v^*]$. All of these pairs represent *chromatic* or chroma information, but they are all numerically and functionally different. Video $[U, V]$ components are neither directly based upon, nor superseded by, any of the CIE colour spaces.

$Y'UV$ and $Y'IQ$ are intermediate quantities toward the formation of composite NTSC, PAL, and S-video. Neither $Y'UV$ nor $Y'IQ$ has a standard component interface, and neither is appropriate when the components are kept separate. Unfortunately, the $Y'UV$ nomenclature has come to be used rather loosely, and some people use $Y'UV$ to denote *any* scaling of $B'-Y'$ and $R'-Y'$.

For a discussion of primary chromaticities, see page 290.

The coding systems described in this chapter can be applied to various *RGB* primary sets – EBU 3213, SMPTE RP 145 (or potentially even BT.709). BT.601 does not specify primary chromaticities: SMPTE RP 145 primaries are implicit in 480*i* SD, and EBU 3213 primaries are implicit in 576*i* SD. However, virtually all of modern consumer receivers interpret content – whether SD or HD – according to BT.709 primaries. As I write, program content created in North America is mastered with SMPTE primaries (contrary to the spirit and letter of ITU-R, SMPTE, and ATSC standards) and content created in Europe is mastered to EBU primaries. However, all of this content is displayed in the consumer domain using BT.709 primaries. We look forward to the day when content creators actually master using the BT.709 colour space.

The equations for [Y', $B'-Y'$, $R'-Y'$], $Y'P_\mathrm{B}P_\mathrm{R}$, and $Y'C_\mathrm{B}C_\mathrm{R}$ can be based upon either the BT.601 luma coefficients of SD or the BT.709 coefficients of HD. The equations and figures of this chapter are based upon the BT.601 coefficients. Unfortunately, the luma coefficients that have been standardized for HD are different from those of BT.601. Concerning the HD luma coefficients, see *BT.709 luma* on page 346; for details of HD colour difference components, see the following chapter, *Component video colour coding for HD,* on page 369.

Chroma components are properly ordered $B'-Y'$ then $R'-Y'$; or P_B then P_R; or C_B then C_R. Blue associates with U, and red with V; U and V are ordered alphabetically. The subscripts in $C_\mathrm{B}C_\mathrm{R}$ and $P_\mathrm{B}P_\mathrm{R}$ are often written in lowercase. In my opinion, this compromises readability, so I write them in uppercase. The B in C_B serves as a tag, not a variable, so I set it in Roman type (that is, upright, not italic). Authors with great attention to detail sometimes "prime" $C_\mathrm{B}C_\mathrm{R}$ and $P_\mathrm{B}P_\mathrm{R}$ to indicate their nonlinear origin, but no standard or deployed image coding system has employed linear-light colour differences, nor would that be sensible for perceptual reasons, so I omit the primes.

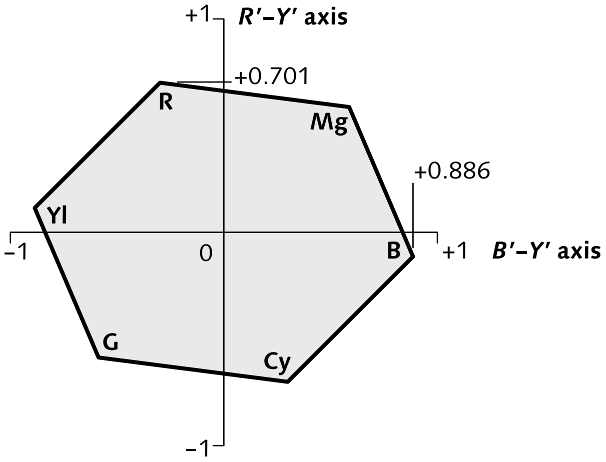


Figure 29.1 *B′–Y′*, *R′–Y′* **components for SD**

B′–Y′, *R′–Y′* components for SD

To obtain [*Y′*, *B′–Y′*, *R′–Y′*] components from *R′G′B′*, for BT.601 luma, use this matrix equation:

$$\begin{bmatrix} {}^{601}Y' \\ B'-{}^{601}Y' \\ R'-{}^{601}Y' \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.299 & -0.587 & 0.886 \\ 0.701 & -0.587 & -0.114 \end{bmatrix} \cdot \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \qquad \text{Eq 29.1}$$

Figure 29.1 shows a plot of the [*B′–Y′*, *R′–Y′*] colour difference plane.

ITU-R Rec. BT.601-5, *Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios*.

As I described on page 346, the BT.601 luma coefficients are used for SD. With these coefficients, the *B′–Y′* component reaches its positive maximum at pure blue ($R' = 0$, $G' = 0$, $B' = 1$; $Y' = 0.114$; $B'-Y' = +0.886$) and its negative maximum at pure yellow ($B'-Y' = -0.886$). Analogously, the extrema of *R′–Y′* take values ±0.701, at pure red and cyan. These are inconvenient values for both digital and analog systems. The P_BP_R, C_BC_R, and *UV* colour difference components all involve versions of [*Y′*, *B′–Y′*, *R′–Y′*] that are scaled to place the extrema of the component values at more convenient values.

P_BP_R components for SD

P_B and P_R denote colour difference components having excursions nominally identical to the excursion of the

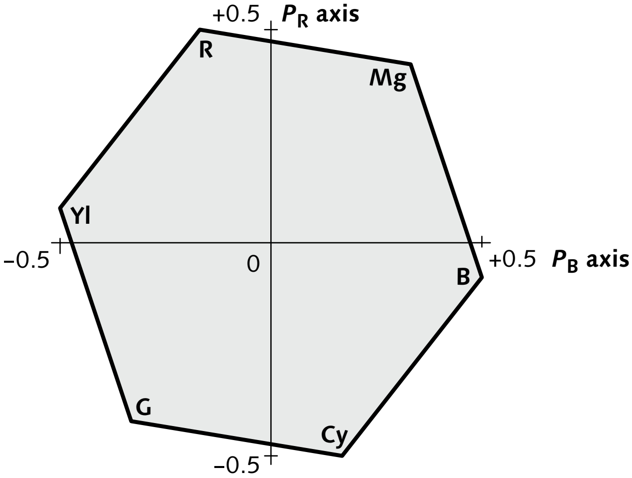


Figure 29.2 P_BP_R **components for SD**

accompanying luma component. For BT.601 luma, the equations are these:

Eq 29.2

$$\begin{aligned} P_B &= \frac{0.5}{1-0.114}\left(B' - {}^{601}Y'\right) &= \frac{1}{1.772}\left(B' - {}^{601}Y'\right) &\quad \approx 0.564\left(B' - {}^{601}Y'\right) \\ P_R &= \frac{0.5}{1-0.299}\left(R' - {}^{601}Y'\right) &= \frac{1}{1.402}\left(R' - {}^{601}Y'\right) &\quad \approx 0.713\left(R' - {}^{601}Y'\right) \end{aligned}$$

These scale factors were chosen to limit the excursion of *each* colour difference component to the range −0.5 to +0.5 with respect to unity luma excursion: 0.114 in the first expression above is the luma coefficient of blue, and 0.299 in the second is for red. Figure 29.2 above shows a plot of the [P_B, P_R] plane.

Expressed in matrix form, the $B'-Y'$ and $R'-Y'$ rows of Equation 29.1 are scaled by ${}^{0.5}\!/_{0.886}$ and ${}^{0.5}\!/_{0.701}$. To encode from $R'G'B'$ where reference black is zero and reference white is unity:

Eq 29.3

$$\begin{bmatrix} {}^{601}Y' \\ P_B \\ P_R \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.168736 & -0.331264 & 0.5 \\ 0.5 & -0.418688 & -0.081312 \end{bmatrix} \cdot \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}$$

The first row of Equation 29.3 comprises the luma coefficients; these sum to unity. The second and third rows each sum to zero, a necessity for colour difference components. The two entries of 0.5 reflect the reference excursions of P_B and P_R, at the blue and red primaries [0, 0, 1] and [1, 0, 0]. The reference excursion is

±0.5; the peak excursion may be slightly larger, to accommodate analog undershoot and overshoot. There are no standards for how much analog footroom and headroom should be provided.

The inverse, decoding matrix is this:

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.402 \\ 1 & -0.344136 & -0.714136 \\ 1 & 1.772 & 0 \end{bmatrix} \cdot \begin{bmatrix} {}^{601}Y' \\ P_B \\ P_R \end{bmatrix} \qquad \text{Eq 29.4}$$

See Table 28.2A on page 353; *Component analog $Y'P_BP_R$ interface, EBU N10*, on page 453; and *Component analog $Y'P_BP_R$ interface, industry standard*, on page 455.

$Y'P_BP_R$ is employed by 480*i* and 576*i* component analog video equipment such as that from Sony and Panasonic, where P_B and P_R are conveyed with roughly half the bandwidth of luma. Unfortunately, three different analog interface level standards are used: $Y'P_BP_R$ is ambiguous with respect to electrical interface.

P_B and P_R are properly written in that order, as I described on page 358. The *P* stands for *parallel*, stemming from a failed effort within SMPTE to standardize a parallel electrical interface for component analog video. In C_BC_R, which I will now describe, *C* stands for *chroma*. The C_BC_R notation predated P_BP_R.

C_BC_R components for SD

A straightforward scaling of $Y'P_BP_R$ components would have been suitable for digital interface. Scaling of luma to the range [0 … 255] would have been feasible; this "full-range" scaling of luma is used in JPEG/JFIF used in computing, as I will describe on page 365. However, for studio applications it is necessary to provide signal-processing footroom and headroom to accommodate ringing from analog and digital filters, and to accommodate signals from misadjusted analog equipment.

For an 8-bit interface, luma could have been scaled to an excursion of 224; $B'-Y'$ and $R'-Y'$ could have been scaled to ±112. This would have left 32 codes of footroom and headroom for each component. Although sensible, that approach was not taken when BT.601 was adopted in 1984. Instead – and unfortunately, in my opinion – different excursions were standardized for luma and chroma. Eight-bit luma excursion was standardized at 219; chroma excursion was standardized at 224. Each colour difference component has as excursion $^{224}/_{219}$ that of luma. Since video component amplitudes are usually referenced to luma excursion, this condition is more clearly stated the opposite way: In $Y'C_BC_R$, each colour difference component has $^{224}/_{219}$ the excursion of the luma component. The notation C_BC_R distinguishes this set from P_BP_R, where the luma and chroma excursions are nominally identical: Conceptually, $Y'P_BP_R$ and $Y'C_BC_R$ differ only in scaling.

Historically, $Y'P_BP_R$ scaling was used at analog interfaces, and $Y'C_BC_R$ was used at digital interfaces. Nowadays so many different scale factors and offsets are in use in both the analog and digital domains that the dual nomenclature is more a hindrance than a help.

The $Y'P_BP_R$ and $Y'C_BC_R$ scaling discrepancy is unfortunate enough, but it is compounded by "full-swing" (or "full-range") $Y'C_BC_R$ used in JPEG/JFIF, scaled similarly but not identically to $Y'P_BP_R$; see page 365. Confusion is also compounded by the EBU referring in Technical Standard N10-1998 to C_BC_R analog colour difference components, when they are properly denoted P_BP_R.

To provide footroom to accommodate luma signals that go slightly negative, an offset is added to luma at a $Y'C_BC_R$ interface. At an 8-bit interface, an offset of +16 is added; this places black at code 16 and white at code 235. At an 8-bit interface, codes 0 and 255 are used for synchronization purposes; these codes are prohibited from video data. Codes 1 through 15 are interpreted as signal levels $-^{15}/_{219}$ through $-^{1}/_{219}$ (respectively), relative to unity luma excursion; codes 236 through 254 are interpreted as signal levels $^{220}/_{219}$ through $^{238}/_{219}$ (respectively), relative to unity excursion. Unfortunately, luma footroom and headroom are asymmetrical.

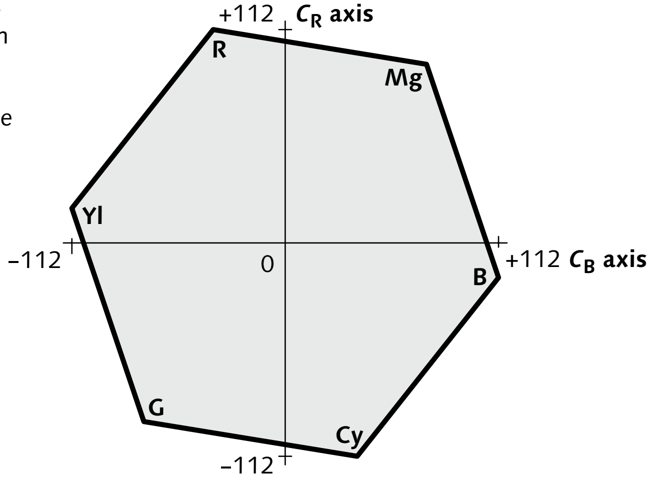
C_BC_R colour difference components are conveyed in offset binary form: An offset of +128 is added. In studio $Y'C_BC_R$, chroma reference levels are 16 and 240, and codes 0 and 255 are prohibited from chroma data.

Reference white and black codes of the 10-bit interface have trailing zeros "to the left" of the least significant bit of the 8-bit representation. "Widening" from 8-bit to higher precision is properly accomplished by shifting left, *not* by multiplying by $^{879}/_{219}$. "Narrowing" to 8 bits is properly accomplished by rounding then shifting right.

BT.601 provides for 10-bit components; 10-bit studio video equipment is now commonplace. At a 10-bit interface, the 8-bit interface levels and prohibited codes are maintained; extra bits are appended as least-significant bits (LSBs) to provide increased precision. The prohibited codes respect the 8-bit interface: Codes having all 8 most-significant bits either all zeros or all ones are prohibited from video data across a 10-bit interface.

For signal-processing arithmetic operations such as gain adjustment, Y', C_B, and C_R must be zero for black: The interface offsets must be removed. For 8-bit luma arithmetic, it is convenient to place reference black at code 0 and reference white at code 219. Colour difference signals are most conveniently handled in two's complement form, scaled so that reference colour

Figure 29.3 **$C_B C_R$ components for SD** are shown in their mathematical form. The range outside [–112 … +112] is available for undershoot and overshoot. At an 8-bit interface, an offset of +128 is added to each colour difference component.



difference signals (at pure yellow, cyan, red, and blue) are ±112. Figure 29.3 above shows the $C_B C_R$ colour difference plane scaled in this manner, without offsets.

As far as I am concerned, the offsets should be treated as an interface feature. Most descriptions of $Y'C_BC_R$, though – including SMPTE and ITU standards – take the $Y'C_BC_R$ notation to include the offset. In the equations to follow, I colour the offset terms. If your goal is to compute abstract, mathematical quantities suitable for signal processing with signed numbers, omit these offset terms. If you are concerned with interfacing unsigned values, include them.

These equations form BT.601 $Y'C_BC_R$ components from [Y', B'–Y', R'–Y'] components ranging [0 … +1]:

The numerical values used in this equation, and in those to follow, are based on the BT.601 luma coefficients. The coefficients for HD are, unfortunately, different. See *BT.601 luma*, on page 346.

$$\begin{aligned} {}^{601}_{219}Y' &= 16 + \left(219 \cdot {}^{601}Y'\right) \\ C_B &= 128 + \frac{112}{0.886}\left(B' - {}^{601}Y'\right) \\ C_R &= 128 + \frac{112}{0.701}\left(R' - {}^{601}Y'\right) \end{aligned} \qquad \text{Eq 29.5}$$

To extend Equation 29.5 to 10 bits, append to each of Y', C_B, and C_R two low-order bits having binary weights $\frac{1}{2}$ and $\frac{1}{4}$. To extend $Y'C_BC_R$ beyond 10 bits, continue the sequence with LSBs weighted $\frac{1}{8}$, $\frac{1}{16}$, and so on. If you prefer to express these quantities as whole numbers, without fractional bits, multiply Equation 29.5 (and all of the equations to follow) by 2^{k-8}, where $8 \le k$ denotes the bit depth.

To obtain 8-bit BT.601 $Y'C_BC_R$ from $R'G'B'$ ranging 0 to 1, scale the rows of the matrix in Equation 29.3 by the factors 219, 224, and 224, corresponding to the excursions of each of Y', C_B, and C_R, respectively:

Eq 29.6

$$\begin{bmatrix} {}^{601}_{219}Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \cdot \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}$$

Summing the top row of this matrix yields 219, the luma excursion. The lower two rows sum to zero. The two entries of 112 reflect the positive C_B and C_R extrema, at the blue and red primaries.

To recover $R'G'B'$ in the range [0…+1] from 8-bit BT.601 $Y'C_BC_R$, invert Equation 29.6:

Eq 29.7

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} 0.00456621 & 0 & 0.00625893 \\ 0.00456621 & -0.00153396 & -0.00318811 \\ 0.00456621 & 0.00791071 & 0 \end{bmatrix} \cdot \left(\begin{bmatrix} {}^{601}_{219}Y' \\ C_B \\ C_R \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \right)$$

You can determine the excursion that an encoding matrix is designed to produce – often 1, 219, 255, or 256 – by summing the coefficients in the top row. In Equation 29.8, the sum is 256. If you find an unexpected sum, suspect an error in the matrix.

This matrix contains entries larger than 256; the corresponding multipliers will need capability for more than 8 bits.

When rounding the matrix coefficients, take care to preserve the intended row sums, in this case, [1, 0, 0]. You must take care to prevent overflow due to roundoff error or other conditions: Use saturating arithmetic.

At the interface, after adding the offsets, clip all three components to the range 1 through 254 inclusive, to avoid the prohibited codes 0 and 255.

$Y'C_BC_R$ from studio *RGB*

In studio equipment, 8-bit $R'G'B'$ components usually have the same 219 excursion as the luma component of $Y'C_BC_R$. To encode 8-bit BT.601 $Y'C_BC_R$ from $R'G'B'$ in the range [0…219], scale the encoding matrix of Equation 29.6 by $^{256}/_{219}$:

Eq 29.8

$$\begin{bmatrix} {}^{601}_{219}Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix} 76.544 & 150.272 & 29.184 \\ -43.366 & -85.136 & 128.502 \\ 128.502 & -107.604 & -20.898 \end{bmatrix} \cdot \begin{bmatrix} {}_{219}R' \\ {}_{219}G' \\ {}_{219}B' \end{bmatrix}$$

For implementation in binary arithmetic, the multiplication by $^{1}/_{256}$ can be accomplished by shifting. To decode to $R'G'B'$ in the range [0…219] from 8-bit BT.601 $Y'C_BC_R$, invert Equation 29.8:

Eq 29.9

$$\begin{bmatrix} {}_{219}R' \\ {}_{219}G' \\ {}_{219}B' \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 256 & 0 & 350.901 \\ 256 & -86.132 & -178.738 \\ 256 & 443.506 & 0 \end{bmatrix} \cdot \left(\begin{bmatrix} {}^{601}_{219}Y' \\ C_B \\ C_R \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \right)$$

The entries of 256 in this matrix indicate that the corresponding component can simply be added; there is no need for a multiplication operation. These transforms assume that the $R'G'B'$ components incorporate gamma correction, such as that specified by BT.709; see page 333.

$Y'C_BC_R$ from computer RGB

In computing it is conventional to use 8-bit $R'G'B'$ components, with no headroom and no footroom: Black is at code 0 and white is at 255. To encode 8-bit BT.601 $Y'C_BC_R$ from $R'G'B'$ in this range, scale the matrix of Equation 29.6 by $^{256}/_{255}$:

Eq 29.10

$$\begin{bmatrix} {}^{601}_{219}Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix} 65.738 & 129.057 & 25.064 \\ -37.945 & -74.494 & 112.439 \\ 112.439 & -94.154 & -18.285 \end{bmatrix} \cdot \begin{bmatrix} {}_{255}R' \\ {}_{255}G' \\ {}_{255}B' \end{bmatrix}$$

To decode $R'G'B'$ in the range [0…255] from 8-bit BT.601 $Y'C_BC_R$, use the transform of Equation 29.11:

Eq 29.11

$$\begin{bmatrix} {}_{255}R' \\ {}_{255}G' \\ {}_{255}B' \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 298.082 & 0 & 408.583 \\ 298.082 & -100.291 & -208.120 \\ 298.082 & 516.411 & 0 \end{bmatrix} \cdot \left(\begin{bmatrix} {}^{601}_{219}Y' \\ C_B \\ C_R \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \right)$$

BT.601 $Y'C_BC_R$ uses the extremes of the coding range to handle signal overshoot and undershoot. Clipping is required when decoding to an $R'G'B'$ range that has no headroom or footroom.

"Full-swing" $Y'C_BC_R$

The $Y'C_BC_R$ coding used in JPEG/JFIF stillframes in computing conventionally uses "full-swing" coding with no footroom and no headroom. Luma (Y') is scaled to an excursion of 255 and represented in 8 bits: Black is at code 0 and white is at code 255. Obviously, luma codes 0 and 255 are not prohibited! Colour difference

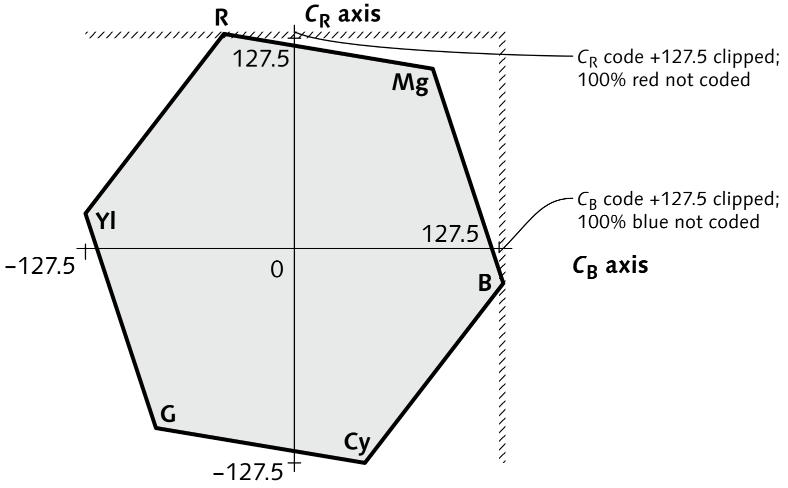


Figure 29.5 **C_BC_R "full-range" components** used in JPEG/JFIF are shown. C_B and C_R are scaled to ±127.5; however, they are encoded into a two's complement range of –128 to +127. Chroma codes +127.5 are clipped; fully saturated blue and red cannot be preserved. No provision is made for undershoot or overshoot. The accompanying luma signal ranges 0 through 255.

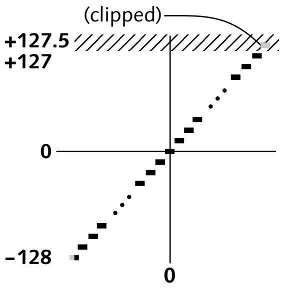


Figure 29.4 A **"full-swing" C_BC_R quantizer** is used in JPEG/JFIF. Code +127.5 – required to represent pure blue or red – is clipped.

components are scaled to an excursion of ±127.5; each colour difference component nominally has the same excursion as luma. However, an offset of +128 is applied (instead of the +127.5 that you might expect), apparently so that pure grey has an integer code value. The +128 offset causes pure blue and pure red to take post-offset values of 255.5, which clip. Figure 29.4 shows the transfer function of the colour difference quantizer, emphasizing that pre-offset chroma code +127.5 (pure blue, or pure red) causes clipping. As a consequence, pure blue and pure red are liable to fail to make the "round-trip" accurately through JPEG/JFIF compression and decompression. Figure 29.5 above shows the full-range C_BC_R colour difference plane.

To encode from *R'G'B'* in the range [0…255] into 8-bit $Y'C_BC_R$, with luma in the range [0…255] and C_B and C_R each ranging ±128, use this transform:

Eq 29.12

$$\begin{bmatrix} {}^{601}_{255}Y' \\ C_B \\ C_R \end{bmatrix} = \frac{1}{256}\begin{bmatrix} 76.544 & 150.272 & 29.184 \\ -43.366 & -85.136 & 128.502 \\ 128.502 & -107.604 & -20.898 \end{bmatrix} \bullet \begin{bmatrix} {}_{255}R' \\ {}_{255}G' \\ {}_{255}B' \end{bmatrix}$$

To decode into *R'G'B'* in the range [0…255] from full-range 8-bit $Y'C_BC_R$, use the transform in Equation 29.13:

Eq 29.13

$$\begin{bmatrix} {}_{255}R' \\ {}_{255}G' \\ {}_{255}B' \end{bmatrix} = \frac{1}{256}\begin{bmatrix} 256 & 0 & 357.510 \\ 256 & -87.755 & -182.105 \\ 256 & 451.860 & 0 \end{bmatrix} \bullet \begin{bmatrix} {}^{601}_{255}Y' \\ C_B \\ C_R \end{bmatrix}$$

$Y'UV$, $Y'IQ$ confusion

I have detailed $Y'P_BP_R$ and $Y'C_BC_R$. These are both based on $[B'-Y', R'-Y']$ components, but they have different scale factors suitable for component analog and component digital interface, respectively.

Colour differences pairs $[U, V]$ and $[I, Q]$ are also based on $B'-Y'$ and $R'-Y'$, but have yet another set of scale factors. UV scaling – or IQ scaling and rotation – is appropriate only when the signals are destined for composite encoding, as in NTSC or PAL.

Unfortunately, the notation $Y'UV$ – or worse, YUV – is sometimes loosely applied to *any* form of colour difference coding based on $[B'-Y', R'-Y']$. Do not be misled by video equipment having connectors labelled $Y'UV$ or Y', $B'-Y'$, $R'-Y'$, or these symbols without primes, or by JPEG being described as utilizing $Y'UV$ coding. In fact the analog connectors convey signals with $Y'P_BP_R$ scaling, and the JPEG standard itself specifies what I would denote ${}^{601}_{255}Y'C_BC_R$.

When the term $Y'UV$ (or YUV) is encountered in a computer graphics or image-processing context, usually BT.601 $Y'C_BC_R$ is meant, but beware!

- Any image data supposedly coded to the original 1953 NTSC primaries is suspect, because it has been roughly four decades since any equipment using these primaries has been built.
- Generally no mention is made of the transfer function of the underlying $R'G'B'$ components, and no account is taken of the nonlinear formation of luma.

When the term $Y'IQ$ (or YIQ) is encountered, beware!

- Image data supposedly coded in $Y'IQ$ is suspect since no analog or digital interface for $Y'IQ$ components has ever been standardized.
- NTSC encoders and decoders built since 1970 have been based upon $Y'UV$ components, not $Y'IQ$. Contrary to much published information, $Y'IQ$ components have not been used for "NTSC" for about 4 decades.

Component video colour coding for HD

30

In the previous chapter, *Component video colour coding for SD*, I detailed various component colour coding systems that use the luma coefficients specified in BT.601. Unfortunately, for no good technical reason, BT.709 for HD standardized different luma coefficients. Deployment of HD requires upconversion and downconversion capabilities both at the studio and at consumers' premises; this situation will persist for a few decades. Owing to this aspect of conversion between HD and SD, if you want to be an HD expert, you have to be an SD expert as well!

Before BT.709 was established, SMPTE 240M-1988 for 1035*i*30 HD standardized luma coefficients based upon the SMPTE RP 145 primaries. Equipment deployed between about 1988 and 1997 used the 240M parameters, but SMPTE 240M is now obsolete. For details, see the first edition of this book.

Today's computer imaging systems – for still frames, desktop video, and other applications – typically use the BT.601 parameters, independent of the image's pixel count ("resolution independence"). In computer systems that perform HD editing, it is highly desirable that all of the content on the same timeline uses the same colour coding, but there's no simple answer whether BT.601 or BT.709 coding should be used. Generally, it is sensible to retain the BT.601 coefficients.

In this chapter, I assume that you're familiar with the concepts of *Luma and colour differences*, described on page 335. I will detail these component sets:

- $B'-Y'$, $R'-Y'$ components, the basis for P_BP_R and C_BC_R
- P_BP_R components, used for analog interfaces
- C_BC_R components, used for digital interfaces

$B'-Y'$, $R'-Y'$ components for BT.709 HD

The $B'-Y'$ component reaches its positive maximum at blue ($R'=0$, $G'=0$, $B'=1$). With BT.709 luma coefficients, the maximum of $B'-Y' = +0.9278$ occurs at

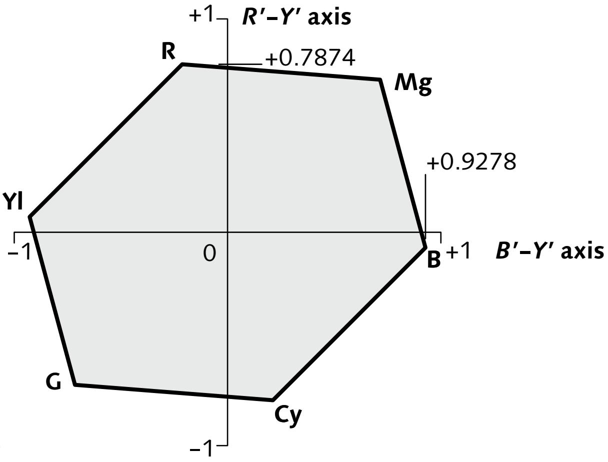


Figure 30.1 *B′–Y′*, *R′–Y′* **components for BT.709 HD**

$Y' = 0.0722$. The $B'-Y'$ component reaches its negative maximum at yellow ($B'-Y' = -0.9278$). Analogously, the extrema of $R'-Y'$ occur at red and cyan at values ±0.7874 (see Figure 30.1 above). These are inconvenient values for both digital and analog systems. The ${}^{709}Y'P_BP_R$ and ${}^{709}Y'C_BC_R$ systems to be described both employ versions of $[Y', B'-Y', R'-Y']$ that are scaled to place the extrema of the component values at more convenient values.

To obtain $[Y', B'-Y', R'-Y']$, from $R'G'B'$, for BT.709 luma coefficients, use this matrix equation:

$$\begin{bmatrix} {}^{709}Y' \\ B'-{}^{709}Y' \\ R'-{}^{709}Y' \end{bmatrix} = \begin{bmatrix} 0.2126 & 0.7152 & 0.0722 \\ -0.2126 & -0.7152 & 0.9278 \\ 0.7874 & -0.7152 & -0.0722 \end{bmatrix} \bullet \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \qquad \text{Eq 30.1}$$

P_BP_R components for BT.709 HD

If two colour difference components are to be formed having excursions identical to luma, then P_B and P_R colour difference components are used. For BT.709 luma, the equations are these:

Eq 30.2

$$\begin{aligned} {}^{709}P_B &= \frac{0.5}{1-0.0722}\left(B'-{}^{709}Y'\right) = \frac{1}{1.8556}\left(B'-{}^{709}Y'\right) \approx 0.5389\left(B'-{}^{709}Y'\right) \\ {}^{709}P_R &= \frac{0.5}{1-0.2126}\left(R'-{}^{709}Y'\right) = \frac{1}{1.5748}\left(R'-{}^{709}Y'\right) \approx 0.6350\left(R'-{}^{709}Y'\right) \end{aligned}$$

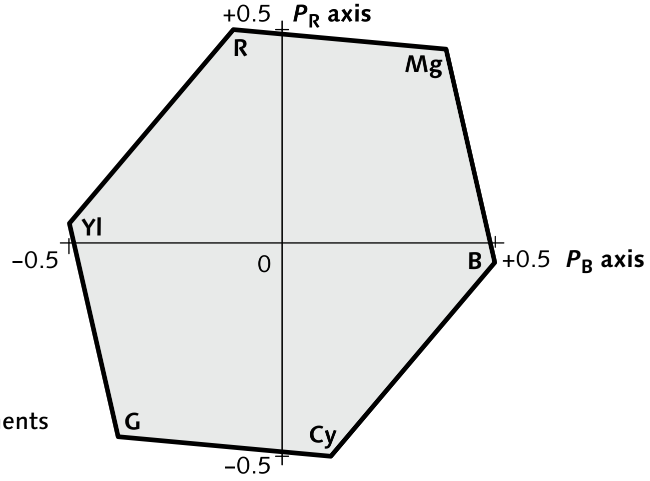


Figure 30.2 $P_B P_R$ components for BT.709 HD

These scale factors limit the excursion of each colour difference component to the range ±0.5 with respect to unity luma excursion: 0.0722 in the first expression above is the luma coefficient of blue, and 0.2126 in the second is for red. At an HD analog interface, luma ranges from 0 mV (black) to 700 mV (white), and P_B and P_R analog components range ±350 mV. Figure 30.2 above shows a plot of the $[P_B, P_R]$ plane.

Expressed in matrix form, the $B'-Y'$ and $R'-Y'$ rows of Equation 30.1 are scaled by $^{0.5}/_{0.9278}$ and $^{0.5}/_{0.7874}$. To encode from $R'G'B'$ where reference black is zero and reference white is unity:

Eq 30.3

$$\begin{bmatrix} {}^{709}Y' \\ P_B \\ P_R \end{bmatrix} = \begin{bmatrix} 0.2126 & 0.7152 & 0.0722 \\ -0.114572 & -0.385428 & 0.5 \\ 0.5 & -0.454153 & -0.045847 \end{bmatrix} \bullet \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}$$

The inverse, decoding matrix is this:

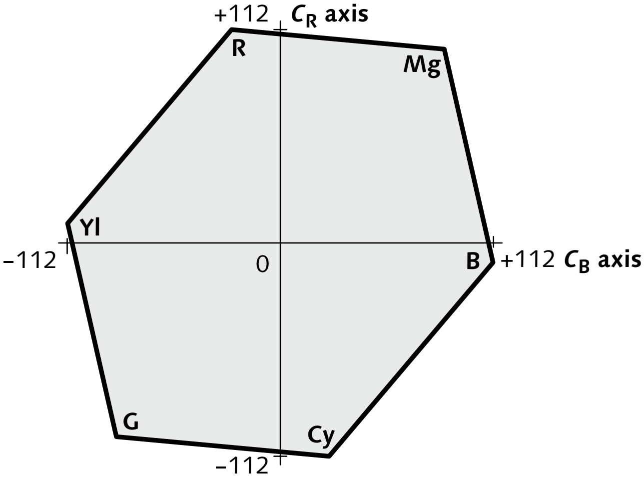
$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.5748 \\ 1 & -0.187324 & -0.468124 \\ 1 & 1.8556 & 0 \end{bmatrix} \bullet \begin{bmatrix} {}^{709}Y' \\ P_B \\ P_R \end{bmatrix}$$

Eq 30.4

$C_B C_R$ components for BT.709 HD

${}^{709}Y'C_BC_R$ coding is used in component digital HD equipment. In 8-bit systems, luma has an excursion of 219. Colour differences C_B and C_R are coded in 8-bit

Figure 30.3 $C_B C_R$ **components for BT.709 HD** are shown referenced to 8-bit processing levels. At an 8-bit interface, an offset of +128 is added to each component.



offset binary form, with excursions of ±112. The $[C_B, C_R]$ plane of HD is plotted in Figure 30.3.

In 8-bit systems, a luma offset of +16 is added at the interface, placing black at code 16 and white at code 235; an offset of +128 is added to C_B and C_R, yielding a range of 16 through 240 inclusive. (Following the convention of the previous chapter, in the equations to follow I write the offset terms in colour.) HD standards provide for 10-bit components, and 10-bit studio video equipment is commonplace. In a 10-bit interface, the 8-bit interface levels and prohibited codes are maintained; the extra two bits are appended as least-significant bits to provide increased precision.

To form $^{709}Y'C_BC_R$ from $[Y', B'-Y', R'-Y']$ components in the range [0…+1], use these equations:

Eq 30.5

$$\begin{aligned} {}^{709}_{219}Y' &= 16 + \left(219 \cdot {}^{709}Y'\right) \\ C_B &= 128 + \frac{112}{0.9278}\left(B' - {}^{709}Y'\right) \\ C_R &= 128 + \frac{112}{0.7874}\left(R' - {}^{709}Y'\right) \end{aligned}$$

To obtain $^{709}Y'C_BC_R$ from $R'G'B'$ ranging 0 to 1, scale the rows of the matrix in Equation 30.3 by the factors

[219, 224, 224], corresponding to the excursions of each of the components:

Eq 30.6

$$\begin{bmatrix} {}^{709}_{219}Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 46.559 & 156.629 & 15.812 \\ -25.664 & -86.336 & 112 \\ 112 & -101.730 & -10.270 \end{bmatrix} \cdot \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix}$$

Summing the first row of the matrix yields 219, the luma excursion from black to white. The two entries of 112 reflect the positive C_BC_R extrema at blue and red.

To recover $R'G'B'$ in the range [0…+1] from ${}^{709}Y'C_BC_R$, use the inverse of Equation 30.6:

Eq 30.7

$$\begin{bmatrix} {}_{219}R' \\ {}_{219}G' \\ {}_{219}B' \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 256 & 0 & 394.150 \\ 256 & -46.885 & -117.165 \\ 256 & 464.430 & 0 \end{bmatrix} \cdot \left(\begin{bmatrix} {}^{709}_{219}Y' \\ C_B \\ C_R \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \right)$$

The ${}^{709}Y'C_BC_R$ components are integers in 8 bits; reconstructed $R'G'B'$ is scaled to the range [0…+1].

Figure 28.2 (on page 339) illustrated that when $R'G'B'$ components are transformed to luma and colour differences, the unit $R'G'B'$ cube occupies only a small fraction of the volume of the enclosing cube. In digital video, only about ¼ of $Y'C_BC_R$ codewords correspond to $R'G'B'$ values between zero and unity. Certain signal-processing operations (such as filtering) may produce $Y'C_BC_R$ codewords that lie outside the *RGB*-legal cube. These codewords cause no difficulty in the $Y'C_BC_R$ domain, but potentially present a problem when decoded to $R'G'B'$. Generally, $R'G'B'$ values are clipped between 0 and 1.

C_BC_R components for xvYCC

xvYCC refers to an IEC standard. x.v.Color and x.v.Colour are Sony's trademarks for the scheme.

One method of extending the colour gamut of an $R'G'B'$ system is to allow components to excurse below zero and above unity. In *Wide-gamut reproduction,* on page 312, I explained one approach. The xvYCC scheme is based upon BT.709 primaries, but enables the *RGB* tristimulus components to excurse from –¼ to +4⁄3.

Concerning Pointer, see the marginal note on page 312.

When transformed to BT.709 $Y'C_BC_R$, all of the real surface colours documented by Pointer – that is, all the colours in Pointer's gamut – produce values that are $Y'C_BC_R$-valid. Though BT.1361 was needed to specify the $R'G'B'$ representation of wide-gamut colours, no special provisions are necessary to carry those colours across a ${}^{709}Y'C_BC_R$ interface. The notation "xvYCC $Y'C_BC_R$," or ${}^{\mathrm{xvYCC}}Y'C_BC_R$, makes it explicit that codewords outside the unit $R'G'B'$ cube are to be interpreted as wide-gamut colours, instead of being treated as RGB-illegal.

There is an SD version of xvYCC, using the BT.601 luma coefficients. That scheme will almost certainly never see any deployment.

Studio equipment conforming to BT.1361 is not yet deployed, and is not anticipated for several years. Wide-gamut acquisition and production equipment will begin to replace film over the next decade or so; however, wide-gamut consumer displays are not expected in that time frame. When these begin to be deployed, it is unlikely that they will all have the same gamut; electronics associated with each display will have to process the colour signals according to the properties of each display. In the long term, gamut mapping strategies comparable to those in the desktop colour management community will have to be deployed.

$Y'C_BC_R$ from studio RGB

In studio equipment, 8-bit $R'G'B'$ components usually use the same 219 excursion as the luma component of $Y'C_BC_R$. To encode $Y'C_BC_R$ from $R'G'B'$ in the range [0…219] using 8-bit binary arithmetic, scale the encoding matrix of Equation 30.6 by $^{256}/_{219}$:

Eq 30.8

$$\begin{bmatrix} {}^{709}_{219}Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix} 54.426 & 183.091 & 18.483 \\ -30.000 & -100.922 & 130.922 \\ 130.922 & -118.918 & -12.005 \end{bmatrix} \cdot \begin{bmatrix} {}_{219}R' \\ {}_{219}G' \\ {}_{219}B' \end{bmatrix}$$

To decode to $R'G'B'$ in the range [0…219] from BT.709 $Y'C_BC_R$ using 8-bit binary arithmetic:

Eq 30.9

$$\begin{bmatrix} {}_{219}R' \\ {}_{219}G' \\ {}_{219}B' \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 256 & 0 & 394.150 \\ 256 & -46.885 & -117.165 \\ 256 & 464.430 & 0 \end{bmatrix} \cdot \left(\begin{bmatrix} {}^{709}_{219}Y' \\ C_B \\ C_R \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \right)$$

$Y'C_BC_R$ from computer *RGB*

In computing it is conventional to use 8-bit $R'G'B'$ components, with no headroom or footroom: Black is at code 0 and white is at 255. To encode $Y'C_BC_R$ from

$R'G'B'$ in the range [0…255] using 8-bit binary arithmetic, the matrix of Equation 30.6 is scaled by $^{256}\!/_{255}$:

Eq 30.10

$$\begin{bmatrix} {}_{219}^{709}Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix} 46.742 & 157.243 & 15.874 \\ -25.765 & -86.674 & 112.439 \\ 112.439 & -102.129 & -10.310 \end{bmatrix} \cdot \begin{bmatrix} {}_{255}R' \\ {}_{255}G' \\ {}_{255}B' \end{bmatrix}$$

To decode $R'G'B'$ in the range [0…255] from BT.709 $Y'C_BC_R$ using 8-bit binary arithmetic:

Eq 30.11

$$\begin{bmatrix} {}_{255}R' \\ {}_{255}G' \\ {}_{255}B' \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 298.082 & 0 & 458.942 \\ 298.082 & -54.592 & -136.425 \\ 298.082 & 540.775 & 0 \end{bmatrix} \cdot \left(\begin{bmatrix} {}_{219}^{709}Y' \\ C_B \\ C_R \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \right)$$

Conversions between HD and SD

The differences among the EBU, SMPTE, and BT.709 primaries are negligible for practical purposes. New equipment should be designed to BT.709. Also, SD and HD have effectively converged to the transfer function specified in BT.709. Consequently, $R'G'B'$ coding uses essentially identical parameters worldwide, for SD and HD. (The sRGB standard for desktop computing uses the primaries of BT.709, but uses a different transfer function.)

Unfortunately, as I have mentioned, the luma coefficients differ dramatically between SD and HD. This wouldn't matter if HD systems were isolated! However, in practice, SD is upconverted and HD is downconverted, both at the studio and at consumers' premises. Serious colour reproduction errors arise if differences among luma coefficients are not taken into account in conversions.

In principle, downconversion can be accomplished by decoding $^{709}Y'C_BC_R$ to $R'G'B'$ using a suitable 3×3 matrix (such as that in Equation 30.7, on page 373), then encoding $R'G'B'$ to $^{601}Y'C_BC_R$ using another 3×3 matrix (such as that in Equation 29.6, on page 364).

The two 3×3 matrices can be combined so that the conversion can take place in one step:

Eq 30.12

$$\begin{bmatrix} {}^{601}_{219}Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 1 & 0.099312 & 0.191700 \\ 0 & 0.989854 & -0.110653 \\ 0 & -0.072453 & 0.983398 \end{bmatrix} \bullet \begin{bmatrix} {}^{709}_{219}Y' \\ C_B \\ C_R \end{bmatrix}$$

Equations 30.12 and 30.13 are written without interface offsets of +16 for luma and +128 for C_B and C_B: If the offsets are present, remove them, transform, then reapply them.

In the first row of the matrix, the coefficient 0.099312 adds about one-tenth of BT.709's C_B into BT.601's luma. This is a consequence of BT.709's blue luma coefficient being just 0.0722, compared to 0.114 for BT.601. The coefficient 0.1917 adds about one-fifth of BT.709's C_R into BT.601's luma; this is a consequence of BT.709's red luma coefficient being 0.2126, compared to 0.299 for BT.601. Clearly, failure to perform this colour transform produces large colour errors.

To convert from SD to HD, the matrix of Equation 30.12 is inverted:

Eq 30.13

$$\begin{bmatrix} {}^{709}_{219}Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 1 & -0.115550 & -0.207938 \\ 0 & 1.018640 & 0.114618 \\ 0 & 0.075049 & 1.025327 \end{bmatrix} \bullet \begin{bmatrix} {}^{601}_{219}Y' \\ C_B \\ C_R \end{bmatrix}$$

Unfortunately, to upconvert or downconvert a subsampled representation such as 4:2:2 or 4:2:0 requires chroma interpolation, colour transformation, then chroma subsampling. This is computationally intensive.

Colour coding standards

ITU-R Rec. BT.709 defines $Y'P_BP_R$ for component analog HD and $Y'C_BC_R$ for component digital HD. The parameters of $Y'P_BP_R$ and $Y'C_BC_R$ for the 1280×720 and 1920×1080 systems are defined by the SMPTE standards cited below.

ITU-R Rec. BT.709, *Basic parameter values for the HDTV standard for the studio and for international programme exchange.*

SMPTE ST 274, *1920×1080 Scanning and Analog and Parallel Digital Interfaces for Multiple Picture Rates.*

SMPTE ST 296, *1280×720 Progressive Image Sample Structure – Analog and Digital Representation and Analog Interface.*

Video signal processing

31

This chapter presents several diverse topics concerning the representation and processing of video signals.

It is ubiquitous in modern computers that integer arithmetic is implemented using the two's complement representation of binary numbers. When the result of an arithmetic operation such as addition or subtraction overflows the fixed bit depth available, two's complement arithmetic ordinarily involves wrapping around – for example, in 16-bit two's complement, taking the largest positive number, 32,767 (or in hexadecimal, $\mathtt{7fff}_h$) and adding one produces the smallest negative number, –32,768 (or in hexadecimal, 8000_h). It is an insidious problem with computer software implementation of video algorithms that wraparound is allowed in integer arithmetic. In video signal processing with integer values, *saturating arithmetic* must be used.

Edge treatment

If an image row of 720 samples is to be processed through a 25-tap FIR filter (such as that of Figure 20.26, on page 216) to produce 720 output samples, any output (result) sample within 12 samples of the left edge or the right edge of the image row will have nonzero filter coefficients associated with input samples beyond the edge of the image.

One approach to this problem is to produce just those output samples – 696 in this example – that can be computed from the available input samples. However, filtering operations are frequently cascaded, particularly in the studio, and it is unacceptable to repeatedly narrow the image width upon application of a sequence of FIR filters. A strategy is necessary to deal with filtering at the edges of the image.

Edge-replication is appropriate for motion-compensated interpolation in video compression: The replicated samples are used as predictions, and are not displayed.

Many digital image-processing (DIP) textbooks suggest padding the area outside the pixel array with copies of the edge samples, replicated as many times as necessary. The assumption is unrealistic for virtually all imaging applications, because if a *small* feature happens to lie at the left edge of the image, upon replication it will effectively turn into a *large* feature and thereby exert undue influence on the filter result – that is, exert undue influence reaching into the interior of the pixel array.

Some textbooks advocate padding the image by mirroring as many left-edge samples as necessary. In the example above, padding would mirror the leftmost 12 image columns. This approach is also unrealistic: In general-purpose imaging, there is no reasonable possibility that the missing content is estimated by mirroring.

Many textbooks consider the image to wrap in a cylinder: Missing samples outside the left-hand edge of the image are copied from the *right-hand* edge of the image! This concept draws from Fourier transform theory, where a finite data set is treated as being cyclic (periodic). This assumption makes the math easy, but is not justified in practice, and the wrapping strategy is even worse than edge-pixel replication.

In video, we treat the image as lying on a field of black: Unavailable samples are taken to be black. With this strategy, repeated lowpass filtering causes the implicit black background to intrude to some extent into the image. In practice, few problems are caused by this intrusion. Video image data nearly always includes some black (or blanking) samples, as I outlined in the discussion of samples per picture width and samples per active line. (See *Scanning parameters,* on page 86.) In studio standards, a region lying within the pixel array is designated as the *clean aperture*, as sketched in Figure 8.4, on page 87. This region is supposed to remain subjectively free from artifacts that originate from filtering at the picture edges.

Transition samples

In *Scanning parameters,* on page 86, I mentioned that it is necessary to avoid an instantaneous transition from

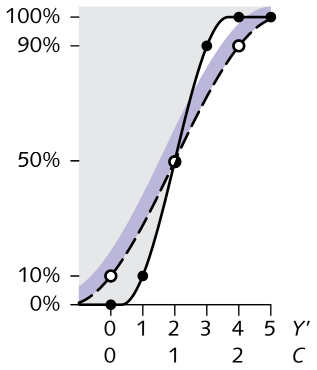


Figure 31.1 **Transition samples.** The solid line, dots (●), and light shading show the luma transition; the dashed line, open circles (○), and colour shading show 4:2:2 chroma limits.

blanking to picture at the start of a line. It is also necessary to avoid an instantaneous transition from picture to blanking at the end of a line. In studio video, the first and the last few active video samples on a line are blanking transition samples. I recommend that the first luma (*Y′*) sample of a line be black, and that this sample be followed by three transition samples clipped to 10%, 50%, and 90% of the full signal amplitude. In 4:2:2, I recommend that the first three colour difference (*C*) samples on a line be transition samples, clipped to 10%, 50%, and 90%. Figure 31.1 sketches the transition samples. The transition values should be applied by clipping, rather than by multiplication, to avoid disturbing the transition samples of a signal that already has a proper blanking transition.

Picture lines

480*i* studio standards historically accommodated up to 487 image rows, as explained in *480i line assignment*, on page 446. 576*i* studio standards provide 574 full lines and two halflines, as explained in *576i line assignment*, on page 458.

Historically, the count of image rows in 480*i* systems was poorly standardized. Various standards specified between 480 and 487 “picture lines.” It is pointless to carry picture on line 21/284 or earlier, because in NTSC transmission this line is reserved for closed caption data: 482 full lines, plus the bottom halfline, now suffice. With 4:2:0 chroma subsampling, as used in JPEG, MPEG-1, and MPEG-2, a multiple of 16 picture lines is required. DCT-based transform compression is now so ubiquitous that a count of 480 lines has become *de rigeur* for 480*i* MPEG video. In 576*i* scanning, a rigid standard of 576 picture lines has always been enforced; fortuitously for MPEG in 576*i*, the number 576 happens to be a multiple of 16.

MPEG-2 accommodates the 1920×1080 image format; however, 1080 is not a multiple of 16. In MPEG-2 coding, the bottom of each 1920×1080 picture is padded with eight image rows containing black to form a 1920×1088 array that is coded. The extra eight lines are discarded upon decoding.

Traditionally, the image array of 480*i* and 576*i* systems had halflines, as sketched in Figures 13.3 and 13.4 on page 132: Halfline blanking was imposed on picture information on the top and bottom lines of each frame. Neither JPEG nor MPEG provides halfline blanking: When halfline-blanked image data is presented to a JPEG or MPEG compressor, the blank image data is compressed. Thankfully, halflines have been abolished from HD.

Active lines (vertically) encompass the picture height. *Active samples* (horizontally) encompass not only the picture width, but also up to about a dozen blanking transition samples.

Studio video standards have no transition samples on the vertical axis: An instantaneous transition from vertical blanking to full picture is implied. However, nonpicture vertical interval information coded like video – such as VITS or VITC – may precede the picture lines in a field or frame. Active lines comprise only picture lines (and exceptionally, in 480*i* systems, closed caption data). L_A excludes vertical interval lines.

Computer display interface standards, such as those from VESA, make no provision for nonpicture (vertical interval) lines other than blanking.

Choice of S_{AL} and S_{PW} parameters

In *Scanning parameters,* on page 86, I characterized two video signal parameters, *samples per active line* (S_{AL}) and *samples per picture width* (S_{PW}). Active sample counts in studio standards have been chosen for the convenience of system design; within a given scanning standard, active sample counts standardized for different sampling frequencies are not exactly proportional to the sampling frequencies.

Historically, "blanking width" was measured instead of picture width. Through the decades, there has been considerable variation in blanking width of studio standards and broadcast standards. Also, blanking width was measured at levels other than 50%, leading to an unfortunate dependency upon frequency response.

HD standards specify that the 50%-points of picture width must lie no further than six samples inside the production aperture.

Most modern video standards do not specify picture width: It is implicit that the picture should be as wide as possible within the production aperture, subject to reasonable blanking transitions. Figure 13.1, on page 130 indicates S_{AL} values typical of studio practice.

For digital terrestrial broadcasting of 480*i* and 480*p*, the ATSC considered the coding of transition samples to be wasteful. Instead of specifying 720 S_{AL}, ATSC established 704 S_{AL}. This created an inconsistency between production standards and broadcast standards: MPEG-2 macroblocks are misaligned between the two.

Computer display interface standards, such as those from VESA, do not accommodate blanking transition samples and have no concept of clean aperture. In these standards, S_{PW} and S_{AL} are equal.

Video levels

I introduced 8-bit studio video levels on page 42. Studio video coding provides headroom and footroom. At an 8-bit interface, luma has reference black at code 16 and reference white at code 235; colour differences are coded in offset binary, with zero at code 128, the negative reference at code 16, and the positive reference at code 240. (It is a nuisance that the positive reference levels differ between luma and chroma.) I use the term *reference* instead of *peak;* the peaks of transient excursions may lie outside the reference levels. This range is known as "studio-swing."

All modern studio interfaces accommodate 10-bit signals, and most equipment today implements 10 bits. In 10-bit systems, the reference levels just mentioned are multiplied by 4; the two LSBs provide additional precision. Reference black is at code 64, and reference white at code 940.

Video levels in 480*i* systems are historically expressed in *IRE units*, sometimes simply called *units*. IRE refers to the Institute of Radio Engineers in the United States, the predecessor of the IEEE. Reference blanking level is defined as 0 IRE; reference white level is 100 IRE. The range between these values is the *picture excursion*.

In an analog interface, sync is coded at voltage level more negative than black; sync is "blacker than black." The ratio of picture excursion to sync amplitude is the *picture:sync ratio*. Two different ratios were standardized: 10:4 is predominant in 480*i* and computing; 7:3 is universal in 576*i* and HD, occasionally used in 480*i*, and rarely used in computing.

Setup (pedestal)

In 480*i systems* with setup, the term *picture excursion* refers to the range from blanking to white, even though strictly speaking the lowest level of the picture signal is 7.5 units, not 0 units.

In 480*i* composite NTSC video in North America, reference black is offset above blanking by 7.5% ($^3/_{40}$) of the picture excursion. *Setup* refers to this offset, expressed as a fraction or percentage of the picture excursion. In a 480*i* system with setup, there are nominally 92.5 IRE units from black to white.

Back porch is described in *Analog horizontal blanking interval*, in Chapter 2 of *Composite NTSC and PAL: Legacy Video Systems*.

Blanking level at an analog interface is established by a back porch clamp. However, in a system with setup, no signal element is present that enables a receiver to accurately recover black level. If an interface has poor

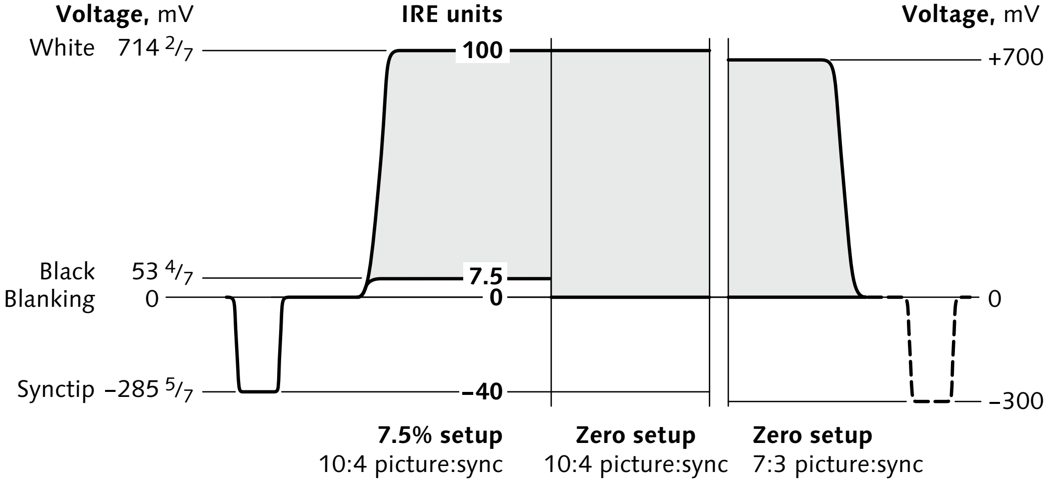


Figure 31.2 **Comparison of 7.5% and zero setup.** The left-hand third shows the video levels of composite 480*i* video, with 7.5% setup and 10:4 picture-to-sync ratio. This coding is used in some studio equipment and in most computer display interfaces. The middle third shows zero setup and 10:4 picture-to-sync, as used in 480*i* video in Japan. EBU N10 component video, 576*i* systems, and HD use zero setup, 700 mV picture, and 300 mV sync, as shown at the right.

tolerance, calibration error, or drift, setup causes problems in maintaining accurate black-level reproduction. Consequently, setup has been abolished from modern video systems: *Zero setup* is a feature of EBU N10 component video, all variants of 576*i* video, and HD. In all of these systems, blanking level also serves as the reference level for black.

Consumer video equipment purchased in Japan has zero setup. Some Japanese CE equipment has setup confugurable 0/7.5; sometimes the 0 setting goes by marketing terms like "enhanced black."

480*i* video in Japan originally used setup. However, in about 1987, zero setup was adopted; 10:4 picture-to-sync ratio was retained. Consequently, there are now three level standards for SD analog video interface. Figure 31.2 shows these variations.

The archaic EIA RS-343-A standard specified monochrome operation, 2:1 interlace with 60.00 Hz field rate, 7 μs horizontal blanking, and other parameters that have no place in modern video systems. Unfortunately, most PC graphics display standards have inherited RS-343-A's 10:4 picture-to-sync ratio and 7.5% setup. (Some high-end workstations have zero setup.)

The term *pedestal* refers to the absolute value of the offset from blanking level to black level, in IRE units or millivolts: Composite 480*i* NTSC incorporates a pedestal of 7.5 IRE. Pedestal includes any deliberate offset added to R', G', or B' components, to luma, or to a composite video signal, to achieve a desired technical or æsthetic intent. In Europe, this is termed *lift*. (I prefer the term *black level* to either *pedestal* or *lift*.)

BT.601 to computing

The coding difference between computer graphics and studio video necessitates image data conversion at the interface. Figure 31.3 overleaf shows the transfer function that converts 8-bit BT.601 studio $R'G'B'$ into computer $R'G'B'$. The footroom and headroom regions of BT.601 are clipped, and the output signal omits 36 code values. This coding difference between computer graphics and studio video is one of many challenges in taking studio video into the computer domain.

Enhancement

This section and several subsequent sections discuss enhancement, median filtering, coring, chroma transition improvement (CTI), and scan-velocity modulation (SVM). Each of these operations superficially resembles FIR filtering: A "window" involving a small set of neighboring samples slides over the input data. For each new input sample, the filtering operation delivers one output sample that has been subject to some fixed time delay with respect to the input. Unlike FIR filtering, with the exception of the most benign forms of enhancement these operations are nonlinear. They cannot, in general, be undone.

The term "enhancement" is widely used in image processing and video. It has no precise meaning. Evidently, the goal of enhancement is to improve, in some sense, the quality of an image. In principle, this can be done only with knowledge of the process or processes that degraded the image's quality. In practice, it is extremely rare to have access to any history of the processes to which image data has been subject, so no systematic approach to enhancement is possible.

In some applications, it may be known that image data has been subject to processes that have introduced specific degradations or artifacts. In these cases, enhancement may refer to techniques designed to reduce these degradations. A common example involves degraded frequency response due to aperture

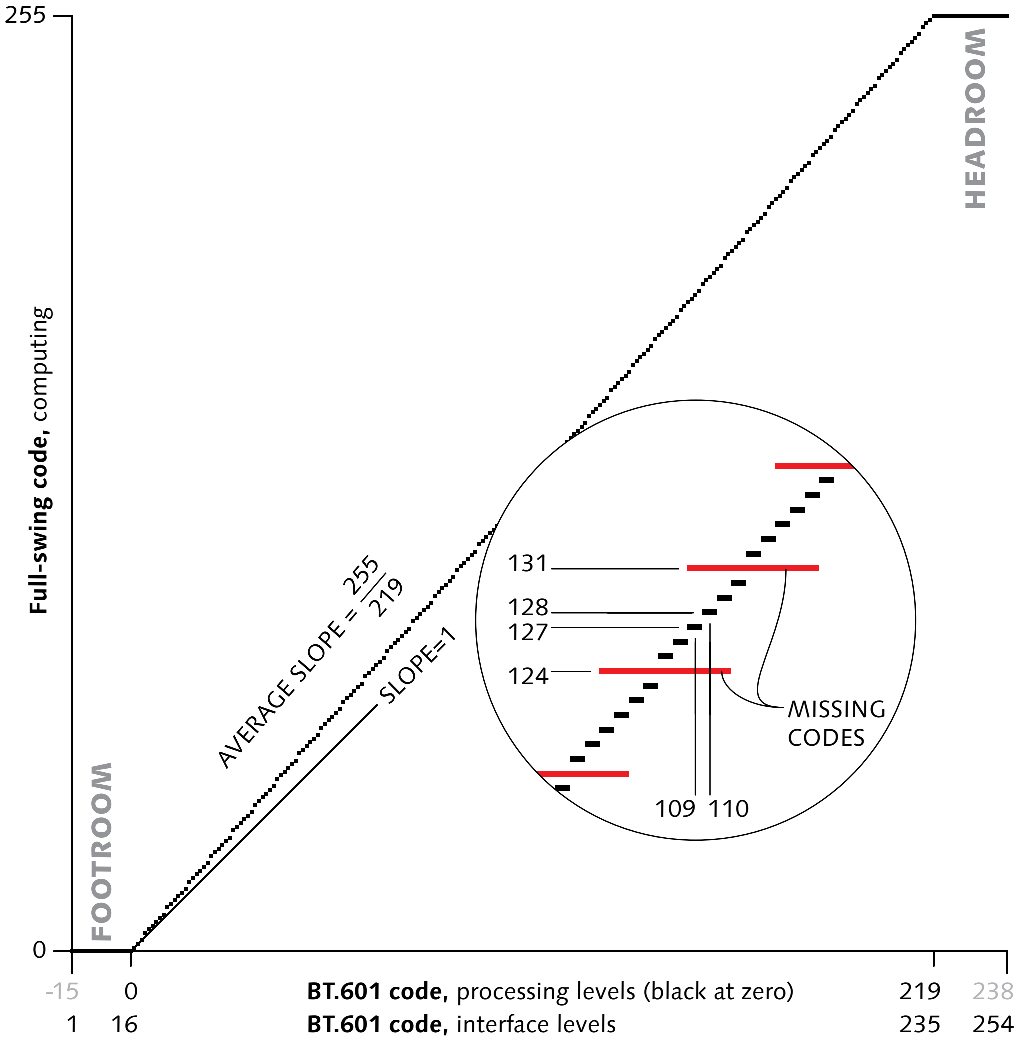


Figure 31.3 **The 8-bit BT.601 to full-range (computer) *R'G'B'* conversion** involves multiplying by a scale factor of $^{255}/_{219}$ (about 1.16), to account for the difference in range. This causes the footroom and headroom regions of the studio video signal to be clipped, and causes the output signal to be missing several code values. The detail shows the situation at mid-scale; the transfer function is symmetrically disposed around input pair [109, 110] and output pair [127, 128]. This graph shows a linear relationship from black to white. The linear relationship is suitable in computer systems where a ramp is loaded into the lookup table (LUT) between the framebuffer and the display; in that case, *R'G'B'* data is displayed on the computer display comparably to the way *R'G'B'* is displayed in video; see *Gamma,* on page 315.

effects. Enhancement in this case, also known as *aperture correction*, is accomplished by some degree of high-pass filtering, either in the horizontal direction, the vertical direction, or both. Compensation of loss of detail (MTF) should be done in the linear-light domain; however, it is sometimes done in the gamma-corrected domain. Historically, vertical aperture correction in interlaced tube cameras (vidicons and plumbicons) was done in the interlaced domain.

The SHARPNESS control in consumer receivers effects horizontal "enhancement" on the luma signal.

More generally, enhancement is liable to involve nonlinear processes that are based on some assumptions about the properties of the image data. Unless signal flow is extremely well controlled, there is a huge danger in using such operations: Upon receiving image data that has *not* been subject to the expected process, "enhancement" is liable to degrade the image, rather than improve it. For this reason, I am generally very strongly opposed to "enhancement."

Median filtering

In the rare case of an even-order median filter, the output is the average of the central two samples after sorting.

A median filter is a nonlinear filter in which each output sample is computed as the median value of the input samples under the window – that is, the result is the middle value after the input values have been sorted. Ordinarily, an odd number of taps is used. Median filtering often involves a horizontal window with 3 taps; occasionally, 5 or even 7 taps are used. Sometimes spatial median filters are used (for example, 3×3).

Any isolated extreme value, such as a large-valued sample due to impulse noise, will never appear in the output sequence of a median filter: Median filtering can be useful to reduce noise. However, a legitimate extreme value will not be included! I urge you to use great caution in imposing median filtering: If your filter is presented with image data whose statistics are not what you expect, you are very likely to degrade the image instead of improving it.

Coring

Coring is a technique widely presumed to reduce noise. The assumption (often incorrect) is made that any high-frequency signal components having low magnitude are noise. The input signal is separated into low- and high-frequency components using complementary filters. The

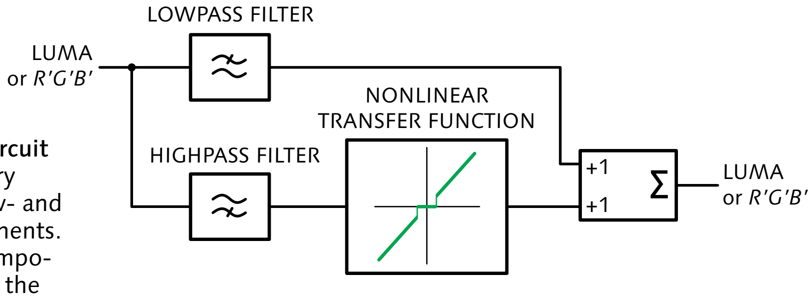


Figure 31.4 **A coring circuit** includes complementary filters that separate low- and high-frequency components. The high-frequency components are processed by the nonlinear transfer function in the sketch.

low-frequency component is passed to the output. The magnitude of the high-frequency component is estimated, and the magnitude is subject to a thresholding operation. If the magnitude is below threshold, then the high-frequency component is discarded; otherwise, it is passed to the output through summation with the low-frequency component. Coring can be implemented by the block diagram shown in Figure 31.4 above.

Like median filtering, coring depends upon the statistical properties of the image data. If the image is a flat-shaded cartoon having large areas of uniform colour with rapid transitions between them, then coring will eliminate noise below a certain magnitude. However, if the input is *not* a cartoon, you run the risk that coring will cause it to look like one! In a close-up of a face, skin texture produces a low-magnitude, high-frequency component that is *not* noise. If coring eliminates this component, the face will take on the texture of plastic.

Coring is liable to introduce spatial artifacts into an image. Consider an image containing a Persian carpet that recedes into the distance. The carpet's pattern will produce a fairly low spatial frequency in the foreground (at the bottom of the image); as the pattern recedes into the background, the spatial frequency of the pattern becomes higher and its magnitude becomes lower. If this image is subject to coring, beyond a certain distance, coring will cause the pattern to vanish. The viewer will perceive a sudden transition from the pattern of the carpet to no pattern at all. The viewer may conclude that beyond a certain distance there is a different carpet, or no carpet at all.

Chroma transition improvement (CTI)

Colour-under VCRs exhibit very poor colour difference bandwidth (evidenced as poor chroma resolution in the horizontal direction). A localized change in luma may be faithfully reproduced, but the accompanying change in colour difference components will be spread horizontally. If you assume that coloured areas tend to be uniformly coloured, one way of improving image quality is to detect localized changes in luma, and use that information to effect repositioning of colour difference information. Techniques to accomplish this are collectively known as *chroma transition improvement* (CTI).

If you use CTI, you run the risk of introducing excessive emphasis on edges. Also, CTI operates only on the horizontal dimension: Excessive CTI is liable to become visible owing to perceptible (or even objectionable) differences between the horizontal and vertical characteristics of the image. CTI works well on cartoons, and on certain other types of images. However, it should be used cautiously.

Mixing and keying

Mixing video signals together to create a transition, or a layered effect – for example, to mix or wipe – is called *compositing.* In America, a piece of equipment (with a control surface) that performs such effects is a *production switcher.* In Europe, the equipment – or the person that operates it! – is called a *vision mixer*.

Accomplishing mix, wipe, or key effects in hardware requires synchronous video signals – that is, signals whose timing matches perfectly in the vertical and horizontal domains.

Porter, Thomas, and Tom Duff (1984), "Compositing digital images," in *Computer Graphics,* **18** (3): 253–259 (July, Proc. SIGGRAPH). The terms *composite* and *compositing* are overused in video!

SMPTE RP 157, *Key Signals.*

Keying (or *compositing,* or *blending*) refers to superimposing a foreground (FG, or *fill video*) image over a background (BG) image. Keying is normally controlled by a *key* (or *matte*) signal, represented like luma, that indicates the opacity of the accompanying foreground image data, coded between black (0, fully transparent) and white (1, fully opaque). In computer graphics, the key signal (data) is called *alpha* (α), and the operation is called *compositing.*

Eq 31.1

$$R = \alpha \cdot FG + \left(1 - \alpha\right) \cdot BG$$

The keying (or compositing) operation is performed as in Equation 31.1. Foreground image data that has been premultiplied by the key is called *shaped* in video, or *associated, integral,* or *premultiplied* in computer graphics. Foreground image data that has not been premultiplied by the key is called *unshaped* in video, or *unassociated* or *nonpremultiplied* in computer graphics.

The key signal is sometimes called *linear key:* The modifier *linear* does not refer to linear light, but to a key signal representing opacity with more than just the two levels fully transparent and fully opaque. In keying or compositing, the compositing operation of Equation 31.1 is applied directly without any transfer function applied to the key signal.

Historically, keying was accomplished in the gamma domain. However, proper simulation of the physics of blending requires keying in the linear-light domain; such an approach is now widely practiced in software systems (sometimes by setting an option denoted something like "blend in gamma = 1.0 space"); an increasing number of hardware-based video switchers are now capable of linear-light blending.

The most difficult part of keying is extracting ("pulling") the matte. For review from a computer graphics perspective, see SMITH, ALVY RAY, and JAMES F. BLINN (1996), "Blue screen matting," in *Computer Graphics* (Proc. SIGGRAPH), 259–268.

The multiplication of foreground and background data in keying is equivalent to modulation: This can produce signal components above half the sampling rate, thereby producing alias components. Aliasing can be avoided by upsampling the foreground, background, and key signals; performing the keying operation at twice the video sampling rate; then suitably filtering and downsampling the result. Most keyers operate directly at the video sampling rate without upsampling or downsampling, and consequently exhibit some aliasing.

Figure 31.5 **This matte image example** shows the typical matte polarity. See CHUANG, YUNG-YU et al. (2002), "Video matting of complex scenes," in *ACM Transactions on Graphics* **21** (3) (Proc. SIGGRAPH), 243–248 (July).

Figure 31.5 is a matte image representative of work by Yung-Yu Chuang and his colleagues at the University of Washington.

In order for a compositing operation to mimic the mixing of light in an actual scene, keying should be performed on foreground and background in the linear-light domain. However, keying in video has historically been performed in the gamma-corrected domain.

Frame, field, line, and sample rates

32

This chapter outlines the field, frame, line, and sampling rates of 480*i* video, 576*i* video, and HD.

The standard sampling frequency for component SD is exactly 13.5 MHz. This rate produces an integer number of samples per line in both 480*i* and 576*i*. HD standards at integer frame rates specify multiples of 13.5 MHz: 720*p*, 1080*i*30, and 1080*p*30 systems sample at 74.25 MHz. HD standards also permit operation at $^{1000}/_{1001}$ times that rate.

Field rate

Television systems originated with field rates based on the local AC power line frequency: 60 Hz for North America, and 50 Hz for Europe.

In the 1940s and 1950s, coupling of the ripple of a receiver's power supply into circuitry – such as video amplifiers and high-voltage supplies – caused display luminance to pulsate. If the vertical scanning frequency was different from the power line frequency, interference caused artifacts called *hum bars*, at the difference in frequency – the beat frequency – between the two. Their visibility was minimized by choosing a field rate the same as the power line frequency, so as to make the hum bars stationary. There was no requirement to have an exact frequency match, or to lock the phase: As long as the pattern was stationary, or drifting very slowly, it was not objectionable. The power supply interactions that were once responsible for hum bars no longer exist in modern circuitry, but the vertical scan rates that were standardized remain with us.

A second reason to lock television scanning to power line frequency concerns image capture. Many light sources – for example, fluorescent lamps – flash at twice the power line frequency. If a camera operates at a picture rate unrelated to the flash rate, then the captured image is liable to contain artifacts owing to the flashing illumination. Various countermeasures can overcome these artifacts; however, the simplest approach to prevent them is to capture at a multiple or submultiple of the power line frequency.

Line rate

The total number of raster lines chosen for the 525-line television is the product of a few small integers: 525 is $7 \times 5^2 \times 3$. The choice of small integer factors arose from the use of vacuum tube divider circuits to derive the field rate from the line rate: Such dividers were stable only for small divisors. The total number of scan lines per frame is odd. Equivalently, the field rate is an odd multiple of half the line rate. The 2:1 relationship generates the 2:1 interlace that I introduced in *Interlaced format,* on page 88. These factors combined to give monochrome 525/60 television a line rate of $30 \times (7 \times 5^2 \times 3)$, or exactly 15.750 kHz.

$$\frac{525 \cdot 60}{2} = 15750$$

The flyback transformer scheme was the precursor to modern switched-mode power supplies (SMPS).

For 525-line receivers, a scheme was invented to develop high voltage for the picture tube using a transformer operating at the horizontal scanning frequency, 15.750 kHz, rather than the AC line frequency. This approach permitted a lightweight transformer, which became known as the *flyback transformer*. (The scheme is still used today; it can be considered as a precursor to the switch-mode power supply.) The flyback transformer was a complex component, and it was tuned to the horizontal frequency.

When European engineers started designing receivers, it was a practical necessity to fix the field rate at 50 Hz to match the local power line frequency. Rather than develop flyback transformers from scratch, European engineers imported them from North America! Horizontal frequency was thereby constrained to a narrow range around 15.750 kHz. The total line count was chosen as 625, that is, 5^4. Monochrome 625-line television had a line rate of 5^6, or exactly 15.625 kHz; that rate was unchanged with the addition of colour.

Sound subcarrier

In about 1941, the first NTSC recognized that visibility of sound-related patterns in the picture could be minimized by making the picture line rate and the sound subcarrier rest frequency coherent. In monochrome 525/60 television the sound subcarrier was placed at exactly $^{2000}/_{7}$ (i.e., $285\,^{5}/_{7}$) times the line rate – that is, at 4.5 MHz. Sound in conventional television is frequency modulated, and with an analog sound modulator even perfect silence cannot be guaranteed to generate an FM carrier of exactly 4.5 MHz. Nonetheless, making the FM sound carrier average out to 4.5 MHz was considered to have some value.

Addition of composite colour

NTSC and PAL colour coding both employ the frequency-interleaving technique to achieve compatibility with monochrome systems. With frequency interleaving, the colour subcarrier frequency is chosen to alternate phase line by line, so as to minimize the visibility of encoded colour on a monochrome receiver. The line-to-line phase relationship makes it possible to accurately separate chroma from luma in an NTSC decoder that incorporates a comb filter (although a cheaper notch filter can be used instead).

NTSC colour subcarrier

HAZELTINE CORPORATION (1956), *Principles of Color Television*, by the Hazeltine Laboratories staff, compiled and edited by KNOX MCILWAIN and CHARLES E. DEAN (New York: Wiley).

In 1953, the second NTSC decided to choose a colour subcarrier frequency of approximately 3.6 MHz. They recognized that any nonlinearity in the processing of the composite colour signal with sound – such as limiting in the *intermediate frequency* (IF) stages of a receiver – would result in intermodulation distortion between the sound subcarrier and the colour subcarrier. The difference, or *beat frequency,* between the two subcarriers, about 920 kHz, falls in the luminance bandwidth and could potentially have been quite visible.

The NTSC recognized that the visibility of this pattern could be minimized if the beat frequency was line-interlaced. Since the colour subcarrier is necessarily an odd multiple of half the line rate, the sound subcarrier had to be made an integer multiple of the line rate.

$$f_{\mathrm{SC,NTSC}} = \frac{455}{2} f_{\mathrm{H},480i}$$

The NTSC decided that the colour subcarrier should be exactly $^{455}/_{2}$ times the line rate. Line interlace of the beat could be achieved by increasing the sound-to-line rate ratio (previously $285\frac{5}{7}$) by the fraction $^{1001}/_{1000}$ to the next integer (286).

Setting broadcast standards in the U.S. was (and remains) the responsibility of the Federal Communications Commission. The FCC could have allowed the sound subcarrier rest frequency to be increased by the fraction $^{1001}/_{1000}$ – that is, increased by 4.5 kHz to about 4.5045 MHz. Had the FCC made this choice, the colour subcarrier in NTSC would have been exactly 3.583125 MHz; the original 525/60 line and field rates would have been unchanged; we would have retained exactly 60 frames per second – and NTSC would have no dropframes! Since sound is frequency modulated, the sound carrier was never crystal-stable at the subcarrier frequency anyway – not even during absolute silence – and the tolerance of the rest frequency was already reasonably large (±1 kHz). The deviation of the sound subcarrier was – and remains – 25 kHz, so a change of 4.5 kHz could easily have been accommodated by the intercarrier sound systems of the day.

$$525 \times \left(\frac{60}{2}\,\text{Hz}\right) \times \frac{1000}{1001} \times \frac{455}{2}$$
$$= \frac{315}{88}\,\text{MHz}$$
$$\approx 3.5795\overline{45}\,\text{MHz}$$

However, the FCC refused to alter the sound subcarrier. Instead, the colour/sound constraint was met by reducing both the line rate and field rate by the fraction $^{1001}/_{1000}$, to about 15.734 kHz and 59.94 Hz. The colour subcarrier was established as $3.5795\overline{45}$ MHz. What was denoted 525/60 scanning became 525/59.94, though unfortunately the 525/60 notation is still used loosely to refer to 525/59.94.

$$525 \times \left(\frac{60}{2}\,\text{Hz}\right) \times \frac{1000}{1001} \times 455 \times 2$$
$$= \frac{315}{22}\,\text{MHz}$$
$$\approx 14.3181\overline{81}\,\text{MHz}$$

Prior to the emergence of framestore synchronizers in the 1980s, every major broadcast network in the United States had an atomic clock to provide 5 MHz, followed by a rate multiplier of $^{63}/_{22}$ to derive its master $14.3181\overline{81}$ MHz clock.

The factors of 1001 are 7, 11, and 13. This numerical relationship was known in ancient times: The book *1001 Arabian Nights* is based on it. The numbers 7, 11, and 13 are considered to be very unlucky. Unfortunately the field rate of $^{60}/_{1.001}$, about 59.94 Hz, means that 60 fields consume slightly more than one second: Counting 30 fields per second does not agree with clock time. Dropframe timecode was invented to alleviate this difficulty; see *Timecode,* on page 399.

NTSC sync generators historically used a master oscillator of $14.3181\overline{81}$ MHz. This clock was divided by 4 to obtain the colour subcarrier, and simultaneously divided by 7 to obtain a precursor of line rate.

576*i* PAL colour subcarrier

In 576*i* PAL, the colour subcarrier frequency is based on an odd multiple of one-quarter the line rate, using the factor $^{1135}\!/_{4}$. The odd multiple of one-quarter, combined with the line-to-line alternation of the phase of the *V* colour difference component, causes the *U* and *V* colour components to occupy separate parts of the composite signal spectrum. This makes the PAL signal immune to the hue errors that result when an NTSC signal is subject to differential phase distortion.

$$625 \times \left(\frac{50}{2}\,\mathrm{Hz} \right) \times \left(\frac{1135}{4} + \frac{1}{625} \right)$$
$$= 4.433618750\ \mathrm{MHz}$$

In standard PAL-B, PAL-G, PAL-H, and PAL-I, an offset of +25 Hz is added to the basic subcarrier frequency so as to minimize the visibility of the Hanover bar effect. The 25 Hz offset means that the phase relationship of subcarrier to horizontal advances exactly +0.576° each line. Consequently, subcarrier-locked sampling in PAL is not line-locked: The subcarrier phase, modulo 90°, of vertically aligned samples is not identical! The introduction of the +25 Hz offset destroyed the simple integer ratio between subcarrier and line rate: The ratio is quite complex, as shown in the margin. The prime factor 64,489 is fairly impenetrable to digital techniques.

$$\frac{1135}{4} + \frac{1}{625} = \frac{709379}{2500} = \frac{11 \times 64489}{2^2 \times 5^3}$$

$4f_{SC}$ sampling

The earliest digital television equipment sampled composite NTSC or PAL video signals. It was convenient for composite digital NTSC equipment to operate at a sampling frequency of exactly four times the colour subcarrier frequency, or about 14.318 MHz, denoted $4f_{SC}$.

Any significant processing of a picture, such as repositioning, resizing, rotating, and so on, requires that the signal be represented in components. For this reason, component video equipment is preferred in production and postproduction. But $4f_{SC}$ equipment has half the data rate of BT.601 equipment; $4f_{SC}$ equipment is cheaper than component equipment, and dominated SD broadcast operations for many years.

Sampling NTSC at $4f_{SC}$ gives 910 samples per total line (S_{TL}). A count of 768 samples (3×2^8) encompasses the active samples of a line, including the blanking transitions. A count of 512 (2^9) lines is just slightly more than the number of nonblanked lines in 480*i* scanning.

The numbers 768 and 512 were convenient for early memory systems: 512 is the ninth power of 2, and 768 is 3 times the eighth power of 2. In the early days of digital television, this combination – 768 and 512 – led to very simple memory and addressing circuits for framestores. The importance of this special combination of 768 and 512 is now irrelevant: Framestore systems today have well ovver a single frame of memory; memory devices have much higher capacities; and total memory capacity is now a more important constraint than active sample and line counts. In any case, the binary numbers 768 and 512 were never any help in the design of 576*i* framestores.

Common sampling rate

The designers of the NTSC and PAL systems chose video parameters based on simple integer ratios. When component digital sampling became feasible it came as a surprise that the ratio of line duration of 480*i* and 576*i* systems turned out to be the ratio of 144 to 143, derived as shown in Table 32.1.

$f_{\mathrm{H},480i}$	:	$f_{\mathrm{H},576i}$
$525 \times \frac{60}{2} \times \frac{1000}{1001}$	:	$625 \times \frac{50}{2}$
$7 \times 5^2 \times 3 \times \frac{5 \cdot 3 \cdot 2^2}{2} \times \frac{5^3 \cdot 2^3}{13 \cdot 11 \cdot 7}$	:	$5^4 \times \frac{5^2 \times 2}{2}$
$3 \times 3 \times 2^4$	:	13×11
144	:	143

Table 32.1 **Derivation of 13.5 MHz common sampling rate**

The lowest common sampling frequency corresponding to these factors is 2.25 MHz, half of the now-familiar NTSC sound subcarrier frequency of 4.5 MHz. Any multiple of 2.25 MHz could have been used as the basis for line-locked sampling of both 480*i* and 576*i*. The most practical sampling frequency is 6 times 2.25 MHz, or 13.5 MHz; this multiplier is a compromise between a rate high enough to ease the design of analog antialiasing filters and low enough to minimize data rate and memory requirements.

ITU-R Rec. BT.601-5, *Studio encoding parameters of digital television for standard 4:3 and widescreen 16:9 aspect ratios*.

At 13.5 MHz, 480*i* video has 858 samples per total line, and 576*i* video has 864 S_{TL}. The blanking tolerances between NTSC and PAL accommodated a choice of 720 samples per active line (S_{AL}) in both systems. Standardization of this number of active samples resulted in a high degree of commonality in the design of SD video processing equipment, since only the difference in active line counts needed to be accommodated to serve both 50 Hz and 60 Hz markets. Also the technically difficult problem of standards conversion was eased somewhat with a common sampling frequency, since horizontal interpolation became unnecessary. However, blanking had to be treated differently in the two systems to meet studio interchange standards.

Numerology of HD scanning

Figure 32.1 gives a graphic representation of the development of the magic numbers in HD. At the upper left is the AC power line frequency in North America, along with the factors of 525 (all small integers: $7 \cdot 5^2 \cdot 3$). Next to that is indicated the AC power frequency in Europe, and the factors of 625 (also, all small integers: 5^4).

The addition of colour to the NTSC system introduced the ratio $^{1000}/_{1001}$, and led to the 525/59.94 system.

Figure 32.1 indicates 575 image rows in 625/50 systems; this constitutes 287 full lines, plus a halfline, in each field. Counting each halfline as a full line, the total is 576.

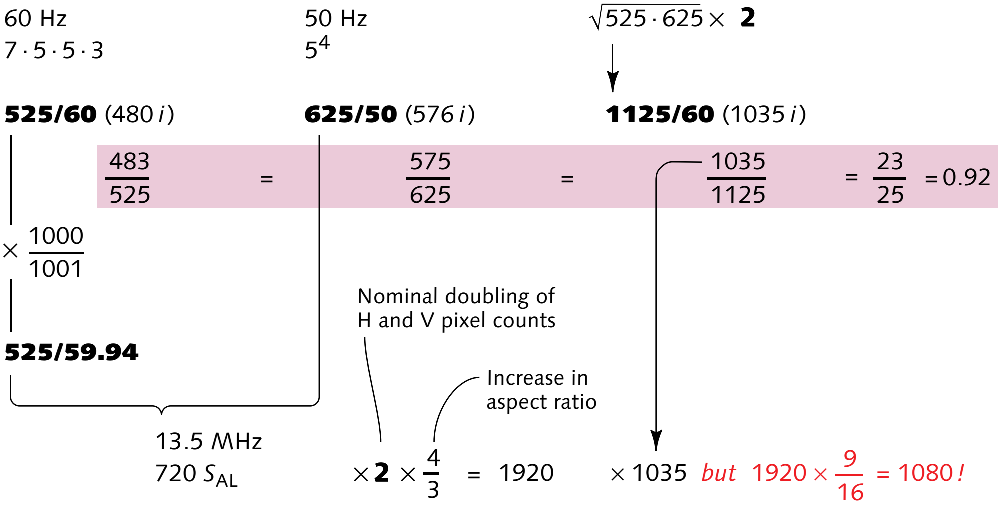


Figure 32.1 **Numerology of HD scanning**

Incidentally, $^{2000}/_{1125}$ equals $^{16}/_{9}$.

HD was originally conceived at NHK as having twice the horizontal and twice the vertical resolution of conventional television: At the top right is the conceptual origin of the total number of HD scanning lines as twice the geometric mean of 525 and 625. North America would have preferred twice 525 and Europe twice 625. The designers choose a total line count of 1125 (i.e., $5^3 \times 3^2$), a compromise that was thought to be politically acceptable on both sides of the Atlantic Ocean.

Underneath the scanning designations 525/60, 625/50, and 1125/60 in Figure 32.1 is a grey bar containing the ratio of image rows to total scanning lines in each system. The count of lines per total vertical (L_T) for each of these systems is the fraction $^{23}/_{25}$ (92%) of the total. This led to NHK's original choice of 1035 image rows for 1125/60 HD.

The desire for a common sampling frequency for component digital video led to the synthesis of line rates of 480*i* and 576*i* into a common sampling frequency, 13.5 MHz, and a common count of samples per active line (S_{AL}), 720. For HD, the active pixel count was doubled to increase the horizontal resolution; then multiplied by the $^4/_3$ increase in aspect ratio (from 4:3 to 16:9), netting 1920.

An image array having dimensions 1920×1035 results from these choices, and SMPTE standardized that as 240M in 1988. However, in about 1991 it became clear that the 1920×1035 structure had a sample pitch unequal in the horizontal and vertical dimensions – nonsquare sampling. The degree of inequality was small – just 4% – but for many applications any departure from equal spacing imposes a burden. In about 1995, the standard was adapted to achieve square sampling by choosing a count of image rows $^9/_{16}$ times 1920, that is, 1080 rows. SMPTE, and subsequently ATSC, enshrined square sampling in the 1920×1080 image array. The system has about two million pixels per frame; the exact number is very slightly less than 2^{21}, a neat fit into binary-sized memory components.

NHK planned to operate the 1920×1035 system at a frame rate of exactly 30 Hz ("30.00 Hz") and early 1035*i* equipment operated only at that rate. However, the discrepancy of about one frame time every 33.367 seconds between 1035*i*30.00 and 480*i*29.97 is a big nuisance in standards conversion. To ease this problem, and to ease engineering difficulties associated with digital audio sample rates, 1080*i* HD standards accommodate both 29.97 Hz and 30 Hz frame rates.

While NHK and others were developing 1125/60 interlaced HD, progressive-scan systems having nearly identical pixel rate were being developed by other organizations, mainly in the United States. The technology of the day permitted a pixel rate of about 60 megapixels per second, whether scanning was interlace or progressive. With interlace scanning, 60 Mpx/s at 30 Hz frame rate allows a two-megapixel image structure. With progressive scanning, 60 Mpx/s at 60 Hz frame rate allows just one megapixel. Partitioning one megapixel into a square lattice yields an image structure of 1280×720; this led to the 720*p* family of standards.

In the mid-2000s, the digital cinema community took advantage of HD equipment and infrastructure, and adopted 1080 image rows for the "2 K" standard. However, 2048 image columns were chosen. That choice produced a new aspect ratio, about 1.896, never before used for movies. The standard 1.85 cinema aspect ratio would have been achieved with 1998 image columns, but apparently some members of the community were fearful that 1998 could not be claimed to be "2 K." The choice of a number somewhat greater than 1920 seems to be motivated by the short term desire to distinguish digital cinema from HD, politically if not technically. The 64 additional pixels on each edge can hardly be argued as increasing resolution. Super-HD has been demonstrated with $2 \cdot 1920$ or 3840 image rows, but "4 K" D-cinema has 4096 image columns. Ultra-HD is proposed with $4 \cdot 1920$ or 7680 image rows but presumably "8 K" D-cinema will offer 8192. It seems to me that the divergence of the $2^K \cdot 1920$ HD-related image formats and the power-of-two D-cinema formats can't be sustained. HD formats, leveraging a connection to consumer volumes, are likely to win in the end.

Audio rates

Digital audio has two standard sample rates: 48 kHz, for professional applications, and 44.1 kHz, for consumer applications. In the standardization of digital audio, manufacturers decided to adopt two different standards in order that professional and consumer equipment could be differentiated! That goal failed miserably, and now the dichotomy in sample rates is a major nuisance in video and audio production.

$$3 \cdot 588 \cdot 25\ \text{Hz} = 44100\ \text{Hz}$$

$$3 \cdot 490 \cdot \frac{30\ \text{Hz}}{1.001} = \frac{44100}{1.001}\ \text{kHz} \approx 44.056\ \text{kHz}$$

The 44.1 kHz sampling rate for consumer digital audio originated from an early PCM recording system (Sony PCM-1600) that recorded three (16-bit) stereo sample pairs on 588 active lines of a 625/50 videocassette recorder. In 525/59.94 countries, the original rate was $^{44100}/_{1.001}$: Three 16-bit stereo sample pairs were recorded on each of 490 active lines per frame. Eventually, the 44.1 kHz rate was standardized worldwide. In $^{1}/_{50}$ s, there are 882 samples, an integer. In $^{1}/_{59.94}$ s, there are exactly 1471.47 samples: the noninteger ratio causes havoc.

The professional audio sampling rate was chosen to be exactly 48 kHz. The time interval of one video picture at 50 fields or frames per second corresponds to exactly 960 audio samples at 48 kHz. An AES/EBU audio frame comprises 192 left/right sample pairs, with 16 bits in each sample: In 50 Hz video standards, a video frame occupies exactly the same time interval as five audio frames. In video at 59.94 fields or frames per second, the timing relationships are unfortunate. There are $1601\,^{3}/_{5}$ audio sample intervals in one picture time: This noninteger number is very inconvenient. The timing relationship of audio samples to video pictures aligns just once every five pictures. There are $^{1001}/_{240}$ (i.e., $4.17083\overline{3}$) AES/EBU audio frames in a video picture time.

Timecode 33

This chapter gives technical details concerning timecode, as used in video, film, audio recording, editing, and sequencing equipment.

Introduction

Timecode systems assign a number to each frame of video to allow each frame to be uniquely identified. Time data is coded in binary-coded decimal (BCD) digits in the form HH:MM:SS:FF, in the range 00:00:00:00 to 23:59:59:29. There are timecode variants for 24, 25, 29.97, and 30 frames per second. Timecode data is digitally recorded with the associated image. *Burnt-in timecode* (BITC) refers to a recording with timecode numbers keyed over the picture content (that is, embedded in the image data).

SMPTE ST 262, *Binary Groups of Time and Control Codes – Storage and Transmission of Data Control Codes.*

SMPTE RP 169, *Television, Audio and Film Time and Control Code – Auxiliary Time Address Data in Binary Group – Dialect Specification of Directory Index Locations.*

In addition to approximately 32 bits required for eight-digit time data, timecode systems accommodate an additional 32 *user bits* per frame. User bits may convey one of several types of information: a second timecode stream (such as a timecode from an original recording); a stream of ASCII/ISO characters; motion picture production data, as specified in SMPTE ST 262; auxiliary BCD numerical information, such as tape reel number; or nonstandard information. A group of 4 user bits is referred to as a *binary group*. The information portion of timecode thus totals 64 bits per frame.

A number of synchronization bits are appended to the 64 information bits of timecode in order to convey timecode through a recording channel. Sixteen synchronization bits are appended to form 80-bit *linear timecode* (LTC). Eighteen sync bits and 8 CRC bits are appended to form 90-bit *vertical interval timecode* (VITC) that can be inserted into a video signal.

No BCD digit can contain all ones, so the all-ones code is available for other purposes. The highest possible value of certain tens digits is less than 8, so the high-order bits of certain timecode digits are available for use as flags; these flag bits are described on page 403.

The colourframe flag is asserted when the least significant bit of the timecode frame number is intentionally locked to the colourframe sequence of the associated video – in 480*i* systems, locked to Colourframes A and B of SMPTE 170M.

See *NTSC two-frame sequence*, in Chapter 7 of *Composite NTSC and PAL: Legacy Video Systems*.

Dropframe timecode

In 25 Hz video, such as in 576*i* video systems, and in 24 Hz film, there is an exact integer number of frames in each second. In these systems, timecode has an exact correspondence with clock time.

During the transition from monochrome to colour television in the United States, certain interference constraints needed to be satisfied among the horizontal scanning, sound, and colour frequencies. These constraints were resolved by reducing the 60.00 Hz field rate of monochrome television by a factor of exactly $^{1000}/_{1001}$ to create the colour NTSC field rate of about 59.94 Hz. This leads to a noninteger number of frames per second in 29.97 Hz or 59.94 Hz systems. The *dropframe* (DF) mechanism can be used to compensate timecode to obtain a very close approximation to clock time. Dropframes are not required or permitted when operating at exact integer numbers of frames per second. Dropframe timecode is optional in 29.97 Hz or 59.94 Hz systems; operation with a straight counting sequence is called *nondropframe* (NDF).

See *Frame, field, line, and sample rates*, on page 389.

In consumer 480*i*29.97 DV, dropframe timecode is mandatory.

Counting frames at the NTSC frame rate of 29.97 Hz is slower than realtime by the factor $^{1000}/_{1001}$, which – in nondropframe code – would result in an apparent cumulative error of about +3.6 seconds in an hour. To make timecode correspond to clock time, approximately once every 1000 frames a frame number is dropped – that is, omitted from the counting sequence. Of course, it is only the *number* that is dropped, not the video frame! Frame numbers are dropped in pairs in

The final field in an hour of 29.97 Hz video has DF code *hh*:59:59;29 and NDF code *hh*:59:56:23.

order to maintain the relationship of timecode (even or odd frame number) to NTSC colourframe (A or B).

Dropping a pair of frames every $66\frac{2}{3}$ seconds – that is, at an interval of 1 minute, 6 seconds, and 20 frames – would result in dropping the codes indicated in Figure 33.1 in the margin. Although this sequence is not easily recognizable, it repeats after exactly ten minutes! This is a consequence of the ratios of the numbers: Two frames in 2000 accumulates 18 frames in 18000, and there are 18000 intervals of $\frac{1}{30}$ second in 10 minutes (30 frames, times 60 seconds, times 10 minutes). To produce a sequence that is easy to compute and easy to remember, instead of dropping numbers strictly periodically, this rule was adopted: *Drop frame numbers 00:00 and 00:01 at the start of every minute, except the tenth minute.* In effect, a dropped pair that is due is delayed until the beginning of the next minute.

hh:mm: ss: ff
xx:x0:00:00
xx:x1:06:20
xx:x2:13:10
xx:x3:20:00
xx:x4:26:20
xx:x5:33:10
xx:x6:40:00
xx:x7:46:20
xx:x8:53:10

Figure 33.1 **Periodic dropped timecode numbers**

Figure 33.2 depicts the convention that has emerged to represent field identification and of the use of dropframe code in timecode displays.

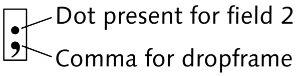


Figure 33.2 **Timecode as displayed,** or represented in ASCII, has the final delimiter (separating seconds from frames) selected from colon, semicolon, period, or comma, to indicate dropframe code and field 2.

Dropframe does not achieve a perfect match to clock time, just a very good match: Counting dropframe code at $\frac{30}{1.001}$ frames per second results in timecode that is about 86.4 ms late (slow) over 24 hours. If the residual error were to accumulate, after 11 or 12 days timecode would fall about one second later than clock time. If a timecode sequence is to be maintained longer than 24 hours, timecode should be jammed daily to reference clock time at a suitable moment. No standard recommends when this should take place; however, the usual technique is to insert duplicate timecode numbers 00:00:00;00 and 00:00:00;01. Editing equipment treats the duplicate codes as a timecode interruption.

Editing

SMPTE ST 258, *Television – Transfer of Edit Decision Lists.*

Timecode is basic to video editing. An edit is denoted by its *in point* (the timecode of the first field or frame to be recorded) and its *out point* (the timecode of the first field or frame beyond the recording). An edited sequence can be described by the list of edits used to produce it: Each entry in an *edit decision list* (EDL) contains the in and out points of the edited material, the in and out points of the source, and tape reel number or other source and transition identification.

An edited program is ordinarily associated with continuous "nonbroken" timecode. Editing equipment historically treated the boundary between 23:59:59:29 and 00:00:00:00 as a timecode discontinuity, and an edit at that point (such as starting a new program) was problematic. Consequently, it is conventional to start a main program segment with timecode 01:00:00:00. On videotape, it was conventional to include 1.5 minutes of "bars and tone" leader starting at timecode 00:58:30:00. (In Europe, it is common to start a program at 10:00:00:00.)

Linear timecode (LTC)

Timecode was historically recorded on studio videotape and audiotape recorders on longitudinal tracks having characteristics similar or identical to those of audio tracks. This became known as *longitudinal timecode* (LTC). The word *longitudinal* became unfashionable, and LTC was renamed *linear timecode* – thankfully, it retains its acronym. LTC was historically transported in the studio as an audio signal pair interfaced through three-pin XLR connector.s

Vertical interval timecode (VITC)

SMPTE RP 164, *Location of Vertical Interval Time Code.*

Due to the limitations of stationary head magnetic recording, longitudinal timecode from a VTR could not be read at very slow speeds or with the tape stopped. Vertical interval timecode (VITC) was coded digitally as a pulse stream onto one or two lines in the vertical interval; the scheme overcomes the disadvantage that LTC cannot be read with videotape stopped or moving slowly.

Timecode structure

Table 33.1A at the top of the facing page illustrates the structure of timecode data. The information bits include a flag bit *polarity/field* (whose function differs between LTC and VITC), and three *binary group flags* BGF_0, BGF_1, and BGF_2; they are interpreted in Table 33.1B and Table 33.1C opposite. A clumsy error was made when SMPTE standards were adapted to 25 Hz timecode: The positions of flag bits BGF_0, BGF_2, and

<table>
<tr><th>row</th><th>col 7</th><th>6</th><th>5</th><th>4</th><th>3</th><th>2</th><th>1</th><th>⇐0</th></tr>
<tr><td>0</td><td colspan="4">1st binary group</td><td colspan="4">Frame units 0–9</td></tr>
<tr><td>8</td><td colspan="4">2nd binary group
(or character 0)</td><td>Colour frame flag</td><td>Drop-frame flag</td><td colspan="2">Frame tens 0–2</td></tr>
<tr><td>16</td><td colspan="4">3rd binary group</td><td colspan="4">Second units 0–9</td></tr>
<tr><td>24</td><td colspan="4">4th binary group
(or character 1)</td><td>Polarity /Field {BGF_0}</td><td colspan="3">Second tens 0–5</td></tr>
<tr><td>32</td><td colspan="4">5th binary group</td><td colspan="4">Minute units 0–9</td></tr>
<tr><td>40</td><td colspan="4">6th binary group
(or character 2)</td><td>BGF_0 {BGF_2}</td><td colspan="3">Minute tens 0–5</td></tr>
<tr><td>48</td><td colspan="4">7th binary group</td><td colspan="4">Hour units 0–9</td></tr>
<tr><td>56</td><td colspan="4">8th binary group
(or character 3)</td><td>BGF_2 {Polarity /Field}</td><td>BGF_1</td><td colspan="2">Hour tens 0–2</td></tr>
</table>

Table 33.1A **Timecode bit assignment table.** The jumbled flag bits of 25 Hz systems are enclosed in braces.

polarity/field were jumbled. Flag bit interpretation unfortunately depends upon whether the timecode is 25 Hz-related. The timecode information bits provide no explicit indication of frame rate; frame rate must be determined after a suitable delay (e.g., 1 second), or from parameters outside timecode.

Dropframe flag	Asserted for dropframe timecode mode in 59.94 Hz systems only
Colourframe flag	Asserted when timecode is locked to the colourframe sequence of the associated video
Polarity (LTC only)	Computed such that the complete 80-bit LTC timecode for a frame contains an even number of zero bits (a.k.a. parity, or biphase mark polarity correction)
Field mark (VITC only)	Asserted for the second field

Table 33.1B **Timecode flag bits**

For details of LTC and VITC encoding, see the first edition of this book.

In LTC, the 64 bits are transmitted serially, followed by 16 LTC sync bits. In VITC, each group of 8 bits is preceded by two VITC sync bits; these ten words are followed by a final pair of VITC sync bits and a CRC.

BGF_2	BGF_1	BGF_0	*User-bit interpretation*
0	0	0	Unspecified characters or data
0	0	1	ISO 646 and ISO 2202 8-bit characters
1	0	1	SMPTE RP 136 data ("page/line")
all other combinations			Unassigned

Table 33.1C **Timecode binary group flags**

Further reading

SMPTE ST 12, *Time and Control Code.*

Timecode for 480*i* television is standardized in SMPTE ST 12. Timecode for 576*i* is described in IEC 60461, with additional information provided in European Broadcasting Union document EBU Tech. N12.

SMPTE RP 136, *Time and Control Codes for 24, 25 or 30 Frame-per-Second Motion-Picture Systems.*

SMPTE RP 136 standardizes magnetic recording of timecode on motion picture film at 24, 25, 29.97, or 30 frames per second. SMPTE ST 262, cited on page 399, standardizes a method of structuring user-bit data.

SMPTE 266M, *4:2:2 Digital Component Systems – Digital Vertical Interval Time Code*.

SMPTE 266M standardizes a version of VITC, *Digital Vertical Interval Timecode* (DVITC), to be transported across a BT.601, 4:2:2 interface.

SMPTE RP 196, *Transmission of LTC and VITC Data as HANC Packets in Serial Digital Television Interfaces*.

SMPTE RP 196 standardizes a mechanism to encode timecode data in ancillary (ANC) packets of the SDI.

2-3 pulldown

34

Motion picture film is intended for display at 24 frames per second. Many television programs, including a large number of prime-time programs, historically originated on film at this rate (or at $^{1000}/_{1001}$ of 24 Hz, that is, 23.976 Hz). This chapter discusses the conversion of film to frame rates higher than 24 Hz.

Except for certain very unusual film formats, motion picture film runs vertically through the camera and the projector. Be wary of anyone who presents a diagram like Figure 34.1 but rotated 90°.

Film has historically been transferred to 29.97 frame per second video using a technique called *2-3 pulldown,* whereby successive film frames are scanned first twice then three times to produce five video fields. The process is then repeated, reversing the roles of the first and second fields. The scheme is sketched in Figure 34.1.

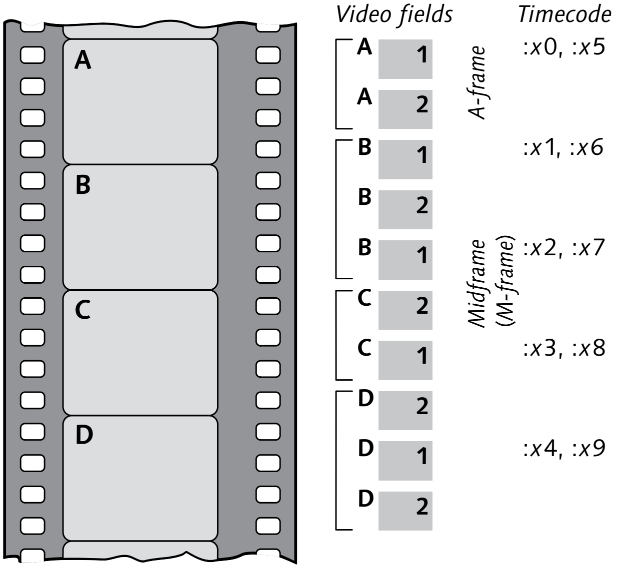


Figure 34.1 **"2-3 pulldown"** refers to transfer of film at about 24 frames per second to video at about 60 fields per second. The first film frame is transferred to two video fields; the second frame is transferred to three. The 2-3 cycle repeats. SMPTE RP 197 denotes film frames as A, B, C, and D. The A-frame is unique in being associated with exactly two fields – first, then second – of a single video frame. ("A-frame" denotes both a film frame *and* a video frame.) According to SMPTE RP 201, the A-frame timecode's frame count should end in 0 or 5.

Telecine is usually pronounced *tell-e-SIN-ee.*

A piece of equipment that performs this film-to-video conversion in realtime is called a *telecine.* (The term *film scanner* ordinarily implies nonrealtime operation.)

What I call *2-3 pulldown* was historically called *3-2 pulldown*. SMPTE standards assign letters A, B, C, and D to sets of four film frames. The A-frame is associated with the frame without a duplicate (redundant) field, so the sequence is best described as *2-3,* not *3-2*.

In 29.97 Hz systems, the film is run 0.1% slow, at about 23.976 Hz, so that the $\frac{5}{2}$ ratio of 2-3 pulldown results in a field rate of exactly 59.94 Hz. Figure 34.1 sketches four film frames; beside the set of film frames is the sequence of video fields produced by 2-3 pulldown. The labels *1* and *2* at the right indicate the first and second fields in an interlaced system.

When a 2-3 sequence is associated with nondrop-frame timecode, it is standard for the A-frames to take timecode numbers ending in 0 and 5.

In a sequence containing 2-3 pulldown, *cadence* refers to the temporal regularity of the A-frames. Careful editing preserves cadence; careless editing disrupts it.

EBU Tech. R62, *Recommended dominant field for 625-line 50-Hz video processing*.

In an interlaced sequence, *field dominance* refers to the field parity (first or second) where temporal coherence is susceptible to interruption due to editing. Video edits can, in principle, be made at any field; however, it is poor practice to make edits anywhere except the beginning of the first field.

Film is transferred to 25 Hz video by the simple expedient of running the film 4% fast, scanning each film frame to two video fields. This is somewhat jokingly referred to as *2-2 pulldown*. The 0.1% speed change of 2-3 pulldown has no significant effect on the accompanying audio; however, the 4% speed change of 2-2 pulldown necessitates audio pitch correction.

I have described transfer to interlaced video. The 2-3 pulldown technique can be used to produce progressive video at 60 Hz. In this case, what I have described as first and second fields are first and second frames.

Engineers are prone to think that better-quality motion results from higher frame rates. Proposals have been made to shoot film for television – even to shoot movies – at 30 frames per second instead of 24. The 24 Hz rate is obviously the world standard for cinema, but it is also uniquely suited to conversion to both 50 Hz systems (through 2-2 pulldown, 4% fast) and 59.94 Hz systems (through 2-3 pulldown, 0.1% slow). Choosing a film frame rate other that 24 Hz would

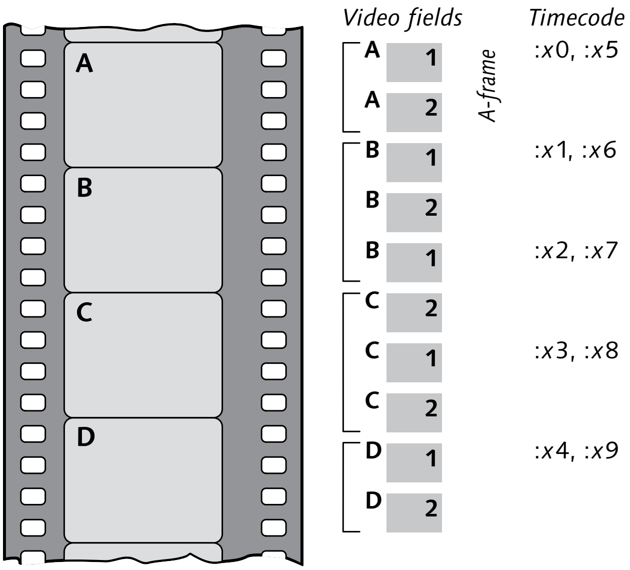


Figure 34.2 **"2-3-3-2 pulldown"** is identical to 2-3 pulldown except that the eighth video field represents film frame C instead of D. The advantage over 2-3 pulldown is that film frames can be reconstructed from intact video frames; no "stitching" of fields is necessary. 35 mm motion picture film runs vertically through the camera and the projector. The scheme is also known as 24*p*A. It is not intended for broadcast; however, if broadcast, the motion impairment is not much worse than 2-3 pulldown.

compromise this widely accepted method of conversion, and make it difficult for film producers to access international markets.

2-3-3-2 pulldown

When film at 23.976 frames per second is subject to 2-3 pulldown, the video "midframe" (timecode x2/x7) has film frames B and C mixed; the following video frame has film frames C and D mixed. (A video frame containing elements from two film frames is sometimes called a *blur frame.*) Although film frames A, B, and D can be simply reconstructed from single video frames, there is no single video frame that contains film frame C.

2-3-3-2 pulldown is denoted *24pA* (for *advanced*) by some manufacturers. When used with DV recording, it is an advantage of 2-3-3-2 that film frames can be reconstructed without decompressing the video bitstream.

This disadvantage is overcome by a minor adaptation to the pulldown sequence, denoted *2-3-3-2 pulldown*, also known as *24pA* (24-frame, progressive, advanced), diagrammed in Figure 34.2. Film frame C can now be reconstructed from the video frame having timecode ending in 3 or 8. A video sequence with 2-3-3-2 pulldown exhibits slight degradation in motion rendition compared to straight 2-3, because the 6 Hz beat frequency of 2-3-3-2 is slightly more visible than the 12 Hz of 2-3 pulldown. Nonetheless, the scheme is popular for desktop video editing.

Conversion of film to different frame rates

When an image sequence originated with 2-3 pulldown is displayed in video, motion portrayal is impaired to a certain degree. The impairment is rarely objectionable. However, if a 2-3 sequence is naively converted to a different frame rate, or if a still frame is extracted from a 2-3 sequence, the resulting impairments are liable to be objectionable. Prior to frame rate conversion from a film original, original film frames need to be reconstructed by weaving together the appropriate pair of fields in a process called *inverse telecine* or *inverse 2-3 pulldown*. Despite the fact that 2-3 pulldown has been used for half a century, no information to aid this "weaving" accompanies the video signal.

A simple method to convert from the 24 Hz film frame rate to any other rate could write successive film lines into a dual-port framebuffer at film scan rate, then read successive lines out of the buffer at video scan rate. But if a scene element is in motion with respect to the camera, this technique won't work. The right portion of Figure 34.3 opposite indicates lines scanned from film being written into a framebuffer. The slanted dashed lines intersect the video scanning; at the vertical coordinate where the lines intersect, the resulting picture switches abruptly from one field to another. This results in an output field that contains spatial discontinuities. (Although this description refers to interlaced scanning, none of these effects are directly related to interlace; exactly the same effects are found in 2-3 pulldown in progressive systems.)

Figure 34.3 shows the vertical-temporal ($V{\cdot}T$) relationships of 2-3 pulldown, with time on the horizontal axis, and vertical dimension of scanning on the vertical axis. Dashed lines represent film sampling; solid lines represent video sampling. Film capture samples the entire picture at the same instant. The staggered sequence introduced by 2-3 pulldown is responsible for the irregular spacing of the film sample lines. In video, sampling is delayed as the scan proceeds down the field; this is reflected in the slant of the lines in the figure.

In deinterlacing for a dedicated display, the output frame rate can be locked to the input video field rate of 59.94 Hz or 50 Hz. But in desktop computing applica-

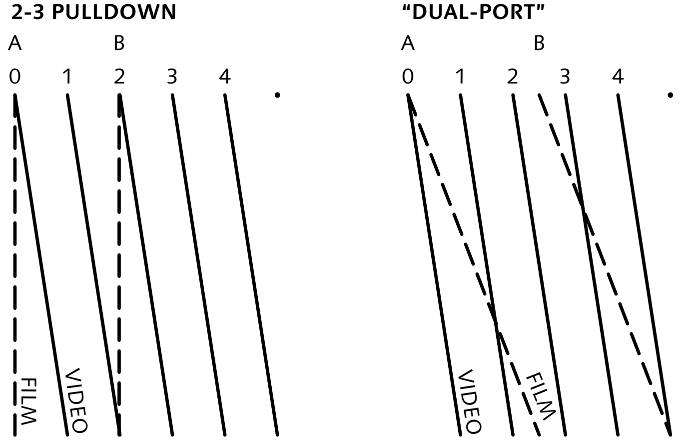


Figure 34.3 **Vertical/temporal relationships of 2-3 pulldown.** Time is on the *x*-axis; vertical displacement is on the *y*-axis. 2-3 pulldown, shown on the left, introduces temporal discontinuities. The "dual-port" approach, on the right, introduces spatial discontinuities.

tions of deinterlacing, the output is generally higher than 60 Hz, and asynchronous to the video rate: The output rate cannot be forced to match the native video rate. In this case, progressive reading from memory is faster than, and asynchronous to, the interlaced writing. If a single framestore is used, at some point the "fast" read pointer will cross the "slow" write pointer. If a "pointer crossing" event occurs on a scan line containing an element in motion, a spatial disturbance will be introduced into the picture. These disturbances can be prevented by using three fields of memory.

Figure 34.4 overleaf shows the effect in the spatial domain. Two intact film frames are shown at the left. The 2-3 pulldown technique introduces temporal irregularity into the video sequence, shown in the center column of five video fields, but the individual images are still intact. The result of the naive framebuffer approach is shown the right column: Spatial discontinuities are introduced into two of the five fields. With conversion from 24 Hz to 60 Hz, depending on the phase alignment of film and video, either two or three discontinuities will be evident to the viewer. With conversion from film at 24 Hz to video at 59.94 Hz, the discontinuities will drift slowly down the screen.

Using a pair of buffers – *double buffering* – and synchronizing the writing and reading of the buffers with the start of the film and video frames, keeps each frame intact and removes the spatial discontinuities. However, the delays involved in this technique reintro-

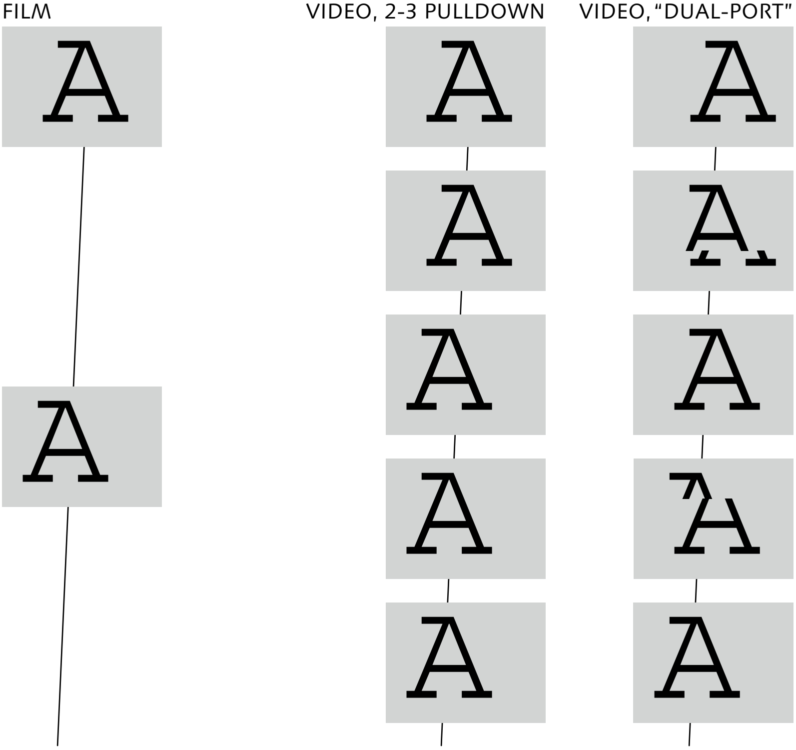


Figure 34.4 **2-3 pulldown, spatial view.** These sketches show the effect on the picture of two schemes to transfer film frames at 24 Hz, shown in the left column, into five video fields at 60 Hz. The center column shows the result of 2-3 pulldown. The right column shows the naive approach of writing into a framebuffer at film rate and reading at video rate.

duce exactly the same temporal stutter as 2-3 pulldown!

Although 2-3 pulldown introduces a temporal artifact into the video stream, acceptable motion portrayal is obtained when the 2-3 video is displayed at video rate. However, if the frame rate is altered a second time, temporal artifacts of the two cascaded conversions may become objectionable, especially when the image is displayed with a wide picture angle.

SMPTE RP 197, *Film-to-Video Transfer List.*

SMPTE RP 201, *Encoding Film Transfer Information Using Vertical Interval Time Code.*

In video that carries a carefully done film transfer, video pictures ending with timecode digits 0 or 5 correspond to A-frames of the film source. Apart from A-frame-locked timecode, there are no standards that convey, along with a video signal, information concerning film origination. Absent locked timecode, the only way to detect 2-3 pulldown is to compare data in successive fields: If two successive first fields contain luma and chroma that are identical, within a certain noise tolerance, and the repeat pattern follows the characteristic 2-3 sequence, then the material can be assumed to have originated from film. Once the original film frames have been identified, conversion to the ultimate display rate can be accomplished with minimal introduction of motion artifacts. Identifying film frames using this method is feasible for dedicated hardware, but it is difficult for today's desktop computers.

Native 24 Hz coding

Traditionally, 2-3 pulldown is imposed in the studio, at the point of transfer from film to video. The repeated fields are redundant, and consume media capacity to no good effect except compatibility with native 60 field-per-second equipment: When movies were recorded on media such as VHS tape, fully 20% of the media capacity was wasted. In the studio, it is difficult to recover original film frames from a 2-3 sequence. Information about the repeated fields is not directly available; decoding equipment typically attempts to reconstruct the sequence by comparing pixel values in successive fields.

Although MPEG-2 itself permits 23.967 Hz or 24 Hz native frame rate, DVD doesn't. Film material on 480*i* DVDs must be coded as 59.94 Hz interlaced. Flags identify film frames to the decoder, which then either inserts 2-3 pulldown for 480*i*29.97 output or – if configured to do so – outputs 480*p*23.976. See *Frame rate and 2-3 pulldown in MPEG*, on page 518.

MPEG-2 coding can handle coding of progressive material at 24 frames per second; it is inefficient to encode a sequence with 2-3 pulldown. Some MPEG-2 encoders are equipped to detect, and remove, 2-3 pulldown, so that 24 progressive frames per second are encoded. However, this process – called *inverse telecine* – is complex and trouble-prone.

Ordinarily, repeated fields in 2-3 pulldown are omitted from an MPEG-2 bitstream. However, the bitstream can include flags that indicate to the decoder that the omitted fields should be repeated at the display; with these flags, a decoder can reconstruct the 2-3 sequence for display at 59.94 Hz. Alternatively, when configured for progressive output, the decoder can reconstruct 23.976 Hz frames and present them directly to suitable display equipment, or frame-triple to 72 Hz (actually, $^{72}/_{1.001}$ or about 71.928 Hz) at the output. All this complexity is avoided in Blu-ray, which can directly code 23.976 Hz progressive frames.

Conversion to other rates

In 2-3 pulldown from 24 Hz to 60 Hz, information from successive film frames is replicated in the fixed sequence {2, 3, 2, 3, 2, 3, …}. The frequency of the repeat pattern – the *beat frequency* – is fairly high: The {2, 3} pattern repeats 12 times per second, so the beat frequency is 12 Hz. The ratio of frame rates in this case is $^5/_2$ – the small integers in this fraction dictate the high beat frequency.

When PCs are to display 29.97 Hz video that originated on film – from sources such as DVD or digital satellite – the situation is more complicated, and motion impairments are more likely to be introduced.

In conversion to a rate that is related to 24 Hz by a ratio of larger integers, the frequency of the repeated pattern falls. For example, converting to 75 Hz involves the fraction $^{25}/_8$; this creates the sequence {3, 3, 3, 3, 3, 3, 3, 4}, which repeats three times per second. Converting to 76 Hz involves the fraction $^{17}/_6$; this creates the sequence {2, 3, 3, 3, 3, 3}, which repeats four times per second. The susceptibility of the human visual system to motion artifacts peaks between 4 and 6 beats per second; the temporal artifacts introduced upon conversion to 75 Hz or 76 Hz are likely to be quite visible. Motion estimation and motion-compensated interpolation could potentially be used to reduce the severity of these conversion artifacts, but these techniques are highly complex, and will remain out of reach for desktop computing for several years.

When 24 Hz material is to be displayed in a computing environment with wide viewing angle and good ambient conditions, the best approach to minimize motion artifacts is to choose a display rate that is an integer multiple of 24 Hz. Displaying at 60 Hz reproduces the situation with video display, and we know well that the motion portrayal is quite acceptable.

Deinterlacing

35

In *Interlaced format,* on page 88, I explained that when a scene element in motion relative to the camera is captured by an interlaced camera, the scene element appears at different positions in the two fields. Reconstruction of progressive frames is necessary for certain image-processing operations, such as upconversion, downconversion, or standards conversion. Also, computer imagery is typically represented in progressive format: Integrating video imagery into computer displays also requires deinterlacing. This chapter outlines deinterlacing techniques.

I will introduce deinterlacing in the spatial domain. Then I will describe the vertical-temporal domain, and outline practical deinterlacing algorithms.

I will discuss the problem of deinterlacing, and the algorithms, in reference to the test scene sketched in Figure 35.1. The test scene comprises a black background, partially occluded by a white disk that is in motion with respect to the camera.

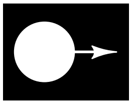


Figure 35.1 **Test scene**

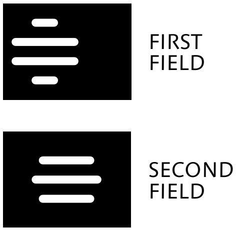


Figure 35.2 **Interlaced capture** samples the position of a football about 60 times per second; frames are produced at half that rate. (A soccer ball takes 50 positions per second.)

Spatial domain

Video captures 50 or 60 unique fields per second. If a scene contains an object in motion with respect to the camera, each field will carry half the spatial information of the object, but the information in the second field will be displaced according to the object's motion. The situation is illustrated in Figure 35.2, which shows the first and second fields, respectively. The example is typical of capture by a CCD camera set for a short exposure time; the example neglects capture blur due to nonzero exposure time at the camera. (For details of

Poynton, Charles (1996), "Motion portrayal, eye tracking, and emerging display technology," in *Proc. 30th SMPTE Advanced Motion Imaging Conference:* 192-202.

temporal characteristics of image acquisition and display, see my SMPTE paper.)

You can think of an interlaced video signal as having its lines in permuted order, compared to a progressive signal. An obvious way to accomplish deinterlacing is to write into two fields of video storage – the first field, then the second – in video order, then read out the assembled frame progressively (in spatial order). This method is sometimes given the sophisticated name *field replication,* or *weave*. This method is quite suitable for a stationary scene, or a scene containing only slow-moving elements. However, image data of the second field is delayed by half the frame time (typically $^1/_{60}$ s or $^1/_{50}$ s) with respect to image data of the first field. If the scene contains an element in fairly rapid motion (such as the disk in our test scene), and image data is interpreted as belonging to the same time instant, then *field tearing* will be introduced: The scene element will be reproduced with jagged edges, either when viewed as a still frame or when viewed in motion. The effect is sketched in Figure 35.3.

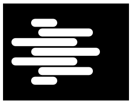


Figure 35.3 **The weave technique** stitches two fields into a frame. When applied to moving objects, it produces the "field tearing," "mouse's teeth," or "zipper" artifact.

Field tearing can be avoided by *intrafield* processing, using only information from a single field of video. The simplest intrafield technique is to replicate each line upon progressive readout. A disadvantage is that this method will reproduce a stationary element with at most half of its potential vertical resolution. Also, line replication introduces a blockiness into the picture, and an apparent downward shift of one image row. The effect is sketched in Figure 35.4.

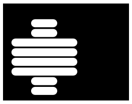


Figure 35.4 **Line replication**

The blockiness of the line replication approach can be avoided by synthesizing information that is apparently located spatially in the opposite field, but located temporally coincident with the same field. This can be accomplished by averaging vertically adjacent samples in one field, to create a synthetic intermediate line, as depicted in Figure 35.5. (In the computer industry, this is called "bob.") The averaging can be done prior to writing into the video memory, or upon reading, depending on which is more efficient for the memory system. Averaging alleviates the disadvantage of blockiness, but does not compensate the loss of vertical resolution. Nonetheless, the method performs well for VHS-grade images, which lack resolution in any case. Rather

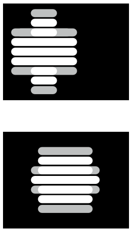


Figure 35.5 **Interfield averaging**

For a modest improvement over 2-tap averaging, use 4 taps with coefficients [$\frac{1}{16}$, $\frac{7}{16}$, $\frac{7}{16}$, $\frac{1}{16}$].

than simply averaging two lines, improved performance can be attained by using longer FIR filters with suitable tap weights; see *Filtering and sampling,* on page 191.

Vertical-temporal domain

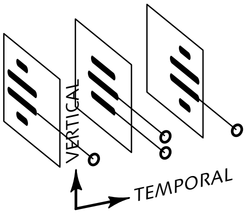


Figure 35.6 ***V·T* development**

Interlace-to-progressive conversion can be considered in the vertical-temporal (*V·T*) domain. Figure 35.6 in the margin sketches the interlaced capture fields of Figure 35.2, in a three-dimensional view. Viewed from the "side," along the axis of the scan lines, the vertical-temporal domain is projected. The temporal samples are at discrete times corresponding to the field instants; the vertical samples are at discrete intervals of space determined by the scan-line pitch. The four open disks of Figure 35.6 represent samples of original picture information that are available at a certain field instant and line number. A calculation on these samples can synthesize the missing sample value at the center of the pattern. In the diagrams to follow, the reconstructed sample will be drawn as a filled disk. (A similar calculation is performed for every sample along the scan line at the given vertical and temporal coordinate: For BT.601 digital video, the calculation is performed 720 times per scan line.)

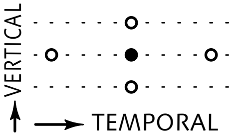


Figure 35.7 ***V·T* domain**

In Figure 35.7, I sketch the vertical-temporal domain, now in a two-dimensional view. Conversion from interlace to progressive involves computing some combination of the four samples indicated by open disks, to synthesize the sample at the center of the four (indicated by the filled disk). Techniques utilizing more than these four samples are possible, but involve more complexity than is justified for desktop video.

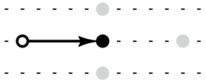


Figure 35.8 **Static lattice in the *V·T* domain** (weave)

In Figure 35.8, I sketch the field replication (or *weave*) technique in the *V·T* domain. The sample to be computed is simply copied from the previous field. The result is correct spatially, but if the corresponding area of the picture contains an element in motion, tearing will be introduced, as indicated in Figure 35.3.

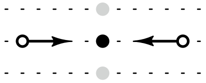


Figure 35.9 **Interframe averaging in the *V·T* domain**

Instead of copying information forward from the previous field, the previous field and the following field can be averaged. This approach is sketched in Figure 35.9. This technique also suffers from a form of field tearing, but it is useful in conjunction with an adaptive approach to be discussed in a moment.

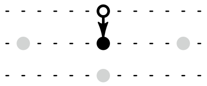


Figure 35.10 **Line replication in the *V*·*T* domain** (“bob“)

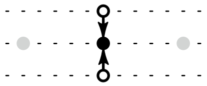


Figure 35.11 **Intrafield averaging in the *V*·*T* domain**

The line replication technique is sketched in the *V*·*T* domain in Figure 35.10. The central sample is simply copied from the line above. Because the copied sample is from the same field, no temporal artifacts are introduced. The line replication technique causes a downward shift of one image row. The shift is evident from Figure 35.4: The disk in the test scene is vertically centered, but in Figure 35.4 it appears off-center.

Intrafield averaging – what some people call the *bob* technique – is sketched in Figure 35.11. The central sample is computed by averaging samples from lines above and below the desired location. The information being averaged originates at the same instant in time, so no temporal artifact is introduced. Also, the one-row downward shift of line replication is avoided. However, the vertical resolution of a static scene is reduced.

Weston, Martin (1988), U.S. Patent 4,789,893, *Interpolating Lines of Video Signals.*

Martin Weston of the BBC found that excellent deinterlacing was possible using two fields and four lines of storage, without adaptivity, using carefully chosen coefficients. His filter coefficients are shown in Table 35.1; the highlighted cell corresponds to the result:

Image row	*Field t*–1	*Field t*	*Field t*+1
i – 4	32		32
i – 3		–27	
i – 2	–119		–119
i –1		539	
i	174	•	174
i+1		539	
i+1	–119		–119
i+1		–27	
i+1	32		32

Table 35.1 **Weston deinterlacer** comprises a vertical-temporal FIR filter having the indicated weights, each divided by 1024. The position marked in red is computed. No adaptivity is used.

Motion adaptivity

Analyzing the conversion in the *V*·*T* domain suggests that an improvement could be made by converting stationary scene elements using the static technique, but converting elements in motion using line averaging. This improvement can be implemented by detecting, for each result pixel, whether that pixel is likely to belong to a scene element in motion. If the element is likely to be in motion, then intrafield averaging is used (avoiding spatial artifacts). If the element is likely to be stationary, then interfield averaging is used (avoiding resolution loss).

Motion can be detected by comparing one field to a previous field. Ideally, a like field would be used – if motion is to be estimated for field 1, then the previous field 1 should be used as a point of reference. However, this approach demands that a full framestore be available for motion detection. Depending on the application, it may suffice to detect motion from the opposite field, using a single field of memory.

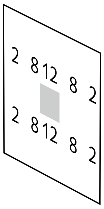


Figure 35.12 **Interstitial spatial filter coefficients**

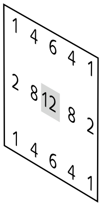


Figure 35.13 **Cosited spatial filter coefficients**

Whether a field or a frame of memory is used to detect motion, it is important to apply a spatial lowpass filter to the available picture information, in order to prevent small details, or noise, from causing abrupt changes in the estimated motion. Figure 35.12 shows the coefficients of a spatial lowpass filter that computes a spatial sample halfway between the scan lines. The shaded square indicates the effective location of the result. This filter requires a linestore (or a dual-ported memory). The weighted sums can be implemented by three cascaded [1, 1] sections, each of which requires a single adder.

A low-pass filtered sample cosited (spatially coincident) with a scan line can be computed using the weights indicated in Figure 35.13. Again, the shaded square indicates the central sample, whose motion is being detected. This filter can also be implemented using just linestores and cascaded [1, 1] sections. The probability of motion is estimated as the absolute value of the difference between the two spatial filter results.

The spatial filters of Figure 35.12 and Figure 35.13 incorporate transverse filters having coefficients [1, 4, 6, 4, 1]. These particular coefficients enable implementation using cascaded [1, 1]-filters. The 2-line spatial filter of Figure 35.12 can be implemented using a linestore, two [1, 4, 6, 4, 1] transverse filters, and an adder. The 3-line spatial filter of Figure 35.13 can be implemented using two linestores, three [1, 4, 6, 4, 1] transverse filters – one of them having its result doubled to implement coefficients 2, 8, 12, 8, 2 – and two adders.

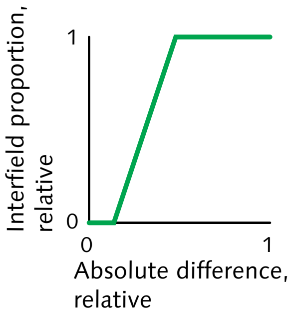


Figure 35.14 **A window function** in deinterlacing

A simple adaptive filter switches from interframe averaging to interfield averaging when the motion estimate exceeds some threshold. However, abrupt switching can result in artifacts: Two neighboring samples may have very similar values, but if one is judged to be stationary and the other judged to be in motion, the samples computed by the deinterlace filter may have dramatically different values. These differences can be visually objectionable. These artifacts can be reduced by mixing proportionally – in other words, fading – between the interframe and interfield averages instead of switching abruptly. Mixing can be controlled by a window function of the motion difference, as sketched in Figure 35.14 in the margin.

Further reading

Bellers and de Haan have written the definitive book on deinterlacing techniques. The book concentrates on techniques patented by Philips and available in VLSI from NXP. A summary of deinterlacing techniques is found in de Haan and Braspenning's chapter in Madisetti's book.

Bellers, Erwin B. and de Haan, Gerard (2000), *De-interlacing: A key technology for scan rate conversion* (Elsevier/North-Holland).

de Haan, Gerard and Braspenning, Ralph (2010), "Video Scanning Format Conversion and Motion Estimation," in Madisetti, Vijay K., *The digital signal processing handbook,* Second edition, Vol. 2 (Boca Raton, Fla., U.S.A.: CRC Press/Taylor & Francis).

Colourbars 36

SD colourbars

ANSI/EIA-189-A, *Encoded Color Bar Signal* (formerly denoted EIA RS-189-A).

SMPTE EG 1, *Alignment Color Bar Test Signal for Television Picture Monitors*.

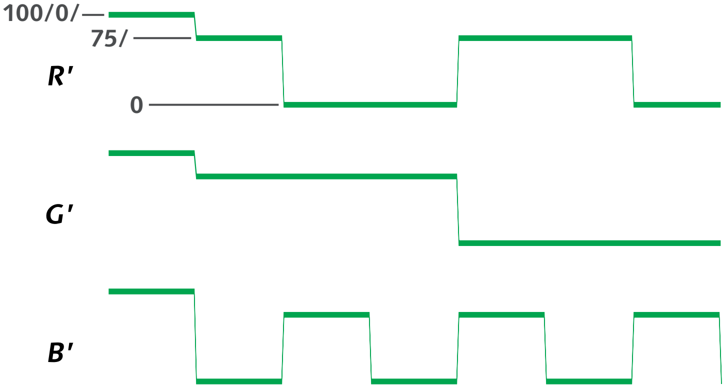
Figure 36.1 below is a sketch of an image produced by the classic SMPTE colourbar test pattern. The upper ⅔ of the image contains a 100% white bar followed by primary and secondary colours of 75% saturation. The narrow, central region contains "reverse bars"; this section enables setting composite NTSC or PAL decoder HUE and CHROMA. The bottom ¼ of the image contains subcarrier frequency at $-I$ phase, a white bar, subcarrier frequency at $+Q$ phase, and (at the right) the PLUGE element, which I will describe in a moment.

PLUGE is pronounced *ploodge*.

Figure 36.2 overleaf shows the $R'G'B'$ components that produce the upper ⅔ of the frame of SMPTE colourbars. Each scan line is based upon a binary sequence of red, green, and blue values either zero or unity. The components are arranged in order of their contributions to luma, so that the eventual luma component decreases from left to right. (The narrow band $^2/_3$ of the way down SMPTE bars has its count

Figure 36.1 **The SMPTE EG 1 SD colourbar test signal** is represented here as an image; however, it is standardized as a signal in the $R'G'B'$ domain. Its colour interpretation depends upon the primary chromaticities in use. The corresponding $Y'C_BC_R$ or $Y'P_BP_R$ waveforms depend upon luma coefficients and scaling.

Figure 36.2 **Colourbar *R′G′B′* primary components** in SMPTE colourbars have amplitude of 75 IRE, denoted 75/0/75/0. A variation denoted 100/0/75/0, whose *R′*, *G′*, and *B′* waveforms are sketched here, places the white bar at 100 IRE. Other variations have different amplitudes for the uncoloured and coloured bars.



sequence reversed and its green component forced to zero.)

In studio equipment, in component video, and in PAL broadcast, the processing, recording, and transmission channels can accommodate all encoded signals that can be produced from mixtures of *R′G′B′* where each component is in the range 0 to 1. The *100%* colourbar signal exercises eight points at these limits.

See the section *UV components* of *Composite NTSC and PAL: Legacy Video Systems*. Strictly speaking, owing to negative AM video modulation, an NTSC transmitter would *under*modulate if presented with composite video level exceeding 120%.

Fully saturated yellow and fully saturated cyan cause a composite PAL signal to reach a peak value $^4/_3$ (133 $^1/_3$%) of reference white. However, an NTSC transmitter's composite signal amplitude is limited to 120% of reference white. If 100% bars were presented to an NTSC transmitter, clipping would result. To avoid clipping, *75%* bars are ordinarily used to test NTSC transmission. The white bar comprises primaries at 100%, but the other bars have their primary components reduced to 75% so as to limit their composite NTSC peak to the level of reference white. The "75% bars" convention was adopted for good reason, but analog NTSC transmitters have been decommissioned; the convention remains in use for no good reason.

R′G′B′ components of 100% colourbars take *R′G′B′* (or *RGB*) values of zero or unity, independent of the chromaticity of the primaries: Owing to differences in primary chromaticities, the exact colours of the bars are not identical among SMPTE, EBU, and HD standards.

SD colourbar notation

ITU-R Rec. BT.471, *Nomenclature and description of colour bar signals*.

I have referred to 100% and 75% colourbars. Several additional variations of colourbars are in use, so many that an international standard is required to denote them. A colourbar signal is denoted by four numbers, all in units (formerly, IRE units), separated by slashes. The first pair of numbers gives the maximum and minimum values (respectively) of the primary components in uncoloured bars – that is, the black or white bars. The second pair gives the maximum and minimum primary values (respectively) in the coloured bars.

The 100% bars, described earlier, are denoted 100/0/100/0. That variation is useful in the studio, in all forms of component video, and in PAL transmission. In legacy 480*i* composite NTSC systems where 7.5% setup is used, *100% bars* refers to 100/7.5/100/7.5. That variation was once useful in the studio. However, as I explained on page 420, terrestrial analog NTSC transmission cannot handle 100% bars. NTSC transmitters are tested using 75% bars with setup, denoted 100/7.5/75/7.5. Japan uses 480*i* video with zero setup; there, 75% bars, denoted 100/0/75/0, are used.

PLUGE element

The lower-right quadrant of the colourbar pattern contains elements produced by *picture line-up generating equipment* (PLUGE); see Figure 36.3. The acronym originates with the "generating equipment," but nowadays PLUGE signifies the signal element. Superimposed on reference black are two elements, one slightly more negative than reference black, the other slightly more positive.

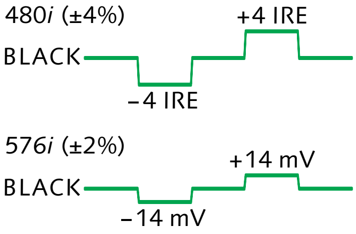


Figure 36.3 **The PLUGE element** of the colourbar signal enables accurate setting of black level. The 14 mV excursion in 576*i* PLUGE is equivalent to ±2 units.

A display's BLACK LEVEL is adjusted so that the first (negative-going) element is just barely indistinguishable from reference black. (The second element should then be barely visible.) Details are found in *Black level setting*, on page 56. The negative-going element of PLUGE cannot be represented in positive *R'G'B'*.

In SD, ±4% PLUGE was standardized in the now-withdrawn SMPTE EG 1. However, SMPTE RP 219 standardizes PLUGE for HD with 8-bit interface codes 44, 64, 84, and 104 – that is, approximately –2%, +2%, and +4%.

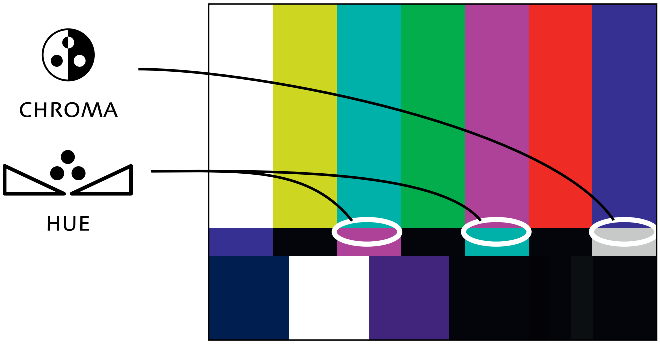


Figure 36.4 **HUE and CHROMA are adjusted** in a composite NTSC decoder using the colourbar test image, using a "blue-only" display. The controls are adjusted until the indicated transitions disappear.

Composite decoder adjustment using colourbars

When composite NTSC or PAL colourbars are decoded, the amount of blue decoded from the white, cyan, magenta, and blue bars should ideally be identical. Any chroma gain (saturation) error will affect the signal decoded from blue, but not the blue decoded from white. Chroma phase error will cause hue errors of opposite direction in the blue decoded from cyan and the blue decoded from magenta. To manually adjust a decoder's HUE and CHROMA (or TINT and COLOUR, or PHASE and SATURATION) controls involves displaying composite SMPTE colourbars, and disabling the red and green components at the decoder output. The amount of blue decoded from each of cyan and magenta is equalized by adjusting the decoder's HUE control. The amount of blue decoded from each of grey and blue is equalized by adjusting CHROMA. The comparison is facilitated by the reversed bar portion of SMPTE colourbars. Figure 36.4 shows a representation of the colourbar image, showing the bars that are visually compared while adjusting HUE and CHROMA.

If the red and green guns at the display cannot be turned off, a similar effect can be accomplished by viewing the CRT through a blue gel filter.

Adjusting HUE and CHROMA controls in this way is only meaningful to compensate errors in NTSC and PAL decoding. In component video such as $R'G'B'$ and $Y'C_BC_R$, no recording or transmission impairment rotates hue or alters chroma (saturation): Using the scheme described above is nonsensical.

–*I*, +*Q*, and PLUGE elements in SD colourbars

The -*I* and +Q elements correspond to *R'G'B'* values of [-0.3824, 0.1088, 0.4427] and [0.2483, -0.2589, 0.6817], respectively; and to 10-bit $^{601}Y'C_BC_R$ values [0, 228, –244] and [0, 345, 159]. To produce *RGB*-legal codes having the same hue and saturation as -*I* and +Q, and having minimum luma, use *R'G'B'* values [0, 0.2456, 0.412545] and [0.253605, 0, 0.470286], respectively. See SMPTE RP 219.

The lower-left quadrant of the SMPTE colourbar pattern contains subcarrier frequency components at –*I* and +*Q* phase. These elements were designed to exercise the encoding and decoding axes of the original NTSC chroma modulation method (circa 1953). Encoding and decoding on *I* and *Q* axes fell into disuse around 1970, being replaced by encoding and decoding on the *B'–Y'* and *R'–Y'* axes, so the utility of this portion of the signal is now lost. The historical –*I* and +*Q* elements contain high chroma resting upon black. These combinations correspond to illegal mixtures of *R'G'B'* where one component is dramatically negative; consequently, the –*I* and +*Q* elements are not representable in the positive *R'G'B'* domain.

The -*I* element, the +*Q* element, and the negative-going element of PLUGE are generated synthetically. None of these elements represents picture information; none is useful in evaluating pictures; and none can be generated in – or survive transit through – the positive *R'G'B'* domain between 0 and 1 (or 0 and 255, or even 16 through 235 or 64 through 940 if footroom is clipped).

In fact, it is not just the –*I*, +*Q*, and PLUGE elements of colourbars that are synthetic: The entire signal is generated synthetically! We call it the colourbar *signal*, not the colourbar *image*, because the –*I*, +*Q*, and PLUGE elements cannot be represented in nonnegative *RGB* components. The colourbar signal represents values of *R'G'B'* as if they came from a gamma-corrected camera and were inserted prior to an encoder. *R'G'B'* values of colourbars are implicitly gamma-corrected.

HD colourbars

SMPTE RP 219, *High-Definition, Standard-Definition Compatible Color Bar Signal*.

Figure 36.1 overleaf is a sketch of an image produced by the SMPTE colourbar test pattern for HD.

Figure 36.5 **The SMPTE RP 219 SD colourbar test signal** is used for HD.

Part 4

Studio standards

Reference display and viewing conditions

37

Historically, the video industry has been lax in setting standards for studio video reference displays. Over the next few years, as fixed-pixel displays (FPDs) reach studio quality levels and see deployment as reference displays, that situation has to change. As I write, standards groups within EBU, SMPTE, and ITU-R are working to remedy these deficiencies. However, given the absence today of official standards, I am writing the remainder of this chapter as if it is the missing standard. This chapter attempts to codify current practice in 2011, as exemplified by CRT "broadcast video monitors." In my view, LCDs are not yet studio-grade. Changes will be necessary to accommodate mastering of wide-gamut colour and/or high dynamic range.

Introduction

The reference display converts an *R'G'B'* video signal to light – characterized by CIE tristimulus values according to CIE 15 – as if through an additive *RGB* process. (A reference display need not physically use additive mixing providing that it behaves as an additive *RGB* device as far as its operation is observed externally.)

Signal interface

R'G'B' video signal values herein are normalized to reference black at value 0 and reference white at value 1. Signal values below 0 are clipped to 0. Signal values up to the fraction $^{955}/_{876}$ of reference white are accommodated; signal values beyond $^{955}/_{876}$ shall be clipped to $^{955}/_{876}$. (Various video standards define mappings into integers; for example, 10-bit digital video encoding according to ITU-R BT.709 includes scaling by 876 and offset of +64; reference black is represented as *R'G'B'* signal code 64, reference white as signal code 940, and peak white as signal code 1019.)

Reference primaries, black, and white

Reference red, green, blue, black, and white tristimuli shall have the chromaticities specified by BT.709. The reference white signal [1, 1, 1] shall display luminance of 100 $\text{cd}\cdot\text{m}^{-2}$. The reference black signal [0, 0, 0] shall display relative luminance denoted β preferably 0.0003 but not exceeding 0.001, of the reference white luminance. Reference white and reference black shall be measured according to ITU-R BT.815.

Reference EOCF

Across the range of each $R'G'B'$ signal component value (denoted V) from reference black to peak white, reference tristimulus value of the associated primary (denoted T), relative to reference white, computed according to SMPTE RP 177 shall be:

Eq 37.1

$$T = \overset{\text{GAIN}}{(1-\beta)}\,\left(\text{MAX}[0,\ \overset{\text{OFFSET}}{b+(1-b)}\cdot V]\right)^{\overset{\text{GAMMA}}{2.4}} + \overset{\text{OFFSET}}{\beta},\ b \approx 0.035,\ \beta \le 0.0003$$

The parameter b takes a value of approximately 0.035, to be determined according to operational practice of setting BLACK LEVEL. (The gain factor of $(1-\beta)$ establishes reference white at unity. Resulting peak white luminance is approximately 122 $\text{cd}\cdot\text{m}^{-2}$. The power function exponent 2.4 conforms to BT.1886.)

Reference viewing conditions

The environment of the reference display – including the display surface and its surround – shall have illuminance of 2π lx (approximately 6 lx) at CIE D_{65} chromaticity. The reference display shall be surrounded by approximately Lambertian neutral grey having diffuse reflectance factor of 0.5. (These conditions yield 1% very dim surround.) The surround should extend horizontally across an angle corresponding to three picture widths and vertically across an angle corresponding to three picture heights.

References

CIE 15:2004, *Colorimetry,* Third Edition (2004).

ITU-R BT.709-5, *Parameter values for the HDTV standards for production and international programme exchange* (2002).

ITU-R BT.815-1, *Specification of a signal for measurement of the contrast ratio of displays* (1994-07).

ITU-R BT.1886, *Reference electro-optical transfer function for flatpanel displays used in HDTV studio production* (2011).

SMPTE RP 177, *Derivation of Basic Television Color Equations* (1993).

SDI and HD-SDI interfaces

38

Composite $4f_{SC}$ digital interfaces are obsolete. For details about them, consult the first edition of this book.

This chapter describes digital interfaces for uncompressed and compressed SD and HD. Tables 38.1 and 38.2 summarize video signal levels.

Interface	*Ref. black*	*Ref. white*
Abstract signal, mathematical	0	1
Abstract signal, units ("IRE")	0	100
Analog NTSC [mV]	$53\,{}^{4}/_{7}$	$714\,{}^{2}/_{7}$
Analog NTSC-J [mV]	0	$714\,{}^{2}/_{7}$
Analog PAL [mV]	0	700
Analog VGA [mV], zero setup	0	700
7.5-percent setup	$53\,{}^{4}/_{7}$	$714\,{}^{2}/_{7}$

Table 38.1 **Analog video levels** in several interfaces are summarized.

Interface	*–Peak non-SDI black*	*–Peak SDI black*	*Ref. black*	*Ref. white*	*+Peak SDI white*	*+Peak non-SDI white*
8-bit computing ("IT," e.g., sRGB)			0	255		
Studio video interface, 8-bit ("CE")	0	1	16	235	254	255
10-bit	0	4	64	940	1019	1023
Studio video processing, 8-bit	–16	–15	0	219	254	255
10-bit	–64	–60	0	876	955	959
Digital cinema interface, 12-bit	0	16 [a]	0	3960 [b]	3960	3960

Table 38.2 **Digital video levels** in several interfaces are summarized.

[a] True reference black in digital cinema cannot be conveyed across an HD-SDI interface: The minimum interface code yields black tristimulus value of $({}^{16}/_{3960})^{2.6}$, or about 0.000 000 6, negligibly different from ideal black.

[b] Peak white code is indicated for digital cinema as 3960: This is for the coded luminance (Y') channel. The other two channels (X' and Z') have peak values 3794 and 3890 respectively.

Component digital SD interface (BT.601)

ITU-R Rec. BT.601-5, *Studio encoding parameters of digital television for standard 4:3 and widescreen 16:9 aspect ratios*.

ITU-R Rec. BT.601, adopted in 1984, specifies abstract coding parameters (including 4:2:2 chroma subsampling) for 480*i*29.97 and 576*i*25 SD. Luma is sampled at 13.5 MHz; C_B and C_R colour difference components are horizontally subsampled by a factor of 2:1 with respect to luma – that is, sampled at 6.75 MHz each. Samples are multiplexed in the sequence $\{C_B, Y_0', C_R, Y_1'\}$. Sync information and optional ancillary data is multiplexed; 10-bit words at 27 MW/s are then serialized for a total bit rate of 270 Mb/s. The external interface is called the *serial digital interface* (SDI); it uses coaxial cable and BNC connectors.

BT.601 originated with 8-bit components, anticipating 10 bits.

Recall from page 124 that in 4:2:2 BT.601, C_B and C_R are *cosited* – each is centered on the same location as Y_j', where j is even; chroma samples are absent when j is odd.

Sampling at 13.5 MHz produces a whole number of samples per total line (S_{TL}) in 480*i* systems (with 858 S_{TL}) and 576*i* systems (with 864 S_{TL}). Both 480*i* and 576*i* have 720 active luma samples per line (S_{AL}). In uncompressed, 8-bit BT.601 video, the active samples consume about 20 MB/s.

The notation *4:2:2* originated as a reference to the chroma subsampling scheme that I outlined on page 124. During the 1980s, 4:2:2 denoted a specific SD component digital video interface standard incorporating 4:2:2 chroma subsampling. In the 1990s, the 4:2:2 chroma subsampling format was adopted for HD; as a result, the notation *4:2:2* came to be independent of image size.

The notations BT.601 and BT.656 have fallen into disuse in studio video. However, desktop video hardware designers often use "Rec. 601" or "BT.601" to denote a parallel interface having separate wires for vertical and horizontal sync signalling, and "Rec. 656" or "BT.656" to denote a parallel interface wherein vertical and horizontal sync are represented by embedded TRS codes.

Figure 38.1 at the top of the facing page shows the luma (or R', G', or B') waveform of a single scan line of 480*i* component video. The time axis shows sample counts at the BT.601 rate of 13.5 MHz; divide the sample number by 13.5 to derive time in microseconds. Amplitude is shown in millivolts (according to EBU Tech. N10 levels), and in 8-bit BT.601 digital interface code values.

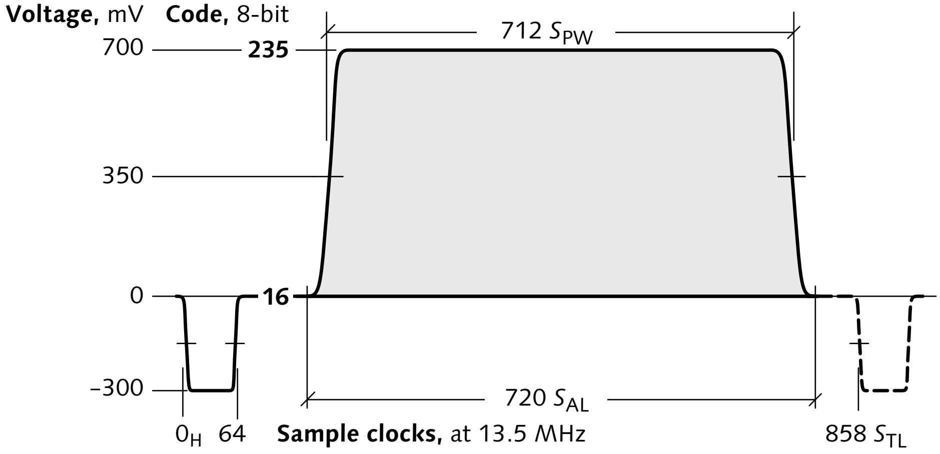
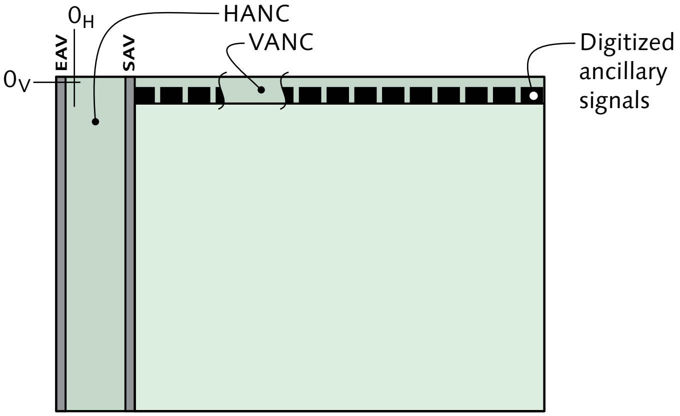


Figure 38.1 **Scan-line waveform for 480*i*29.97, 4:2:2 component luma.** EBU Tech. N10 analog levels are shown; however, these levels are rarely used in 480*i*. In analog video, sync is *blacker-than-black*, at –300 mV. (In digital video, sync is not coded as a signal level.) This sketch shows 8-bit interface levels (in bold); black is at code 16 and white is at code 235. The 720 active samples contain picture information; the remaining 138 sample intervals of the 858 comprise horizontal blanking.

Digital video interfaces convey active video framed in *timing reference signal* (TRS) sequences including *start of active video* (SAV) and *end of active video* (EAV). Ancillary data (ANC) and digitized ancillary signals are permitted in regions not occupied by active video. Figure 38.2 below shows the raster diagram of Chapter 8, augmented with EAV, SAV, and the HANC and VANC regions. Details will be presented in *SDI and HD-SDI sync, TRS, and ancillary data,* on page 433.

Figure 38.2 **The BT.656 component digital interface** uses EAV to signal the start of each horizontal blanking interval, and SAV to signal the start of active video. Between EAV and SAV, ancillary data (HANC) can be carried. In a nonpicture line, the region between SAV and EAV can carry ancillary data (VANC). Digitized ancillary signals may be carried in lines other than those that convey either VANC or analog sync.



Serial digital interface (SDI)

SMPTE 259M, *10-Bit 4:2:2 Component and 4f_{SC} Composite Digital Signals – Serial Digital Interface*.

Serial digital interface (SDI) refers to a family of interfaces standardized by SMPTE. The BT.601 or 4f_{SC} data stream is serialized, then subjected to a scrambling technique. SMPTE ST 259 standardizes several interfaces, denoted by letters A through D as follows:

- Composite 4f_{SC} NTSC video, about 143 Mb/s
- Composite 4f_{SC} PAL video, about 177 Mb/s
- BT.601 4:2:2 component video, 270 Mb/s (This interface is standardized in BT.656.)
- BT.601 4:2:2 component video sampled at 18 MHz to achieve 16:9 aspect ratio, 360 Mb/s

All but scheme C are now obsolete.

SDI is standardized for electrical transmission through coaxial cable, and for transmission through optical fiber. The SDI electrical interface uses ECL levels, 75 Ω impedance, BNC connectors, and coaxial cable. Electrical and mechanical parameters are specified in SMPTE standards and in BT.656; see *SDI coding* on page 439. Fiber-optic interfaces for digital SD, specified in SMPTE 297M, are straightforward adaptations of the serial versions of BT.656.

Component digital HD-SDI

The basic coding parameters of HD systems are standardized in BT.709. Various scanning systems are detailed in several SMPTE standards referenced in Table 15.2, on page 145.

Component SD, composite 4f_{SC} NTSC, and composite 4f_{SC} PAL all have different sample rates and different serial interface bit rates. In HD, a uniform sample rate of 74.25 MHz is adopted (modified by the ratio $^{1000}/_{1001}$ in applications where compatibility with 59.94 Hz frame rate is required). A serial interface bit rate of 20 times the sampling rate is used. Variations of the same standard accommodate mainstream 1080*i*30, 1080*p*24, and 720*p*60 scanning; 1080*p*30; and the obsolete 1035*i*30 system. The integer picture rates 24, 30, and 60 can be modified by the fraction $^{1000}/_{1001}$, giving rates of 23.976 Hz, 29.97 Hz, and 59.94 Hz.

The 23.976 Hz, 29.97 Hz, and 59.94 Hz frame rates are associated with a sampling rate of:

$$\frac{74.25}{1.001} \approx 74.176 \text{ Mpx/s}$$

The corresponding HD-SDI serial interface bit rate is:

$$\frac{1.485}{1.001} \approx 1.483 \text{ Gb/s}$$

The SDI interface at 270 Mb/s has been adapted to HD by scaling the bit rate by a factor of 5.5, yielding a fixed bit rate of 1.485 Gb/s. The sampling rate and serial bit rate for 23.976 Hz, 29.97 Hz, and 59.94 Hz

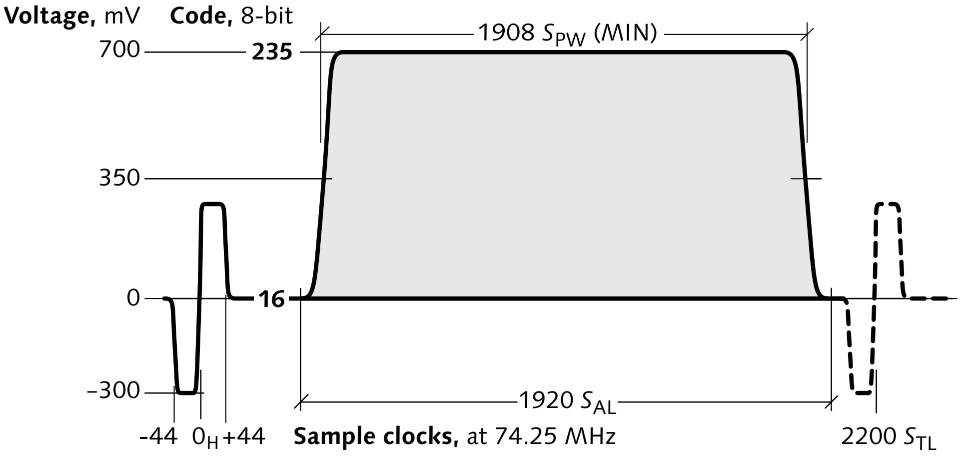


Figure 38.3 **Scan-line waveform for 1080*i*30 HD component luma.** Analog trilevel sync is shown, excursing ±300 mV. (In digital video, sync is not coded as a signal level.) At an 8-bit interface, black is represented by code 16 and white by 235. The indicated 1920 active samples contain picture information; the remaining sample intervals of the 2200 total comprise horizontal blanking.

interfaces are indicated in the margin. This interface is standardized for $Y'C_BC_R$, subsampled 4:2:2. Dual-link HD-SDI can be used to convey R'G'B'A, 4:4:4:4.

HD-SDI accommodates 1080*i*25 and 1080*p*25 variants that might find use in Europe. This is accomplished by placing the 1920×1080 image array in a scanning system having 25 Hz rate. S_{TL} is altered from the 30 Hz standard to form an 1125/25 raster.

See Figure 15.2, on page 144.

The standard HD analog interfaces use trilevel sync, instead of the bilevel sync that is used for analog SD. Figure 38.3 above shows the scan-line waveform, including trilevel sync, for 1080*i*30 HD.

SMPTE ST 292, *Bit-Serial Digital Interface for High-Definition Television Systems*.

The HD-SDI interface is standardized in SMPTE ST 292. Fiber-optic interfaces for digital HD are also specified in SMPTE ST 292.

SDI and HD-SDI sync, TRS, and ancillary data

Along with picture information, a streaming interface needs to convey information about which time instants – or which digital samples – are associated with the start of each frame and the start of each line. In digital video, this information is conveyed by *timing reference signals* (TRS) that I will explain in this chapter.

SDI has the capacity to transmit ancillary (ANC) data. The *serial data transport interface* (SDTI) resembles SDI, but has no uncompressed active video – instead, the full data capacity of the link is dedicated to carrying ANC packets. Compressed digital video can be conveyed in these packets. SDTI will be described on page 441.

The IEEE 1394/DV interface, described on page 167, and the DVB-ASI interface, to be described on page 443, have no ancillary data and do not use TRS.

Standard serial interfaces transmit 10-bit samples; a transmitter must present all 10 bits at the interface (even if the two LSBs are zero).

I use the subscript h to denote a hexadecimal (base 16) integer. Sample values in this chapter are expressed in 10 bits.

TRS and ANC sequences are introduced by 10-bit codewords 0 and $3FF_h$. Stemming from legacy parallel interfaces such as SMPTE RP 125 and EBU Tech. 3246, a receiver must ignore the two LSBs in identifying TRS and ANC. Apart from their use to delimit TRS and ANC, codewords 0, 1, 2, 3, and $3FC_h$, $3FD_h$, $3FE_h$, and $3FF_h$ are prohibited from digital video data.

TRS in 4:2:2 SD-SDI

In *Component digital SD interface (BT.601)*, on page 430, I explained that 4:2:2 samples are multiplexed in the sequence $\{C_B, Y_0', C_R, Y_1'\}$ onto the SDI. BT.601 defines the abstract signal coding parameters; BT.656 defines the interface.

Active luma samples are numbered from zero to $S_{AL}-1$; active chroma samples are numbered from zero to $(S_{AL}/2)-1$. The interface transmits two words for each luma sample clock: Even-numbered words convey chroma samples; odd-numbered words convey luma samples. The sample structure aligns with 0_H: If analog sync were digitized, a particular digitized luma sample would precisely reflect the 50% value of sync.

In 4:2:2 video, a four-word TRS sequence immediately precedes active video, indicating *start* of *active video* (SAV). SAV is followed by C_B sample zero. Immediately following the last active sample of the line is another four-word TRS sequence, *end* of *active video* (EAV). The TRS sequence comprises a word of all ones (codeword $3FF_h$), a word of all zeros, another word of all zeros, and finally a word including flag bits F (Field), V (Vertical), H (Horizontal), P_3, P_2, P_1, and P_0 (Parity). SAV is indicated by $H=0$; EAV has $H=1$.

The *V* and *H* bits are asserted during the corresponding blanking intervals. The *F* bit denotes field, not frame.

Table 38.3 shows the elements of TRS.

Word	*Value*	*MSB*									*LSB*
		9	*8*	*7*	*6*	*5*	*4*	*3*	*2*	*1*	*0*
0	$\mathrm{3FF}_h$	1	1	1	1	1	1	1	1	1	1
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0
3		1	F	V	H	P_3	P_2	P_1	P_0	0	0

Table 38.3 **Timing reference sequence (TRS)** for 4:2:2 comprises 4 codewords. Start of active video (SAV) is indicated by $H=0$; end of active video (EAV) has $H=1$. For compatibility with 8-bit equipment, the 2 LSBs are ignored in decoding TRS.

Value		*F*	*V*	*H*	$P_3 = V \oplus H$	$P_2 = F \oplus H$	$P_1 = F \oplus V$	$P_0 = F \oplus V \oplus H$		
$\mathrm{200}_h$	1	0	0	0	0	0	0	0	0	0
$\mathrm{274}_h$	1	0	0	1	1	1	0	1	0	0
$\mathrm{2AC}_h$	1	0	1	0	1	0	1	1	0	0
$\mathrm{2DB}_h$	1	0	1	1	0	1	1	0	0	0
$\mathrm{31C}_h$	1	1	0	0	0	1	1	1	0	0
$\mathrm{368}_h$	1	1	0	1	1	0	1	0	0	0
$\mathrm{380}_h$	1	1	1	0	1	1	0	0	0	0
$\mathrm{3C4}_h$	1	1	1	1	0	0	0	1	0	0

Table 38.4 **Protection bits for SAV and EAV** are computed as the exclusive-or (⊕) of various combinations of *F*, *V*, and *H*. The code can correct 1-bit errors, and can detect 2-bit errors. The error-correction capability was arguably useful for the parallel interface. However, it is useless for SDI, because a single-bit error in the SDI bitstream, when descrambled, corrupts up to 5 bits.

The *F* and *V* bits change state in the EAV prior to the start of the associated line; rather than calling it EAV, you might call it *start of horizontal interval.* In interlaced systems, *F* is asserted during the second field. In 480*i* systems, *F* changes at lines 4 and 266; in other scanning systems, including 576*i* and HD, *F* changes state at line 1. In progressive systems, *F* is always zero (except in 483*p*59.94, where *F* encodes frame parity.)

The vertical blanking (*V*) bit is zero in every line that is defined by the associated scanning standard to contain active (picture) video; it is asserted elsewhere – that is, in the vertical interval.

The *F*, *V*, and *H* bits are protected by parity bits P_3, P_2, P_1, and P_0, formed as indicated in Table 38.4 by an exclusive-or across two or three of *F*, *V*, and *H*.

In BT.601-4 (1994) and in SMPTE RP 125-1992, in 480*i* systems, an SAV with $V=0$ could occur prior to the first active (picture) line – as early as line 10 or line 273. To be compatible with legacy equipment, do not rely upon the 1-to-0 transition of *V*.

SMPTE standards are inconsistent in their numbering of words outside the active region. EAV functions as the start of a digital line with regard to state changes to the F and V bits, so I number words from 0 at EAV. In this scheme, SAV starts at word $S_{TL}-S_{AL}-4$. Another reason for numbering EAV as word 0 is that the proposed SMPTE standard for HD-SDTI anticipates a scheme to advance the timing of SAV codes. Word and sample numbering is strictly notational: Neither word nor sample numbers appear at the interface.

SMPTE ST 348, *High Data-Rate Serial Data Transport Interface (HD-SDTI).*

The horizontal blanking interval at the interface, from EAV to SAV, can contain ancillary data (HANC). In each active (picture) line, the interval from SAV to EAV contains active video. Outside the active picture lines, the interval between SAV and EAV can be used for ancillary data (VANC) packets. If a line outside the active picture is not carrying VANC, and the line isn't associated with analog sync elements, the interval from SAV to EAV can carry a digitized ancillary signal coded like active video. Intervals not used for EAV, SAV, active video, digitized ancillary signals, or ancillary (ANC) data are filled by alternating codes {chroma 200_h, luma 40_h}, which, in active picture, would represent blanking.

TRS in HD-SDI

HD-SDI is similar to 4:2:2 SD SDI; however, the single link carries two logical streams, one carrying chroma, the other carrying luma. Each stream has TRS sequences; independent ANC packets can be carried in each stream. The two streams are word multiplexed; the multiplexed stream is serialized and scrambled. Four words indicated in Table 38.5 at the top of the facing page are appended to each EAV. Each bit 9 is the complement of bit 8. Words 4 and 5 convey line number (LN0, LN1). Words 6 and 7 provide CRC protection for the stream's active video. Each stream has a CRC generator that implements a characteristic function $x^{18}+x^5+x^4+1$. Each generator is reset to zero immediately after SAV, and accumulates words up to and including LN1.

Word	Value	MSB 9	8	7	6	5	4	3	2	1	LSB 0
4	*LN0*	$\overline{L}_6$	L_6	L_5	L_4	L_3	L_2	L_1	L_0	**0**	**0**
5	*LN1*	**1**	**0**	**0**	**0**	L_{10}	L_9	L_8	L_7	**0**	**0**
6	*CR0*	$\overline{CRC}_8$	CRC_8	CRC_7	CRC_6	CRC_5	CRC_4	CRC_3	CRC_2	CRC_1	CRC_0
7	*CR1*	$\overline{CRC}_{17}$	CRC_{17}	CRC_{16}	CRC_{15}	CRC_{14}	CRC_{13}	CRC_{12}	CRC_{11}	CRC_{10}	CRC_9

Table 38.5 **Line number and CRC in HD-SDI** comprises four words immediately following EAV; the package is denoted EAV+LN+CRC. Bit 9 of each word is the complement of bit 8. Line number is coded in 11 bits, L_{10} through L_0, conveyed in two words. The CRC covers the first active sample through the line number. An HD-SDI interface conveys two streams, each including EAV+LN+CRC and SAV sequences; one stream carries chroma-aligned words, the other carries luma-aligned words.

Analog sync and digital/analog timing relationships

In analog interlaced video, 0_V denotes the start of either field. In digital video, 0_V for the second field is unimportant; some people use 0_V to denote the start of a frame.

In analog video, sync is conveyed by video levels "blacker than black." Line sync is achieved by associating, with every scan line, a line sync (horizontal) datum denoted 0_H (pronounced *zero-H*) defined at the midpoint of the leading (falling) edge of sync. Field and frame sync is achieved by associating, with every field, a vertical sync datum denoted 0_V (pronounced *zero-V*).

Sync separation recovers the significant timing instants associated with an analog video signal. *Genlock* reconstructs a sampling clock.

The relationship between the digital and analog domains is established by the position of 0_H with respect to some TRS element. Table 38.6 overleaf summarizes several standards for digital representation of component 4:2:2 video; the rightmost column gives the number of luma sample intervals between the first word of EAV and the 0_H sample (if it were digitized).

Ancillary data

To determine whether a line that is a candidate for a digitized ancillary signal actually contains such a signal, examine every chroma/luma pair for {200_h, 40_h}: If any pair is unequal to these values, a digitized ancillary signal is present.

In 4:2:2 SD, and in HD, ancillary data is permitted immediately after any EAV (HANC), or immediately after SAV (VANC) on a line containing neither active picture nor digitized ancillary data. (In 576*i*, ancillary data is limited to lines 20 through 22 and 333 through 335.) An ancillary packet is introduced by an ancillary data flag (ADF) comprising the three-word sequence {0, $3FF_h$, $3FF_h$}.

System	*AR*	*Scanning*	*Standard*	S_{TL}	S_{AL}	*EAV to* 0_H
483*i*29.97		525/59.94/2:1	SMPTE 125M, BT.601	858	720	12
576*i*25		625/50/2:1	EBU 3246, BT.601	864	720	16
483*i*29.97	16:9	525/59.94/2:1	SMPTE 267M, BT.601	1144	960	16
483*p*59.94	16:9	525/59.94/1:1	SMPTE 293M	1144	960	16
576*i*25	16:9	625/50/2:1	EBU 3246, BT.601	1152	960	21
720p60	16:9	750/60/1:1	SMPTE ST 296	1650	1280	110
1080*i*30	16:9	1125/60/2:1	SMPTE ST 274	2200	1920	88
1080*p*30	16:9	1125/30/1:1	SMPTE ST 274	2200	1920	88
1080*p*25	16:9	1125/25/1:1	SMPTE ST 274	2640	1920	192
1080*p*24	16:9	1125/24/1:1	SMPTE ST 274	2750	1920	192

Table 38.6 **Digital to analog timing relationships.** for several scanning standards are summarized. The rightmost column relates TRS to 0_H; it gives the count of luma sample intervals from EAV word 0 ($3FF_h$) to the 0_H sample (if it were digitized).

SMPTE ST 291, *Ancillary Data Packet and Space Formatting.*

An ANC packet must not interfere with active video, or with any TRS, SAV, or EAV. Multiple ANC packets are allowed, provided that they are contiguous. Certain ANC regions are reserved for certain purposes; consult SMPTE ST 291.

An ancillary packet comprises the three-word (4:2:2) or one-word ($4f_{SC}$) ADF, followed by these elements:

- A one-word *data ID* (DID)
- A one-word *data block number* (DBN) or *secondary DID* (SDID)
- A one-word *data count* (DC), from 0 to 255
- Zero to 255 *user data words* (UDW)
- A one-word checksum (CS)

Each header word – DID, DBN/SDID, and DC – carries an 8-bit value. Bit 8 of each header word is parity, asserted if an odd number of bits 7 through 0 is set, and deasserted if an even number of bits is set. Bit 9 is coded as the complement of bit 8. (Codewords having 8 MSBs all-zero or all-one are thereby avoided; this prevents collision with the 0_h and $3FF_h$ codes used to introduce TRS and ANC sequences.)

Two types of ANC packet are differentiated by bit 7 of the DID word. If DID_7 is asserted, the packet is Type 1; DID is followed by data block number (DBN). There are 128 DID codes available for Type 1 packets.

The DBN value indicates continuity: If zero, it is inactive; otherwise, it counts packets within each DID from 1 through 255, modulo 255.

If DID_7 is negated, the packet is Type 2: The DID is followed by a secondary data ID (SDID), giving $127 \cdot 255$ (i.e., 32,385) ID codes for Type 2 packets.

Three DID values, 004_h, 008_h, and $00C_h$ indicate a Type 2 ANC packet coded with 8-bit data; other DID values in the range 001_h through $00F_h$ are prohibited. DID 80_h marks a packet for deletion. DID 84_h marks the last ANC packet in a VANC or HANC region.

The data count (DC) word contains a value from 0 through 255 (protected by two parity bits), indicating the count of words in the user data area. The DC word spans all ten bits of the interface. Even if an 8-bit DID is indicated, SMPTE standards imply that the two least significant bits of the DC word are meaningful. (If they were not, then the count of user data words could not be uniquely determined.)

The checksum (CS) word provides integrity checking for the contents of an ancillary packet. In every word from DID through the last word of UDW, the MSB is masked out (to zero); these values are summed modulo 512. The 9-bit sum is transmitted in bits 8 through 0 of CS; bit 9 is coded as the complement of bit 8.

SDI coding

In the serial interface it is necessary for a receiver to recover the clock from the coded bitstream. The coded bitstream must therefore contain significant power at the coded bit rate. Coaxial cable attenuates high-frequency information; equalization is necessary to overcome this loss. Because equalizers involve high-frequency AC circuits, the coded bitstream should contain little power at very low frequencies: The code must be *DC-free*. To enable economical equalizers, the frequency range required for correct recovery of the signal should be as small as possible. A ratio of the highest to lowest frequency components of about 2:1 – where the coded signal is contained in one octave of bandwidth – is desirable. These considerations argue for a high clock rate. But it is obviously desirable to have

Details of the application of SDI in the studio are found in Chapter 7 of Robin, Michael, and Michel Poulin (2000), *Digital Television Fundamentals: Design and Installation of Video and Audio Systems,* Second Edition (New York: McGraw-Hill).

a low clock rate so as to minimize cost. The choice of a clock rate is a compromise between these demands.

SDI uses scrambled coding, where the data stream is serialized, then passed through a shift register arrangement with exclusive-or taps implementing a characteristic function $x^9 + x^4 + 1$.

Scrambling techniques using a single scrambler are well known. But the SDI and HD-SDI scrambler has a second-stage scrambler, whose characteristic function is $x + 1$. The two cascaded stages offer improved performance over a conventional single-stage scrambler. The scrambling technique is self-synchronizing; there is no need for initialization.

The data rate at the interface is the word rate times the number of bits per word. It is standard to serialize 10-bit data; when coding 8-bit video, the two LSBs are forced to zero.

No provision is made to avoid data sequences that would result, after serialization and scrambling, in serial bit sequences with long runs of zeros or long runs of ones. But a long run of zeros or ones provides no signal transitions to enable a receiver to recover the clock! In practice, such *pathological sequences* are rare.

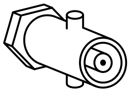


Figure 38.4
BNC connector

SDI is standardized for electrical transmission at ECL levels through coaxial cable, using a BNC connector as depicted in Figure 38.4. Distances between 200 m and 400 m are practical. SDI is also standardized for transmission through optical fiber. Fiber-optic interfaces for digital SD are straightforward adaptations of SDI.

HD-SDI coding

SMPTE ST 292, *Bit-Serial Digital Interface for High-Definition Television Systems.*

The SDI interface at 270 Mb/s was adapted to HD by scaling the bit rate by a factor of 5.5, yielding a bit rate of 1.485 Gb/s (or in 24.976, 29.97, or 59.94 Hz systems, $^{1.485}/_{1.001}$ Gb/s) – call these both “1.5 Gb/s.” HD-SDI is standardized in SMPTE ST 292. The interface is modeled after SD SDI, but there are two significant changes:

- Chroma and luma are encoded in separate streams, each with its own TRS sequence. The streams are multiplexed, then scrambled.
- Coded line number and a CRC are appended to the EAV portion of the TRS sequence (giving what is called EAV+LN+CRC).

In 1080*i* and 1080*p* standards documents, samples are numbered with respect to 0_H (unlike SD standards documents, where samples are numbered with respect to the zeroth active sample of the line).

Further developments doubled the data rate of the 1.5 Gb/s interface to about 3 Gb/s (sometimes termed 3G-SDI). That rate is sufficient to carry 1080*i*30 at 4:4:4:4 (that is, no chroma subsampling and an alpha component) or 1080*p*60 at 4:2:2.

Interfaces for compressed video

Compressed digital video interfaces are impractical in the studio owing to the diversity of compression systems, and because compressed interfaces would require decompression capabilities in signal processing and display equipment. Compressed 4:2:2 digital video studio equipment is usually interconnected through uncompressed SDI interfaces.

Compressed interfaces can be used to transfer video into nonlinear editing systems, and to "dub" (duplicate) between VTRs sharing the same compression system. Compressed video can be interfaced directly using *serial data transport interface* (SDTI), to be described in a moment. The DVB ASI interface is widely used to convey MPEG-2 transport streams in network or transmission applications (but not in production). The IEEE 1394/DV interface, sometimes called *FireWire* or *i.LINK*, is widely used in the consumer electronics arena, and is beginning to be deployed in broadcast applications.

SDTI

SMPTE ST 305.2, *Serial Data Transport Interface*.

SMPTE has standardized a derivative of SDI, *serial data transport interface* (SDTI), that transmits arbitrary data packets in place of uncompressed active video. SDTI can be used to transport DV25, DV50, DV100, and Sony MPEG IMX compressed datastreams. Despite DV bitstreams being standardized, different manufacturers have chosen incompatible techniques to wrap their compressed video data into SDTI streams. This renders SDTI useful only for interconnection of equipment from a single manufacturer.

Switching and mixing

SMPTE RP 168, *Definition of Vertical Interval Switching Point for Synchronous Video Switching*.

Switching or editing between video sources – "cutting" – is done in the vertical interval, so that each frame of the resulting video remains intact, without any switching transients. When switching between two signals in a hardware switcher, if the output signal is to be made continuous across the instant of switching, the input signals must be synchronous – the 0_V instants of both signals must match precisely in time. To prevent switching transients from disturbing vertical sync elements, switching is done somewhat later than 0_V; see SMPTE RP 168.

Timing in digital facilities

FIFO: First in, first out.

Modern digital video equipment has, at each input, a buffer that functions as a FIFO. The buffer at each input accommodates an advance of timing at that input (with respect to reference video) of up to several line times. Timing a digital facility involves advancing each signal source so that signals from all sources arrive in time at the inputs of the facility's main switcher. This timing need not be exact; it suffices to guarantee that no buffer overruns or underruns. When a routing switcher switches among SDI streams, a timing error of several dozen samples is tolerable; downstream equipment will recover timing within one or two lines after the instant of switching.

When a studio needs to accommodate an asynchronous video input – one whose frame rate is within tolerance, but whose phase cannot be referenced to house sync, such as a satellite feed – then a *framestore synchronizer* is used. This device contains a frame of memory that functions as a FIFO buffer for video. An input signal with arbitrary timing is written into the memory with timing based upon its own sync elements. The synchronizer accepts a reference video signal; the memory is read out at rates locked to the sync elements of the reference video. (Provisions are made to adjust SYSTEM PHASE – that is, the timing of the output signal with respect to the reference video.) An asynchronous signal is thereby delayed up to one frame time, perhaps even a little more, so as to match the local reference. The signal can then be used as if it were a local source.

Some video switchers incorporate digital video effects (DVE) capability; a DVE unit necessarily includes a framestore.

Studio video devices commonly incorporate framestores, and exhibit latency of a field, a frame, or more. Low-level timing of such equipment is accomplished by introducing time advance so that 0_V appears at the correct instant. However, even if video content is timed correctly with respect to 0_V, it may be late by a frame, or in a very large facility, by several frames. Attention must be paid to delaying audio by a similar time interval, to avoid lip-sync problems.

ASI

ETSI EN 50083-9, *Cable networks for television signals, sound signals and interactive services – Part 9: Interfaces for CATV/SMATV headends and similar professional equipment for DVB/MPEG-2 transport streams*.Standards are not clear on whether transformer coupling is required or whether capacitive coupling suffices.

Within a broadcast facility, an MPEG-2 transport stream can be serialized onto a dedicated *asynchronous serial interface* (ASI). A serialized ASI stream for broadcast has a payload bit rate of around 20 Mb/s; however, the ASI interface bit rate is 270 Mb/s, chosen so that SDI distribution infrastructure can be used. The ASI interface uses BNC connectors and coaxial cable. ASI is polarity sensitive (unlike SDI), though modern ASI receivers typically detect and correct polarity inversion.

Some people write 8B/10B; however, the elements involved are bits, not bytes, so lowercase *b* is apt.

Although the SDI physical layer is used, the serialized ASI stream has no TRS codes and the interface does not use SDI scrambling. Instead, channel data is encoded according to the 8b/10b scheme borrowed from Fibre Channel standards. (ASI interface data rate is therefore at most 216 Mb/s.) An 8b/10b bitstream never has more than four consecutive 0s or 1s, so clock recovery is simple. An 8b/10b encoder minimizes low frequency ("DC") content on the media.

The *synchronous serial interface* (SSI) was standardized by SMPTE for the purpose of conveying MPEG transport streams between equipment, but SSI has largely fallen into disuse.

Since the ASI payload rate is typically far lower than the channel capacity, stuffing codes are inserted to occupy idle time. Stuffing codes – Fibre Channel *comma* codes, denoted *K28.5* – are inserted either at the earliest opportunity (bytewise, "spaced byte mode"), or at the completion of the current packet (packetwise, "burst mode"). MPEG packets are separated by at least two comma codes.

It is increasingly common to convey transport streams using IP protocols across Ethernet.

Summary of digital interfaces

Table 38.7 summarizes SD and HD digital interface standards.

ITU-R Rec. BT.656, *Interfaces for digital component video signals in 525-line and 625-line television systems operating at the 4:2:2 level of Recommendation ITU-R BT.601*.

SMPTE 125M, *Component Video Signal 4:2:2 – Bit-Parallel Digital Interface*.

SMPTE 259M, *10-Bit 4:2:2 Component and* $4f_{SC}$ *Composite Digital Signals – Serial Digital Interface*.

SMPTE 267M, *Bit-Parallel Digital Interface – Component Video Signal 4:2:2 16×9 Aspect Ratio*.

SMPTE ST 292, *Bit-Serial Digital Interface for High-Definition Television Systems*.

SMPTE ST 297, *Serial Digital Fiber Transmission System for ANSI/SMPTE 259M Signals*.

Table 38.7 **SD and HD interface standards**

480*i* component video

39

This chapter details the scanning, timing, sync structure, and picture structure of 480*i*29.97 (525/59.94/2:1) video. The scanning and timing information in this chapter applies to all variants of 480*i* video, both analog and digital. The sync information relates to component analog, composite analog, and composite digital systems.

Frame rate

$$f_{FR} = \frac{30}{1.001} \approx 29.97\ \text{Hz}$$

480*i* video represents stationary or moving two-dimensional images sampled temporally at a constant rate of $^{30}/_{1.001}$ frames per second. For studio video, the tolerance on frame rate is normally ±10 ppm. In practice the tolerance applies to a master clock at a high frequency, but for purposes of computation and standards writing, it is convenient to reference the tolerance to the frame rate.

Interlace

$$f_{H} = \frac{9}{0.572} \approx 15.734\ \text{kHz}$$

It is confusing to refer to fields as *odd* and *even*. Use *first field* and *second field* instead.

A frame is conveyed as a sequence of 525 horizontal raster lines of equal duration, uniformly scanned top to bottom and left to right. Scanning has 2:1 interlace to form a *first* field and a *second* field; scan lines in the second field are displaced vertically by half the vertical sampling pitch, and delayed temporally by half the frame time, from scanning lines in the first field. In MPEG-2 terms, the first field is the bottom field.

Lines are numbered consecutively throughout the frame, starting at 1.

Table 39.1 480*i* line assignment

EQ	Equalization pulse
BR	Broad pulse
CC	Closed caption
[*n*]	Line number relative to start of second field (deprecated)

Line number, first field ($F=0$)	*Line number, second field* ($F=1$)	*V*	*Contents, left half*	*Contents, right half*
	266 [3]		EQ	BR
4			BR	BR
	267 [4]		BR	BR
5			BR	BR
	268 [5]		BR	BR
6			BR	BR
	269 [6]		BR	EQ
7			EQ	EQ
	270 [7]		EQ	EQ
8			EQ	EQ
	271 [8]		EQ	EQ
9			EQ	EQ
	272 [9]		EQ	none
10–19		$V=0$§	Vertical interval video (10 lines)	
	273–282 [10–19]		Vertical interval video (10 lines)	
20		$V=0$ (487 lines)	Vertical interval video	
	283 [20]		Vertical interval video	
21			CC	
	284 [21]		CC	
22			Picture	
	285 [22]		Picture	
23–261			Picture (239 lines)	◊
	286–524 [23–261]		Picture (239 lines)	
262			Picture	
	525 [262]		Picture	
263			Picture	EQ
	‡1		EQ	EQ
‡264 [1]			EQ	EQ
	‡2		EQ	EQ
‡265 [2]			EQ	EQ
	‡3		EQ	EQ

§ $V=0$ in RP 125-1992, and in the 480*i* version of BT.601-4 (1994); in later standards, $V=1$ for these lines.

◊ The thick vertical bar at the right indicates lines carried in 480*i or* 480*p* MPEG-2 according to SMPTE RP 202. (The vertical center of the picture is located midway between lines 404 and 142.) Unfortunately, 480*i* DV systems digitize a range one image row up from this.

‡ In analog terminology, lines 1 through 3 are considered part of the first field; lines 264 and 265 are considered part of the second field.

Table 39.1 opposite shows the vertical structure of a frame in 480*i* video, and indicates the assignment of line numbers and their content.

Concerning closed captions, see ANSI/EIA/CEA-608-B, *Line 21 Data Services*.

In legacy equipment, the picture may start as early as line 20 or line 283. However, video on lines 21 and 284 is liable to be replaced by line 21 closed caption data upon NTSC transmission, so it is pointless to provide more than 483 picture lines in the studio. With the wide use of 480*i* DV and MPEG-2 systems, I argue that it is pointless to provide more than 480 lines; however, 483 lines were broadcast in analog NTSC.

For details concerning VITC line assignment, see SMPTE RP 164, *Location of Vertical Interval Timecode.*

Lines 10 through 21 and 273 through 284 may carry ancillary ("vertical interval") signals either related or unrelated to the picture. If *vertical interval timecode* (VITC) is used, it should be located on line 14 (277); a second, redundant copy can be placed on line 16 (279). Failing line 14, line 18 (281) is suggested. See *Vertical interval timecode (VITC),* on page 402.

Line sync

Horizontal events are referenced to an instant in time denoted 0_H. In the analog domain, 0_H is defined by the 50%-point of the leading (negative-going) edge of each line sync pulse. In a component digital interface, the correspondence between sync and the digital information is determined by a *timing reference signal* (TRS) conveyed across the interface. (See *SDI and HD-SDI sync, TRS, and ancillary data,* on page 433.)

In an analog interface, every line commences at 0_H with the negative-going edge of a sync pulse. With the exception of vertical sync lines, which I will describe in a moment, each line commences with a *normal* sync pulse, to be described. Each line that commences with normal sync may contain video information. Every line that commences with a sync pulse other than normal sync maintains blanking level, except for the intervals occupied by sync pulses.

Field/frame sync

To define vertical sync, the frame is divided into intervals of halfline duration. Each halfline either contains no sync information or commences with the assertion of a sync pulse having one of three durations, each having a tolerance of ±0.100 µs:

EIA and FCC standards in the United States rounded the equalization pulse duration to two digits, to 2.3 μs, slightly less than the theoretical value of 2.35 μs. Equipment is usually designed to the letter of the regulation, rather than its intent.

- A *normal* sync pulse having a duration of 4.7 μs
- An *equalization* pulse having half the duration of a normal sync pulse
- A *broad* pulse, having a duration of half the line time less the duration of a normal sync pulse

Each set of 525 halflines in the field commences with a vertical sync sequence as follows:

- Six preequalization pulses
- Six broad pulses
- Six postequalization pulses

Line 263 commences with a normal sync pulse and has an equalization pulse halfway through the line. Line 272 commences with an equalization pulse and remains at blanking with no sync pulse halfway through the line. For analog details, see Chapter 2, *Analog SD sync, genlock, and interface,* of *Composite NTSC and PAL: Legacy Video Systems*.

Vertical events are referenced to an instant in time denoted 0_V. In the analog domain, 0_V is defined by the first equalization pulse coincident with 0_H. Line number 1 is signalled by 0_V; lines count in interlaced time order (not spatial order) throughout the frame. 480*i* systems are exceptional in identifying 0_V and line 1 at the first equalization pulse: In 576*i*, and in HD, 0_V and line 1 are marked at the first broad pulse.

Historically, in the analog domain, 0_V was defined for each field by the 50%-point of the first equalization pulse; lines were numbered from 1 to 263 in the first field and from 1 to 262 in the second field. In the digital domain, the first field contains 262 lines and the second field contains 263 lines.

Figure 39.1 opposite shows details of the sync structure; this waveform diagram is the analog of Table 39.1, *480i line assignment,* on page 446.

When sync is represented in analog or digitized $4f_{SC}$ form, a raised-cosine transition having a risetime (from 10% to 90%) of 140±20 ns is imposed; the midpoint of the transition is coincident with the idealized sync.

SMPTE RP 168, *Definition of Vertical Interval Switching Point for Synchronous Video Switching*.

Switching between video sources is performed in the vertical interval, to avoid disruption of sync or picture. Switching occurs 30±5 μs after 0_H of the first normal line of each field. In 480*i* systems, switching occurs midway through line 10. (If field 2 were dominant, switching would occur midway through line 273.)

R'G'B' EOCF and primaries

Picture information is referenced to linear-light primary red, green, and blue (*RGB*) tristimulus values, represented in abstract terms in the range 0 (reference black) to +1 (reference white).

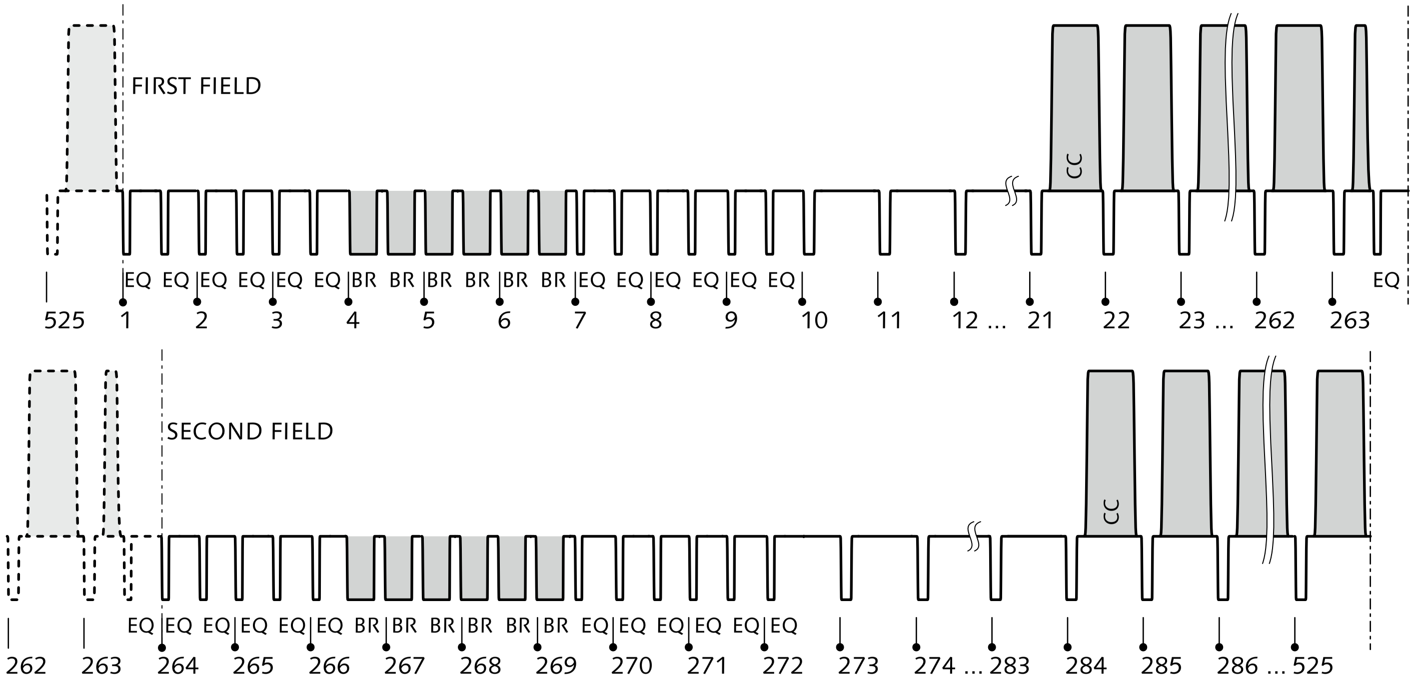


Figure 39.1 **480*i* raster, vertical.** This drawing shows the waveforms of the first and second fields, detailing the vertical sync intervals. CC indicates a line that may contain EIA 608 ("line 21") closed caption data.

Historical NTSC standards documents indicate a "precorrection" with a $^1/_{2.2}$-power (approximately 0.45), but in practice the BT.1886 EOCF is used. See *Gamma in video* on page 318.

Three nonlinear primary components R', G', and B' are computed such that the intended image appearance is obtained on the reference display in the reference viewing conditions; see *Reference display and viewing conditions,* on page 427.

In the default power-up state of a camera, the nonlinear primary components are computed from the camera's *RGB* tristimulus estimates according to the opto-electronic conversion function of *BT.709 OECF*, described on page 320; this process is loosely called *gamma correction*.

The colorimetric properties of the display primaries are supposed to conform to *BT.709 primaries* described on page 290: DTV transmission standards call for BT.709, and modern consumer displays use BT.709. However, production and mastering of 480*i* content historically used SMPTE primary chromaticities, not BT.709 (see *SMPTE RP 145 primaries*, Table 26.5 on page 293).

Luma (Y')

Luma in 480*i* systems is computed as a weighted sum of nonlinear R', G', and B' primary components, according to the luma coefficients of BT.601, as detailed in *BT.601 luma,* on page 346:

$$^{601}Y' = 0.299\,R' + 0.587\,G' + 0.114\,B' \qquad \text{Eq 39.1}$$

The luma component Y', being a weighted sum of nonlinear $R'G'B'$ components, has no simple relationship with CIE relative luminance (Y) used in colour science. Video encoding specifications typically place no upper bound on luma bandwidth (though transmission standards may).

Picture center, aspect ratio, and blanking

SMPTE RP 187, *Center, Aspect Ratio and Blanking of Video Images.*

$$\frac{S_{TL} - S_{EAV\text{-}0_H} + 0.5\left(S_{AL} - 1\right)}{S_{TL}} = \frac{858 - 736 + 0.5\left(720 - 1\right)}{858} = \frac{321}{572}$$

The center of the picture is located midway between the central two of the 720 active samples of BT.601, at the fraction $^{321}/_{572}$ between 0_H instants. Concerning the vertical center, see Table 39.1, on page 446.

In 4:3 systems, the aspect ratio is defined to be 4:3 with respect to a *clean aperture* pixel array, 708 samples wide at a sampling rate of 13.5 MHz, and 480 lines high.

In *Transition samples*, on page 378, I mentioned that it is necessary to avoid, at the start of a line, an instantaneous transition from blanking to picture information. SMPTE standards call for picture information to have a risetime of 140±20 ns. For 480*i* or 576*i* video, a blanking transition is best implemented as a three-sample sequence where the video signal is limited in turn to 10%, 50%, and 90% of its full excursion.

No studio standard addresses square sampling of 480*i* video. I recommend using a sample rate of $780 f_H$, that is, 12 $^3/_{11}$ MHz (i.e., $12.272\overline{727}$ MHz). I recommend using 648 samples – or, failing that, 644 or 640 – centered as mentioned above.

SMPTE RP 202, *Video Alignment for MPEG-2 Coding.*

When MPEG-2 with 480 or 512 image rows is used in the studio, the bottom image row corresponds to line 525 (as indicated in Table 39.1). The bottom left-hand halfline (on line 263) is not among the coded image rows. Unfortunately, 480*i* DV systems digitize a range one image row up from this.

Halfline blanking

Most component video equipment treats the top and bottom lines of both fields as integral lines; blanking of halflines is assumed to be imposed at the time of conversion to analog. In composite equipment and analog equipment, halfline blanking must be imposed.

$$30.593 \approx \frac{63.55\overline{5}}{2} - \frac{732-716}{13.5}$$

In the composite and analog domains, video information at the bottom of the picture, on the left half of line 263, should terminate 30.593 µs after 0_H. This timing is comparable to blanking at the end of a full line, but preceding the midpoint between 0_H instants instead of preceding the 0_H instant itself.

$$41.259 \approx \frac{63.55\overline{5}}{2} + \frac{858-732+2}{13.5}$$

Historically, in the composite and analog domains, a right halfline at the top of the picture – such as picture on line 284 – commenced about 41 µs after 0_H. This timing is comparable to blanking at the start of a full line, but following the midpoint between 0_H instants instead of following the 0_H instant itself. However, in NTSC broadcast, line 284 must remain available for closed captioning (along with line 21). So, it is now pointless for studio equipment to carry the traditional right-hand halfline of picture on line 284: Picture should be considered to comprise 482 full lines, plus a left-hand halfline on line 263.

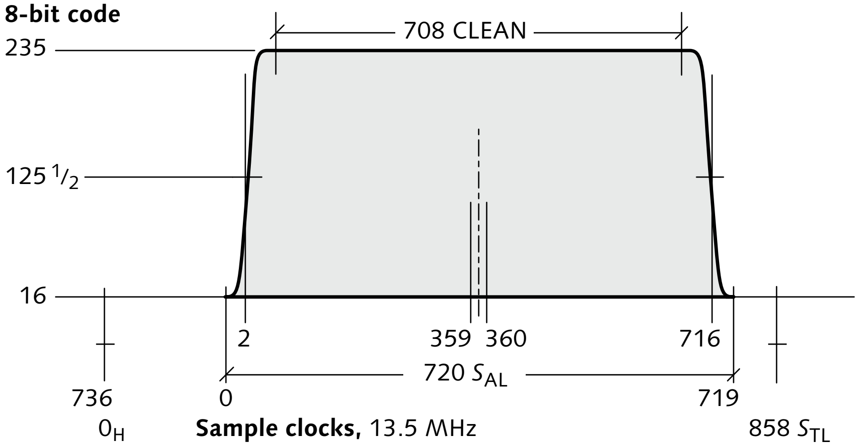


Figure 39.2 **480*i* component digital 4:2:2 luma waveform**

Halfline blanking has been abolished from progressive scan video, and from JPEG, MPEG, and HD.

Component digital 4:2:2 interface

SDI was introduced on page 432. Eight-bit SD interfaces between digital ICs are often described as "601" (where horizontal and vertical sync signals are conveyed on dedicated wires) or "656" (where sync is embedded as TRS codes).

The C_B and C_R colour difference components of digital video are formed by scaling B'-Y' and R'-Y' components, as described in *C_BC_R components for SD* on page 361. $Y'C_BC_R$ signals are usually conveyed through the *serial digital interface* (SDI), which I introduced on page 432. *R'G'B'* 4:4:4 (or *R'G'B'A* 4:4:4:4) components can be conveyed across a dual-link interface using two SDI channels; alternatively, the single-link 540 Mb/s SDI interface of SMPTE 344M can be used.

In 13.5 MHz sampling of 480*i*, the sample located 16 sample clock intervals after EAV corresponds to the line sync datum (0_H): If digitized, that sample would take the 50% value of analog sync.

Figure 39.2 above shows a waveform drawing of luma in a 480*i* component digital 4:2:2 system.

Component analog *R'G'B'* interface

A component analog 480*i* *R'G'B'* interface is based on nonlinear *R'*, *G'*, and *B'* signals. Analog *R'*,*G'*, *B'* signals are conveyed as voltage, with a range of 1 V from synctip to reference white. (Transient excursions slightly above reference white are permitted.)

SMPTE 253M, *Three-Channel RGB Analog Video Interface.*

In studio systems, analog component *R'G'B'* signals ideally have zero setup, so zero in Equation 39.2 corresponds to 0 V_{DC}. According to SMPTE 253M, unity corresponds to 700 mV. Sync is added to the green component according to Equation 39.2, where *sync* and *active* are taken to be unity when asserted and zero otherwise:

$$G'_{\mathrm{sync}} = \frac{7}{10}\left(active \cdot G'\right) + \frac{3}{10}\left(-sync\right) \qquad \text{Eq 39.2}$$

Sadly, the SMPTE *R'G'B'* analog interface is unpopular, and "NTSC-related" levels are usually used, either with or without setup.

Some systems, such as 480*i* studio video in Japan, use a picture-to-sync ratio of 10:4 and zero setup. In this case, unity in Equation 39.3 corresponds to $^5\!/_7$ V, about 714 mV:

$$V'_{\mathrm{sync}} = \frac{5}{7}\left(active \cdot V'\right) + \frac{2}{7}\left(-sync\right) \qquad \text{Eq 39.3}$$

Many systems – such as computer framebuffers using the levels of the archaic EIA RS-343-A standard – code component video similarly to composite video, with 10:4 picture-to-sync ratio and 7.5% setup:

$$V'_{\mathrm{sync}} = \frac{3}{56} active + \frac{37}{56}\left(active \cdot V'\right) + \frac{2}{7}\left(-sync\right) \qquad \text{Eq 39.4}$$

Component analog $Y'P_BP_R$ interface, EBU N10

The P_B and P_R scale factors are appropriate only for component analog interfaces. For details concerning scale factors in component digital systems, see C_BC_R *components for SD*, on page 361. For details concerning scale factors in composite analog or digital NTSC or PAL, see *UV components*, in Chapter 5 of *Composite NTSC and PAL: Legacy Video Systems*.

The P_B and P_R colour difference components of analog video are formed by scaling $B'-Y'$ and $R'-Y'$ components, as described in P_BP_R *components for SD* on page 359. Wideband P_B and P_R components are theoretically possible but very rarely used; normally, P_B and P_R are lowpass filtered to half the bandwidth of luma.

Component $Y'P_BP_R$ signals in 480*i* are sometimes interfaced with zero setup, with levels according to the EBU Tech. N10 standard. Zero (reference blanking level) for Y', P_B, and P_R corresponds to a level of 0 V_{DC}, and unity corresponds to 700 mV. Sync is added to the luma

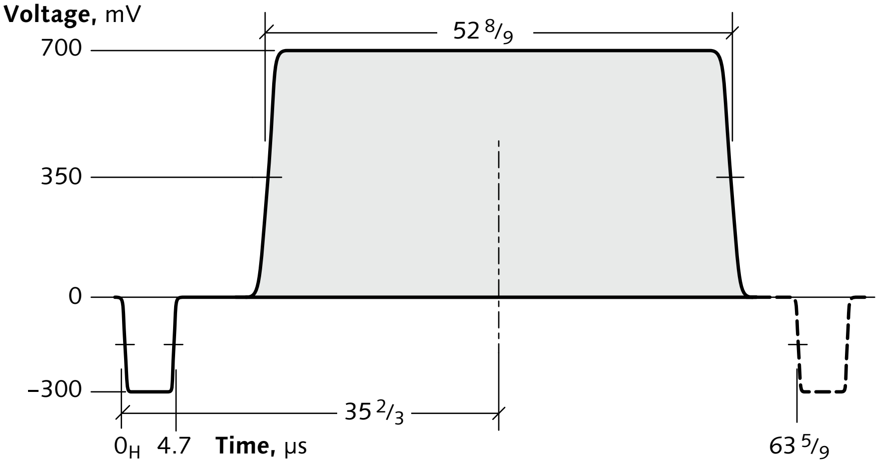


Figure 39.3 **480*i* component analog luma waveform** with SMPTE levels and zero setup.

component; *sync* is taken to be unity when asserted and zero otherwise:

$$Y'_{\text{sync}} = \frac{7}{10} Y' + \frac{3}{10}\left(-sync\right) \qquad \text{Eq 39.5}$$

Figure 39.3 shows a waveform drawing of luma in a 480*i* component analog interface according to the EBU Tech. N10 standard. In North America, the levels of EBU N10 mysteriously became known as "SMPTE levels," or "SMPTE/EBU N10 interface," even though N10 is solely an EBU standard and SMPTE failed to standardize a component analog luma/colour difference interface.

EIA/CEA-770.2, Standard Definition TV Analog Component Video Interface.

CEA has standardized the 700 mV, zero-setup levels for use by consumer electronics devices such as DVD players and set-top boxes, for 480*i* and 480*p* formats at 4:3 and 16:9 aspect ratios.

CEA/CEDIA-863-B, *Connection Color Codes for Home Theater Systems.*

In 2011, about a quarter of a century after the introduction of component analog video interfaces, CEA (with CEDIA, Custom Electronic Design and Installation Association) standardized the colours of the connectors to be used: green for Y', blue for C_B, and red for C_R. (In the consumer domain, composite NTSC or PAL video is typically carried on a wire having yellow connectors.)

Component analog $Y'P_BP_R$ interface, industry standard

Unfortunately, equipment from two manufacturers was deployed before SMPTE reached agreement on a standard component video analog interface for studio use. Although it is sometimes available as an option, the SMPTE standard is rarely used in 480*i*. Instead, two "industry" standards are in use: Sony and Panasonic. Ideally the $Y'P_BP_R$ nomenclature would signify that luma has zero setup, and that colour difference components have the same excursion (from black to white) as luma. However, both of the industry standards use setup, and neither gives the colour difference components the same excursion as luma.

Details are found in *Luma/colour difference component sets*, on page 352.

Sony SD equipment utilized 10:4 picture-to-sync ratio (roughly 714 mV luma, 286 mV sync) with 7.5% setup on luma (giving a picture excursion of $660\,{}^5\!/_7$ mV). Colour differences range ${}^4\!/_3$ times ±350 mV, that is, $\pm 466\,{}^2\!/_3$ mV. (75% colourbars have a P_BP_R excursion of ±350 mV.)

Panasonic SD equipment utilized 7:3 picture-to-sync ratio (exactly 700 mV luma, 300 mV sync) with 7.5% setup on luma (giving a picture excursion of 647.5 mV). Colour differences are scaled by the ${}^{37}\!/_{40}$ setup fraction, for an excursion of ±323.75 mV.

576*i* component video 40

This chapter details the scanning, timing, sync structure, and picture structure of 576*i*25 (625/50/2:1) video. The scanning and timing information here applies to all variants of 576*i*25 video, both analog and digital. The sync information relates to component analog, composite analog, and composite digital systems. I assume that you are familiar with 480*i* component video, described on page 445.

Frame rate

576*i* video represents stationary or moving two-dimensional images sampled temporally at a constant rate of 25 frames per second. For studio video, the tolerance on frame rate is normally ±4 ppm. In practice the tolerance applies to a master clock at a high frequency, but for purposes of computation and standards writing it is convenient to reference the tolerance to the frame rate.

Interlace

The derived line rate is 15.625 kHz.

It is confusing to refer to fields in 576*i* as *odd* and *even*. Use *first field* and *second field* instead.

A frame comprises a total of 625 horizontal raster lines of equal duration, uniformly scanned top to bottom and left to right with 2:1 interlace to form a *first* field and a *second* field. Scanning lines in the second field are displaced vertically by half the vertical sampling pitch, and delayed temporally by half the frame time, from scanning lines in the first field. In MPEG-2 terms, the first field is the top field.

Lines are numbered consecutively throughout the frame, starting at 1.

Table 40.1 **576*i* line assignment**

EQ	Equalization pulse
BR	Broad pulse
‡	Burst suppressed if –135° phase
§	Burst suppressed unconditionally
¶	In $4f_{SC}$ PAL, line recommended for 1137 S_{TL} (“reset”)

Line number, first field (F = 0)	*Line number, second field (F = 1)*	*V*	*Contents, left half*	*Contents, right half*
	¶313		EQ	BR
1			BR	BR
	314		BR	BR
2			BR	BR
	315		BR	BR
3			BR	EQ
	316		EQ	EQ
4			EQ	EQ
	317		EQ	EQ
5			EQ	EQ
	318		EQ	none
‡6			Vertical interval video	
	‡319		Vertical interval video	
7–18			Vertical interval video	
	320–331		Vertical interval video	
19			VITC	
	332		VITC	
†20			Vertical interval video	
	†333		Vertical interval video	
†21			VITC	
	†334		VITC	
†22			Quiet	
	†335		Quiet	
23		$V = 0$ (576 lines)	WSS	Picture ◊
	336–622		Picture (287 lines)	
24–‡310			Picture (287 lines)	
	§623		Picture	EQ
311			EQ	EQ
	624		EQ	EQ
312			EQ	EQ
	¶625		EQ	EQ

† VANC is permitted only on lines 20 through 22 and 333 through 335.

◊ The thick vertical bar at the right indicates lines carried in 576*i or* 576*p* MPEG-2 according to SMPTE RP 202. (The vertical center of the picture is located midway between lines 479 and 167.) Unfortunately, 576*i* DV systems digitize a range one image row up from this.

For details concerning VITC in 576*i*, see EBU Technical Standard N12, *Time-and-control codes for television recording.*

Table 40.1 opposite shows the vertical structure of a frame in 576*i* video, and indicates the assignment of line numbers and their content.

Lines 6 through 21 and 319 through 334 may carry ancillary ("vertical interval") signals either related or unrelated to the picture. If *vertical interval timecode* (VITC) is used, redundant copies should be placed on lines 19 (332) and 21 (334); see *Vertical interval timecode (VITC),* on page 402.

Line sync

Horizontal events are referenced to an instant in time denoted 0_H. In the analog domain, 0_H is defined by the 50%-point of the leading (negative-going) edge of each line sync pulse. In a component digital interface, the correspondence between sync and the digital information is determined by a *timing reference signal* (TRS) conveyed across the interface. (See *SDI and HD-SDI sync, TRS, and ancillary data,* on page 433.)

In an analog interface, every line commences at 0_H with the negative-going edge of a sync pulse. With the exception of the vertical sync lines of each field, each line commences with the assertion of a *normal* sync pulse, to be described. Each line that commences with normal sync may contain video information. Every line that commences with a sync pulse *other* than normal sync maintains blanking level, here denoted zero, except for the interval(s) occupied by sync pulses.

Analog field/frame sync

To define vertical sync, the frame is divided into intervals of halfline duration. Each halfline either contains no sync information, or commences with the assertion of a sync pulse having one of three durations, each having a tolerance of ±0.100 μs:

- A *normal* sync pulse having a duration of 4.7 μs
- An *equalization* pulse having half the duration of a normal sync pulse
- A *broad* pulse having a duration of half the line time less the duration of a normal sync pulse

Each set of 625 halflines in the frame is associated with a vertical sync sequence, as follows:

Line 623 commences with a normal sync pulse and has an equalization pulse halfway through the line. Line 318 commences with an equalization pulse and remains at blanking with no sync pulse halfway through the line.

- Five preequalization pulses
- Five broad pulses
- Five postequalization pulses

In analog sync, line 1 and 0_V are defined by the first broad pulse coincident with 0_H; see Figure 2.3 in Chapter 2 of *Composite NTSC and PAL: Legacy Video Systems*. (This differs from the 480*i* convention.)

Figure 40.1 opposite shows the vertical sync structure of 576*i* analog video. This waveform diagram is the analog of Table 40.1, *576i line assignment,* on page 458.

See Table 13.1, on page 132, and Figure Figure 2.2 and Figure 2.3 in Chapter 2 of *Composite NTSC and PAL: Legacy Video Systems*.

Sync in 576*i* systems has several differences from 480*i* sync. There are five preequalization, broad, and postequalization pulses per field (instead of six of each). The frame is defined to start with the field containing the top line of the picture, actually a right-hand halfline. (In 480*i* scanning, the first picture line of a frame is a full line, and the right-hand halfline at the top of the picture is in the second field.)

In 576*i* systems, lines are numbered starting with the first broad sync pulse: preequalization pulses are counted at the end of one field instead of the beginning of the next. This could be considered to be solely a nomenclature issue, but because line numbers are encoded in digital video interfaces, the issue is substantive. In 576*i* systems, lines are always numbered throughout the frame.

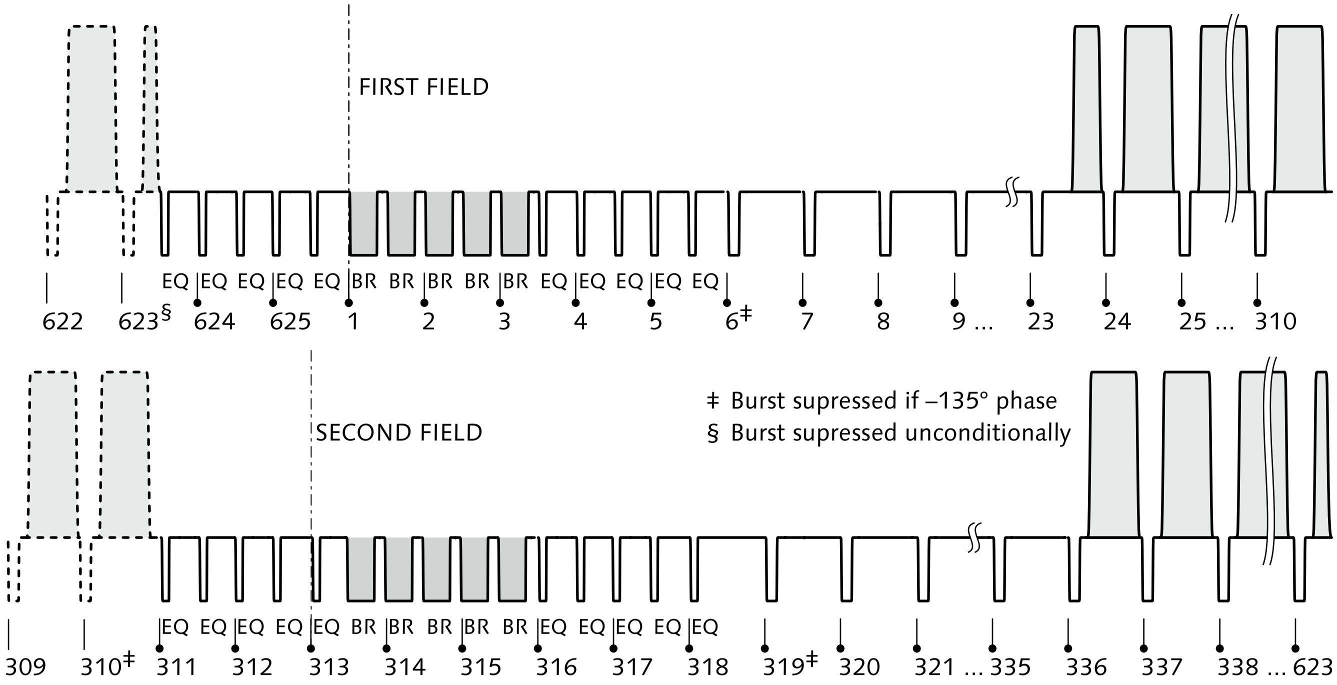
When sync is represented in analog or digitized form, a raised-cosine transition having a risetime (from 10% to 90%) of 200±20 ns is imposed, where the midpoint of the transition is coincident with the idealized sync.

SMPTE RP 168, *Definition of Vertical Interval Switching Point for Synchronous Video Switching*.

Switching between video sources is performed in the vertical interval, to avoid disruption of sync or picture. Switching occurs 30±5 µs after 0_H of the first normal line of each field. In 576*i* systems, switching occurs midway through line 6. (If field 2 were dominant, switching would occur midway through line 319.)

R'G'B' EOCF and primaries

Picture information is referenced to linear-light primary red, green, and blue (*RGB*) tristimulus values, represented in abstract terms in the range 0 (reference black) to +1 (reference white).



‡ Burst supressed if –135° phase
§ Burst supressed unconditionally

Figure 40.1 **576*i* raster, vertical.** This drawing shows the waveforms of the first and second fields, detailing vertical sync intervals. The first field comprises 312 lines, and the second field comprises 313 lines.

Historical PAL standards documents indicate a "precorrection" (OECF) with a $^1\!/_{2.8}$-power (approximately 0.36); however, that value is unrealistic. In practice, the BT.1886 EOCF value of $^1\!/_{2.4}$ is used. See *Gamma in video* on page 318.

Three nonlinear primary components *R'*, *G'*, and *B'* are computed such that the intended image appearance is obtained on the reference display in the reference viewing conditions (see *Reference display and viewing conditions,* on page 427).

In the default power-up state of a camera, the nonlinear primary components are computed from the camera's *RGB* tristimulus estimates according to the opto-electronic conversion function of the BT.709 OECF, described on page 320; this process is loosely called *gamma correction*.

The colorimetric properties of the display primaries are supposed to conform to the BT.709 primaries described on page 290: DTV transmission standards call for BT.709, and modern consumer displays use BT.709. However, production and mastering of 576*i* content historically used EBU primary chromaticities, not BT.709 (see *EBU Tech. 3213 primaries*, Table 26.4 on page 293).

Luma (*Y'*)

Luma in 576*i* systems is computed as a weighted sum of nonlinear *R'*, *G'*, and *B'* primary components according to the luma coefficients of BT.601:

$$^{601}Y' = 0.299\,R' + 0.587\,G' + 0.114\,B' \qquad \text{Eq 40.1}$$

The luma component *Y'*, being a weighted sum of nonlinear *R'G'B'* components, has no simple relationship with the CIE relative luminance (*Y*) used in colour science. Video encoding specifications typically place no upper bound on luma bandwidth (though transmission standards may).

Picture center, aspect ratio, and blanking

SMPTE RP 187, *Center, Aspect Ratio and Blanking of Video Images.*

The center of the picture is located midway between the central pair of the 720 active samples of BT.601, at the fraction $^{983}\!/_{1728}$ between 0_H instants. Concerning the vertical center, see Table 40.1, on page 458.

$$\frac{S_{TL} - S_{EAV-0_H} + 0.5\left(S_{AL} - 1\right)}{S_{TL}} = \frac{864 - 732 + 0.5\left(720 - 1\right)}{864} = \frac{983}{1728}$$

Aspect ratio is defined as 4:3 with respect to a *clean aperture* pixel array, 690 samples wide at a sampling rate of 13.5 MHz, and 566 lines high. Blanking transitions should not intrude into the clean aperture.

In the composite and analog domains, video information on the left-hand halfline of line 623 terminates 30.350±0.1 µs after 0_H. Video information on the right-hand halfline of line 23 commences 42.500±0.1 μs after 0_H.

No studio standard addresses square sampling of 576*i* video. I recommend using a sample rate of $944f_H$, that is, 14.75 MHz. I recommend using 768 active samples, centered as mentioned above.

SMPTE RP 202, *Video Alignment for MPEG-2 Coding.*

When MPEG-2 with 576 or 608 image rows is used in the studio, the bottom image row corresponds to line 623 (as indicated in Table 40.1). The bottom left-hand halfline (on line 623) is among the coded image rows. The right-hand half of this line will be blank when presented to the MPEG encoder; upon decoding, it may contain artifacts. Unfortunately, 576*i* DV systems digitize a range one image row up from this.

Component digital 4:2:2 interface

The C_B and C_R colour difference components of digital video are formed by scaling B'-Y' and R'-Y' components, as described in *C_BC_R components for SD* on page 361. $Y'C_BC_R$ signals were once conveyed through the parallel digital interface specified in Rec. 656 and EBU Tech. 3246; nowadays, the *serial digital interface* (SDI) is used.

SDI was introduced on page 432. Mechanical and electrical details were presented on page 439.

In 13.5 MHz sampling of 576*i*, sample 732 corresponds to the line sync datum, 0_H. If digitized, that sample would take the 50% value of analog sync. SMPTE RP 187 specifies that samples 8 and 710 correspond to the 50%-points of picture width. For flat-panel displays, EBU suggests that the central 702 samples contain active video.

The choice of 720 active samples for BT.601 accommodates the blanking requirements of both 480*i* and 576*i* analog video: 720 samples are sufficient to accommodate the necessary transition samples for either system; see page 378.

Unfortunately, the blanking tolerances between 480*i* and 576*i* do not permit a single choice of blanking transition samples: The narrowest possible picture width in 480*i* is several samples too wide to meet 576*i* tolerances.

Figure 40.2 overleaf shows a waveform drawing of luma in a 576*i* component digital 4:2:2 system.

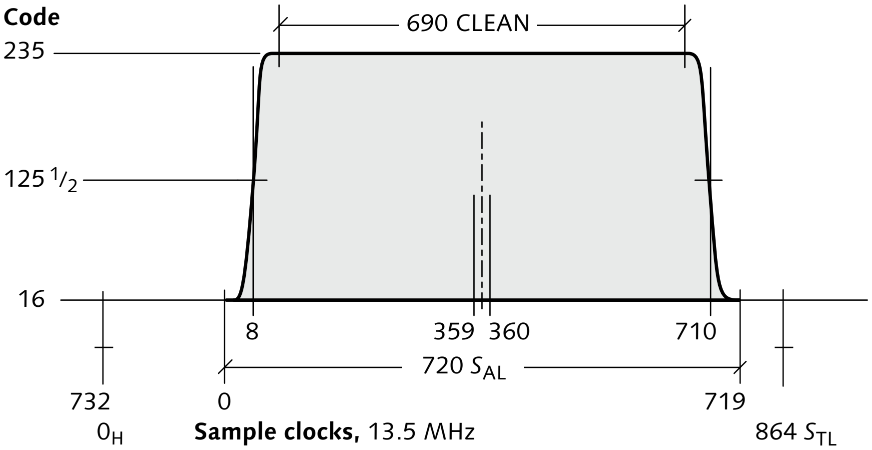


Figure 40.2 **576*i* component digital 4:2:2 luma waveform**

Component analog 576*i* interface

EBU Tech. N20, *Parallel interface for analogue component video signals in GRB form.*

A component analog 576*i* $R'G'B'$ interface is based on nonlinear R', G', and B' signals conveyed as voltage, with a range of 1 V_{PP} from synctip to reference white. Transient excursions slightly outside the reference black-to-white range are permitted. A video signal of zero – blanking level, equal to reference black – corresponds to a level of 0 V_{DC}. A video signal of unity corresponds to 700 mV.

Sync is added to the green component according to Equation 40.2, where *sync* and *active* are taken to be unity when asserted and zero otherwise:

$$G'_{\text{sync}} = \frac{7}{10}\left(\mathit{active} \cdot G'\right) + \frac{3}{10}\left(-\mathit{sync}\right) \qquad \text{Eq 40.2}$$

The excursion of the G' signal from synctip to reference white is 1 V_{PP}. Levels in 576*i* systems are usually specified in millivolts, not the IRE units common in 480*i* systems. If IRE units were used, 1 IRE would equal 7 mV.

EBU Tech. N10, *Parallel interface for analogue component video signals.*

Analog luma (Y') is carried as a voltage ranging 1 V_{PP} from synctip to reference white. Luma signal of zero – blanking level, equal to reference black – corresponds to a level of 0 V_{DC}. Luma signal of unity corresponds to

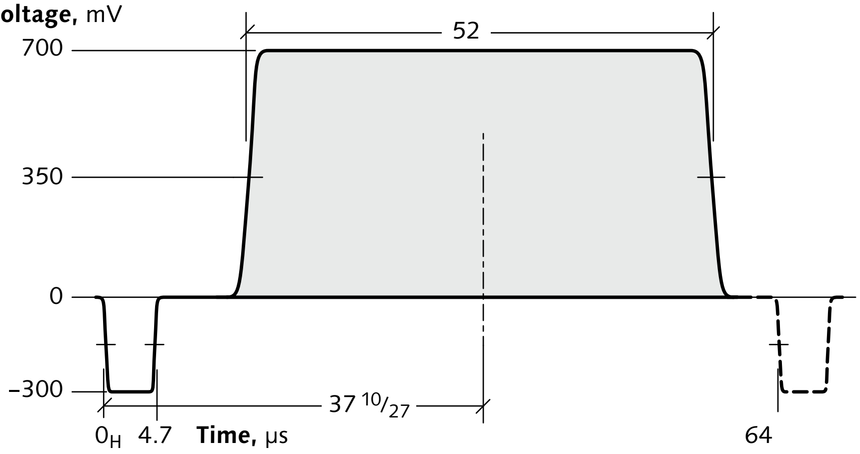


Figure 40.3 **576*i* component analog luma waveform**

700 mV. Sync is added to the luma component according to Equation 40.3:

$$Y'_{\text{sync}} = \frac{7}{10}Y' + \frac{3}{10}\left(-sync\right) \qquad \text{Eq 40.3}$$

The picture excursion of the *Y'* signal is 700 mV. Figure 40.3 above shows a waveform drawing of luma in a 576*i* component analog interface.

The P_B and P_R colour difference components are formed by scaling *B'-Y'* and *R'-Y'* components, as described in *PBPR components for SD on page 359*. Although it is possible in theory to have wideband P_B and P_R components, in practice they are lowpass filtered to about half the bandwidth of luma. P_B and P_R compnents are carried as voltage with reference excursion ±350 mV.

1280×720 HD 41

SMPTE ST 296, *1280×720 Progressive Image Sample Structure – Analog and Digital Representation and Analog Interface.*

This chapter details the scanning, timing, and sync structure of 1280×720 video, also called 720*p.* The scanning and timing information in this chapter applies to all variants of 720*p* video, both analog and digital.

Scanning

$$\frac{24}{1.001} \approx 23.976$$

$$\frac{30}{1.001} \approx 29.97$$

$$\frac{60}{1.001} \approx 59.94$$

720*p* video represents stationary or moving two-dimensional images sampled temporally at a constant rate of $^{24}/_{1.001}$, 24, 25, $^{30}/_{1.001}$, 30, 50, $^{60}/_{1.001}$, or 60 frames per second. The sampling rate is 74.25 MHz (modified by the ratio $^{1000}/_{1001}$ in 720*p*59.94, 720*p*29.97, and 720*p*23.976). All of these systems have 750 total lines (L_T). The number of samples per total line (S_{TL}) is adapted to achieve the desired frame rate. Table 41.1 below summarizes the scanning parameters.

A frame comprises a total of 750 horizontal raster lines of equal duration, uniformly progressively scanned top to bottom and left to right, numbered consecutively

System	f_S *[MHz]*	S_{TL}
720*p*60	74.25	1650
720*p*59.94	$^{74.25}/_{1.001}$	1650
720*p*50	74.25	1980
720*p*30	74.25	3300
720*p*29.97	$^{74.25}/_{1.001}$	3300
720*p*25	74.25	3960
720*p*24	74.25	4125
720*p*23.976	$^{74.25}/_{1.001}$	4125

Table 41.1 **720*p* scanning parameters** are summarized.

tri	Trilevel pulse
BR	Broad pulse

The vertical center of the picture is located midway between lines 385 and 386.

Line number	*Contents*
1–5	tri/BR (5 lines)
6	Blanking
7–25 (19 lines)	Blanking/Ancillary
26–745 (720 lines)	Picture [Clean aperture 702 lines]
746–750 (5 lines)	Blanking

Table 41.2 **1280×720 line assignment**

starting at 1. Of the 750 total lines, 720 contain picture. Table 41.2 above shows the assignment of line numbers and their content.

For studio video, the tolerance on frame rate is normally ±10 ppm. In practice the tolerance applies to a master clock at a high frequency, but for purposes of computation and standards writing it is convenient to reference the tolerance to the frame rate.

At a digital interface, video information is identified by a *timing reference signal* (TRS) conveyed across the interface. (See *SDI and HD-SDI sync, TRS, and ancillary data,* on page 433.) The last active line of a frame is terminated by EAV where the *V*-bit becomes asserted. That EAV marks the start of line 746; line 1 of the next frame starts on the fifth following EAV.

Analog sync

0_H precedes the first word of SAV by 256 clocks.

0_H follows the first word of EAV by $S_{TL} - 1280 - 260$ clocks.

Horizontal events are referenced to 0_H, defined by the zero-crossing of trilevel sync. Digital samples and analog timing are related such that the first (zeroth) sample of active video follows the 0_H instant by 260 reference clock intervals.

At an analog interface, each line commences with a trilevel sync pulse. Trilevel sync comprises a negative portion asserted to −300±6 mV during the 40 reference clock intervals preceding 0_H, and a positive portion asserted to +300±6 mV during the 40 reference clock intervals after 0_H. The risetime of each transition is 4±1.5 reference clock intervals.

Vertical sync in the analog domain is signaled by *broad pulses,* one each on lines 1 through 5. Each broad pulse is asserted to –300±6 mV, with timing identical to active video – that is, to the production aperture's picture width. The risetime of each transition is 4±1.5 reference clock intervals. Line 1 can be detected as the first broad pulse of a frame – that is, by a line without a broad pulse followed by a line with one.

Lines 7 through 25 do not convey picture information. They may convey ancillary or other signals either related or unrelated to the picture.

Analog signal timing is defined by the digital standard; the digital sampling frequency defines reference time intervals used to define analog timing.

Figure 41.1 overleaf shows details of the sync structure; this waveform diagram is the analog of Table 41.2.

Picture center, aspect ratio, and blanking

The center of the picture is located midway between the central two of the 1280 active samples – that is, between samples 639 and 640 – and midway between the central two 720 picture lines – that is, between lines 385 and 386.

The aspect ratio is defined to be 16:9 with respect to the production aperture of 1280×720.

In *Transition samples,* on page 378, I mentioned that it is necessary to avoid, at the start of a line, an instantaneous transition from blanking to picture information. A *clean aperture* pixel array 1248 samples wide and 702 lines high, centered on the production aperture, should remain subjectively uncontaminated by edge transients.

R'G'B' EOCF and primaries

Picture information is referenced to linear-light primary red, green, and blue (*RGB*) tristimulus values, represented in abstract terms in the range 0 (reference black) to +1 (reference white).

Three nonlinear primary components *R'*, *G'*, and *B'* are computed such that the intended image appearance is obtained on the reference display in the reference viewing conditions; see *Reference display and viewing conditions,* on page 427.

In the default power-up state of a camera, the nonlinear primary components are computed according to the opto-electronic conversion function of *BT.709*

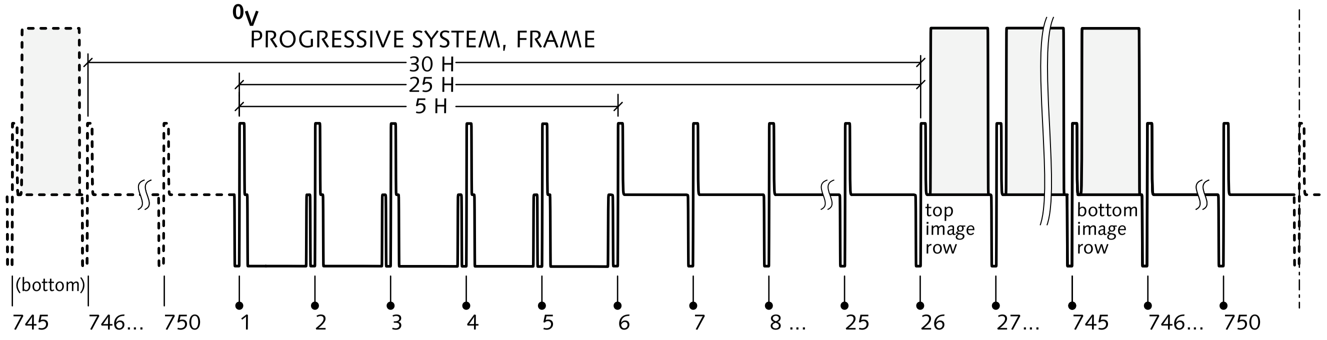


Figure 41.1 **720*p* raster, vertical**

OECF, described on page 320; this process is loosely called *gamma correction.*

The colorimetric properties of the primary estimates are supposed to conform to *BT.709 primaries* described on page 290: DTV transmission standards call for BT.709, and modern consumer displays use BT.709. However, as I write in 2011, nearly all 50 Hz program material is created and mastered using EBU primaries (see *EBU Tech. 3213 primaries*, Table 26.4 on page 293), and nearly all 60 Hz program material is created and mastered using SMPTE primaries (see *SMPTE RP 145 primaries*, Table 26.5 on page 293). I expect the situation in mastering to change upon the introduction of new studio display technologies such as OLEDs – but among content creators and broadcasters, old habits die hard.

Luma (Y')

Luma is a weighted sum of nonlinear R', G', and B' components according to the BT.709 luma coefficients:

$$^{709}Y' = 0.2126\,R' + 0.7152\,G' + 0.0722\,B' \qquad \text{Eq 41.1}$$

The luma component Y', being a weighted sum of nonlinear $R'G'B'$ components, has no simple relationship with the CIE relative luminance (Y) used in colour science. The formulation of luma in HD differs from that of SD; see *SD and HD luma chaos,* on page 350. Video encoding specifications typically place no upper bound on luma bandwidth. Video encoding specifications typically place no upper bound on luma bandwidth.

Component digital 4:2:2 interface

For details of the analog interface, see *Component analog HD interface,* on page 485.

$Y'C_BC_R$ components are formed by scaling Y', $B'-Y'$, and $R'-Y'$ components, as described in *C_BC_R components for BT.709 HD* on page 371. TRS is inserted as described in *SDI and HD-SDI sync, TRS, and ancillary data* on page 433. The HD-SDI interface is described in *HD-SDI coding,* on page 440. It is standard to subsample according to the 4:2:2 scheme (sketched in the third column of Figure 12.1, on page 124). Image quality wouldn't suffer if subsampling were 4:2:0 – that is, if colour differences were subsampled vertically as well as horizontally – but this would be inconvenient for hardware design and interface.

1920×1080 HD

42

SMPTE ST 274, *1920×1080 Scanning and Analog and Parallel Digital Interfaces for Multiple Picture Rates.*

This chapter details a family of high-definition television systems standardized in SMPTE ST 274. The systems have an image format of 1920×1080, an aspect ratio of 16:9, and square sampling.

SMPTE ST 274 represents agreement on colorimetry according to the international standard BT.709. Previous revisions of SMPTE ST 274 allowed use of 240M colour and luma parameters on an "interim" basis. It will take some time for manufacturers and users to complete the transition to the new standard.

HD equipment based on 1125/60 scanning and 1920×1035 image format, with nonsquare sampling, according to SMPTE 240M, was deployed for several years. That system is obsolete. (It can now be considered as a 1035*i* variant of 1080*i*30 or 1080*i*29.94 having 1035 picture lines instead of 1080, nonsquare sampling, and slightly different colorimetry.)

Scanning

1920×1080 video represents stationary or moving two-dimensional images sampled temporally at a constant rate of $^{24}/_{1.001}$, 24, 25, $^{30}/_{1.001}$, 30, 50, $^{60}/_{1.001}$, or 60 frames per second. The base sampling rate is 74.25 MHz; this rate is modified by the ratio $^{1000}/_{1001}$ for systems having noninteger frame rate, and doubled to 148.5 MHz (possibly times $^{1000}/_{1001}$) for progressive systems at frame rates higher than 30 Hz. All of these systems have 1125 total lines (L_T); the number of samples per total line (S_{TL}) is adapted to achieve the desired frame rate. Table 42.1 above summarizes the scanning parameters.

$$\frac{24}{1.001} \approx 23.976$$

$$\frac{30}{1.001} \approx 29.97$$

$$\frac{60}{1.001} \approx 59.94$$

For studio video, the tolerance on frame rate is normally ±10 ppm. In practice the tolerance applies to a master clock at a high frequency, but for purposes of computation and standards writing it is convenient to reference the tolerance to the frame rate.

A frame comprises a total of 1125 horizontal raster lines of equal duration, uniformly scanned top to

System	f_S [*MHz*]	S_{TL}
1080*p*60¶	148.5	2200
1080*p*59.94¶	$148.5/1.001$	2200
1080*p*50¶	148.5	2640
1080*i*30	74.25	2200
1080*i*29.97	$74.25/1.001$	2200
1080*i*25	74.25	2640
1080*p*30	74.25	2200
1080*p*29.97	$74.25/1.001$	2200
1080*p*25	74.25	2640
1080*p*24	74.25	2750
1080*p*23.976	$74.25/1.001$	2750

Table 42.1 **1920×1080 scanning parameters** are summarized. 1080*p* systems, marked ¶, are not allowed for ATSC broadcast.

bottom and left to right, numbered consecutively starting at 1. Of the 1125 total lines, 1080 contain picture. Table 42.2 opposite indicates the assignment of line numbers and their content.

A progressive system conveys 1080 active picture lines per frame in order top to bottom.

An interlaced system scans a frame as a *first* field then a *second* field. The scan lines of each field have half the vertical spatial sampling density of the frame. Scanning lines in the second field are displaced vertically by the vertical sampling pitch, and delayed temporally by half the frame time, from scanning lines in the first field. The first field conveys 540 active picture lines, starting with the top picture line of the frame. The second field conveys 540 active picture lines, ending with the bottom picture line of the frame.

At a digital interface, video information is identified by a *timing reference signal* (TRS) conveyed across the interface. (See *SDI and HD-SDI sync, TRS, and ancillary data,* on page 433.) In progressive systems, the last active line of a frame is terminated by EAV where the *V*-bit becomes asserted. That EAV marks the start of line 1122; line 1 of the next frame starts on the fourth following EAV. In interlaced systems, the last active line of a field is terminated by EAV where the *V*-bit becomes asserted. In the first field, that EAV marks the

Line number, progressive	*Line number, first field ($F=0$)*	*Line number, second field ($F=1$)*	*V*	*Contents, left half*	*Contents, right half*
		563			
1	1				
		564			
2	2				
		565			
3	3				
		566			
4	4				
		567			
5	5				
		568			
6	6				
7–41 (35 lines)	7–20 (14 lines)			vertical interval video	
		569–583 (15 lines)		vertical interval video	
42–1121 (1080 lines)	21–560 (540 lines)		$V=0$ (1080 lines)	picture	
		584–1123 (540 lines)		picture	
1122–1125 (4 lines)	561–562 (2 lines)	1124–1125 (2 lines)			

Table 42.2 **1080*i* and 1080*p* line assignment**

start of line 561. In the second field, that EAV marks the start of line 1124; line 1 of the next frame starts on the second following EAV.

Analog sync

At an analog interface, each line commences with a trilevel sync pulse. The zero-crossing of trilevel sync defines the line sync datum 0_H, to which horizontal events are referenced. Digital samples and analog timing are related such that the first (zeroth) sample of

0_H precedes the first word of SAV by 192 clocks.

0_H follows the first word of EAV by S_{TL} – 1920 –192 clocks.

active video follows the 0_H instant by 192 reference clock intervals.

Trilevel sync comprises a negative portion asserted to –300±6 mV during the 44 reference clock intervals preceding 0_H, and a positive portion asserted to +300±6 mV during the 44 reference clock intervals after 0_H. The risetime of each transition is 4±1.5 reference clock intervals.

Details of horizontal timing are shown in Figure 42.1 opposite.

Vertical sync in the analog domain is signaled by *broad pulses,* whose structure differs between progressive and interlaced systems.

A progressive system has five broad pulses per frame, one each on lines 1 through 5. Each broad pulse is asserted to –300±6 mV, 132 reference clock intervals after 0_H, and deasserted 2112 reference clock intervals after 0_H. (Deassertion coincides with the end of active video – that is, with the right-hand edge of the production aperture.) The risetime of each transition is 4±1.5 reference clock intervals. Line 1 is defined by the first broad pulse of a frame – that is, by a line with a broad pulse preceded by a line without one. Line 6 has a second trilevel pulse whose zero-crossing is $S_{TL}/2$ reference clock intervals after 0_H. This pulse is reminiscent of an equalization pulse in analog SD.

In an interlaced system, several lines in the vertical interval have a second trilevel sync pulse whose zero-crossing is at $S_{TL}/2$ reference clock intervals after 0_H. An interlaced system has ten broad pulses per field, in the arrangement indicated in Table 42.2. Each broad pulse is asserted to –300±6 mV, 132 reference clock intervals after the zero-crossing of the immediately preceding trilevel sync, and is deasserted 880 reference clock intervals later. The risetime at each transition is 4±1.5 reference clock intervals. Line 1 can be decoded as the first broad pulse in a left-hand halfline – that is, by detecting a normal line (with no broad pulse and no mid-line trilevel sync) followed by a broad pulse immediately after 0_H. Each broad pulse is preceded by a trilevel pulse whose zero-crossing is either at 0_H or delayed $S_{TL}/2$ reference clock intervals from 0_H.

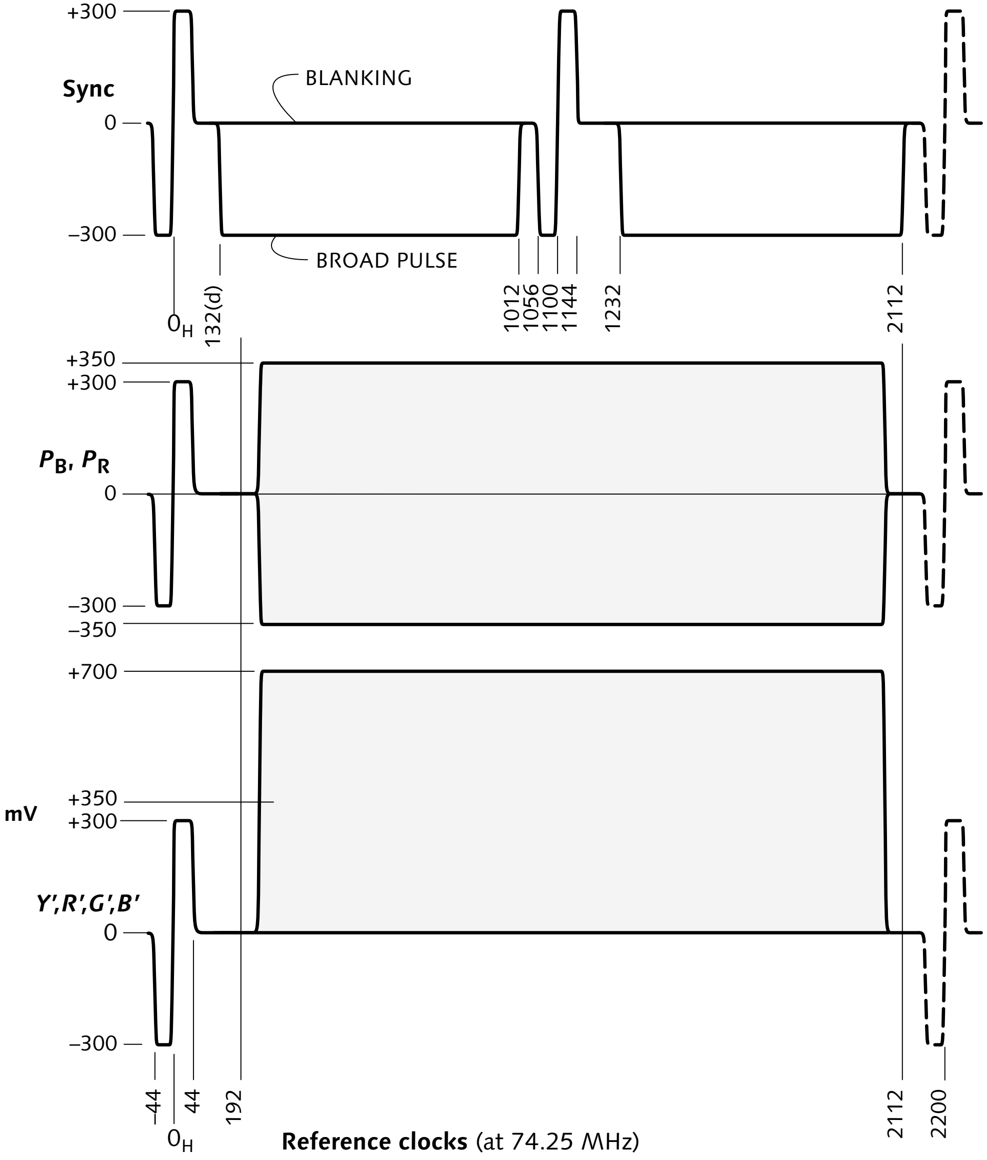


Figure 42.1 **1080*i*30 analog line details.** Time intervals are shown as intervals of a reference clock at 74.25 MHz, with reference clock zero defined at 0_H. To obtain sample numbers relative to the zeroth sample of active video, add $S_{AL} + 88$ (modulo S_{TL}) to these counts.

Vertical interval video lines do not convey picture information. They may convey ancillary or other signals either related or unrelated to the picture.

Analog signal timing is defined by the digital standard; the digital sampling frequency defines reference time intervals used to define analog timing.

Figure 42.2 shows details of the vertical sync structure; this waveform diagram is the analog of Table 42.2, on page 475.

Picture center, aspect ratio, and blanking

The center of the picture is located midway between the central two of the 1920 active samples (between samples 959 and 960), and midway between the central two 1080 picture lines (between lines 581 and 582 in a progressive system, and between lines 290 and 853 in an interlaced system).

The aspect ratio is defined to be 16:9 with respect to the production aperture of 1920×1080.

In *Transition samples*, on page 378, I mentioned that it is necessary to avoid, at the start of a line, an instantaneous transition from blanking to picture information. A *clean aperture* pixel array 1888 samples wide and 1062 lines high, centered on the production aperture, should remain subjectively uncontaminated by edge transients.

R'G'B' EOCF and primaries

Picture information is referenced to linear-light primary red, green, and blue (*RGB*) tristimulus values, represented in abstract terms in the range 0 (reference black) to +1 (reference white).

Three nonlinear primary components *R'*, *G'*, and *B'* are computed such that the intended image appearance is obtained on the reference display in the reference viewing conditions; see *Reference display and viewing conditions*, on page 427.

In the default power-up state of a camera, the nonlinear primary components are computed according to the opto-electronic conversion function of *BT.709 OECF*, described on page 320; this process is loosely called *gamma correction.*

The colorimetric properties of the display primaries are supposed to conform to *BT.709 primaries* described

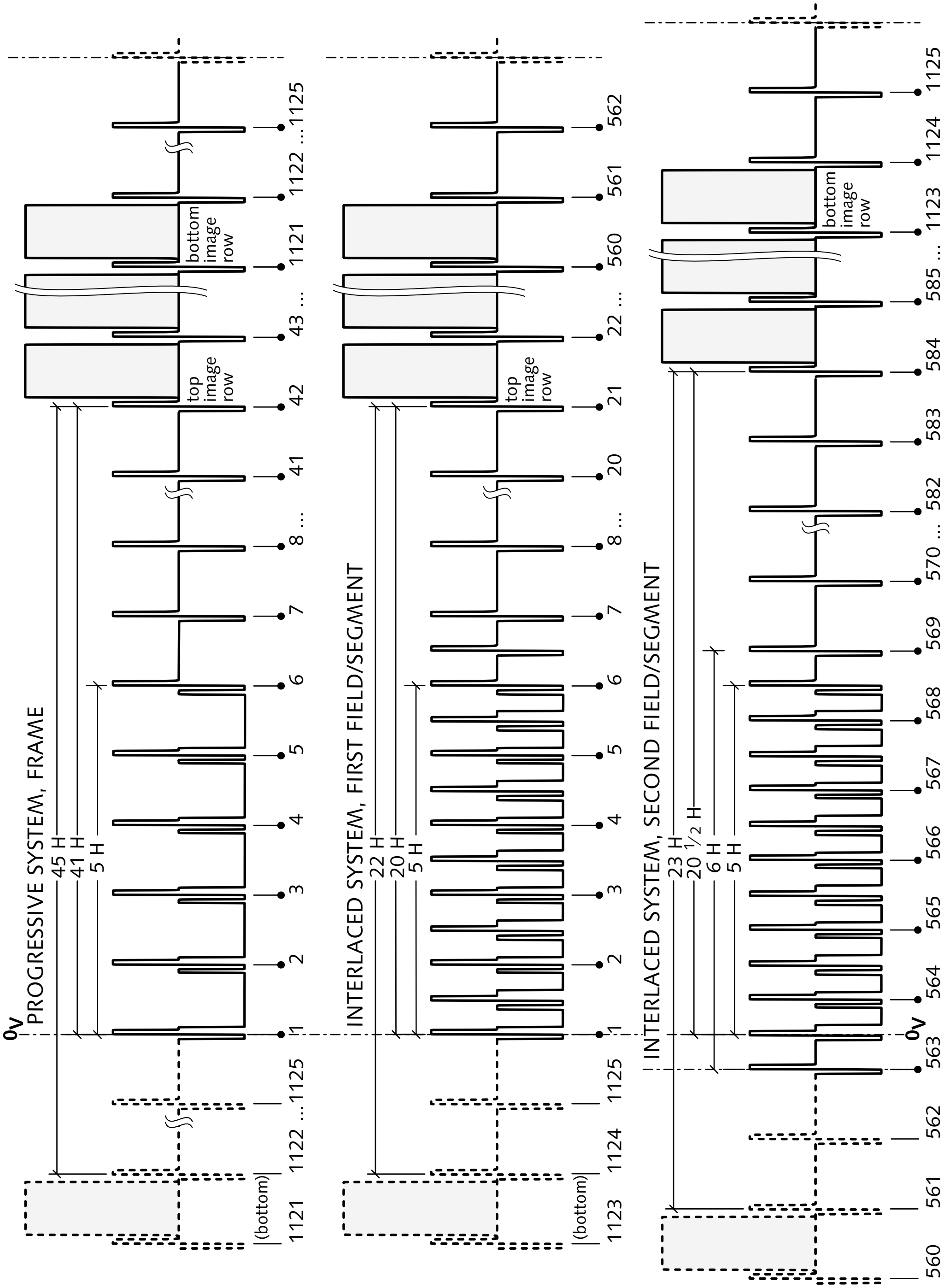


Figure 42.2 **1080*i* and 1080*p* vertical blanking interval**

on page 290: DTV transmission standards call for BT.709, and modern consumer displays use BT.709. However, as I write in 2011, nearly all 50 Hz program material is created and mastered using EBU primaries (see *EBU Tech. 3213 primaries*, Table 26.4 on page 293), and nearly all 60 Hz program material is created and mastered using SMPTE primaries (see *SMPTE RP 145 primaries*, Table 26.5 on page 293). I expect the situation in mastering to change upon the introduction of new studio display technologies such as OLEDs – but among content creators and broadcasters, old habits die hard.

Luma (Y')

Luma is computed as a weighted sum of nonlinear R', G', and B' primary components according to the luma coefficients introduced in *BT.709 luma,* on page 346:

$$^{709}Y' = 0.2126\,R' + 0.7152\,G' + 0.0722\,B' \qquad \text{Eq 42.1}$$

The luma component Y', being a weighted sum of nonlinear $R'G'B'$ components, has no simple relationship with the CIE relative luminance (Y) used in colour science. The formulation of luma in HD differs from that of SD; see *SD and HD luma chaos,* on page 350. Video encoding specifications typically place no upper bound on luma bandwidth.

Component digital 4:2:2 interface

For details of the analog interface, see *Component analog HD interface,* on page 485.

$Y'C_BC_R$ components are formed by scaling $B'-Y'$ and $R'-Y'$ components, as described in *C_BC_R components for BT.709 HD* on page 371. The HD-SDI interface is described *HD-SDI coding,* on page 440.

HD videotape

43

For SD, refer to *Composite NTSC and PAL: Legacy Video Systems.*

Since publication of the first edition of this book there has been a dramatic decline in the use of videotape. Much legacy SD videotape equipment is still in use, but SD requirements are now fairly easily met by hard drive and flash media. Videotape is still in fairly wide use for HD, and still retains a few advantages over hard drive and flash media. Also, several compression systems originally devised for videotape have successfully made the transition into hard drive and flash media.

Table 43.1 summarizes digital videotape formats for HD.

Notation	*Method*	*Tape*	*Data rate [Mb/s]*	*Notes*
D-5 HD (HD-D5, D-15)	Component compressed, 4:2:2 HD	12.65 mm (½ inch, VHS-derived)	270	(Panasonic)
D-6	Uncompressed 4:2:2 HD	19 mm	1188	
HDCAM (D-11)	Component compressed, M-JPEG-like, 1440×1080, 3:1:1 HD	12.7 mm (Beta-derived)	135	(Sony)
DVCPRO HD (D-12)	Component compressed, DV100, 1280×1080 (or 960×720), 4:2:2 HD	6.35 mm	100	DV100
HDCAM SR (D-16)	Component compressed, MPEG-4 SStP, 1920×1080, 4:2:2 (Lite), 4:2:2 (SQ), 4:4:4 (HQ)	12.65 mm (Beta-derived)	 220 440 880	(Sony)

Table 43.1 **Digital videotape formats for HD.** The D-6 format is “uncompressed” (although subject to chroma subsampling); all of the other formats use compression.

D-5 HD (HD-D5, D-15)

The SMPTE D-5 standard defined recording of uncompressed SD at 270 Mb/s. Panasonic adapted the D-5 format to HD by equipping it with a motion-JPEG codec having a compression ratio of about 5:1; the coding is quite similar to that of DV. This variant of the D-5 VTR is denoted D-5 HD or HD-D5; after its introduction, it was standardized by SMPTE as D-15.

The D-5 HD system records 720*p* or 1080*i* video at either 59.94 Hz or 60.00 Hz, with 4:2:2 chroma subsampling.

MALVAR, HENRIQUE S. (1992), *Signal Processing with Lapped Transforms* (Norwood, Mass.: Artech House).

One difference from DV is that D-5 HD compression uses a lapped (overlapped) transform, where the rightmost column of samples in an 8×8 block overlaps the leftmost column of the next block to the right. Upon decoding, the redundantly coded columns are reconstructed by averaging appropriate samples from two neighboring blocks. This scheme reduces blocking artifacts compared to the nonlapped transform used in DV.

D-6

SMPTE D-6 defined a videotape format for uncompressed, 8-bit, 4:2:2 HD, at a bit rate of 1.188 Gb/s. This equipment had superb performance and was very expensive; it is now obsolete.

HDCAM (D-11)

Downsampling of 1920×1080 to 1440×1080 means that luma sample aspect ratio is effectively $4/3$ (and chroma subsampling is equivalent to 4:1:1) relative to the original luma array.

HDCAM was developed and commercialized by Sony. Interlaced or progressive HD video at any of several frame rates is downsampled to 1440×1080 image format and subject to 3:1:1 chroma subsampling. Compression is M-JPEG-style, comparable to that of DV, using a lapped (overlapped) transform (like D-5 HD), coded at about 135 Mb/s. HDCAM uses a tape cassette derived from Betacam, with $1/2$-inch tape. The videotape recording scheme was standardized as SMPTE D-11.

DVCPRO HD (D-12)

Downsampling of 1920×1080 to 1280×1080 means that luma sample aspect ratio is effectively $3/2$ (and chroma subsampling is equivalent to 3:1:1) relative to the original 1920×1080 luma array.

The DV25/DV50 standards were adapted to 100 Mb/s to accommodate 1080*i*30 HD signals downsampled to 1280×1080 image format (or 720*p*60 downsampled to 960×720), with 4:2:2 chroma subsampling in the downsampled domain. The compression scheme is denoted DV100. The compression scheme is sometimes called DVCPRO100; digital videotape recorders were introduced by Panasonic as DVCPRO HD, sharing many of the mechanical and signal processing elements of the DV25 and DV50 DVTRs. The videotape recording scheme was later standardized as SMPTE D-12.

HDCAM SR (D-16)

HDCAM SR compression conforms to MPEG-4 Part 2 *simple studio profile* (SStP). HDCAM SR records 1920×1080, either $Y'C_BC_R$ 4:2:2 at 440 Mb/s (SQ mode), or $R'G'B'$ 4:4:4 at 880 Mb/s (HQ mode). HDCAM SR equipment can also record or play two independent $Y'C_BC_R$ 4:2:2 streams in 880 Mb/s mode. An additional mode, SR Lite, was added having a data rate of 220 Mb/s. ("Lite" refers to the data rate requirement; the compression is heavy.)

Component analog HD interface

44

Component analog HD signals are interfaced in the analog domain as $Y'P_BP_R$ signals conveyed as voltage. Luma ranges 1 V_{PP} from synctip to reference white. A luma signal of zero – blanking level, equal to reference black – corresponds to a level of 0 V_{DC}. A luma signal of unity corresponds to 700 mV. Transient excursions slightly outside the reference black-to-white range are permitted.

Sync is added to luma according to Equation 44.1, where *sync* and *active* are taken to be unity when asserted and zero otherwise:

$$Y'_{\text{sync}} = \frac{7}{10}Y' + \frac{3}{10}\left(-sync\right) \qquad \text{Eq 44.1}$$

The excursion of the Y' signal from synctip to reference white is 1 V_{PP}. One IRE unit corresponds to 7 mV.

Figure 44.1 overleaf shows a waveform drawing of luma in a 720*p*60 component analog interface; Figure 44.2 overleaf shows a waveform drawing of luma in 1080*i*30.

P_B and P_R colour difference components are formed by scaling B'-Y' and R'-Y' components, as described in *P_BP_R components for BT.709 HD* on page 370. Although it is possible in theory to have wideband P_B and P_R components, in practice they are lowpass filtered to about half the bandwidth of luma. P_B and P_R compnents are carried as voltage with reference excursion ±350 mV.

SMPTE standards do not specify tolerances on time-coincidence of R', G', and B' components at an analog interface. I recommend that the components be time-

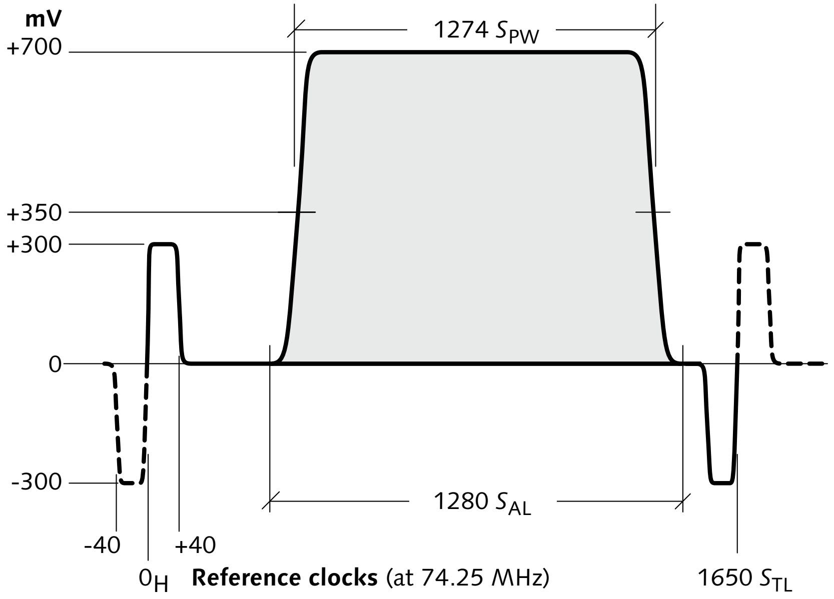


Figure 44.1 **720*p*60 component analog luma waveform.** Time intervals are shown as intervals of a reference clock at 74.25 MHz.

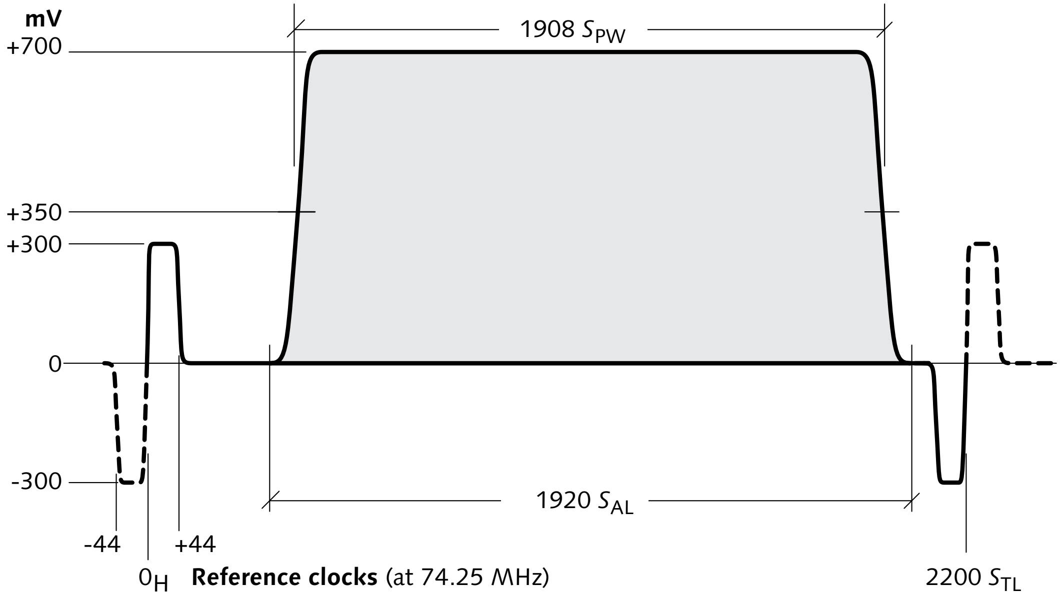


Figure 44.2 **1080*i*30/1080*p*30 component analog luma waveform.** Time intervals are shown as intervals of a reference clock at 74.25 MHz.

coincident with each other, and with sync, within ¼ of a sample clock – for 720*p*60 and 1080*i*30, time-coincident within ±3.4 ns.

EIA/CEA-770.3, *High Definition TV Analog Component Video Interface.*

EIA/CEA has standardized the 720*p*60 and 1080*i*30 systems as described here, including these $Y'P_BP_R$ levels, for use in consumer HD electronics devices.

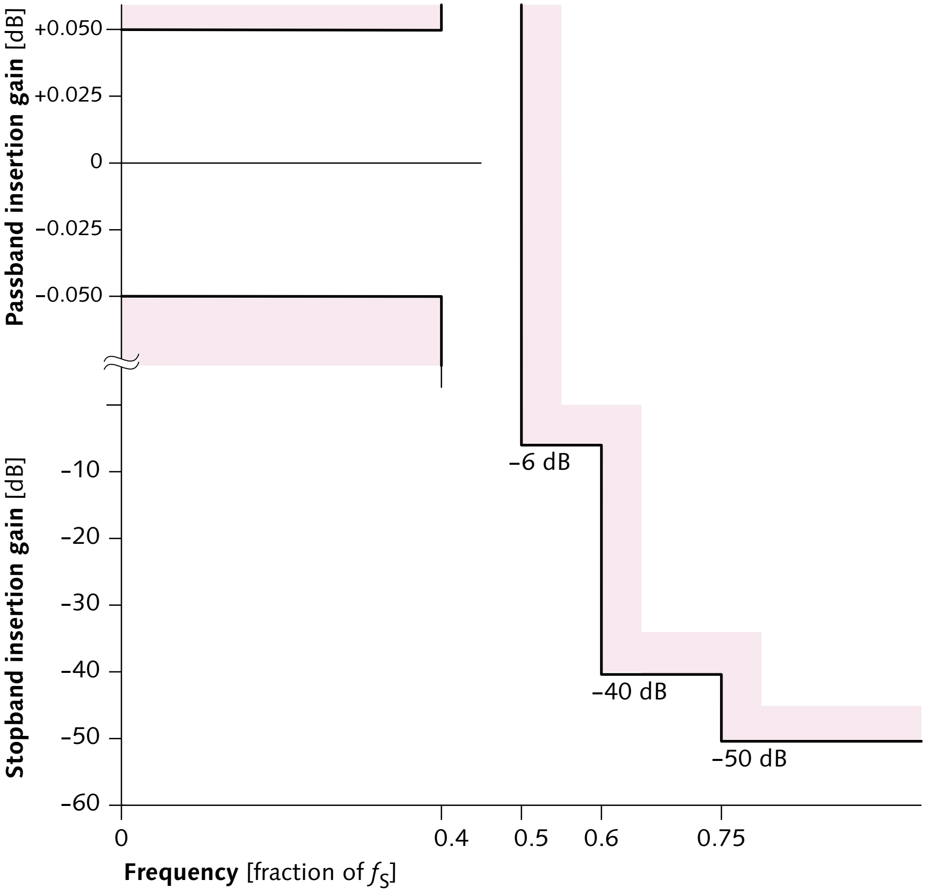
Pre- and postfiltering characteristics

Component Y', R', G', or B' signals in 720*p*59.94 or 720*p*60 – and for that mattter, 1080*i*30 – have a nominal passband of 30 MHz; colour difference components have a nominal passband of 15 MHz.

Figure 44.3 overleaf, *Filter template for Y′ and R′G′B′ components*, depicts filter characteristics for pre- and postfiltering of R', G', B', and Y' component signals. Analog P_B and P_R colour difference component signals are pre- and postfiltered using the same template scaled by a factor of two on the frequency axis. The characteristics are frequency-scaled from the template included in BT.601, with a few modifications.

Amplitude ripple tolerance in the passband is ±0.05 dB with respect to insertion loss at 100 kHz. Insertion loss is 6 dB or more at half the sampling rate of the Y', R', G', and B' components.

SMPTE ST 274 includes luma and colour difference templates in “Annex B (Informative).” In standards lingo, the word *informative* signals that the information is distributed with the standard, but compliance with the annex is not required to claim conformance with the standard.



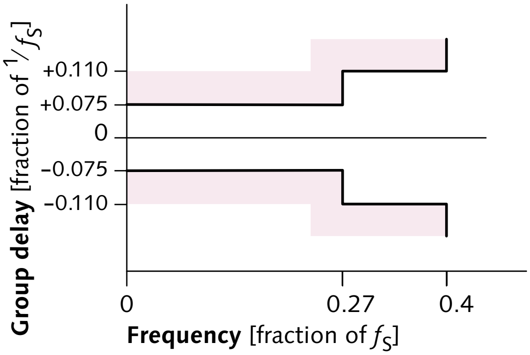


Figure 44.3 **Filter template for *Y'* and *R'G'B'* components**

Part 5

Video compression

JPEG and motion-JPEG (M-JPEG) compression 45

This chapter describes *JPEG*, a standard for lossy compression of still images. JPEG is based upon the discrete cosine transform (DCT). JPEG is rarely used directly in video, but it forms the basis of M-JPEG (used in desktop video editing) and DV compression. Also, JPEG techniques form the core of MPEG.

Motion-JPEG (M-JPEG) refers to the use of a JPEG-like algorithm to compress each field or frame in a sequence of video fields or frames. M-JPEG systems use the methods of JPEG, but rarely (if ever) conform to the ISO/IEC JPEG standard. DV is a specific type of M-JPEG, which is well standardized; it is described in the following chapter, *DV compression,* on page 505. The *I-frame-only* variant of MPEG-2 is conceptually equivalent to M-JPEG, but again has a well-respected standard; see *MPEG-2 video compression,* on page 513.

ISO/IEC 10918, *Information Technology – Digital compression and coding of continuous-tone still images.*

The JPEG standard, cited in the margin, defines four modes: *sequential, hierarchical, progressive,* and *lossless*. The JPEG standard accommodates DCT coefficients having from 2 to 16 bits, and accommodates two different entropy coders (*Huffman* and *arithmetic*). *Baseline* refers to a defined subset of JPEG's sequential mode that is restricted to 8-bit coefficients and restricted to Huffman coding. Only baseline JPEG is commercially important; JPEG's other modes are mainly of academic interest, and won't be discussed here.

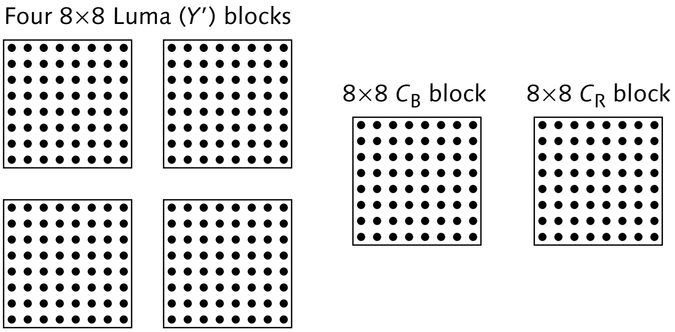


Figure 45.1 **A JPEG 4:2:0 minimum coded unit** (MCU) comprises six 8×8 blocks: four luma blocks, a block of C_B, and a block of C_R. The six constituent blocks result from nonlinear $R'G'B'$ data being matrixed to $Y'C_BC_R$, then subsampled according to the 4:2:0 scheme; chroma subsampling is effectively the first stage of compression. The blocks are processed independently.

JPEG blocks and MCUs

An 8×8 array of sample data is known in JPEG terminology as a *block*. Prior to JPEG compression of a colour image, normally the nonlinear $R'G'B'$ data is matrixed to $Y'C_BC_R$, then subsampled 4:2:0. According to the JPEG standard (and the JFIF standard, to be described), other colour subsampling schemes are possible; strangely, different subsampling ratios are permitted for C_B and C_R. However, only 4:2:0 is widely deployed, and the remainder of this discussion assumes 4:2:0. Four 8×8 luma blocks, an 8×8 block of C_B, and an 8×8 block of C_R are known in JPEG terminology as a *minimum coded unit* (MCU); this corresponds to a *macroblock* in DV or MPEG terminology. The 4:2:0 macroblock arrangement is shown in Figure 45.1 above.

In MPEG, a *macroblock* is the area covered by a 16×16 array of luma samples. In DV, a macroblock comprises the Y', C_B, and C_R blocks covered by an 8×8 array (block) of chroma samples. In JPEG, an MCU comprises those blocks covered by the minimum-sized tiling of Y', C_B, and C_R blocks. For 4:2:0 subsampling, all of these definitions are equivalent; they differ for 4:1:1 and 4:2:2 (or for JPEG's other rarely used patterns).

In desktop graphics, saving JPEG at high quality may cause individual $R'G'B'$ channels (components) to be compressed without subsampling.

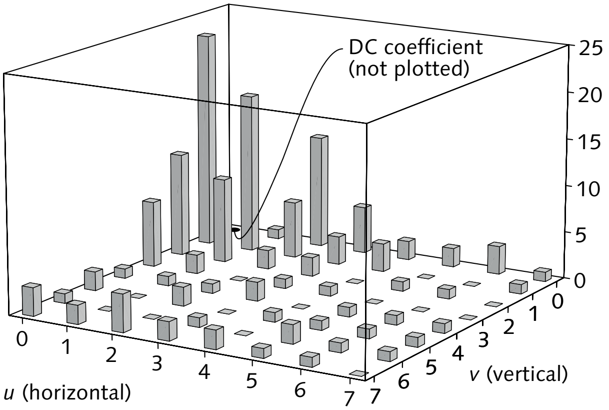
The luma and colour difference blocks are processed independently by JPEG, using virtually the identical algorithm. The only significant difference is that the quantizer matrix and the VLE tables used for chroma blocks are usually different from the quantizer matrix and VLE tables used for luma blocks.

Quantizer matrices and VLE tables will be described in the example starting on page 496.

As explained in *Spatial frequency domain* on page 238, typical images are dominated by power at low spatial frequencies. In Figure 45.4, on page 496, I present an example 8×8 array of luma samples from an image. In Figure 45.2 at the top of the facing page, I show an 8×8 array of the spatial frequencies computed from this luma array through the DCT. The [0, 0] entry (the *DC term*), at the upper left-hand corner of that array represents power at zero frequency. That entry typically contains quite a large value; it is not

I use zero-origin array indexing.

Figure 45.2 **The DCT concentrates image power** at low spatial frequencies. In Figure 45.4, on page 496, I give an example 8×8 array of luma samples from an image. The magnitudes of the spatial frequency coefficients after the DCT transform are shown in this plot. Most of the image power is collected in the [0, 0] (DC) coefficient, whose value is so large that it is omitted from this plot. Only a handful of other (AC) coefficients are much greater than zero.



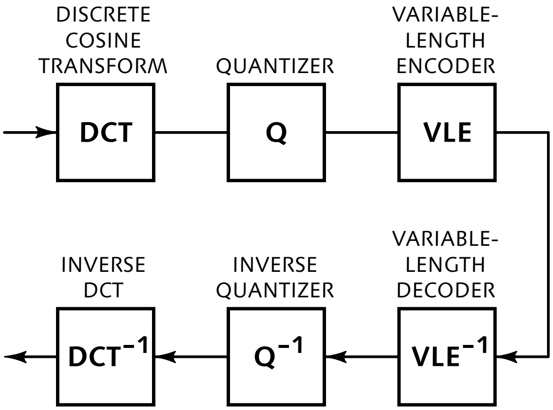
plotted here. Coefficients near that one tend to have fairly high values; coefficients tend to decrease in value further away from [0, 0]. Depending upon the image data, a few isolated high-frequency coefficients may have high values.

This typical distribution of image power, in the spatial frequency domain, represents the redundancy present in the image. The redundancy is reduced by coding the image in that domain, instead of coding the sample values of the image directly.

In addition to its benefit of removing redundancy from typical image data, representation in spatial frequency has another advantage. The lightness sensitivity of the visual system depends upon spatial frequency: We are more sensitive to low spatial frequencies than high, as can be seen from the graph in Figure 23.5, on page 252. Information at high spatial frequencies can be degraded to a large degree, without having any objectionable (or perhaps even perceptible) effect on image quality. Once image data is transformed by the DCT, high-order coefficients can be approximated – that is, coarsely quantized – to discard data corresponding to spatial frequency components that have little contribution to the perceived quality of the image.

In principle, the DCT algorithm could be applied to any block size, from 2×2 up to the size of the whole image, perhaps 512×512. (DCT is most efficient when applied to a matrix whose dimensions are powers of

Figure 45.3 **The JPEG block diagram** shows the encoder (at the top), which performs the *discrete cosine transform* (DCT). The DCT is followed by a *quantizer* (Q), then a *variable-length encoder* (VLE). The decoder (at the bottom) performs the inverse of each of these operations, in reverse order.



two.) The choice of 8×8 blocks of luma for the application of DCT in video represents a compromise between a block size small enough to minimize storage and processing overheads, but large enough to effectively exploit image redundancy.

The DCT operation discards picture information to which vision is insensitive. Surprisingly, though, the JPEG standard itself makes no reference to perceptual uniformity. Because JPEG's goal is to represent visually important information, it is important that so-called *RGB* values presented to the JPEG algorithm are first subject to a nonlinear transform such as that outlined in *Perceptual uniformity,* on page 8, that mimics vision.

JPEG block diagram

The JPEG block diagram in Figure 45.3 shows, at the top, the three main blocks of a JPEG encoder: the *discrete cosine transform* (DCT) computation (sometimes called *forward DCT*, FDCT), *quantization* (Q), and *variable-length encoding* (VLE). The decoder (at the bottom of Figure 45.3) performs the inverse of each of these operations, in reverse order. The inverse DCT is sometimes denoted *IDCT*; inverse quantization is sometimes called *dequantization*, and sometimes denoted *IQ.*

Inverse quantization (IQ) has no relation to the historical NTSC *IQ* colour difference components.

Owing to the eight-line-high vertical transform, eight lines of image memory are required in the DCT subsystem of the encoder, and in the IDCT (DCT^{-1}) subsystem of the decoder. When the DCT is implemented in separable form, as is almost always the case, this is called *transpose* memory.

Level shifting

The DCT formulation in JPEG is intended for signed sample values. In ordinary hardware or firmware, the DCT is implemented in fixed-point, two's complement arithmetic. Standard video interfaces use offset binary representation, so each luma or colour difference sample is *level shifted* prior to DCT by subtracting 2^{k-1}, where k is the number of bits in use.

Discrete cosine transform (DCT)

The 8×8 forward DCT (FDCT) takes an 8×8 array of 64 sample values (denoted $\boldsymbol{f}$, whose elements are $f_{i,j}$), and produces an 8×8 array of 64 transform coefficients (denoted $\boldsymbol{F}$, whose elements are $F_{u,v}$). The FDCT is expressed by this equation:

Eq 45.1

$$F_{u,v} = \frac{1}{4} C(u) C(v) \sum_{i=0}^{7} \sum_{j=0}^{7} f_{i,j} \cos\left[\frac{(2i+1)u\pi}{16}\right] \cos\left[\frac{(2j+1)v\pi}{16}\right];$$

$$C(w) = \begin{cases} \dfrac{1}{\sqrt{2}}; & w = 0 \\ 1; & w = 1, 2, \ldots, 7 \end{cases}$$

The cosine terms need not be computed on-the-fly; they can be precomputed and stored in tables.

The inverse transform – the IDCT, or DCT^{-1} – is this:

Eq 45.2

$$f_{i,j} = \frac{1}{4} \sum_{u=0}^{7} \sum_{v=0}^{7} C(u) C(v) F_{u,v} \cos\left[\frac{(2i+1)u\pi}{16}\right] \cos\left[\frac{(2j+1)v\pi}{16}\right]$$

The forward and inverse transforms involve nearly identical arithmetic: The complexity of encoding and decoding is very similar. The DCT is its own inverse (within a scale factor), so performing the DCT on the transform coefficients would perfectly reconstruct the original samples, subject only to the roundoff error in the DCT and IDCT.

If implemented directly according to these equations, an 8×8 DCT requires 64 multiply operations (and 49 additions) for each of the 64 result coefficients, for a total of 4096 multiplies, an average of 8 multiplication operations per pixel. However, the DCT is *separable:* an 8×8 DCT can be computed as eight 8×1 horizontal transforms followed by eight 1×8 vertical transforms. This optimization, combined with other optimizations comparable to those of the fast Fourier transform (FFT), greatly reduces computational complexity: A fully optimized 8×8 DCT requires as few as 11 multiplies for each 8 samples (or in an IDCT, transform coefficients).

JPEG encoding example

I will illustrate JPEG encoding by walking through a numerical example. Figure 45.4 represents an 8×8 array of luma samples from an image, prior to level shifting:

Figure 45.4 **An 8×8 array of luma samples** from an image is shown. This 8×8 array is known in JPEG terminology as a *block*.

$$\boldsymbol{f} = \begin{bmatrix} 139 & 144 & 149 & 153 & 155 & 155 & 155 & 155 \\ 144 & 151 & 153 & 156 & 159 & 156 & 156 & 156 \\ 150 & 155 & 160 & 163 & 158 & 156 & 156 & 156 \\ 159 & 161 & 162 & 160 & 160 & 159 & 159 & 159 \\ 159 & 160 & 161 & 162 & 162 & 155 & 155 & 155 \\ 161 & 161 & 161 & 161 & 160 & 157 & 157 & 157 \\ 162 & 162 & 161 & 163 & 162 & 157 & 157 & 157 \\ 162 & 162 & 161 & 161 & 163 & 158 & 158 & 158 \end{bmatrix}$$

The result of computing the DCT, rounded to integers, is shown in Figure 45.5:

Figure 45.5 **The DCT tends to concentrate** the power of the image block into low-frequency DCT coefficients (those in the upper left-hand corner of the matrix). No information is lost at this stage. The DCT is its own inverse, within a scale factor, so performing the DCT on these transform coefficients would reconstruct the original samples (subject only to roundoff error).

$$\boldsymbol{F} = \begin{bmatrix} 1260 & -1 & -12 & -5 & 2 & -2 & -3 & 1 \\ -23 & -17 & -6 & -3 & -3 & 0 & 0 & 1 \\ -11 & -9 & -2 & 2 & 0 & -1 & -1 & 0 \\ -7 & -2 & 0 & 1 & 1 & 0 & 0 & 0 \\ -1 & -1 & 1 & 2 & 0 & -1 & 1 & 1 \\ 2 & 0 & 2 & 0 & -1 & 1 & 1 & -1 \\ -1 & 0 & 0 & -1 & 0 & 2 & 1 & -1 \\ -3 & 2 & -4 & -2 & 2 & 1 & -1 & 0 \end{bmatrix}$$

This example shows that image power is concentrated into low-frequency transform coefficients – that is, those coefficients in the upper left-hand corner of the DCT matrix. No information is lost at this stage. The DCT is its own inverse, so performing the DCT a second time would perfectly reconstruct the original samples, subject only to the roundoff error in the DCT and IDCT.

In MPEG-2, DC terms can be coded with 8, 9, or 10 bits – or, in 4:2:2 profile, 11 bits – of precision.

As expressed in Equation 45.1, the arithmetic of an 8×8 DCT effectively causes the coefficient values to be multiplied by a factor of 8 relative to the orig-

$$
\boldsymbol{Q} = \begin{bmatrix}
16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\
12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\
14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\
14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\
18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\
24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\
49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\
72 & 92 & 95 & 98 & 112 & 100 & 103 & 99
\end{bmatrix}
$$

Figure 45.6 **A typical JPEG quantizer matrix** reflects the visual system's poor sensitivity to high spatial frequencies. Transform coefficients can be approximated, to some degree, without introducing noticeable impairments. The quantizer matrix codes a step size for each spatial frequency. Each transform coefficient is divided by the corresponding quantizer value; the remainder (or fraction) is discarded. Discarding the fraction is what makes JPEG lossy.

inal sample values. The value 1260 in the [0, 0] entry – the *DC* coefficient, or term – is $\frac{1}{8}$ of the sum of the original sample values. (All of the other coefficients are referred to as *AC*.)

The human visual system is not very sensitive to information at high spatial frequencies. Information at high spatial frequencies can be discarded, to some degree, without introducing noticeable impairments. JPEG uses a *quantizer matrix* (***Q***), which codes a step size for each of the 64 spatial frequencies. In the quantization step of compression, each transform coefficient is divided by the corresponding quantizer value (step size) entry in the ***Q*** matrix. The remainder (fraction) after division is discarded.

It is not the DCT itself, but the discarding of the fraction after quantization of the transform coefficients, that makes JPEG lossy!

In MPEG, default quantizer matrices are standardized, but they can be overridden by matrices conveyed in the bitstream.

JPEG has no standard or default quantizer matrix; however, sample matrices given in a nonnormative appendix are often used. Typically, there are two matrices, one for luma and one for colour differences.

An example ***Q*** matrix is shown in Figure 45.6 above. Its entries form a radially symmetric version of Figure 23.5, on page 252. The [0, 0] entry in the quantizer matrix is relatively small (here, 16), so the DC term is finely quantized. Further from [0, 0], the entries get larger, and the quantization becomes more coarse. Owing to the large step sizes associated with the high-order coefficients, they can be represented by fewer bits.

In the JPEG and MPEG standards, and in most JPEG-like schemes, each entry in the quantizer matrix takes a value between 1 and 255.

At first glance, the large step size associated with the DC coefficient (here, $Q_{0,0} = 16$) looks worrisome: With 8-bit data ranging from –127 to +128, owing to the divisor of 16, you might expect this quantized coefficient to be be represented with just 4 bits. However, as mentioned earlier, the arithmetic of Equation 45.1 scales the coefficients by 8 with respect to the sample values, so a quantizer value of 16 corresponds to 7 bits of precision when referenced to the sample values.

DCT coefficients after quantization, and after discarding the quotient fractions, are shown in Figure 45.7:

Figure 45.7 **DCT coefficients after quantization** are shown. Most of the high-frequency information in this block – DCT entries at the right and the bottom of the matrix – are quantized to zero. The nonzero coefficients have small magnitudes.

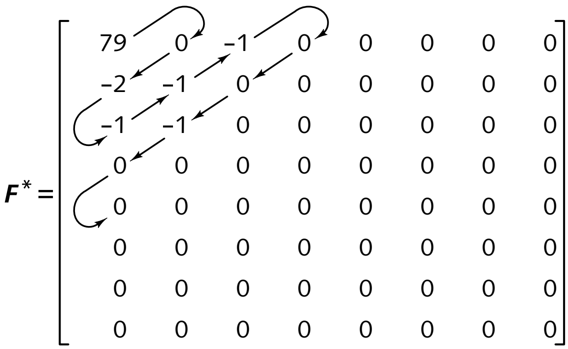
$$\boldsymbol{F}^* = \begin{bmatrix} 79 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Most of the high-frequency information in this block – the DCT entries at the right and the bottom of the matrix – are quantized to zero. Apart from the DC term, the nonzero coefficients have small magnitudes.

Following quantization, the quantized coefficients are rearranged according to the likely distribution of image power in the block. This is accomplished by *zigzag scanning*, sketched in Figure 45.8 at the top of the facing page.

Once rearranged, the quantized coefficients are represented in a one-dimensional string; an *end of block* (EOB) code marks the location in the string where all

Figure 45.8 **Zigzag scanning** is used to rearrange quantized coefficients according to the likely distribution of image power in the block.



$$F^* = \begin{bmatrix} 79 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

succeeding coefficients are zero, as sketched in Figure 45.9:

79	0	–2	–1	–1	–1	0	0	–1	EOB

Figure 45.9 **Zigzag-scanned coefficient string**

In the usual case that just a few high-order quantized coefficients are nonzero, zigzag reordering tends to produce strings of repeating zeros. Additional compression can be accomplished by using *variable-length encoding* (VLE, also known as *Huffman* coding). Variable-length encoding is a lossless process that takes advantage of the statistics of the "run length" (the count of zero codes) and the "level" (absolute value, or magnitude) of the following transform coefficient.

In JPEG and MPEG terminology, the magnitude (absolute value) of a coefficient is called its *level*.

The DC term is treated specially: It is differentially coded. The first DC term is coded directly (using a DC VLE table), but successive DC terms are coded as differences from that. In essence, the previous DC term is used as a predictor for the current term. Separate predictors are maintained for Y', C_B, and C_R.

Zero AC coefficients are collapsed, and the string is represented in {run length, level} pairs, as shown in Figure 45.10:

{1: –2}, {0: –1}, {0: –1}, {0: –1}, {2: –1}, EOB

Figure 45.10 **VLE {run length, level} pairs**

A JPEG encoder has one or more VLE tables that map the set of {run length, level} pairs to variable-length bitstrings; pairs with high probability are assigned short bitstrings. JPEG has no standard VLE tables; however, sample tables given in a nonnormative appendix are often used. Typically, there are two tables, one for luma and one for colour differences. The tables used for an

MPEG's VLE tables are standardized; they do not need to be transmitted with each sequence or each picture.

image are included at the head of the JPEG bitstream, and thereby conveyed to the decoder.

JPEG decoding

Decompression is achieved by performing the inverse of the encoder operations, in reverse order. Figure 45.11 shows the matrix of differences between original sample values and reconstructed sample values for this example – the reconstruction error. As is typical of JPEG, the original sample values are not perfectly reconstructed. However, discarding information according to the spatial frequency response of human vision ensures that the errors introduced during compression will not be too perceptible.

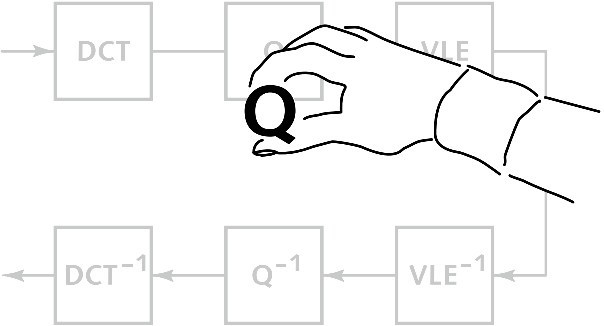
JPEG performance is loosely characterized by the error between the original image data and the reconstructed data. Metrics such as mean-squared error (MSE) are used to objectify this measure; however, MSE (and other engineering and mathematical measures) don't necessarily correlate well with subjective performance. In practice, we take care to choose quantizer matrices according to the properties of perception. Imperfect recovery of the original image data after JPEG decompression effectively adds noise to the image. Imperfect reconstruction of the DC term can lead to JPEG's 8×8 blocks becoming visible – the JPEG *blocking artifact.*

The lossiness of JPEG, and its compression, come almost entirely from the quantizer step. The DCT itself may introduce a small amount of roundoff error; the inverse DCT may also introduce a slight roundoff error. The variable-length encoding and decoding processes are perfectly lossless.

Figure 45.11 **Reconstruction error** is shown in this matrix of differences between original sample values and reconstructed sample values. Original sample values are not perfectly reconstructed, but discarding information according to the spatial frequency response of human vision ensures that the errors will not be too perceptible.

$$\varepsilon = \begin{bmatrix} -5 & -2 & 0 & 1 & 1 & -1 & -1 & -1 \\ -4 & 1 & 1 & 2 & 3 & 0 & 0 & 0 \\ -5 & -1 & 3 & 5 & 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & -2 & -1 & 0 & 2 & 4 \\ -4 & -3 & -3 & -1 & 0 & -5 & -3 & -1 \\ -2 & -2 & -3 & -3 & -2 & -3 & -1 & 0 \\ 2 & 1 & -1 & 1 & 0 & -4 & -2 & -1 \\ 4 & 3 & 0 & 0 & 1 & -3 & -1 & 0 \end{bmatrix}$$

Figure 45.12 **Compression ratio control in JPEG** is effected by altering the quantizer matrix: The larger the entries in the quantizer matrix, the higher the compression ratio. The higher the compression ratio, the higher the reconstruction error. At some point, compression artifacts will become visible.



Compression ratio control

The larger the entries in the quantizer matrix, the higher the compression ratio. Compression ratio control in JPEG can be achieved by altering the quantizer matrix, as suggested by the manual control sketched in Figure 45.12. Larger step sizes give higher compression ratios, but image quality is liable to suffer if the step sizes get too big. Smaller step sizes give better quality, at the expense of poorer compression ratio. There is no easy way to predict, in advance of actually performing the compression, how many bytes of compressed data will result from a particular image.

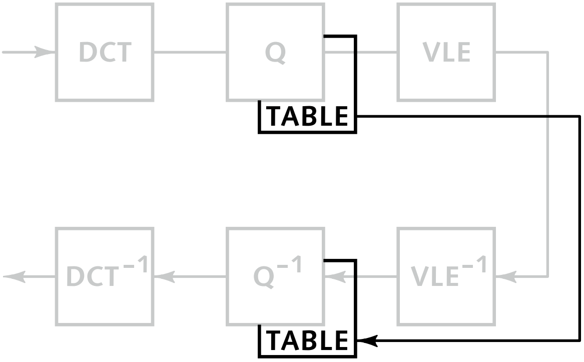
In ISO JPEG, the quantizer matrix is directly conveyed in the bitstream. In the DV adaptation of JPEG, several quantizer matrices are defined in the standard; the bitstream indicates which one to use.

The quantizer matrix could, in principle, be chosen adaptively to maximize the performance for a particular image. However, this isn't practical. JPEG encoders for still images generally offer a choice of several compression settings, each associated with a fixed quantizer that is chosen by the system designer.

Because different quantizer matrices may be associated with different images, the quantizer matrix must be conveyed to the decoder, as sketched in Figure 45.13, either as part of the file, or through a side channel. In colour images, separate quantizers are typically used for the luma and chroma components. In stillframe applications, the overhead of this operation is small. In a realtime system, the overhead of conveying quantizer matrices with every frame, or even within a frame, is a burden.

A modified approach to compression ratio control is adopted in many forms of M-JPEG (and also, as you will see in the next chapter, in MPEG): Reference luma and chroma quantizer matrices are established, and all of

Figure 45.13 **Because the quantizer is adjustable,** the quantizer matrix must be conveyed through a side channel to the decoder. In colour images, separate quantizers are used for the luma and chroma components.



The notation *mquant* is found in the ITU-T H.261 standard for teleconferencing; *mquant* (or *MQUANT*) is not found in JPEG or MPEG documents, but is used informally.

their entries are scaled up and down by a single numerical parameter, the *quantizer scale factor* (QSF, sometimes denoted *Mquant*). QSF can be varied to accomplish rate control.

As mentioned earlier, JPEG ordinarily uses luma/chroma coding with 4:2:0 chroma subsampling. However, the JPEG standard accommodates *R'G'B'* image data without subsampling, and also accommodates four-channel image data (such as CMYK, used in print) without subsampling. These schemes are inapplicable to video.

JPEG/JFIF

The ISO/IEC standard for JPEG defines a bitstream, consistent with the original expectation that JPEG would be used across communication links. To apply the JPEG technique to computer files, a small amount of supplementary information is required; in addition, it is necessary to encode the ISO/IEC *bit*stream into a *byte*stream. The de facto standard for single-image JPEG files is the *JPEG File Interchange Format* (JFIF), adopted by an industry group led by C-Cube. A JFIF file encapsulates a JPEG bitstream, along with a small amount of supplementary data. If you are presented with an image data file described as JPEG, it is almost certainly an ISO/IEC JPEG bitstream in a JFIF wrapper.

Hamilton, Eric (1992), *JPEG File Interchange Format*, Version 1.02 (Milpitas, Calif.: C-Cube Microsystems). This informal document was endorsed by ECMA and was published in June 2009 as ECMA TR/98 having the same title.

The JPEG standard itself implies that JPEG could be applied to linear-light *RGB* data. However, JPEG has poor visual performance unless applied to perceptually coded image data, that is, to gamma-corrected *R'G'B'*.

The ISO/IEC JPEG standard itself seems to suggest that the technique can be applied to arbitrary *RGB* data. The standard itself fails to mention primary chromaticities, white point, transfer function, or gamma correction. If accurate colour is to be achieved, then means outside the standard must be employed to convey these parameters. Prior to Mac OS X 10.6, there were two classes of JPEG/JFIF files, PC and Macintosh. Files that were created on classic Macintosh conformed to the default Macintosh coding, where *R'G'B'* codes were expected to be raised to the 1.52 power to produce display tristimulus. There was no reliable way to distinguish the two classes of files. Files created on PCs, and modern Macs, are interpreted in sRGB coding.

Motion-JPEG (M-JPEG)

Motion-JPEG (M-JPEG) refers to the use of a JPEG-like algorithm to compress every picture in a sequence of video fields or frames. I say "JPEG-like": The algorithms used have all of the general features of the algorithm standardized by JPEG, including DCT, quantization, zigzag scanning, and variable-length encoding. However, ISO JPEG bitstreams are not typically produced, and some systems add algorithmic features outside of the JPEG standard. Various M-JPEG systems are widely used in desktop video editing; however, there are no well established standards, and compressed video files typically cannot be interchanged between M-JPEG systems.

In studio applications, file interchange is a practical necessity, and two approaches have emerged. Both are functionally equivalent to M-JPEG, but have firm standards.

The first approach is DV compression, developed for consumer digital recording on videotape, but also used in desktop video editing. DV compression is described in the following chapter.

The second approach is MPEG-2 video compression, described in Chapter 47, on page 513. MPEG-2 was developed to exploit interframe coherence to achieve much higher compression ratios than M-JPEG, and is intended mainly for video distribution. However, the I-picture-only variant of MPEG-2 (sometimes called *I-frame-only*) is functionally equivalent to M-JPEG, and is being used for studio editing.

A few studio DVTR formats, such as Digital Betacam and HD-D5, use M-JPEG-style compression, but are not intimately related to any of JPEG, DV, or MPEG.

Further reading

Clarke describes the theory of transform coding of stilll images. Rabbani and Jones have written an excellent introduction to the mathematics of still image compression. Symes provides an approachable introduction to video compression.

Clarke, R.J. (1985), *Transform Coding of Images* (Boston: Academic Press).

Rabbani, Majid, and Paul W. Jones (1991), *Digital Image Compression Techniques* (Bellingham, Wash.: SPIE).

Symes, Peter (2003), *Digital Video Compression* (New York: McGraw-Hill).

DV compression

46

DV denotes the compression and data packing scheme introduced for consumer *digital video cassette* (DVC) recorders and later adapted for professional use. DV compression uses discrete cosine transform (DCT), quantization, and variable-length encoding (VLE) comparable to JPEG; however, DV does not *conform* to the JPEG standard: Optimizations have been made to accommodate interlaced scanning, constant bit-rate (CBR) operation, and other features related to videotape recording. Interlace is handled by allowing the encoder to dynamically choose between frame and field modes. Constant bit-rate is achieved by dynamically altering the quantization matrices to avoid exceeding the available capacity.

Concerning DVC recording of SD, see *DV recording*, in Chapter 9 of *Composite NTSC and PAL: Legacy Video Systems*. DVCPRO and DVCPRO 50 have been standardized as SMPTE D-7. JVC's Digital-S was standardized as D-9. See *Professional DV variants*, on page 512.

Consumer DVC has a data rate of 25 Mb/s. I call this *DV25*. DV25 coding was adopted for studio use in D-7 (DVCPRO) and DVCAM; it was extended to 50 Mb/s (*DV50*), used in the DVCPRO50 and D-9 systems, and then to 100 Mb/s (*DV100*), used in D-11 (DVCPRO HD).

DV25 and DV50 compress either 480*i*29.97 or 576*i*25 SD video according to BT.601. DV100 compresses HD video according to BT.709. Chroma subsampling in DV is a smorgasbord; see Table 46.1.

	480*i*	576*i*	HD
DV25 (consumer, DVCAM)	4:1:1	4:2:0	
DV25 (DVCPRO, D-7)	4:1:1	4:1:1	
DV50 (DVCPRO50, D-9)	4:2:2	4:2:2	
DV100 (DVCPRO HD, D-12)			4:2:2

Table 46.1 **DV chroma subsampling**

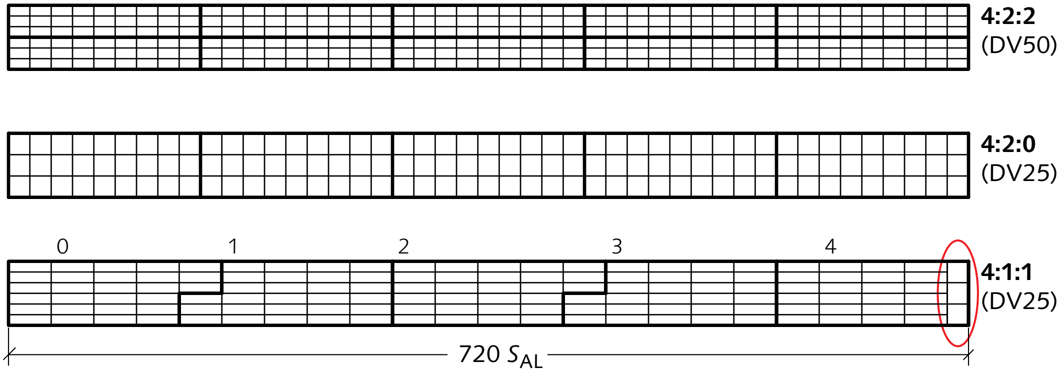


Figure 46.1 **DV superblocks** are shown for three chroma subsampling schemes – 4:2:2 (top), 4:2:0 (middle), and 4:1:1 (bottom). Thin lines enclose *macroblocks.* Each macroblock contains an 8×8 block of C_B, an 8×8 block of C_R, and two or more 8×8 blocks of luma. Thick lines enclose *superblocks;* each superblock comprises 27 macroblocks. A 4:1:1 macroblock is about four times wider than it is tall.

DV chroma subsampling

Different versions of DV use different subsampling schemes: 4:2:2, 4:1:1, 4:2:0, and even 3:1:1 and 3:1:0! In DV, a macroblock comprises an 8×8 block of C_B samples, an 8×8 block of C_R samples, and the requisite number and arrangement of 8×8 blocks of luma samples.

SMPTE 314M defines DV25 and DV50 for studio use. The Blue Book, and IEC standards, use the word *decimated* instead of *discarded.*
IEC 61834-1, cited in Chapter 9, *Videotape recording,* of *Composite NTSC and PAL: Legacy Video Systems*, prescribes the subsampling schemes for consumer DV.

SMPTE 314M declares that in subsampling 4:2:2 to 4:1:1, "every other pixel is discarded." Obviously, high image quality requires that proper filtering be performed before discarding samples. In DV, C_B and C_R samples coincide with luma both horizontally and vertically. However, in the 4:2:0 scheme used in 576*i* consumer equipment, C_R samples are not sited at the same locations as C_B samples. Instead, C_B and C_R samples are sited in line-alternate vertical positions throughout each field: Each C_B sample is centered two image rows below an associated C_R sample.

The DV format originated for consumer applications. Consumer DV25 aggregates 27 macroblocks into a *superblock* (SB), covering 48 image rows, whose arrangement is depicted in Figure 46.1 above. In 480*i*29.97 DV25, there are 50 superblocks per frame; in 576*i*25 DV25, there are 60 superblocks per frame.

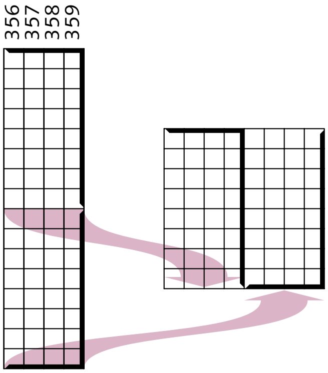


Figure 46.2 **Chroma samples in 4:1:1 DV** require special treatment at the right-hand edge of the picture, in the region circled in Figure 46.1. Here I show chroma as if square-sampled; actually, each 4:1:1 chroma sample is about four times wider than tall.

For consumer 576*i*25 DV25, and the 576*i*25 variant of DVCAM, 4:2:0 chroma subsampling was chosen, allegedly to offer some compatibility with DVB. The 4:2:0 superblock structure is shown in the middle sketch of Figure 46.1; in 4:2:0, each superblock comprises a regular array of 9×3 macroblocks.

For consumer 480*i*29.97 DV25, 4:1:1 subsampling was chosen. This was sufficient to achieve NTSC chroma bandwidth, and avoided conflict between subsampling and interlace. BT.601 video adapted to 4:1:1 subsampling has 180 active chroma samples per line; however, 180 isn't divisible by 8, so a regular superblock structure wasn't possible. To form 8×8 chroma blocks, C_B and C_R samples at the right-hand edge of the image are treated strangely, as detailed in Figure 46.2 in the margin: Pairs of vertically adjacent 4×8 arrays of chroma samples from the right end of each line are abutted horizontally to form 8×8 blocks.

The studio version of DV25 (used in D-7) uses 4:1:1 chroma subsampling in both 480*i*29.97 and 576*i*25; the strange 4:1:1 superblock structure is used.

DV50 has twice the data rate of DV25. DV50 uses 4:2:2 chroma subsampling for both 480*i*29.97 and 576*i*25. Owing to 4:2:2 subsampling, each macroblock has four blocks (instead of the six blocks of either 4:1:1 or 4:2:0); a frame has twice as many macroblocks as DV25, and twice as many superblocks. The 4:2:2 superblock structure is depicted in the top sketch of Figure 46.1. Today's DV50 recorders are implemented using two DV25 encoders, processing alternate 24-row bands of the image; they record the resulting two channels to parallel tracks.

DV frame/field modes

ISO JPEG assumes a progressive source image: All image rows are processed in spatial order. In video, the two fields of an interlaced frame can be woven together and processed as if progressive, where every 8×8 block has alternate rows taken from opposite fields. Providing interfield motion is limited, this scheme works well. In the presence of interfield motion, though, the high vertical frequency AC coefficients produced by the DCT will have high magnitude. This produces low quality for a given data rate. Higher quality can be obtained if the fields are coded separately. DV dynamically adapts to the degree of interfield motion by allowing the compressor to choose, on a block-by-block basis, whether to use *8-8-DCT mode* or *2-4-8-DCT mode*.

See *weave*, on page 414.

In 8-8-DCT ("frame") mode, opposite fields are woven together and subject to a single DCT as an 8×8 block.

This scheme involves two 8×4 DCTs; it should be called *2-8-4-DCT!* MPEG-2 has a similar scheme: An encoder can choose *frame DCT coding* or *field DCT coding* for each macroblock in an interlaced MPEG-2 sequence.

In 2-4-8-DCT ("field") mode, two 8×4 arrays of samples from like fields are formed. The element-by-element sums of the two 8×4 arrays, and their element-by-element differences, are subject to separate DCTs. The sum matrix is associated with a DC term; the difference array has only AC terms. Each of the sum and difference DCT arrays is zigzag scanned, and the coefficients are alternated for joint VLE encoding.

DV standards do not dictate how an encoder is to choose between 8-8-DCT and 2-4-8-DCT modes. Typically, luma differences are analyzed in the spatial domain to detect interfield motion, and 2-4-8-DCT mode is chosen if interfield motion is significant.

Picture-in-shuttle in DV

It is a practical necessity that a VCR recover a usable picture in fast forward and rewind ("shuttle"). When a digital VCR is operating in shuttle mode, isolated sync blocks are read from the tape. In DV, recovering a usable picture in shuttle mode is made possible by having each sync block correspond to one *coded macroblock* (CM); an entire macroblock can thereby be reconstructed individually from an isolated sync block. This scheme precludes predictive coding of the DCT's DC terms: DC terms are represented directly in each CM.

The correspondence of sync blocks with coded macroblocks seems to require that *every* macroblock be coded into a fixed number of bits. In the DV system, every macroblock is associated with a 77-byte coded macroblock, and a sync block contains one CM. Each CM contains the DC term, and several AC terms, of each block's DCT. However, that's not the whole story.

DV overflow scheme

When a portion of an image is devoid of detail, a few low-frequency AC coefficients of its DCT may have significant magnitude, but nearly all high-frequency AC terms will have very small magnitude. When a portion of an image has a lot of fine detail, many high-frequency AC terms will have large magnitude. To reproduce the image accurately requires that these terms be recorded. Generally, increasing amounts of detail require increased data capacity.

It is unusual for an image to contain detail everywhere; typically, complexity is spatially concentrated in an image. A compression algorithm should adapt to the spatial distribution of image detail, by allocating bits where they are needed. If fine detail is distributed throughout an image – in a full-frame image of the leaves of a tree, perhaps – then even quite large reconstruction errors are likely to be imperceptible.

The requirement for picture-in-shuttle seems to preclude allocation of data capacity to the regions where more bits are required. However, DV implements an overflow scheme whereby the bits resulting from compression of a handful of macroblocks are shared among a handful of CMs. Should the VLE-coded AC coefficients for a complex block (augmented by a 4-bit EOB) require more bits than the fixed capacity assigned to that block in the CM, the overflow bits "spill" into other blocks whose capacity was not filled. Overflow data first spills into space that might remain in other blocks of the same CM; any remaining bits spill into available space in any of four other CMs associated with diverse regions of the image. The set of five coded macroblocks that share overflow space is called a *segment;* a segment has a fixed capacity of 385 bytes. The macroblocks of a segment are spatially distributed throughout the frame: No two macroblocks are taken from the same row or the same column of superblocks. This distribution exploits the statistical likelihood that only one or two of the macroblocks will be complex.

CMs in a segment are denoted *a* through *e*.

When a DV VCR is in shuttle mode, it is likely that a single CM will be recovered individually, without any of the other four CMs of its segment. Overflow data for the macroblock is quite likely to be missing. However, overflow data is limited to high-frequency AC coefficients. In shuttle playback, missing overflow coefficients are replaced by zero. This causes loss of picture detail in the reconstructed picture; however, the absence of these coefficients does not seriously degrade picture quality, and in any event users do not expect the same picture quality in shuttle mode as in normal playback.

DV quantization

The main challenge of DV encoding is to determine suitable quantization matrices for a segment's AC coefficients, such that when all of the quantized coefficients are subject to variable-length encoding, the VLE-coded coefficients just neatly fit in the available space. The goal is to quantize the AC terms as finely as possible, without exceeding the capacity of a segment. In essence, this is a form of rate control. Quantization of a segment takes place after the DCT, using this algorithm:

- First, each block in the segment is assigned to a *class* from 0 (fine) to 3 (coarse), representing the block's spatial complexity. DV standards provide a table that suggests how an encoder can assign a class number according to the magnitude of the largest AC term of a block; however, use of that table is not mandatory.
- Then, up to 15 trial quantization passes are made, to determine a *quantization number* (QNO) from 0 (coarse) to 15 (fine). Class number and quantization number are combined to determine a quantization matrix according to tables in the standard. For each trial QNO, DCT coefficients are quantized and zigzag scanned. Nonzero AC coefficient are identified; {run length, level} pairs are computed for each nonzero coefficient; and the required number of variable-length-encoded bits is accumulated. Quantization is eased by the fact that the entries in the quantization matrices are all powers of two (1, 2, 4, 8, 16); each coefficient quantization operation is merely a binary shift. The lookup and assembly of VLE-coded bitstream need not be performed at this stage; it suffices for now to accumulate the bit count.

What DV standards call *level* is the magnitude – that is, the absolute value – of the AC coefficient. Sign is coded separately.

The final QNO for the segment is the one that produces the largest number of bits not exceeding the capacity of the segment – for DV25, 500 bits for luma AC coefficients (including four 4-bit EOBs), and 340 bits for chroma AC coefficients (including two 4-bit EOBs).

For a more elaborate description, and the quantization tables, see Symes, cited on page 535.

Once the segment's QNO is determined, VLE coding takes place, and the CMs of the segment are assembled. Each CM starts with one byte containing its QNO and error concealment *status* (STA) bits. Each block includes its DC coefficient, its mode (8-8-DCT or 2-4-8-DCT), and its class. Finally, the VLE-coded AC coefficients are distributed in a deterministic three-pass algorithm – first to the associated block, then to unused space in other blocks of the same CM (if space is available), and finally to unused space in other CMs of the segment. QNO has been chosen such that sufficient space for all coefficients is guaranteed to be available: Every bit of every coefficient will be stored somewhere within the segment.

This scheme is described as *three-pass;* however, the first pass is trivial

Each CM comprises 77 bytes, including by a 1-byte header. In DV25, a CM includes four coded luma blocks and two coded chroma blocks:

- A coded luma block totals 14 bytes, and includes a 9-bit DC term, one mode bit, and a 2-bit class number. One hundred bits are available for AC coefficients.
- A coded chroma block totals 10 bytes, and includes a 9-bit DC term, one mode bit, and a 2-bit class number. Sixty-eight bits are available for AC coefficients.

For 4:2:2 subsampling in DV50, a CM has four blocks, not six; space that in 4:1:1 or 4:2:0 would be allocated to luma blocks is available for overflow data. For 3:1:0 subsampling (used in SDL, to be described in a moment), a CM has eight blocks, not six: Each luma block has 10 bytes; each chroma block has 8 bytes.

DV digital interface (DIF)

The superblocks that I have mentioned form the basis for digital interface of DV bitstreams. A 3-byte ID is prepended to each 77-byte coded macroblock to form an 80-byte *digital interface* (DIF) *block*.

A coded DV25 superblock is represented by 135 video DIF blocks. That is augmented with several nonvideo DIF blocks to form a *DIF sequence* of 150 DIF blocks:

- 1 header DIF block
- 2 subcode DIF blocks
- 3 VAUX DIF blocks
- 9 audio DIF blocks
- 135 video DIF blocks

Realtime DV25 video requires 10 or 12 DIF sequences – that is, about 1500 or 1800 DIF blocks – per second. DV50, and DV100 systems have comparable structures, but different data rates.

Once packaged in DIF sequences, DV bitstreams can be conveyed across the IEEE 1394 interface, also known as *FireWire* and *i.LINK*, that I described on page 167. IEEE 1394 is suitable for consumer use, and is widely used in desktop video. For professional applications, DIF sequence bitstreams can be transported across various interfaces including the SDTI interface that I introduced on page 441. (The 3-byte ID is unused in DV-over-SDTI.) DIF sequences can be stored in files – for example, QuickTime or MXF files.

Consumer DV recording

DV25 was widely adopted for consumer SD recording on MiniDV cassettes. Several schemes to extend DV to HD were described in the first edition of this book; none of these were commercialized. Consumer HD recording on MiniDV uses the HDV system, outlined on page 161.

Professional DV variants

DVCPRO and DVCAM data rate is 25 Mb/s, identical to consumer DV; however, the physical magnetic tape is more robust and more suitable for professional use.

DV technology was introduced for consumer videotape recording, and was widely deployed in consumer camcorders and desktop video editing systems. DV videotape technology was adapted to professional videotape recorders – first for SD as DVCPRO (D-7) and DVCAM, then for HD as DVCPRO HD (D-12). The DV compression system made the transition into products using hard disk drive and flash media.

DVCPRO50 DVTRs are standardized in SMPTE's D-7 series; DV50 SD bitstreams are recorded onto 6 mm tape in DVC-style cassettes.

DV50, also for SD, has twice the data rate, twice as many macroblocks per second (or per frame), and twice as many superblocks per second (or per frame) as DV25. DV50 uses 4:2:2 subsampling; a CM contains just two luma blocks instead of four. Space that in DV25 would have been allocated to AC terms of the other two luma blocks is available for overflow AC terms. The corresponding DC terms, mode bits, and class bits are reserved. In DV50, the first four bits of DV25's AC coefficient spaces are filled with EOB symbols.

Panasonic introduced the DVCPRO HD DVTR format, which was subsequently standardized as SMPTE D-12. See page 482.

DV100 doubles the DV50 data rate to 100 Mb/s, and accommodates HD image formats and 4:2:2 chroma subsampling: 1280×1080 downsampled from 1080*i* or 1080*p*, or 960×720 downsampled from 720*p*.

MPEG-2 video compression

47

I assume that you are familiar with *Introduction to video compression,* on page 147, and with JPEG, M-JPEG, and DV, described in the preceding two chapters.

The DCT-based intrafield or intraframe compression at the heart of M-JPEG is suitable for video production; however, for distribution, dramatically higher compression ratios can be obtained by using interframe coding. MPEG-2 video compression exploits temporal coherence – the statistical likelihood that successive pictures in a video sequence are very similar. MPEG-2's intended application ranges from below SD to beyond HD; the intended bit rate ranges from about 1.5 Mb/s to well over 20 Mb/s. MPEG-2 also defines audio compression, and provides for the transport of video with associated audio.

ISO/IEC 13818-1, *Generic coding of moving pictures and associated audio information: Systems* [MPEG-2], also published as ITU-T H.220.0.

ISO/IEC 13818-2, *Generic coding of moving pictures and associated audio information: Video* [MPEG-2], also published as ITU-T H.262.

MPEG-2 refers to a suite of standards, promulgated jointly by ISO/IEC and ITU-T. The suite starts with Part 1: *Systems* and Part 2: *Video,* cited in the margin, which are jointly published by ISO, IEC, and ITU-T. Six other parts are jointly published by ISO and IEC – Part 3: *Audio;* Part 4: *Conformance testing;* Part 5: *Software simulation;* Part 6: *Extensions for DSM-CC;* Part 7: *Advanced Audio Coding* (*AAC*); Part 9: *Extension for real time interface for systems decoders;* and Part 10: *Conformance extensions for Digital Storage Media Command and Control* (*DSM-CC*). The projected Part 8, for 10-bit video, was discontinued. MPEG-2 standards were first issued in 1996; subsequently, several corrigenda and amendments have been issued.

MPEG-2 specifies exactly what constitutes a legal bitstream: A legal ("conformant") encoder generates only legal bitstreams; a legal decoder correctly decodes any legal bitstream. MPEG-2 does *not* standardize how an encoder accomplishes compression!

The MPEG-2 standard implicitly defines exactly how a decoder reconstructs pictures data from a coded bitstream, without dictating the implementation of the decoder. MPEG-2 explicitly avoids specifying what it calls the "display process" – how reconstructed pictures are displayed. Most MPEG-2 decoder implementations have flexible output formats; however, MPEG-2 decoder equipment is ordinarily designed to output a specific raster standard.

An MPEG-2 bitstream may represent interlaced or progressive pictures. Typical decoder equipment outputs either interlace or progressive signals. Certain decoder equipment has the capability to switch between the two output formats. Because interlaced scanning remains dominant in consumer electronics – both in SD and in HD – a decoder system must be capable of producing an interlaced signal from a progressive sequence. Also, it is a practical necessity for an MPEG-2 decoder to have spatial resampling capability: If an HD MPEG-2 decoder is presented with an SD sequence, consumers would complain if reconstructed pictures were not upconverted for display in HD.

MPEG-2 profiles and levels

MPEG-2 specifies several algorithmic features – such as arbitrary frame rate, and 4:4:4 chroma subsampling – that are not permitted in any standard profile. These features are unlikely to see commercialization.

An MPEG-2 bitstream can potentially invoke many algorithmic features – some practitioners call them "tools" – at a decoder. Also, a bitstream can reflect many possible parameter values. The MPEG-2 standard classifies bitstreams and decoders in a matrix of *profiles* and *levels*.

Profiles constrain the algorithmic features potentially used by an encoder, present in a bitstream, or implemented in a decoder. The higher the profile, the more complexity is required of the decoder. MPEG-2 defines six profiles: *Simple* (SP), *Main* (MP), *4:2:2* (422P), *SNR*, *Spatial* (Spt), *High* (HP), and *Multiview* (MVP).

Levels place restrictions on parameter values used by an encoder or decoder. The higher the level, the more memory or data throughput is required of a decoder. MPEG-2 defines four levels: *Low* (LL), *Main* (ML), *High-1440* (H14), and *High* (HL).

A profile and level combination is indicated by profile and level separated by an at sign – for example, MP@ML or MP@HL. The SNR, Spatial, High, and

Profile *@Level*	**MPEG-1 CPB**	Simple (no B pictures)	Main (MP)	4:2:2 (422P)
High (HL)			1920×1152 60 Hz 80 Mb/s	1920×1088 60 Hz‡ 300 Mb/s‡
High-1440 (H14)			1440×1152 60 Hz 47 Mb/s	
Main (ML)		720×576 30 Hz 15 Mb/s	720×576 30 Hz 15 Mb/s	720×608 30 Hz 50 Mb/s
Low (LL)			352×288 30 Hz 4 Mb/s	
MPEG-1 CPB †*max 99 Kpx*	768×576† 30 Hz 1.856 Mb/s			

Table 47.1 **MPEG-2 profiles,** here arranged in columns, specify algorithmic features. (I exclude SNR, Spt, HP, and MVP.) **MPEG-2 levels,** here arranged in rows, constrain parameter values. Each entry gives maximum picture size, frame rate, and bit rate. The two shaded entries are commercially dominant: Main profile at main level (MP@ML) is used for SD distribution; main profile at high level (MP@HL) is used for HD distribution. SMPTE 308M places constraints on GoP structure for 422P@HL. Any compliant MPEG-2 decoder must decode an MPEG-1 *constrained-parameters bitstream* (CPB); the constrained parameters effectively constitute a profile/level combination.

Multiview profiles have no relevance to video production or distribution, and are unlikely to see commercial deployment. I won't discuss them further.

The profile and level combinations defined by MPEG-2 – excluding SNR, Spt, HP, and MVP – are summarized in Table 47.1 above. Excepting 422P, the combinations have a hierarchical relationship: A decoder claiming conformance to any profile must be capable of decoding all profiles to its left in Table 47.1; also, a decoder claiming conformance to any level must be capable of decoding all lower levels. Exceptionally, a simple profile at main level (SP@ML) decoder must be capable of decoding main profile at low level (MP@LL).

Every compliant MPEG-2 decoder must be capable of decoding an MPEG-1 *constrained-parameters bitstream* (CPB). I include MPEG-1 CPB at the lower left of Table 47.1, as if it were both a profile and a level, to emphasize this MPEG-2 conformance requirement.

Profile@Level	Image columns (N_C)	Image rows (N_R)	Frame rate, Hz	Luma rate [samples/s]	Bit rate [Mb/s]	VBV size [KBytes]
422P@HL	1920	1088	60	62,668,800	300	5,760
MP@HL	1920	1088	60	62,668,800	80	1,194
MP@H-14	1440	1088	60	47,001,600	60	896
422P@ML	720	608	60	11,059,200	50	1,152
MP@ML	720	576	30	10,368,000	15	224
MP@LL	352	288	30	3,041,280	4	58

Table 47.2 **MPEG-2 main and 4:2:2 profiles** are summarized. MP@ML and MP@HL are shaded to emphasize their commercial significance. The DVD-video specification requires MP@ML compliance, and imposes additional constraints. ATSC standards for 720*p*, 1080*p*, and 1080*i* HD require MP@HL compliance, and impose additional constraints.

The simple profile has no B-pictures. Prohibition of B-pictures minimizes encoding latency, and minimizes buffer storage at the decoder. However, the simple profile lacks the compression efficiency of B-pictures.

Of the eight combinations in Table 47.1, only two are commercially important to television. MP@ML is used for SD distribution, and for DVD, at rates from about 2 Mb/s to about 6 Mb/s. MP@HL is used for HD distribution, usually between 10 Mb/s and 20 Mb/s.

422P@ML allows 608 lines at 25 Hz frame rate, but is limited to 512 lines at 29.97 and 30 Hz frame rates.

The 4:2:2 profile allows 4:2:2 chroma subsampling; it is intended for use in television production. The major reason for a separate 4:2:2 profile is that main profile *disallows* 4:2:2 chroma subsampling. MPEG-2's high profile allows 4:2:2 subsampling, but to require high-profile conformance would oblige a decoder to handle SNR and spatial scalability. 422P@ML is used in the studio, as Sony MPEG IMX, at bit rates between 30 Mb/s and 50 Mb/s. Some numerical parameter limits of main and 4:2:2 profiles are presented in Table 47.2 above.

SMPTE 308M, *Television – MPEG-2 4:2:2 Profile at High Level*.

MPEG-2 defines 4:2:2 profile at high level (422P@HL). In addition to MPEG-2's requirements for 422P@HL, SMPTE 308M imposes these restrictions on

GoP structures permitted at high data rates, as shown in Table 47.3:

Bit rate	*Interlace?*	*GoP structure allowed*
0 to 175	any	any
175 to 230	any	I-only, IP, or IB
230 to 300	interlaced	I-only
	progressive	I-only, IP, or IB

Table 47.3 **GoP restrictions in SMPTE 308M**

I have presented the profile and level constraints of MPEG-2 itself. Certain applications of MPEG-2 – such as ATSC broadcasting, and DVD – impose restrictions beyond those of MPEG's profiles and levels. For example, MPEG-2 allows a frame rate of 24 Hz, but that rate is disallowed for DVD. (Movies originating at 24 frames per second are ordinarily coded onto 480*i* DVD at 29.97 Hz, but signalling 2-3 pulldown.)

Picture structure

Concerning the distinction between luminance and luma, see Appendix A, on page 567.

Each frame in MPEG-2 is coded with a fixed number of image columns (S_{AL}, called *horizontal size* in MPEG) and image rows (L_A, called *vertical size*) of luma samples. I use the term *luma;* MPEG documents use *luminance,* but that term is technically incorrect in MPEG's context.

A frame in MPEG-2 has either square sampling, or 4:3, 16:9, or 2.21:1 picture aspect ratio – that is, nonsquare sampling is permitted only at 4:3, 16:9, or 2.21:1. (MPEG writes aspect ratio unconventionally, as as *height:width*.) Table 47.4 presents MPEG-2's *aspect ratio information* field. The 2.21:1 value is not permitted in any defined profile.

Code	*Aspect ratio*
0000	Forbidden
0001	Square sampling
0010	4:3
0011	16:9
0100	2.21:1
0101 … 1111	Reserved

Table 47.4 **MPEG-2 *aspect ratio information***. The 2.21:1 aspect ratio is not permitted in any defined profile.

MPEG-2 accommodates both progressive and interlaced material. An image having N_C columns and N_R rows of luma samples can be coded directly as a *frame-structured* picture, as depicted at the left of Figure 47.1 overleaf. In a frame-structured picture, all of the luma and chroma samples of a frame are taken to originate at the same time, and are intended for display at the same time. A flag *progressive sequence* asserts that a sequence contains only frame-structured pictures.

Alternatively, a video frame (typically originated from an interlaced source) may be coded as a pair of *field-structured* pictures – a *top-field* picture and a *bottom-*

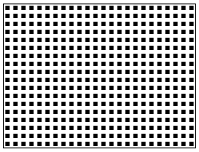


Figure 47.1 **An MPEG-2 frame picture** contains an array of luma samples N_C columns by N_R rows. It is implicit that an MPEG-2 frame picture occupies the entire frame time.

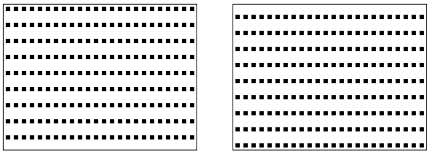


Figure 47.2 **An MPEG-2 field picture pair** contains a *top-field* picture and a *bottom-field* picture, each N_C columns by $N_R/2$ rows. The samples are vertically offset. Here I show a pair ordered {top, bottom}; alternatively, it could be {bottom, top}. Concerning the relation between top and bottom fields and video standards, see *Interlacing in MPEG-2,* on page 132.

field picture – each having N_C columns and $N_R/2$ rows as depicted by Figure 47.2. The two fields are time-offset by half the frame time, and are intended for interlaced display. Field pictures always come in pairs having opposite parity (top/bottom). Both pictures in a field pair must have the same *picture coding type* (I, P, or B), except that an I-field may be followed by a P-field (in which case the pair functions as an I-frame, and may be termed an *IP-frame*).

I, P, and B picture coding types were introduced on page 153.

Frame rate and 2-3 pulldown in MPEG

The defined profiles of MPEG-2 provide for the display frame rates shown in Table 47.5 in the margin. Frame rate is constant within a video *sequence* (to be defined on page 533). Unfortunately, it is unspecified how long a decoder may take to adapt to a change in frame rate.

In a sequence of frame-structured pictures, provisions are made to include, in the MPEG-2 bitstream, information to enable the display process to impose 2-3 pulldown upon display. Frames in such a sequence are coded as frame-structured pictures; in each frame, both fields are associated with the same instant in time. The flag *repeat first field* may accompany a picture; if that flag is set, then an interlaced display is expected to display the first field, the second field, and then the first field again – that is, to impose 2-3 pulldown. The frame rate in the bitstream specifies the display rate after 2-3 processing. I sketched a 2-3 sequence of four film frames in Figure 34.1, on page 405; on DVD, that sequence would be coded as the set of four progressive MPEG-2 frames flagged as indicated in Table 47.6 at the top of the facing page.

Code	*Frame rate*
0000	Forbidden
0001	$24/1.001$
0010	24
0011	25
0100	$30/1.001$
0101	30
0110	50
0111	$60/1.001$
1000	60
1001 … 1111	Reserved

Table 47.5 **MPEG-2 *frame rate code***

A frame-coded picture can code a top/bottom pair or a bottom/top pair – that is, a frame picture may correspond to a video frame, or may straddle two video frames. The latter case accommodates M-frames in 2-3 pulldown.

Film frame	*Top first field (TFF)*	*Repeat first field (RFF)*
A	0	0
B	0	1
C	1	0
D	1	1

Table 47.6 **2-3 pulldown sequence in MPEG-2.** For the definitions of film frames A through D, see *2-3 pulldown,* on page 405.

Luma and chroma sampling structures

MPEG-2 accommodates 4:2:2 chroma subsampling, suitable for studio applications, and 4:2:0 chroma subsampling, suitable for video distribution. Unlike DV, C_B and C_R sample pairs are spatially coincident. The MPEG-2 standard includes 4:4:4 chroma format, but it isn't permitted in any defined profile, so is highly unlikely to be commercialized.

There is no vertical subsampling in 4:2:2 – in this case, subsampling and interlace do not interact. 4:2:2 chroma for both frame-structured and field-structured pictures is depicted in the third (BT.601) column of Figure 12.1, on page 124.

4:2:0 chroma subsampling in a frame-structured picture is depicted in the rightmost column of Figure 12.1; C_B and C_R samples are centered vertically midway between luma samples in the frame, and are cosited horizontally.

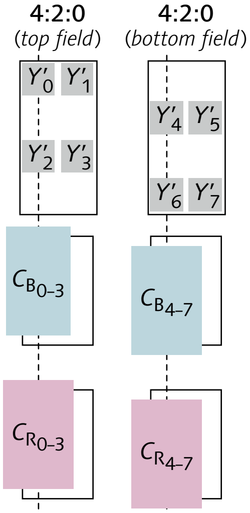


Figure 47.3 **Chroma subsampling in field-structured pictures**

4:2:0 chroma subsampling in a field-structured picture is depicted in Figure 47.3 in the margin. In the top field, chroma samples are centered ¼ of the way vertically between a luma sample in the field and the luma sample immediately below in the same field. (In this example, $C_{B_{0-3}}$ is centered ¼ of the way down from Y'_0 to Y'_2.) In the bottom field, chroma samples are centered ¾ of the way vertically between a luma sample in the field and the luma sample immediately below in the same field. (In this example, $C_{B_{4-7}}$ is centered ¾ of the way down from Y'_4 to Y'_6.) This scheme centers chroma samples at the same locations that they would take in a frame-structured picture; however, alternate rows of chroma samples in the frame are time-offset by half the frame time.

Macroblocks

If *horizontal size* or *vertical size* is not divisible by 16, then the encoder pads the image with a suitable number of black "overhang" samples at the right edge or bottom edge. For example, when coding HD at 1920×1080, an encoder appends 8 rows of black pixels to the image array, to make the row count 1088. Upon decoding, these samples are cropped prior to display. The overhang regions are retained in the reference framestores.

At the core of MPEG compression is the DCT coding of 8×8 blocks of sample values (in I-pictures, as in JPEG), or 8×8 blocks of prediction errors – *residuals,* or *residues* (in P- and B-pictures). To simplify the implementation of subsampled chroma, the same DCT and block coding scheme is used for both luma and chroma. When combined with 4:2:0 chroma subsampling, two 8×8 block of chroma samples are associated with a 16×16 block of luma. This leads to the tiling of a field or frame into units of 16×16 luma samples. Each such unit is a *macroblock* (MB). Macroblocks lie a on 16×16 grid aligned with the upper-left luma sample of the image.

Each macroblock comprises four 8×8 luma blocks, accompanied by the requisite number and arrangement of 8×8 C_B blocks and 8×8 C_R blocks, depending upon *chroma format*. In the usual 4:2:0 chroma format, a macroblock comprises six blocks: four luma blocks, a C_B block, and a C_R block. In the 4:2:2 chroma format, a macroblock comprises eight blocks: four luma blocks, two blocks of C_B, and two blocks of C_R.

Picture coding types – I, P, B

MPEG-2 and H.264 use the term *reference picture*. Some people say *anchor picture*.

I-pictures, P-pictures, and B-pictures were introduced on page 152. Coded I-picture and P-picture data are used to reconstruct *reference pictures* – fields or frames that are available for constructing predictions. An MPEG decoder maintains two reference framestores, one past and one future. An encoder also maintains two reference framestores, reconstructed as if by a decoder; these track the contents of the decoder's reference framestores. The simple profile has no B-pictures; a single reference framestore suffices.

Each I-picture is coded independently of any other picture. When an I-picture is reconstructed by a decoder, it is displayed. Additionally, it is stored as a reference frame so as to be available as a predictor. I-pictures are compressed using a JPEG-like algorithm, using perceptually based quantization matrices.

Each P-picture is coded using the past reference picture as a predictor. Residuals are compressed using the same JPEG-like algorithm that is used for I-pictures, but typically with quite different quantization matrices.

When a decoder reconstructs a P-picture, it is displayed; additionally, the picture is written into a reference frame so as to be available for subsequent predictions.

Each B-picture contains elements that are *bipredicted* from one or both reference frames. The encoder computes, compresses, and transmits residuals. The decoder reconstructs a B-picture, displays it, then discards it: No B-picture is used for prediction.

Each reference picture is associated with a full frame of storage. When a decoder reconstructs a reference *field* (an I-field or a P-field), half the lines of the reference framestore are written; the other half retains the contents of the previous reference field. After the first field of a field pair has been reconstructed, it is available as a predictor for the second field. (The first field of the previous reference frame is no longer available.)

Prediction

In Figure 16.1, on page 152, I sketched a naïve interpicture coding scheme. For any scene element that moves more than few pixels from one video frame to the next, the naïve scheme is liable to produce large interpicture difference values. Motion can be more effectively coded by having the encoder form motion-compensated predictions. The encoder also produces motion vectors; these are used to displace a region of a reference picture to improve the prediction of the current picture relative to an undisplaced prediction. The residuals are then compressed using DCT, quantized, and VLE-encoded.

At a decoder, predictions are formed from the reference picture(s), based upon the transmitted motion vectors and prediction modes. Residuals are recovered from the bitstream by VLE decoding, inverse quantization, and inverse DCT. Finally, the decoded residual is added to the prediction to form the reconstructed picture. If the decoder is reconstructing an I-picture or a P-picture, the reconstructed picture is written to the appropriate portion (or the entirety) of a reference frame.

Inverse quantization is sometimes denoted *IQ*, not to be confused with *IQ* colour difference components.

The obvious way for an encoder to form forward interpicture differences is to subtract the current source picture from the reference picture. (The reference picture would have been subject to motion-compensated interpolation, according to the encoder's motion estimate.) Starting from an intra coded picture, the decoder would then accumulate interpicture differences. However, MPEG involves lossy compression: Both the I-picture starting point of a GoP and each set of decoded interpicture differences are subject to reconstruction errors. With the naïve scheme of computing interpicture differences, reconstruction errors would accumulate at the decoder. To alleviate this potential source of decoder error, the encoder incorporates a decoder. The interpicture difference is formed by subtracting the current source picture from the previous reference picture *as a decoder will reconstruct it.* Reconstruction errors are thereby brought "inside the loop," and are prevented from accumulating.

A prediction region in a reference frame is rarely aligned to a 16-luma-sample macroblock grid; it is not properly referred to as a *macroblock*. Some authors fail to make the distinction between macroblocks and prediction regions; other authors use the term *prediction macroblocks* for prediction regions.

The prediction model used by MPEG-2 is blockwise translation of 16×16 blocks of luma samples (along with the associated chroma samples): A macroblock of the current picture is predicted from a like-sized region of a reconstructed reference picture. The choice of 16×16 region size was a compromise between the desire for a large region (to effectively exploit spatial coherence, and to amortize motion vector overhead across a fairly large number of samples), and a small region (to efficiently code small scene elements in motion).

In a *closed GoP*, no B-picture is permitted to use forward prediction to the I-picture that starts the next GoP. See the caption to Figure 16.5, on page 155.

Macroblocks in a P-picture are typically forward-predicted. However, an encoder can decide that a particular macroblock is best intracoded (that is, not predicted at all). Macroblocks in a B-picture are typically predicted as averages of motion-compensated past and future reference pictures – that is, they are ordinarily bidirectionally predicted. However, an encoder can decide that a particular macroblock in a B-picture is best intracoded, or unidirectionally predicted using either forward or backward prediction. Table 47.7 at the top of the facing page indicates the four macroblock types. The macroblock types allowed in any picture are restricted by the declared picture type, as indicated in Table 47.8.

Each nonintra macroblock in an interlaced sequence can be predicted either by frame prediction (typically chosen by the encoder when there is little motion

Table 47.7 **MPEG macroblock types**		*Prediction*	*Typ. quantizer matrix*
Intra		None – the macroblock is self-contained	Perceptual
Inter (Nonintra)	Backward predictive-coded	Predicts from the future reference picture	Flat
	Forward predictive-coded	Predicts from the past reference picture	Flat
	Bipredictive-coded	Averages predictions from past and future reference pictures	Flat

Table 47.8 **MPEG picture coding types**	*Binary code*	*Reference picture?*	*Permitted macroblock types*
I-picture	001	Yes	Intra
P-picture	010	Yes	Intra Forward predictive-coded
B-picture	011	No	Intra Forward predictive-coded Backward predictive-coded Bipredictive-coded

Table 47.9 **MPEG-2 prediction modes**			*Max. MVs*	
	For	*Description*	*back.*	*fwd.*
Frame prediction	(P, B)-pictures	Predictions are made for the frame, using data from one or two previously reconstructed frames.	1	1
Field prediction	(P, B)-pictures, (P, B)-fields	Predictions are made independently for each field, using data from one or two previously reconstructed fields.	1	1
16×8 motion compensation (16×8 MC)	(P, B)-fields	The upper 16×8 and lower 16×8 regions of the macroblock are predicted separately. (This is completely unrelated to top and bottom fields.)	2	2
Dual prime	P-fields with no intervening B-pictures	Two motion vectors are derived from the transmitted vector and a small *differential motion vector* (DMV, –1, 0, or +1); these are used to form predictions from two reference fields (one top, one bottom), which are averaged to form the predictor.	1	1
Dual prime	P-pictures with no intervening B-pictures	As in dual prime for P-fields (above), but repeated for 2 fields; 4 predictions are made and averaged.	1	1

between the fields), or by field prediction (typically chosen by the encoder when there is significant inter-field motion). This is comparable to field/frame coding in DV, which I described on page 507. Predictors for a field picture must be field predictors. However, predictors for a frame picture may be chosen on a macroblock-by-macroblock basis to be either field predictors or frame predictors. MPEG-2 defines several additional prediction modes, which can be selected on a macroblock-by-macroblock basis. MPEG-2's prediction modes are summarized in Table 47.9.

Motion vectors (MVs)

A *motion vector* identifies a region of 16×16 luma samples in a reference picture that are to be used for prediction. A motion vector refers to a prediction region that is potentially quite distant (spatially) from the region being coded – that is, the motion vector range can be quite large. Even in field pictures, motion vectors are specified in units of frame luma samples. A motion vector can specify integer pixel coordinates, in which case forming the 16×16 prediction is accomplished by merely copying pixels. However, in MPEG, a motion vector can be specified to half-sample precision: If the fractional bit of a motion vector is set, then the prediction is formed by averaging sample values at the neighboring integer coordinates – that is, by linear interpolation. Transmitted motion vector values are halved for use with subsampled chroma. All defined profiles require that no motion vector refers to any sample outside the bounds of the reference frame.

Each macroblock's header contains a count of motion vectors. Motion vectors are themselves predicted! An initial MV is established at the start of a *slice* (see page 534); the motion vector for each successive nonintra macroblock is differentially coded with respect to the previous macroblock in raster-scan order.

Motion vectors are variable-length encoded, so that short vectors – the most likely ones in large areas of translational motion or no motion – are coded compactly. Zero-valued motion vectors are quite likely, so provision is made for compact coding of them.

Intra macroblocks are not predicted, so motion vectors are not necessary for them. However, in certain circumstances *concealment motion vectors* (CMVs) are allowed: If a macroblock is lost owing to transmission error, CMVs allow a decoder to use its prediction facilities to synthesize picture information to conceal the erred macroblock. A CMV would be useless if it were contained in its own macroblock! So, a CMV is associated with the macroblock immediately below.

Coding of a block

Each macroblock is accompanied by a small amount of prediction mode information; zero, one, or more *motion vectors* (MVs); and DCT-coded residuals.

A perverse encoder could use an intra quantizer matrix that isn't perceptually coded.

Each block of an intra macroblock is coded similarly to a block in JPEG. Transform coefficients are quantized with a quantizer matrix that is (ordinarily) perceptually weighted. Provision is made for 8-, 9-, and 10-bit DC coefficients. (In the 422 profile [422P], 11-bit DC coefficients are permitted.) DC coefficients are differentially coded within a slice (to be described on page 534).

In an I-picture, DC terms of the DCT are differentially coded: The DC term for each luma block is used as a predictor for the corresponding DC term of the following macroblock. DC terms for C_B and C_R blocks are similarly predicted.

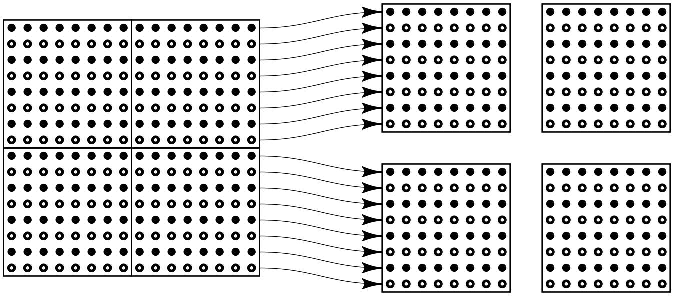
A perverse encoder could use a nonintra quantizer matrix that isn't flat. Separate nonintra quantizer matrices can be provided for luma and chroma.

In principle, residuals in a nonintra macroblock could be encoded directly. In MPEG, they are coded using DCT, for two reasons. First, DCT coding exploits any spatial coherence that may be present in the residual. Second, DCT coding allows use of the same rate control (based upon quantization) and VLE encoding that are already in place for intra macroblocks. The residuals for a nonintra block are dequantized, then added to the motion-compensated values from the reference frame. Because the dequantized transform coefficients are not directly viewed, it is not appropriate to use a perceptually weighted quantizer matrix. By default, the quantizer matrix for nonintra blocks is *flat* – that is, it contains the same value in all entries.

Frame and field DCT types

Luma in a macroblock is partitioned into four blocks according to one of two schemes, *frame DCT coding* or *field DCT coding*. I will describe three cases where frame

Figure 47.4 **The frame DCT type** involves straightforward partitioning of luma samples of each 16×16 macroblock into four 8×8 blocks. This is most efficient for macroblocks of field pictures, native progressive frame pictures, and frame-structured pictures having little interfield motion.



DCT coding is appropriate, and then introduce field DCT coding.

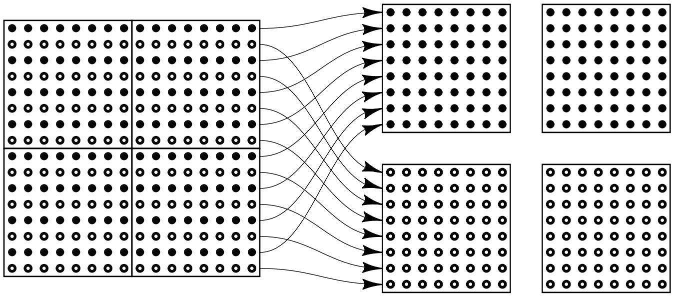
- In a frame-structured picture that originated from a native-progressive source, every macroblock is best predicted by a spatially contiguous 16×16 region of a reference frame. This is *frame DCT coding:* Luma samples of a macroblock are partitioned into 8×8 luma blocks as depicted in Figure 47.4 above.

At first glance it is a paradox that *field*-structured pictures must use *frame* DCT coding!

- In a field-structured picture, alternate image rows of each source frame have been unwoven by the encoder into two fields, each of which is free from interlace effects. Every macroblock in such a picture is best predicted from a spatially contiguous 16×16 region of a reference field (or, if you prefer to think of it this way, from alternate lines of a 16×32 region of a reference frame). This is also frame DCT coding.
- In a frame-structured picture from an interlaced source, a macroblock that contains no scene element in motion is ordinarily best predicted by frame DCT coding.

An alternate approach is necessary in a frame-structured picture from an interlaced source where a macroblock contains a scene element in motion. Such a scene element will take different positions in the first and second fields: A spatially contiguous 16×16 region of a reference picture will form a poor predictor. MPEG-2 provides a way to efficiently code such a macroblock. The scheme involves an alternate partitioning of luma into 8×8 blocks: Luma blocks are formed by collecting alternate rows of the reference frame. The scheme is called *field DCT coding;* it is depicted in Figure 47.5 at the top of the facing page.

Figure 47.5 **The field DCT type** creates four 8×8 luma blocks by collecting alternate image rows. This allows efficient coding of a frame-structured picture from an interlaced source, where there is significant interfield motion. (Comparable unweaving is already implicit in field-structured pictures.)



You might think it a good idea to handle chroma samples in interlaced frame pictures the same way that luma is handled. However, with 4:2:0 subsampling, that would force having either 8×4 chroma blocks or 16×32 macroblocks. Neither of these options is desirable; so, in a frame-structured picture with interfield motion, chroma blocks are generally poorly predicted. Owing to the absence of vertical subsampling in the 4:2:2 chroma format, 4:2:2 sequences are inherently free from such poor chroma prediction.

Zigzag and VLE

Once DCT coefficients are quantized, an encoder scans them in zigzag order. I sketched zigzag scanning in JPEG in Figure 45.8, on page 499. This scan order, depicted in Figure 47.6 overleaf, is also used in MPEG-1.

In addition to the JPEG/MPEG-1 scan order, MPEG-2 provides an alternate scan order optimized for frame-structured pictures from interlaced sources. The alternate scan, sketched in Figure 47.7 overleaf, can be chosen by an encoder on a picture-by-picture basis.

In MPEG terminology, the absolute value of an AC coefficient is its *level*. I prefer to call it *amplitude.* Sign is coded separately.

After zigzag scanning, zero-valued AC coefficients are identified, then {run-length, level} pairs are formed and variable-length encoded. For intra macroblocks, MPEG-2 allows an encoder to choose between two VLE schemes: the scheme first standardized in MPEG-1, and an alternate scheme more suitable for frame-structured pictures with interfield motion.

Block diagrams of an MPEG-2 encoder and decoder system are sketched in Figure 47.8 overleaf.

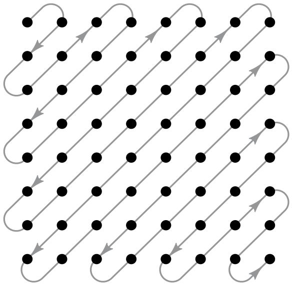


Figure 47.6 **Zigzag *scan*[0]** denotes the scan order used in JPEG and MPEG-1, and available in MPEG-2.

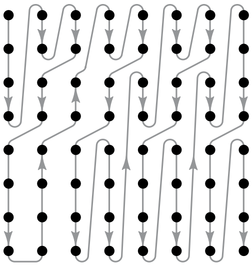


Figure 47.7 **Zigzag *scan*[1]** may be chosen by an MPEG-2 encoder on a picture-by-picture basis.

Refresh

Distributed refresh does not guarantee a deterministic time to complete refresh. See Lookabaugh, cited at the end of this chapter.

Occasional insertion of I-macroblocks is necessary for three main reasons: to establish a reference picture upon channel acquisition; to limit the duration of artifacts introduced by uncorrectable transmission errors; and to limit *drift* (that is, divergence of encoder and decoder predictors due to mistracking between the encoder's IDCT and the decoder's IDCT). MPEG-2 mandates that every macroblock in the frame be refreshed by an intra macroblock before the 132nd P-macroblock. Encoders usually meet this requirement by periodically or intermittently inserting I-pictures. However, I-pictures are not a strict requirement of MPEG-2, and *distributed refresh* – where I-macroblocks are used for refresh, instead of I-pictures – is occasionally used, especially for direct broadcast from satellite (DBS).

A sophisticated encoder examines the source video to detect scene cuts, and adapts its sequence of picture types according to picture content.

Motion estimation

A motion vector must do more than cover motion from one frame to the next: With B-pictures, a motion vector must describe motion from one *reference* frame to the next – that is, from an I-picture or P-picture to the following I-picture or P-picture. As the number of interposed B-pictures increases – as page 155's M value increases – motion vector range must increase. The cost

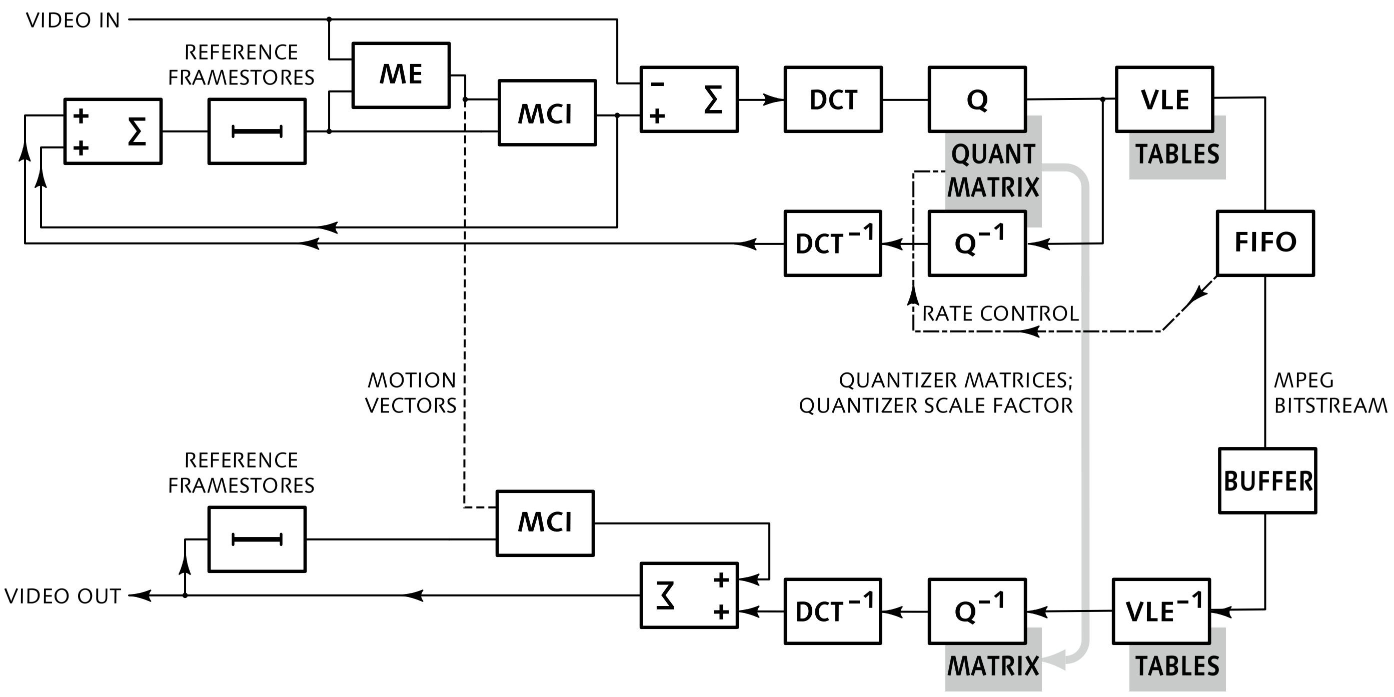


Figure 47.8 **MPEG encoder and decoder** block diagrams are sketched here. The encoder includes a motion estimator (ME); this involves huge computational complexity. Motion vectors (MVs) are incorporated into the bitstream and thereby conveyed to the decoder; the decoder does not need to estimate motion. The encoder effectively contains a copy of the decoder; the encoder's picture difference calculations are based upon reconstructed picture information that will be available at the decoder.

and complexity of motion estimation increases dramatically as search range increases.

The burden of *motion estimation* (ME) falls on the encoder. Motion estimation is very complex and computationally intensive. MPEG-2 allows a huge motion vector range: For MP@ML frame-structured pictures, the 16×16 prediction region can potentially lie anywhere within [–1024…+1023½] luma samples horizontally and [–128…+127½] luma samples vertically from the macroblock being decoded. Elements in the picture header (*f code*) specify the motion vector range used in each picture; this limits the number of bits that need to be allocated to motion vectors for that picture.

Whether an encoder actually searches this extent is not standardized!

The purpose of the motion estimation in MPEG is not exactly to estimate motion in regions of the picture – rather, it is to access a prediction region that minimizes the amount of prediction error (residual) information that needs to be coded. Usually this goal will be achieved by using the best estimate of average motion in the 16×16 macroblock, but not always.
I make this distinction because some video processing algorithms need accurate motion vectors, where the estimated motion is a good match to motion as perceived by a human observer. In many video processing algorithms, such as in temporal resampling used in standards converters, or in deinterlacing, a motion vector is needed for every luma sample, or every few samples. In MPEG, only one or two vectors are needed to predict a macroblock from a 16×16 region in one or two reference pictures.

If the fraction bit of a motion vector is set, then predictions are formed by averaging sample values from neighboring pixels (at integer coordinates). This is straightforward for a decoder. However, for an encoder to *produce* ½-luma-sample motion vectors in both horizontal and vertical axes requires quadruple the computational effort of producing full-sample vectors.

There are three major methods of motion estimation:

- *Block matching*, also called *full search*, involves an exhaustive search for the best match of the target macroblock through some two-dimensional extent of the reference frame. For the large ranges of MPEG-2, full block matching is impractical.

- *Pixel-recursive* (or *pel-recursive*) methods start with a small number of initial guesses at motion, based upon motion estimates from previous frames. The corresponding coordinates in the reference frame are searched, and each guess is refined. The best guess is taken as the final motion vector.

This method is sometimes called "logarithmic," which I consider to be a very poor term in this context.

- *Pyramidal* methods form spatial lowpass-filtered versions of the target macroblock, and of the reference frames; block matches are performed at low resolution. Surrounding the coordinates of the most promising candidates at one resolution level, less severely filtered versions of the reference picture regions are formed, and block matches are performed on those. Successive refinement produces the final motion vector. This technique tends to produce smooth motion-vector fields.

Rate control and buffer management

The fractions 0.6, 0.3, and 0.1 are comparable to the SD luma coefficients of green, red, and blue; so, I use green, red, and blue to designate I-picture, P-picture, and B-picture data respectively in Figures 16.2, 16.3, and 16.4 (page 153).

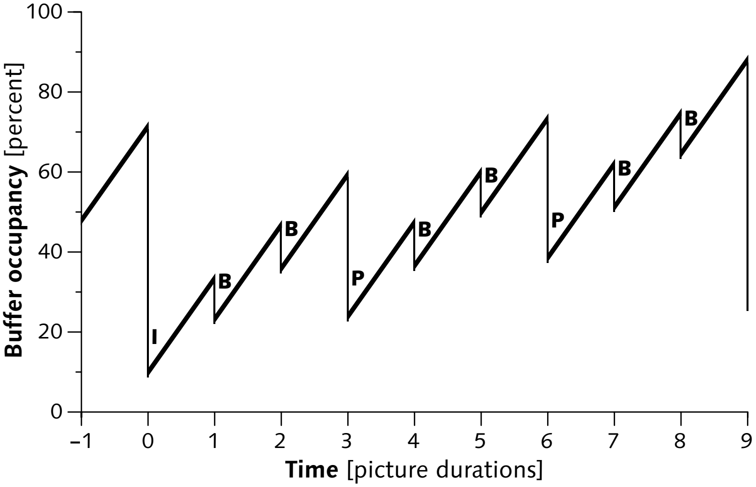
A typical video sequence, encoded by a typical MPEG-2 encoder, produces I-, P-, and B-pictures that consume bits roughly in the ratio 60:30:10. An I-picture requires perhaps six times the number of bits as two B-pictures.

Many applications of MPEG-2, such as DTV, involve a transmission channel with a fixed data rate. This calls for *constant bit rate* (CBR) operation. Other applications of MPEG-2, such as DVD, involve a channel having variable (but limited) data rate. Such applications call for *variable bit rate* (VBR) operation, where the instantaneous bit rate is varied to achieve the desired picture quality for each frame, maximizing storage utilization.

The larger the decoder's buffer size, the more flexibility is available to the encoder to allocate bits among pictures. However, a large buffer is expensive. Each profile/level combination dictates the minimum buffer size that a decoder must implement.

An encoder effects *rate control* by altering the quantization matrices – the perceptually weighted matrix used for intra macroblocks, and the flat matrix used for nonintra macroblocks. MPEG-2 allows quantizer matrices to be included in the bitstream. Additionally, and more importantly, a *quantizer scale code* is transmitted at the slice level, and may be updated at the

Figure 47.9 **Buffer occupancy** in MPEG-2 is managed through an idealized *video buffering verifier* (VBV) that analyzes the output bitstream produced by any encoder. This graph shows buffer occupancy for a typical GoP.



macroblock level. This code determines an overall scale factor that is applied to the quantizer matrices. The encoder quantizes more or less severely to achieve the required bit rate; the quantizer scale code is conveyed to the decoder so that it dequantizes accordingly.

Video display requires a constant number of frames per second. Because an I-picture has a relatively large number of bits, during decoding and display of an I-picture in all but degenerate cases, the decoder's net buffer occupancy decreases. During decoding and display of a B-picture, net buffer occupancy increases. Figure 47.9 shows typical buffer occupancy at the start of a sequence, for a duration of about one GoP.

An MPEG bitstream must be constructed such that the decoder's buffer doesn't overflow: If it did, bits would be lost. The bitstream must also be constructed so that the buffer doesn't underflow: If it did, a picture to be displayed would not be available at the required time.

Buffer management in MPEG-2 is based upon an idealized model of the decoder's buffer: All of the bits associated with each picture are deemed to be extracted from the decoder's buffer at a certain precisely defined instant in time with respect to the bitstream. Every encoder implements a *video buffering verifier* (VBV) that tracks the state of this idealized buffer. Each picture header contains a *VBV delay* field that declares the fullness of the buffer at the start of that picture. After channel acquisition, a decoder waits a corresponding amount of time before starting decoding. (If the decoder did not wait, buffer underflow could result.)

Bitstream syntax

The end product of MPEG-2 video compression is a bitstream partitioned into what MPEG calls a *syntactic hierarchy* having six layers: *sequence, GoP, picture, slice, macroblock,* and *block*. Except for the video sequence layer, which has a *sequence end* element, each syntactic element has a header and no trailer. The sequence, GoP, picture, and slice elements each begin with a 24-bit *start code prefix* comprising 23 zero bits followed by a one bit. A start code establishes byte alignment, and may be preceded by an arbitrary number of zero-stuffing bits. All other datastream elements are constructed so as to avoid the possibility of 23 or more consecutive zero bits.

Video sequence layer

The top layer of the MPEG syntax is the *video sequence*. The sequence header specifies high-level parameters such as bit rate, picture rate, picture size, picture aspect ratio, profile, level, progressive/interlace, and chroma format. The *VBV buffer size* parameter declares the maximum buffer size required within the sequence. The sequence header may specify quantizer matrices. At the encoder's discretion, the sequence header may be retransmitted intermittently or periodically throughout the sequence, to enable rapid channel acquisition by decoders.

The start of each interlaced video sequence establishes an immutable sequence of field pairs, ordered either {top, bottom, …}, typical of 480*i*, or {bottom, top, …}, typical of 576*i* and 1080*i*. Within a sequence, any individual field may be field-coded, and any two adjacent fields may be frame-coded; however, field parity must alternate in strict sequence.

Group of pictures
(GoP header)

The GoP is MPEG's unit of random access. The GoP layer is optional in MPEG-2; however, it is a practical necessity for most applications. A GoP starts with an I-picture. (Additional I-pictures are allowed.) The GoP header contains SMPTE timecode, and *closed GoP* and *broken link* flags.

A GoP header contains 23 bits of coded SMPTE timecode. If present, this applies to the first frame of the GoP (in display order). It is unused within MPEG.

If a GoP is *closed,* no coded B-picture in the GoP may reference the first I-picture of the following GoP. This is inefficient, because the following I-picture ordinarily contains useful prediction information. If a GoP is *open*, or the GoP header is absent, then B-pictures in the GoP may reference the first I-picture of the following GoP. To allow editing of an MPEG bitstream, GoPs must be closed.

A device that splices bitstreams at GoP boundaries can set *broken link;* this signals a decoder to invalidate B-pictures immediately following the GoP's first I-picture.

Picture layer

The picture header specifies picture structure (frame, top field, or bottom field), and picture coding type (I, P, or B). The picture header can specify quantizer matrices and *quantizer scale type*. The *VBV delay* parameter is used for buffer management.

Slice layer

A slice aggregates macroblocks in raster order, left to right and top to bottom. No slice crosses the edge of the picture. All defined profiles have “restricted slice structure,” where slices cover the picture with no gaps or overlaps. The slice header contains the *quantizer scale code*. The slice serves several purposes. First, the slice is the smallest unit of resynchronization in case of uncorrected data transmission error. Second, the slice is the unit of differential coding of intra-macroblock DC terms. Third, the slice is the unit for differential coding of nonintra motion vectors: The first macroblock of a slice has motion vectors coded absolutely, and motion vectors for subsequent macroblocks are coded in terms of successive differences from that.

Macroblock layer

The macroblock is MPEG's unit of motion prediction. Coded macroblock data contains an indication of the macroblock type (intra, forward predicted, backward predicted, or bipredicted); a *quantizer scale code;* 0, 1, or 2 forward motion vectors; and 0, 1, or 2 backward motion vectors. The *coded block pattern* flags provide a compact way to represent blocks that are not coded (owing to being adequately predicted without the need for residuals).

Block layer

Each block is represented in the bitstream by VLE-coded DCT coefficients – a differentially encoded DC coefficient, and zero or more AC coefficients. Each coded block's data is terminated by a 4-bit *end of block* (EOB).

Transport

Various syntax elements of MPEG video or audio are serialized to form an *elementary stream* (ES). MPEG-2 defines a mechanism to divide an ES into packets, forming a *packetized elementary stream* (PES). Each PES pack header contains system-level clock information, packet priority, packet sequence numbering, and (optionally) encryption information. If an MPEG-2 PES is stored in a file, the file conventionally has the extension *m2v;* however, a video PES can't contain audio! More commonly, though, MPEG-2 video and audio are multiplexed, then stored or transported, using *program streams* or *transport streams,* to be discussed in the chapter *MPEG-2 storage and transport,* on page 555.

Further reading

GIBSON, JERRY D., TOBY BERGER, TOM LOOKABAUGH, DAVID LINDBERGH, and RICHARD L. BAKER (1998), *Digital Compression for Multimedia* (San Francisco: Morgan Kaufmann). Lookabaugh's chapter provides an excellent 55-page description of MPEG-2. His chapter also covers MPEG audio.

HASKELL, BARRY G., ATUL PURI, and ARUN N. NETRAVALI (1997), *Digital Video: An Introduction to MPEG-2* (New York: Chapman & Hall).This book fails to distinguish luminance and luma; both are called *luminance* and given the symbol *Y.* See Appendix A, *YUV and luminance considered harmful*, on page 567 of the present book.

MITCHELL, JOAN L., WILLIAM B. PENNEBAKER, CHAD E. FOGG, and DIDIER J. LEGALL (1997), *MPEG Video Compression Standard* (New York: Chapman & Hall).This book concentrates on MPEG-1. Egregiously incorrect information appears concerning chroma subsampling.

SYMES, PETER (2003), *Digital Video Compression* (New York: McGraw-Hill).

WATKINSON, JOHN (1999), *MPEG-2* (Oxford: Focal Press).

H.264 video compression

48

ITU-T H.264, *Advanced Video Coding for Generic Audiovisual Services – Coding of Moving Video*, also published as ISO/IEC 14496-10 (MPEG-4 Part 10), *Advanced Video Coding.*

H.264 is usually pronounced ***H***-dot-***TWO-SIX-FOUR***.

H.264 denotes a codec standardized by ITU-T (under the designation *H.264*) and by ISO/IEC (under the designation *MPEG-4 Part 10*). The Simple Studio Profile (SStP) of MPEG-4 Part 2 is used in HDCAM. That aspect of Part 2, and all of Part 10, are applicable to broadcast-quality video; other than those cases, MPEG-4 is generally not applicable to broadcast-quality video. H.264 was developed by the *Joint Video Team* (JVT), where it was referred to as *Advanced Video Coding* (AVC); its ITU-T nomenclature during development was H.26L. All of these terms were once used to denote what is now, after adoption of the standard, best called H.264.

Compounding 1.06 twelve times yields a factor of two:

$1.06^{12} \approx 2$

H.264 is broadly similar to MPEG-2, but the "low fruit" had been taken. Compression improvements in H.264 are obtained by a dozen or so techniques, each having perhaps 6% improvement in coding efficiency – but a dozen of those cascaded yields twice the efficiency of MPEG-2. (Practitioners claim efficiency as low as 1.5 and as high as 3 times that of MPEG-2.) H.264 spans a wide range of applications, from surveillance video, to video conferencing, to mobile devices, to internet video streaming, to HDTV broadcasting.

H.264 is complicated. The standard (in its 2010-03 edition) comprises 669 pages of very dense description. Implementing an encoder or decoder takes many man-years. Software, firmware, and hardware implementations are commercially available. Even hardware implementations require embedded firmware: H.264 VLSI solutions typically involve one or more embedded RISC processors and quite a bit of associated firmware.

The H.264 features that extend MPEG-2 are described in the remaining sections of this chapter. I assume that you are familiar with *Introduction to video compression,* on page 147, and with JPEG, M-JPEG, DV, and MPEG-2, described in the preceding three chapters.

Like MPEG-2, H.264 specifies exactly what constitutes a conformant bitstream: A conformant ("legal") encoder generates only conformant bitstreams; a legal decoder correctly decodes any conformant bitstream. H.264 effectively standardizes the behaviour of a *decoder,* but does *not* standardize the encoder!

The goal of compression is to reduce data rate while minimizing the visibility of artifacts. The best way – most experts say, the only way – to establish the performance of an encoder is to visually assess the result of compressing and decompressing video streams.

MPEG LA, L.L.C. is not affiliated with MPEG (the standards group). LA apparently stands for *Licensing Administration.* The organization is based in Denver, not Los Angeles.

H.264 is covered by hundreds of patents. Implementors, manufacturers, users, and/or others may or may not be required to take out a licence to the "patent pool" administered by MPEG LA.

Not all features of H.264 are expected to be implemented in every decoder; for example, B-slices (comparable to MPEG-2 B-pictures) are prohibited in the baseline profile. Applications have various bit rates, and decoders can have various levels of resources (e.g., memory); like MPEG-2, a system of profiles and levels determines the minimum requirements.

Algorithmic features, profiles, and levels

Table 48.1 opposite summarizes the algorithmic features of H.264 beyond MPEG-2. The features in the top section are available in all profiles; features in the sections below are profile-dependent.

The features available in the baseline and extended profiles concern robust handling of data conveyed across unreliable channels. These features (and profiles) are generally not of interest for professional video, and they are not permitted in the main and high profiles.

The features of the extended, main, and high profiles offer improved coding efficiency. CABAC improves the performance of variable-length entropy coding.

Fidelity range extensions (FRExt) refers to several algorithmic features incorporated into the high profiles – HiP, Hi10P, Hi422P, and Hi444P – to enable

	Algorithmic feature ("tool") / *Profile*	*Baseline (BP)*	*Extended (XP)*	*Main (MP)*	*High (HiP)*
Features in all profiles	Multiple reference pictures	●	●	●	●
	Flexible motion compensation	●	●	●	●
	I-slices and P-slices	●	●	●	●
	$^1/_4$-pel motion-comp. interpolation	●	●	●	●
	16-bit exact-match integer transform	●	●	●	●
	Unified variable-length coding	●	●	●	●
	(UVLC/Exp-Golomb)	●	●	●	●
	CAVLC	●	●	●	●
	Deblocking filter in-the-loop	●	●	●	●
Set 1	Flexible macroblock ordering (FMO)	●	●		
	Arbitrary slice order (ASO)	●	●		
	Redundant slices (RS)	●	●		
Set 2	Data partitioning		●		
	SI & SP slices		●		
Set 3	B-slices		●	●	●
	Interlaced coding (PicAFF, MBAFF)		●	●	●
	Weighted and offset MC prediction		●	●	●
Set 4	CABAC entropy coding			●	●
FRExt	8×8 luma intra prediction				●
	Increased sample depth				●
	4:4:4 and 4:2:2 chroma subsampling				●
	Inter-picture lossless coding				●
	8×8/4×4 transform adaptivity				●
	Quantization scaling matrices				●
	Separate C_B and C_R QP control				●
	Monochrome (4:0:0)				●

Table 48.1 **H.264 features** are arranged in rows; the columns indicate presence of features in the commercially important profiles.

higher quality video. Hi10P allows 10-bit video; Hi422P permits 4:2:2 chroma subsampling, and Hi444P permits 4:4:4, 12-bit video, and several other features.

Four of H.264's profiles are commercially important: baseline, extended, main, and high. The main and high profiles are relevant to professional video. H.264 has fifteen levels, accommodating images ranging from 176×144 (coded at rates as low as 64 kb/s) to 4 K×2 K (coded at rates as high as 240 Mb/s). Profile and level combinations important to professional video are summarized in Table 48.2 overleaf.

Level	*Typ. image format*	*Typ. frame rate [Hz]*	*Max. bit rate [b/s]*
L1	176×144	15	64 k
L1b	176×144	15	128 k
L1.1	352×288 or 176×144	7.5 or 30	192 k
L1.2	352×288	15	384 k
L1.3	352×288	30	768 k
L2	352×288	30	2 M
L2.1	352×480 or 352×576	30 or 25	4 M
L2.2	SD	15	4 M
L3.0	SD	30 or 50	10 M
L3.1	1280×720	30	14 M
L3.2	1280×720	60	20 M
L4.0	1920×1080	30	20 M
L4.1	1920×1080	30	50 M
L4.2	1920×1080	60	50 M
L5	2048×1024	72 or 30	135 M
L5.1	4096×2048	30	240 M

Table 48.2 **H.264 levels** are summarized.

Baseline and extended profiles

You might imagine a *baseline* profile to be decodable by every decoder. That is not the case in H.264. The baseline profile is intended to address low bit-rate applications that suffer from poor quality transmission. The flexible macroblock ordering (FMO), arbitrary slice order (ASO), and redundant slices (RS) features all contribute to robustness. Other features – in particular, B-slices – are excluded from the baseline profile, so as to achieve low computational complexity. The baseline profile is rarely used (if used at all) in professional video.

You might imagine an *extended* profile to have features beyond those of the main profile. That is not the case in H.264. The extended profile extends the robustness features of the *baseline* profile by including two additional features, data partitioning and SI and SP slices. Two additional features improve coding efficiency: B-slices, and interlaced coding (PicAFF, MBAFF). The extended profile is rarely used in professional video.

High profiles

The original H.264 features were augmented by the Fidelity range extensions (FRExt), which are available in the high profiles.

Ten bit sample depth is available in Hi10P and Hi422P; fourteen bit sample depth is available in Hi444P.

Hi444PP stands for *High 4:4:4 predictive profile.*

Hi422P and Hi444P offer 4:2:2 chroma subsampling: $Y'C_BC_R$ 4:2:2 (loosely, *Y'UV* 4:2:2) can be coded. Hi444P offers 4:4:4 "chroma subsampling" – that is, no subsampling at all.

Hierarchy

The syntax elements in an H.264 bitstream have a hierarchical structure like that of MPEG-2. The bitstream hierarchy of H.264 – the *syntax* hierarchy – is as follows:

- sequence
- picture
- slice
- macroblock
- macroblock partition
- sub-macroblock partition
- block
- sample

The *video coding layer* (VCL) comprises elements at the slice level and below. A *network abstraction layer* (NAL) defines *NAL units* to convey coded data. Information at layers above the VCL – that is, at the sequence and picture levels – is conveyed in *non-VCL NAL units.* The two types of NAL units (VCL and non-VCL) can be transmitted in different streams, for example to achieve higher network robustness, though specification of such transmission mechanisms is outside the scope of H.264.

Supplemental enhancement information (SEI) and *video usability information* (VUI) are "messages" inserted into non-VCL NAL units of the coded bitstream. SEI comprises sequence and picture parameter sets (SPS and PPS). VUI conveys information comparable to the contents of the sequence display extension of MPEG-2.

Multiple reference pictures

VP8 has three reference frames.

MPEG-2 has two reference frames: one in the past, and one in the "future." The "future" frame is available to predict B-pictures that lie earlier in display order.

Multiple reference pictures may be useful to predict "uncovered background" depending upon the encoder's ability to discover it. Use of "future" reference pictures incurs latency, and may be impractical in some applications.

In H.264, multiple reference pictures are allowed – between 4 and 13, depending upon level. If the material being coded has a quick cut to a reverse shot, the encoder can instruct the decoder to retain the picture at the end of the first shot, and use it to predict the picture upon return from the reverse shot. Reference pictures can be addressed in arbitrary order.

Slices

Slices offer a decoder the option of parallelism: No intra prediction crosses a slice boundary. Decoder state effectively resets on slice boundaries, so slices limit the spatial extent of transmission-induced impairments. Slices can be coded redundantly to further mitigate against transmission error.

Spatial intra prediction

In MPEG-2, a macroblock may be coded entirely independently as an I-macroblock, or may exploit temporal prediction and be coded as a P-macroblock. In the development of H.264 it was realized that decoded intra macroblocks above the current one, and those to the left in the same slice, have prediction value in the spatial domain. H.264 implements intra prediction based upon that data, where image data above or to the left is copied directionally in several modes. The prediction can then be refined by transform-coded quantized residuals in the usual way. Intra prediction uses only information from intra-coded macroblocks.

There is also an intra-PCM mode, where I-macroblock pixel data is directly coded, bypassing the transform. The mode is potentially useful at very high data rates.

Flexible motion compensation

In MPEG-2, motion prediction is accomplished in units of 16×16 blocks of luma pixels – that is, macroblocks. The encoder tries to find a 16×16 region of a reference picture that forms a good predictor, then codes the relative coordinates of that block into the data stream as a motion vector.

In H.264, a macroblock can be partitioned into several shapes and sizes for prediction from different regions of a reference picture, even prediction from different reference pictures. An entire macroblock can be predicted from one 16×16 source; alternatively, the macroblock can be partitioned into two 8×16 macroblock partitions, two 16×8 macroblock partitions, or four 8×8 macroblock partitions, all predicted independently. In high profiles, if a macroblock is partitioned into four 8×8 macroblock partitions, each of those can be partitioned into two 4×8 sub-macroblock partitions, two 8×4 sub-macroblock partitions, or four 4×4 sub-macroblock partitions, again all predicted independently. A macroblock can be associated with up to 16 motion vectors.

Quarter-pel motion-compensated interpolation

What is $^1/_4$-pel for luma is $^1/_8$-pel for 4:2:0 chroma.

In MPEG-2, motion vectors can have $^1/_2$-pixel precision with respect to luma samples. In H.264, motion-compensated interpolation can be performed to quarter-pel precision – that is, motion vectors can be encoded in units of $^1/_4$-pel (sometimes called *quarter-pel,* or *Qpel*). The interpolation operation uses simple 6-tap FIR filters, and has the beneficial effect of lowpass-filtering the prediction signal in addition to delivering it at an optimal spatial position.

Weighting and offsetting of MC prediction

MPEG-2 behaves poorly in fades from one picture to another and in fades to black – or, in the case of *Six Feet Under,* fades to white. The DC terms of the transform coefficients are coded reasonably well, but in fade to black all of the AC terms scale down together; that stresses the quantizer. H.264 implements weighting and offsetting of MC prediction, to improve performance in fades and certain other circumstances.

16-bit integer transform

MPEG-2 followed JPEG in using the 8×8 DCT, virtually always implemented in binary fixed-point arithmetic. The theoretical DCT matrix contains irrational numbers; encoders and decoders approximate them in fixed-point binary integers, usually 16-bit. Neither the JPEG nor MPEG-2 standards specify the accuracy of the DCT. The encoder includes a simulation of the decoding process, but owing to different roundoff error in different implementations, the encoder's DCT may not match the decoder's DCT. When a decoded block is used as a prediction, the prediction formed at the decoder may not exactly match the prediction expected by the encoder. We assume that the encoder has more computational resources that the decoder, and is likely to have more accuracy, so we term the problem – perhaps unfairly to the decoder – as *decoder drift.*

H.264's transform is sometimes termed *HCT*, which is either *H.264 cosine transform* or *high correlation transform* depending upon whom you ask.

In H.264, decoder drift is eliminated through use of a transform defined by a matrix of simple binary fractions whose inverse also comprises simple binary fractions. With 8-bit residuals and 16-bit arithmetic, no roundoff error occurs, so no drift occurs.

Quantizer

In MPEG-2, the transform coefficient quantizer levels are uniformly spaced. In H.264, the quantizer has 52 steps that are exponentially spaced: Each step increases the step size by a ratio of 1.122, that is, six steps double the step size. (As a rough guide, increasing quantizer step size by +1 decreases bit rate by about 10%, and doubling halves the bit rate. This heuristic can be used for rate control at an encoder.)

Variable-length coding

Symbol	*Scheme F*	*Scheme V*
A	00	0
B	01	10
C	10	110
D	11	111

Table 48.3 **Two hypothetical coding schemes** mapping symbols (A through D) into a bitstream are sketched. Scheme F allocates a fixed number of bits to each symbol; Scheme V allocates a variable number of bits to symbols.

Suppose you're given sequences of four symbols (A, B, C, and D) to encode into a bitstream. Consider two simple coding schemes set out in Table 48.3 in the margin. Scheme F uses two bits for any of the four symbols. Scheme V uses one, two, or three bits, depending upon the symbol being coded. Both schemes faithfully encode *any* input sequence that is presented – that is, both encodings are lossless. However, if the input contains a lot of As, scheme V emits fewer bits than scheme F. Scheme V exemplifies the basic notion of variable-length coding: It's advantageous to have an encoding that reflects the probabilities of the symbols being coded. In this example, scheme F is well adapted to inputs where A, B, C, and D have equal probabilities. Scheme V is well adapted to probabilities [$^1/_2$, $^1/_4$, $^1/_8$, $^1/_8$] respectively.

In MPEG-2, a few dozen VLC coding schemes were devised for various syntax elements. H.264 required many additional syntax elements, and the developers got tired of constructing *ad hoc* tables. A systematic method, *universal variable-length coding* (UVLC) was

POS	*INT*	*Coded bitstream*
1	0	**1**
2	+1	0**1**0
3	−1	0**1**1
4	+2	00**1**00
5	−2	00**1**01
6	+3	00**1**10
7	−3	00**1**11
8	+4	000**1**000
9	−4	000**1**001
10	+5	000**1**010
11	−5	000**1**011

Table 48.4 **An example of exponential Golomb coding** of positive numbers 1 through 11 or integers ranging ±5 is shown.

INT	*Coded bitstream*
0	**1**
±1	0**1**_s_
±2…±3	00**1**_xs_
±4…±7	000**1**_xxs_
±8…±15	0000**1**_xxxs_
±16…±31	00000**1**_xxxxs_
±32…±63	000000**1**_xxxxxs_
±64…±127	0000000**1**_xxxxxxs_

Table 48.5 **Exp-Golomb coding can be generalized** to signed integers represented in 1 bit, 2 bits, 3 bits, 4 bits, and more, indefinitely. The scheme favours inputs where small numbers are most likely: If inputs ±127 were equally likely, then fixed-length 8-bit two's complement coding would be more efficient.

adopted. It is based upon the *exponential Golomb* scheme, an example of which is sketched in Table 48.4.

Decoding of the positive number (POS) symbols of the example proceeds as follows: If the datastream bit is **1**, the coded value is 1. Otherwise, count leading zero bits, denoting the count N. Consider the following $N+1$ bits (including the leading **1** bit) to be the binary-coded positive number, most-significant bit first.

When used for signed integers (the INT symbols of the example), decode as follows: If the datastream bit is **1**, the coded value is 0. Otherwise, count leading zero bits, denoting the count N. Consider the following N bits (including the leading **1** bit) to be the absolute value of the coded number, expressed in binary, most-significant bit first. The trailing $(N+1)^{\text{th}}$ bit is the sign.

The INT example in Table 48.4 encodes signed integers such as those encountered in motion vector displacements. The code is easily adapted to nonnumeric symbols by simply assigning the required values or symbols to the appropriate number.
Table 48.5 shows how the coding extends to arbitrarily large numbers (or to a set of symbols of arbitrary size).

In H.264, UVLC is used at syntax levels above the transform coefficients, for data such as prediction modes and motion vectors. The UVLC scheme is not used for transform coefficients: either CAVLC or CABAC is used for those.

The POS example of Table 48.4 is constructed for ease of explanation; H.264's *unsigned integer* (ue) codes are the indicated numbers less one. The INT example of Table 48.4 corresponds to H.264's *signed integer* (se) codes.

Context adaptivity

The MPEG-2 designers used their judgement and experience, and the results of many experiments, to set up MPEG-2's VLC tables. However, those tables are static.

Usage of VLC entries by particular source material can be considered as a statistical distribution – that is, VLC table usage depends upon history, upon *context.*

Context adaptivity refers to an encoder dynamically keeping track of the use of table entries, estimating the probability of their use, and changing VLC mapping so that the coded bitstream has a compact representation for the symbols that are likely to be encountered.

Context adaptivity leads to increased complexity at both the encoder and the decoder. In H.264, the basic form of context adaptivity is *context-adaptive variable-length coding* (CAVLC), used at slice level and below.

CABAC

VLC (and its relative, CAVLC) as outlined above are optimal when coding symbols whose probabilities can be expressed as binary fractions, for example, if in the Scheme V example of Table 48.3 the probabilities of [A, B, C, and D] were [$^1/_2$, $^1/_4$, $^1/_8$, $^1/_8$]. In general, symbols being coded don't have probabilities that all lie close to binary fractions. For example, your task may be to code three symbols having probabilities [$^1/_3$, $^1/_3$, $^1/_3$].

noun, *ah-RITH-meh-tik;*
adjective, *are-ith-MEH-tik.*

A technique called *arithmetic coding* can be used to efficiently encode distributions where individual symbols occupy the equivalent of fractions of a bit. A potentially large group of symbols is collected, then coded into what amounts to a single number, where the range of the number is divided into subranges corresponding to the probabilities of the individual symbols.

Like VLC, arithmetic coding can be made context adaptive – hence, *CABAC: context-adaptive binary arithmetic coding.* If that sounds complicated, it is. CABAC can yield 10% or so bit rate improvement; however, it adds complexity to both the encoder and the decoder and it consumes processing and memory resources. CABAC is available in H.264's main and high profiles.

CABAC is part of the main and high profiles of H.264; however, its use may be gated by concerns outside H.264 proper. For example, CABAC is allowed in AVC-Intra 50, but prohibited in AVC-Intra 100.

Deblocking filter

In MPEG-2, it is a problem that the inverse transform tends to produce discontinuities – blocking artifacts –

where two 8×8 blocks abut. Many MPEG-2 decoders include post-processing to mitigate the effects of blocking artifacts, but treatment after the fact ("out of the loop") is invisible to the encoder.

H.264 standardizes an adaptive, in the loop deblocking filter. The filter adapts to picture content, such as edges (which typically cause the worst artifacts). Deblocking is standardized, and it takes place within the encoder's prediction loop.

Buffer control

In MPEG-2, a *video buffer verification* (VBV) value is transmitted from the encoder to the decoder, to ensure that the decoder's buffer tracks that of the encoder without underflowing or overflowing.

In H.264 it's more complicated. Memory occupancy is tracked at both the encoder and the decoder for two hypothetical buffers: the *coded picture buffer* (CPB), representing pictures in the coded bitstream; and the *decoded picture buffer* (DPB), representing pictures after decoding.

Scalable video coding (SVC)

Scalable video coding – defined in Annex G of H.264 – allows conveyance of information structured in a hierarchical manner to allow portions of the bitstream to be extracted at lower bit rate than the complete sequence to enable decoding of pictures with multiple image structures (for sequences encoded with spatial scalability), pictures at multiple picture rates (for sequences encoded with temporal scalability), and/or pictures with multiple levels of image quality (for sequences encoded with SNR/quality scalability).

In a single bitstream, a decoder having limited computational resources can extract the base bitstream to decode a low-level representation. (No data rate advantage accrues in this case.)

Different layers can be separated into different bitstreams. All decoders access the base stream; more capable decoders can access enhancement streams. However, for some applications – like HTTP live streaming – it may be more efficient to encode a single program at several different rates, each in a self-

Profile *@Level*	*High 10 Intra (Hi10Intra)*
L3.2	AVC-Intra 50: CABAC, 4:2:2, 1280×720*p* downsampled to 960×720
L4.0	AVC-Intra 50: CABAC, 4:2:2, 1920×1080 downsampled to 1440×1080
L4.1	AVC-Intra 100: CAVLC, 4:2:2, native 1280×720*p* and 1920×1080

Table 48.6 **AVC-Intra profile/level combinations** are summarized.

contained stream, so that a decoder can simply access a single stream at a suitable rate.

Multiview video coding (MVC)

Multiview video coding – Annex H of H.264 – standardizes features to efficiently code two (or potentially more than two) pictures that are highly spatially correlated. The common application is to code the left and right images of a stereo pair.

MVC adds two new profiles, *multiview high profile* (MHP) and *stereo high profile* (SHP). Video encoded with either of these profiles is backward compatible with H.264 high profile; the decoder sees just the base view. (In the case of SHP, this is typically the left eye.)

AVC-Intra

AVC-Intra is Panasonic's notation for studio-quality H.264 using Hi10Intra profile. Video in 720*p*, 1080*i*, or 1080*p* format at various frame rates is represented in 10-bit $Y'C_BC_R$ 4:2:2 components. In *AVC-Intra 50,* 720*p*, 1080*i*, and 1080*p* video is downsampled and compressed to 50 Mb/s. In *AVC-Intra 100,* 720*p*, 1080*i*, and 1080*p* video at native pixel count are compressed to 100 Mb/s. Table 48.6 summarizes.

Further reading

Richardson, Iain E.G. (2010), *The H.264 Advanced Video Compression Standard* (Chichester, U.K.: Wiley).

Sullivan, Gary J., Pankaj N. Topiwala, and Ajay Luthra (2004), "The H.264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions," in *Proc. SPIE* **5558**: 454–474.

Sullivan, Gary J. and Thomas Wiegand (2005), "Video compression – From concepts to the H.264/AVC standard," in *Proc. IEEE* **93** (1): 18–31.

VP8 compression

49

In 2010, Google acquired a company called On2 that had, over a decade or more, developed a series of proprietary software-based codecs for video distribution. Google made the VP8 codec open-source and used it as the basis for a proposal called *WebM* for web (IP-based) distribution of video to consumers. WebM comprises video encoded by the VP8 codec and audio encoded by the *Vorbis* codec, both wrapped in the *Matroska* file wrapper.

VP3, a distant predecessor to VP8, was made available by On2 as open source. VP3 subsequently developed into *Theora.* On2 licensed VP6 and VP7 to Adobe as the basis for Flash 8 video; subsequently, H.264 was incorporated into Flash 9. On2 licensed VP7 to Skype.

The VP8 codec is broadly based upon the principles of MPEG-2 and H.264 discussed in earlier chapters, although Google intends VP8 to be unencumbered by MPEG-2 and H.264 intellectual property rights (IPR, in this case, patent rights). Patents on elements of VP8 were issued to On2; Google permits their royalty-free use. Google's license to VP8 requires that the user not litigate any IP that addresses VP8 ("mutual nonassert"). There's no guarantee or indemnity that Google's VP8 implementation does not infringe patents not controlled by Google – perhaps even patents in the MPEG-2 or H.264 pools.

IP-based means based upon internet (TCP/IP) protocols. *H.264 IP* means intellectual property (patent) rights associated with H.264.

It is a technical and commercial problem with VP8 that the descriptive standard is not comprehensive: The definitive specification of VP8 is effectively its reference code. In places, there is opaque code that raises the question, should the VP8 "standard" be defined by what was apparently intended, or by what is executed by the code? In the absence of a written standard, implementors are forced to treat the reference code as definitive, even if performance or interoperability suffer.

Algorithmic features

As mentioned earlier, the VP8 codec is broadly based upon the principles of MPEG-2 and H.264. To make the most of what follows, you should be familiar with *Introduction to video compression* (on page 147), and with JPEG/M-JPEG, DV, MPEG-2, and H.264, described in the preceding four chapters.

Google documents refer to $Y'C_BC_R$ as *YUV.*

VP8 codes only progressive, 8-bit, 4:2:0 $Y'C_BC_R$ video. No provision is made for interlace.

Every picture is accompanied by a 1-bit flag *show_frame*, signalling whether to display the frame. That flag can cause a decoded frame to be placed into one of the reference frames but not displayed. Under unusual circumstances, using this mechanism can simulate a B-frame.

VP8 has what it calls *key-frames* (comparable to MPEG-2 I-frames), and *inter-frames* (like MPEG-2 P-frames). VP8 has no B-frames: All decoded frames are potentially available for predictions. A VP8 decoder has three reference frames: the *golden* frame, the *previous* frame, and the *altref* ("alternate reference") frame.

The bitstream is partitioned into *segments.* Within a segment there is a 4-byte frame header, and between one and nine *partitions* denoted I, II, III, and so on. A partition is a sequence of bytes representing aspects of video (akin to the separation of VCL NAL units and non-VCL NAL units in H.264). Partition I conveys prediction modes and motion vectors, per macroblock, in raster order. Partitions beyond I convey quantized transform coefficients (in VP8, sometimes termed *texture*). Macroblock rows can be mapped to a single partition, or to 2, 4, or 8 partitions each of which can be processed in parallel. (Entropy contexts, to be described, are shared among partitions; binary arithmetic *decoding* can be parallelized to some extent, but encoding can't be.)

VP8 subdivides 16×16 macroblocks into *subblocks* of 4×4 pixels. There are 24 subblocks in each $Y'C_BC_R$ 4:2:0 macroblock. Unlike H.264, VP8 has no 8x8 luma blocks. Chroma prediction is performed on 8×8 chroma blocks.

VP8 has no 8×8 intra luma prediction.

VP8 has two luma intra prediction modes – *i16x16* and *i4x4* – which reference previously decoded pixels in the same frame. Using intra prediction precludes parallelism.

The bitstream identifies one of four methods through which the intra prediction for each block can be obtained:

- *V_PRED:* Prediction values are replicated down the block from the row above.

- *H_PRED:* Prediction values are replicated across the block from the column to the left.
- *DC_PRED:* Prediction values are all set to the average value of the row above and the column to the left; this is called "DC" chroma prediction.
- *TM_PRED:* Prediction values are extrapolated from the row above and the column to the left using (fixed) second differences from the upper-left corner. (This mode is roughly comparable to H.264's planar prediction.)

VP8's core transform is a 4×4 DCT approximated by 16-bit integer coefficients. The decoder uses exact 16-bit arithmetic; there is no decoder drift.

Every entry in a Walsh-Hadamard matrix is either +1 or −1.

For the 16×16 luma prediction mode, luma processing involves a second level (Y2) transform: After the 16 luma subblocks have been transformed by the DCT, the 16 DC coefficients are collected and a (twenty-fifth) 4×4 transform is performed on those coefficients. The second-level transform is not a DCT, but a Walsh-Hadamard transform (WHT).

There are six quantizers, each with its own levels. Which quantizer is used depends upon the "plane" (first-order luma, second-order [Y2] luma, or chroma), and whether the coefficient is DC or AC.

Quantizer level is a 7-bit number that indexes an entry in one of the quantization tables. Quantization is potentially region-adaptive: The encoder associates each macroblock with one of four classes; each class has a different quantization parameter set.

VP8 implements a sophisticated arithmetic coding scheme, simpler than CABAC, but having comparable performance and lighter processing load. The encoder constructs estimates of probabilities of various syntax elements and parameter values. A default baseline parameter set is maintained; upon the occurrence of a keyframe, probability distributions are reset to the baseline. Probabilities are updated as each frame is processed; the encoder signals whether upon completion of decoding the updated set is to become the new baseline ("persistent") or is to be discarded ("one-time").

VP8 has an adaptive in-loop deblocking filter having quality and complexity roughly comparable to that of H.264's deblocking filter.

Further reading

Bankoski, Jim, Paul Wilkins, and Yaowu Xu (2011), "Technical overview of VP8, an open source video codec for the web," in *Multimedia and Expo (ICME), 2011 IEEE International Conf.*: 1–6.

Bankoski, Jim, Paul Wilkins, and Yaowu Xu (2011), *VP8 Data Format and Decoding Guide*, IETF Informational RFC. This information is available in a more readable form as Google On2 (2011), *VP8 Data Format and Decoding Guide* (revised 2011-02-04).

Feller, Christian, Juergen Wuenschmann, Thorsten Roll, and Albrecht Rothermel (2011), "The VP8 video codec – overview and comparison to H.264/AVC," in *Consumer Electronics – Berlin (ICCE-Berlin), IEEE International Conf.*: 57–61.

Part 6

Distribution standards

MPEG-2 storage and transport 50

Multimedia encompasses video and audio, potentially accompanied by other elements such as subtitles, coded in a manner suitable for synchronous presentation to the viewer. Many video compression systems are in use; for consumer use, MPEG-2 and H.264 are widely used. Many audio compression systems are in use; in the consumer domain, Dolby Digital (AC-3) and MPEG-1 Level III (MP3) are widely used.

Multimedia broadcasting or distribution requires that the various elements – *essences,* in the lingo of multimedia – are multiplexed into a single file or stream where the video and audio elements can subsequently be synchronized so as to be presented simultaneously.

Some multimedia formats used in PCs use multiple files – for example, one file for video and another for audio. Such schemes effectively push the multiplexing operation to the player software. Such schemes are prone to failure to play one kind of essence, or to have essences fall out of sync.

In multimedia computing, multiplexing is accomplished by structuring the various components into a *container file*. Microsoft's AVI, Apple's QuickTime, and *Matroska* (used in WebM) are examples. Such container formats are fairly well suited for computers, but not usually well suited to broadcast and sometimes even not very well suited to dedicated, high-performance playback from media such as DVD and Blu-ray disc.

In the section *MPEG-4*, on page 159, I briefly discussed the *ISO Base Media File Format*. That format serves as a container format for MPEG-4 Part 2/ASP video. That format is generally agreed to be inapplicable to professional video.

The *Systems* part of the MPEG-1 standard from 1992 established a multiplexing structure. That scheme was extended in MPEG-2, and the MPEG-2 scheme is now widely used in computing, in broadcasting, and in consumer video applications (including consumer camcorders using hard drive or flash media). MPEG-2 Part 1, *Systems,* defines two multiplexing mechanisms, the *program stream* (PS) and the *transport stream* (TS). Both can be regarded as MPEG "containers," whose structure is the subject of the remainder of this chapter.

Elementary stream (ES)

A coder – audio or video – produces a stream of bytes known as an *elementary stream.* The previous chapter outlines the information that is encoded into a video elementary stream. (Audio encoding is outside the scope of this book.)

An elementary stream can contain private streams.

Packetized elementary stream (PES)

An elementary stream is packetized into packets of 188 bytes, the first byte being MPEG's sync byte valued 47_h. Some systems construct 204 byte packets, expecting the channel coder to overwrite the final 16 bytes of each packet; in this case the sync byte will be $B8_h$.

MPEG-2 program stream

An MPEG-2 *program stream* (PS) a relatively simple mechanism to multiplex video and audio of a single program for storage or transmission on relatively error-free media such as computer disks or digital optical media. PS packets are variable-length; packets of 1 KByte or 2 KBytes are typical, though a packet can be as long as 64 KBytes. MPEG-2 program streams are used in applications such as these:

- DVD media uses a strict subset of MPEG-2 program stream encoding; the associated file extension is *vob*.
- The MOD consumer video format is essentially an MPEG-2 MP@ML SD program stream according to DVD conventions. On a computer, such files typically have extensions *mpg* or *mpeg*.

MOD is reported to stand for *MPEG on disk*.

MPEG-2 transport stream

An MPEG-2 *transport stream* (TS) is a part of the MPEG-2 suite of standards that specifies a relatively complex mechanism of multiplexing video and audio for one or more programs into a data stream, typically having short packets, suitable for transmission through error-prone media where relatively powerful forward error-correction (FEC) is required. A transport stream is suitable for applications where a player connects to a transmission in progress (like television), as opposed to reading a file from its beginning. For terrestrial over-the-air (OTA) or cable television, TS packets are expected to be suitably protected; however specification of the FEC and channel coding lies outside the MPEG standards and ordinarily lies within the realm of digital television standards (for example, ATSC standards in North America, and DVB standards in Europe).

ATM: Asynchronous transfer mode, a protocol for high performance networking.

A *transport stream packet* (TSP) comprises 188 bytes – a 4-byte header (whose first byte has the value 47_h), including a 13-bit *packet identifier* (PID), and 184 bytes of payload. Packet size was designed with ATM in mind: One TS packet fits into four ATM cells (48 bytes each). Owing to a lack of external interfaces for program streams, a *single program transport stream* (SPTS) may be used to carry one program. For some applications, a *multiple program transport stream* (MPTS) is used.

Transport stream packets with PID 0 contain the *program association table* (PAT), repeated a few times per second. The PAT lists one or more PIDs of subsequent packets containing *program map tables* (PMTs). A PMT lists PIDs of video and audio elementary streams associated with a single program.

ATSC Standard A/65, *Program and System Information Protocol.*

An ATSC DTV transport stream contains a set of packets implementing the *program and system information protocol* (PSIP). PSIP identifies channels and programs, and conveys time-of-day and station callsign information. PSIP enables a receiver to provide an electronic program guide (EPG).

On a computer, 188-byte transport stream packets typically have a 4-byte timecode appended (resulting in 192-byte packets); a file comprising a sequence of such packets typically has the extension *m2t, m2ts,* or just *ts*.

MPEG-2 transport streams are used in applications such as these:

TOD is reported to stand for *transport stream on disk*.

- The TOD consumer video format (essentially an MPEG-2 MP@HL HD transport stream)
- The BDAV container of Blu-ray
- H.264 compressed video
- AVCHD compressed video (in computing, the file extension *mts* is usual)

System clock

I write 29.97 Hz; expressed exactly, it's $^{30}/_{1.001}$.

Synchronization in MPEG is achieved through a *system clock reference* (SCR). The lowest common multiple of 25 Hz and 29.97 Hz is 30 kHz; In MPEG-2, 90 kHz was chosen as the basis for the *program clock reference* (PCR). A program clock value is represented in 33 bits, sufficient to provide unique PCR values over 24 hours.

27 MHz divided by 90 kHz is 300.

MPEG system timing is based upon a 27 MHz reference clock, expressed by augmenting the PCR by a nine-bit field taking a value from 0 through 299. Table 50.1 in the margin enumerates the number of PCR counts per frame at various frame rates.

Frame rate [Hz]	*PCR counts per frame*
30	3000
29.97	3003
25	3600
24	3750

Table 50.1 **MPEG-2 PCR counts per frame**

Each program stream has a single reference clock. Different programs in an MPTS can have different program clocks, so provision is made for a transport stream to carry multiple independent PCRs.

Further reading

Chen, Xuemin (2002), *Digital Video Transport System* (Springer).

Whitaker, Jerry C. (2003), "DTV Service Multiplex and Transport Systems," Chapter 13.2 in *Standard Handbook of Video and Television Engineering,* Fourth Edition (McGraw-Hill).

Whitaker, Jerry C. (2003), "DTV Program and System Information Protocol," Chapter 13.4 in *Standard Handbook of Video and Television Engineering,* Fourth Edition (McGraw-Hill).

Digital television broadcasting 51

This chapter briefly summarizes digital television broadcasting. Most digital broadcast systems that have been standardized are based upon MPEG-2 compression, described in *MPEG-2 video compression* on page 513. Some cable and satellite systems use H.264.

HDTV transmission systems were conceived to deliver images of about twice the vertical and twice the horizontal resolution of SDTV – that is, about 2 megapixels – in a 6 MHz analog channel. MPEG-2 can compress 2 megapixel images at 30 frames per second to about 20 Mb/s. Modern digital modulation schemes suitable for terrestrial RF transmission have a payload of about 3.5 bits per hertz of channel bandwidth. Combining these numbers, you can see that one HDTV digital signal can be transmitted in the spectrum formerly occupied by one analog NTSC 6 MHz channel.

The basic RF parameters of the 525-line, 60-field-per-second interlaced transmission scheme are basically unchanged since the introduction of black-and-white television in 1941! The modulation scheme requires that potential channels at many locations remain unused, owing to potential interference into other channels. The unused channels were called *taboo.* Digital television transmission takes advantage of half a century of technological improvements in modulation systems. The modulation system chosen allows very low power. This low power has two major consequences: It minimizes interference from digital transmitters into NTSC or PAL, and it allows use, for digital television transmission, of the channels that were formerly taboo. Digital television service is thus overlaid on top of analog service. (In early deployment of HDTV, program material was simulcast on a conventional analog transmitter.)

Japan

NHK SCIENCE AND TECHNICAL RESEARCH LABORATORIES (1993), *High Definition Television: Hi-Vision Technology* (New York: Van Nostrand Reinhold).

HDTV broadcasting based on 1035*i*30.00 scanning and MUSE compression was deployed in Japan in the early 1990s; the system is called *Hi-Vision*. MUSE is a hybrid analog/digital system optimized for direct broadcast from satellite (DBS); it is documented in the book from NHK Labs. MUSE predates the MPEG standards; nowadays, it is generally agreed that Japan adopted analog HDTV transmission standards prematurely.

In 2003, MUSE was superseded by ISDB-T, a terrestrial broadcasting system based upon MPEG-2 video coding and OFDM transmission.

United States

HDTV developers in the United States planned for broadcasters to gradually replace analog SDTV transmission with digital HDTV transmission. Partway through the development of HDTV, it became clear that the same compression and transmission technology that would allow one HDTV channel to be coded and transmitted at about 20 Mb/s would be equally suited to allow five SDTV channels to be coded, multiplexed, and transmitted at 4 Mb/s each! So, what began as high-definition television evolved into digital television (DTV), which encompasses both SDTV and HDTV. Compression is in accordance with MPEG-2 MP@ML for SD and MP@HL for HD, with restrictions specified in ATSC A/53.

ATSC A/53, *Digital Television Standard*.

DTV standards in the United States were developed by the Advanced Television Systems Committee (ATSC). Those standards were adopted by the Federal Communications Commission (FCC), with one significant change: The FCC rejected the set of 18 formats documented in Table 3 of ATSC Standard A/53 (presented as my Table 15.1, on page 143). Bowing to pressure from the computer industry, the FCC deleted that table, but left the rest of the ATSC standards intact. The fact that Table 3 is absent from FCC standards has virtually no practical import: In practice, consumer receivers are obliged to decode the Table 3 formats; they cannot be depended upon to decode anything else.

The FCC's deletion of Table 3 supposedly left the choice of raster standards to the marketplace: In principle, any format compliant with MPEG-2 MP@HL could be used. In practice, no U.S. broadcaster has chosen, and no consumer equipment is guaranteed to implement, any format outside ATSC's Table 3. In practice, DTV decoders conform to MPEG standards, and additionally conform to the restrictions imposed by the ATSC standards.

MPEG-2 specifies upper bounds for picture size at various levels. Surprisingly, ATSC A/53 specifies exact values. MPEG-2 MP@ML allows 720 image columns, but ATSC A/53 does not.

In the United States, DTV audio is standardized with Dolby Digital audio coding (also known as AC-3), a coding scheme not specified in the MPEG-2 standard. Dolby Digital is capable of "5.1" channels: left and right (stereo) channels, a front center channel, left and right surround channels, and a low-frequency effects (LFE) channel (the ".1" in the notation) intended for connection to a "subwoofer." ATSC audio consumes a maximum of 512 kb/s.

EIA-708-B, *Digital Television (DTV) Closed Captioning.*

EIA-708-B standardizes a method for conveying closed caption data (DTVCC).

One or more MPEG-2 program streams, the associated audio streams, ancillary data, and other data is multiplexed into a transport stream with a bit rate of about 19.28 Mb/s. The transport stream is augmented by Reed-Solomon forward error correction: Each 188-byte transport packet is augmented by 16 FEC bytes, resulting in 204-byte packets that are presented to the modulator.

ATSC modulation

$$4.5\,\frac{684}{286} \approx 10.762$$

The ATSC standardized 8-level digital vestigial sideband (8-VSB) modulation, transmitting about 10.762 million 3-bit symbols per second. To enable the receiver to overcome errors introduced in transmission, two forward error-correction (FEC) schemes are concatenated: The outer code is Reed-Solomon (R-S), and the inner code is a simple $^3/_2$ trellis code. Between the R-S and trellis coding stages, data is interleaved. Synchronization information comprising segment and field syncs is added after interleaving and coding. A low-level pilot carrier is inserted 310 kHz above the lower band edge to aid in carrier recovery. Analog techniques are used to upconvert to the UHF broadcast channel.

At a receiver, analog techniques downconvert the UHF broadcast to intermediate frequency (IF), typically 44 MHz. Demodulation is then accomplished digitally. Typically, an analog frequency and phase-locked loop (FPLL) recovers the carrier frequency based upon the pilot carrier. A quadrature demodulator then recovers *I* and *Q* components. The *I* component is converted from analog to digital at 10.76 MHz to recover the bitstream; the *Q* component is processed to effect phase control. The bitstream is then subject to trellis decoding, deinterleaving, R-S decoding, and MPEG-2 demultiplexing.

I and *Q* refer to *in-phase* and *quadrature;* these are the same concepts as in NTSC chroma modulation, but applied to completely different ends.

It is a challenge to design a demodulator that is immune to transmission impairments such as multipath distortion and co-channel interference from NTSC transmitters. An interference rejection filter – a variation of a comb filter – is built into the demodulator; it attenuates the video, chroma, and audio carriers of a potentially interfering NTSC signal. An adaptive equalizer built into the demodulator alleviates the effects of multipath distortion; the field sync component of the signal serves as its reference signal. An adaptive equalizer is typically implemented as an FIR filter whose coefficients are updated dynamically as a function of estimated channel parameters.

ATSC defines a cable mode using 16-VSB without trellis coding, but this mode hasn't been deployed. Digital cable television is detailed in the book by Ciciora and colleagues cited at the end of this chapter.

Cable television has very different channel characteristics than terrestrial broadcast. DTV over cable typically does not use 8-VSB modulation: Quadrature amplitude modulation (QAM) is used instead, with either 64 or 256 levels (64-QAM or 256-QAM).

For DBS, quadrature phase-shift keying (QPSK) is generally used.

Consumer receivers in the United States must accept the diversity of frame rates and raster standards in ATSC's Table 3 (my Table 15.1 on page 143). Although multiscan displays are ubiquitous in computing, both price and performance suffer when a display has to accommodate multiple rates. Most consumer HDTV receivers are designed with displays that operate over a limited range of scanning standards; the wide range of ATSC Table 3 is accommodated by digital resampling. In early deployment of HDTV, many receivers used

1366×768 displays, upconverting 720*p* or downconverting 1080*i* to that format. Today, most consumer HDTV receivers are 1080*i*-native, and convert other formats to 1080*i*.

Europe

In Europe, huge efforts were made in the 1980s and 1990s to develop an HDTV broadcasting system using 1250/50 scanning and a transmission system built upon the MAC transmission technology originally designed for 576*i*. MAC failed in the marketplace. HD-MAC failed also; this was partially a consequence of the commercial failure of MAC, partially because of technical weaknesses of HD-MAC, and partially because HD-MAC did not address worldwide markets.

DVB standards are promulgated by ETSI [www.etsi.org].

HDTV broadcasting in Europe was late, but digital broadcasting of SD is deployed based upon MPEG-2 MP@ML, with 720×576 image structure. The Digital Video Broadcasting (DVB) organization has created a comprehensive set of standards for cable (DVB-C), satellite (DVB-S), and terrestrial (DVB-T) broadcasting. DVB audio conforms to MPEG-2 audio. These standards are promulgated by ETSI.

The RF modulation system chosen for DVB-T is coded orthogonal frequency division multiplexing (COFDM). COFDM uses a large number of subcarriers to spread the information content of a signal evenly across a channel. The subcarriers of COFDM are individually modulated, typically using QPSK or QAM. COFDM exhibits greatly improved immunity to multipath distortion compared to 8-VSB. Also, COFDM accommodates transmission using single-frequency networks (SFNs) where the same bitstream is transmitted from multiple transmitters at different locations but operating on the same frequency.

Further reading

Adams, Michael, (2000), *Opencable Architecture* (Indianapolis, Indiana, U.S.A.: Cisco Press).

ATSC A/54A, (2006), *Recommended Practice: Guide to the Use of the ATSC Digital Television Standard, including Corrigendum No. 1* (Dec.)

Ciciora, Walter, James Farmer, David Large, and Michael Adams (2004), *Modern Cable Television Technology*, Second Edition (San Francisco: Morgan Kaufmann).

Collins, Gerald W. (2001), *Fundamentals of Digital Television Transmission* (New York: Wiley).

Dambacher, Paul (1998), *Digital Terrestrial Television Broadcasting* (Berlin: Springer).

Fischer, Walter (2008), *Digital Video and Audio Broadcasting Technology:* A *Practical Engineering Guide*, Second Edition (Berlin: Springer).

Reimers, Ulrich, ed. (2005), *DVB: The Family of International Standards for Digital Video Broadcasting* (Berlin: Springer).

Weiss, S. Merrill (1997), *Issues in Advanced Television Technology* (Boston: Focal Press).

Whitaker, Jerry C. (2003), "The ATSC DTV system," Chapter 13.1 in *Standard Handbook of Video and Television Engineering,* Fourth Edition (McGraw-Hill).

Appendices

This page intentionally left blank

YUV and *luminance* considered harmful

A

This is a plea for precise terminology. The notation *YUV*, and the term *luminance*, are widespread in digital video, computer graphics, and digital image processing. Actually, digital video almost never uses *Y'UV* colour difference components, and never directly represents the *luminance* of colour science. The common terms are almost always wrong. This note explains why. I urge video engineers and computer graphics specialists to use the correct terms, almost always $Y'C_B C_R$ and *luma*.

Cement *vs.* concrete

I'll demonstrate by analogy why it is important to use correct terms. Next time you're waiting in line for a bus, ask the person next to you in line what building material is used to construct a sidewalk. Chances are that person will answer, "cement."

The correct answer is *concrete*. Cement is calcined lime and clay, in the form of a fine, grey powder. Cement is one ingredient of concrete; the other ingredients are sand, gravel, and water.

In an everyday situation, you need not be precise about which of these terms are used: If you refer to a sidewalk as being constructed of "cement," people will know what you mean. Laypeople are not confused by the term *cement.* Interestingly, experts are not confused either. If a construction superintendent yells out to his foreman, "Get me 50 pounds of cement!" the foreman understands immediately from context whether the superintendent actually wants concrete. However, if you phone your local building material supplier and order 50 pounds of cement, you will certainly not receive 50 pounds of concrete! Laypeople have no trouble with the loose nomenclature, and the experts have little trouble. It is the people in the middle who are liable to become confused. Worse still, they are liable to use a term without realizing that it is ambiguous or wrong!

True CIE luminance

Absolute luminance – symbolized L, and having units of $cd \cdot m^{-2}$, colloquially *nit* – is rarely used in video. We ordinarily use an approximation to *relative* luminance, symbolized Y.

The principles of colour science dictate that true CIE relative luminance – denoted Y – is formed as a weighted sum of linear (tristimulus) *RGB* components. If CIE luminance were transmitted in a video system, the system would conform to the *Principle of Constant Luminance*. But in video we implement an engineering approximation that departs from this principle. It was standardized for NTSC in 1953, and remains standard for all contemporary video systems (both SD and HD), to form *luma*, denoted Y', as a weighted sum of *nonlinear* (gamma-corrected) $R'G'B'$ components:

$$^{601}Y' = 0.299\,R' + 0.587\,G' + 0.114\,B' \qquad \text{Eq A.1}$$

The nonlinear transfer function usually used is roughly comparable to a square root. We use the theoretical coefficients of colour science, but we use them in a block diagram different from the one prescribed by colour science: Gamma correction is applied *before* forming the weighted sum, not after. The "order of operations" is reversed from what you might expect from colour science.

The misinterpretation of luminance

Video engineers in the 1950s recognized that the video quantity Y' was very different from CIE luminance, and that it needed to be distinguished from luminance. They described it by the phrase *the quantity representative of luminance* or *the luminance signal*. They used the symbol Y, but augmented it with a prime to denote the nonlinearity: Y'. Obviously the qualifier "quantity representative of" was cumbersome, and over the decades, it was elided. And over time, the prime symbol was elided as well.

Unfortunately, no new word was invented to supplement *luminance*, to reinforce the distinction between the colour science quantity and the video quantity. Most video engineers nowadays are unfamiliar with colour science, and most do not understand the distinction. Engineers today often carelessly use the word *luminance*, and the symbol *Y*, to refer to the weighted sum of nonlinear (gamma-corrected) *R'G'B'* components.

Pritchard, Dalton H. (1977), "U.S. color television fundamentals – A review," in *SMPTE Journal,* **86** (11): 819–828 (Nov.).

Smith, Alvy Ray (1978), "Color gamut transform pairs," in *Computer Graphics* **12** (2): 12–19 (Aug., Proc. SIGGRAPH).

Foley, James D., and Andries van Dam (1984), *Fundamentals of Interactive Computer Graphics* (Reading, Mass.: Addison-Wesley).

Foley, James D., Andries van Dam, Steven Feiner, and John Hughes (1990), *Computer Graphics: Principles and Practice*, Second Edition (Reading, Mass.: Addison-Wesley). 589 (Section 13.3.3).

The sloppy nomenclature made its way into ostensibly authoritative video references, such as Pritchard's influential SMPTE paper published in 1977.

The computer graphics pioneer Alvy Ray Smith encountered the word *luminance* in his quest to adapt video principles to computer graphics. Smith apparently correlated the use of the term *luminance* with his knowledge of colour science, and understandably – though wrongly – concluded that video "luminance" and colour science luminance were identical. Consequently, video *Y'IQ* was introduced to computer graphics, having its *Y* component alleged to be identical to CIE luminance.

That incorrect interpretation propagated into authoritative computer graphics textbooks. *Computer Graphics: Principles and Practice*, Second Edition, in the section entitled *The YIQ Color Model*, includes this sentence:

> The *Y* component of YIQ is not yellow but luminance, and is defined to be the same as the CIE **Y** primary.

The emphasis is in the original. "Yellow" refers to *CMY;* printing inks were mentioned in the immediately preceding section. "CIE **Y** primary" would be more accurately denoted "CIE *Y* component."

Widespread use of incorrect terminology is not a new phenomenon. The indigenous people of North America were, for many centuries, referred to as "Indians." Why? After his long voyage across what we now call the Atlantic Ocean, when Christopher Columbus finally saw land, he thought it was India.

Contrary to the quoted paragraph, the so-called *Y* component of video – more properly designated with a prime symbol, *Y'* – is *not* the same as CIE luminance. Video *Y'* cannot even be computed from CIE *Y*, unless two other colour components (typically colour difference components based upon *B'–Y'* and *R'–Y'*) are also available. The quoted passage is quite wrong. Apparently, hundreds of thousands of copies of various editions and adaptations of this book have been printed. Confusion is rampant.

Pratt, William K. (1991), *Digital Image Processing*, Second Edition (New York: Wiley), 64. The error is corrected in the third [2001] and fourth [2007] editions.

The error propagated into the digital image-processing community. Pratt's textbook states:

> N.T.S.C. formulated a color coordinate system for transmission composed of three tristimulus values YIQ. The Y tristimulus value is the luminance of a color.

The video quantities are certainly *not* tristimulus values, which are, by CIE's definition, proportional to intensity.

Loose nomenclature on the part of video engineers has misled a generation of digital image processing, computer software, and computer hardware engineers.

The enshrining of luma

The term *luma* was used in video for a long time, without having had a precise interpretation. I campaigned among video engineers, and among computer graphics experts, for adoption of the term *luma* to designate the nonlinear video quantity. The term offers no impediment to video engineers – in fact, it slides off the tongue more easily than *luminance*. By virtue of its being a different word from *luminance*, the word *luma* invites readers from other domains to investigate fully before drawing conclusions about its relationship with luminance.

SMPTE EG 28, *Annotated Glossary of Essential Terms for Electronic Production*.

With the help of Fred Kolb, my campaign succeeded: In 1993, SMPTE adopted Engineering Guideline EG 28, *Annotated Glossary of Essential Terms for Electronic Production*. EG 28 defines the term *luma*, and clarifies the two conflicting interpretations of the term *luminance*. While a SMPTE EG is not quite a SMPTE "Standard," at long last the term has received official recognition. There's no longer any excuse for sloppy use of the term *luminance* by the authors of video engineering papers and books. Had the term *luma* been widespread 20 years ago when A. R. Smith was writing about *YIQ*, or when Foley and van Dam were preparing *Computer Graphics: Principles and Practice*, this whole mess would have been avoided. But EG 28 was unavailable at the time.

It is a shame that today's SMPTE, ISO/IEC, ITU-R, and ITU-T standards persist in using the incorrect word *luminance,* without ever mentioning the ambiguity – even conflict – with the CIE standards of colour science.

Colour difference scale factors

When I say *NTSC* and *PAL*, I refer to colour encoding, not scanning. I do not mean 480*i* and 576*i*, or 525/59.94 and 625/50!

To represent colour, luma is accompanied by two *colour difference* – or *chroma* – components, universally based on *blue minus luma* and *red minus luma*, where blue, red, and luma have all been subject to gamma correction: $B'-Y'$ and $R'-Y'$. Different scale factors are applied to the basic $B'-Y'$ and $R'-Y'$ components for different applications. $Y'P_BP_R$ scale factors are optimized for component analog video. $Y'C_BC_R$ scale factors are optimized for component digital video, such as 4:2:2 studio video, JPEG, and MPEG. Correct use of the $Y'UV$ and $Y'IQ$ scale factors is limited to the formation of composite NTSC and PAL video.

ITU-R Rec. BT.601-5, *Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios*.

$Y'C_BC_R$ scaling as defined by BT.601 is appropriate for component digital video. $Y'C_BC_R$ chroma is almost always subsampled using one of three schemes: 4:2:2, or 4:2:0, or 4:1:1.

$Y'UV$ scaling is properly used only as an intermediate step in the formation of composite NTSC or PAL video signals. $Y'UV$ scaling is not appropriate when the components are kept separate. However, the $Y'UV$ nomenclature is now used rather loosely, and sometimes – particularly in computing – it denotes *any* scaling of $B'-Y'$ and $R'-Y'$.

In about 1991, digital disk recorders (DDRs) were introduced that were able to transfer files across Ethernet using IP protocols. Abekas introduced the extension *yuv* for these files. But the scale factors typically used (then and now) actually correspond to $Y'C_BC_R$. Use of the *yuv* extension reinforced the misleading *YUV* nomenclature.

Chroma components are properly ordered $B'-Y'$ then $R'-Y'$, or C_B then C_R. Blue associates with U, and red with V; U and V are in alphabetic order.

Subsampling is a digital technique, properly performed only on component digital video – that is, on $Y'C_BC_R$. Subsampling is inappropriate for $Y'UV$ in all but very specialized applications (namely, digital encoding of $4f_{SC}$ NTSC or PAL composite video). If you see a system described as *Y'UV 4:2:2*, you have a dilemma. Perhaps the person who wrote the description is unfamiliar with the principles of component video, and the scale factors actually implemented in the equipment (or the software) are correct. But you must allow for the possibility that the engineers who designed or implemented the system used the wrong scale factors! If the wrong equations were used, then colour accuracy will suffer; however, this can be difficult to diagnose.

Hamilton, Eric (1992), *JPEG File Interchange Format*, Version 1.02 (Milpitas, Calif.: C-Cube Microsystems).

Proper $Y'C_BC_R$ scaling is usual in Motion-JPEG, and in MPEG. However, the $Y'C_BC_R$ scaling used in stillframe JPEG/JFIF in computer applications usually uses full-range luma and chroma excursions, without any headroom or footroom. The chroma excursion is $^{256}/_{255}$ of the luma excursion. The scaling is almost exactly that of $Y'P_BP_R$, but is unfortunately described as $Y'C_BC_R$: Now even $Y'C_BC_R$ is ambiguous! It is far too late for proper $Y'C_BC_R$ scaling to be incorporated into JFIF; compressed stillframe and motion imagery in computing is bound to suffer a conversion process.

Y'IQ coding has been obsolete in studio practice for at least three decades. It should now be banished in favour of $Y'C_BC_R$.

Conclusion: A plea

Using the term *luminance* for video *Y'* is tantamount to using the word *cement* instead of *concrete* to describe the primary construction material of a sidewalk. Lay people don't care, and experts can live with it, but people in the middle – in this case, the programmers and engineers who are reimplementing video technology in the computer domain – are liable to draw the wrong conclusions from careless use of terms, and thereby make inaccurate colour. The accurate exchange of images is compromised, and users suffer.

I urge video engineers and computer graphics specialists to avoid the terms *YUV*, *Y'UV*, *YIQ*, *Y'IQ*, and *luminance*, except in the highly specialized situations where those terms are technically correct. The appropriate terms are almost always $Y'C_BC_R$ and *luma*.

Introduction to radiometry and photometry

B

The domain of *radiometry* involves optical power and its spatial and angular distributions. *Photometry* is, in essence, radiometry weighted by the spectral response of vision. These fields involve several subtle concepts, masked by a bewildering array of symbols and units. I strive to sort out some of the confusion, and make some suggestions concerning units and nomenclature.

Table B.1 below summarizes radiometric quantities, symbols, and units (in the left columns) and the corresponding photometric quantities, symbols, and units (to the right). The symbol for a photometric quantity is just the symbol for the corresponding radiometric quantity, with the addition of the subscript v (for *visual*). Some people add the subscript e to the radiometric symbols.

ANSI/IESNA RP-16, *Nomenclature and Definitions for Illuminating Engineering*.

CIE Nº 17.4 (E-1.1) (1987), *International Lighting Vocabulary*, 4th Edition (Vienna, Austria: Commission Internationale de L'Éclairage).

Differentiate flux with respect to	*Radiometric (Radiant)*		*Photometric (Luminous)*	
	Quantity (Symbol)	*Unit*	*Quantity (Symbol)*	*Unit*
–	**radiant flux, power** (Φ, F, P)	watt [W]	**luminous flux** (Φ_v, F_v, P_v)	lumen [lm]
area	**irradiance** (E), **radiant exitance** (M)	$\mathrm{W \cdot m^{-2}}$	**illuminance** (E_v), **luminous exitance** (M_v)	$\mathrm{lm \cdot m^{-2}}$ = lux [lx]
solid angle	**radiant intensity** (I)	$\mathrm{W \cdot sr^{-1}}$	**luminous intensity** (I_v)	$\mathrm{lm \cdot sr^{-1}}$ = candela [cd]
area *and* solid angle	**radiance** (L)	$\mathrm{W \cdot sr^{-1} \cdot m^{-2}}$	**luminance** (L_v)	$\mathrm{cd \cdot m^{-2}}$ = nit [nt]

Table B.1 **Quantities, symbols, and units of radiometry and photometry.** The symbol L_v (or L) is used for absolute luminance (e.g., in photometry). The symbol Y is used for relative luminance (e.g., in colour science and in video).

Radiometry and photometry involve light, in space. No surface is necessary! Light properties may be described *at* a real or imaginary surface; however, they are not properties *of* a surface. Absorptance (α), reflectance (ρ), and transmittance (τ) are intrinsic properties of surfaces, not properties of light.

In what follows, I use the usual physics convention of writing letter symbols in italics and units in Roman type.

Radiometry

Radiometry starts with energy (symbolized Q) at wavelengths between about 100 nm and 1 mm. Energy is expressed in units of joules [J]. A photon's energy (Q_p) is related to its wavelength (λ), here given in meters:

Eq B.1

$$Q_\mathrm{p} = h\frac{c}{\lambda}; \qquad \begin{aligned} &h \approx 6.6260755 \cdot 10^{-34}\ \mathrm{J \cdot s}\ \text{(Planck's constant)} \\ &c \equiv 299792458\ \mathrm{m \cdot s^{-1}}\ \text{(speed of light)} \end{aligned}$$

The rate of flow – or formally, the time-derivative – of radiant energy is power (P), also known as radiant flux (F, or preferably, Φ), expressed in units of watts [W].

Radiant exitance sums emitted and reflected light. The former term *emittance* excludes reflected light.

Some thermal engineers use the term *radiosity* for radiant exitance. In computer graphics, *radiosity* refers to a specialized technique to compute illumination; it does not refer to any particular quantity.

Radiant flux per unit area – that is, flux density – arriving at a point, throughout all directions in a hemisphere, is *irradiance* (E). Irradiance is expressed in units of watts per meter squared [$\mathrm{W \cdot m^{-2}}$ or $\mathrm{W/m^2}$]. Solar irradiance (*insolation*) at noon is about 1 $\mathrm{kW \cdot m^{-2}}$.

Radiant flux per unit area leaving a point, in all directions, is *radiant exitance* (M). Formally, this is the derivative of radiant flux with respect to area. Radiant exitance is expressed in units of watts per meter squared [$\mathrm{W \cdot m^{-2}}$]. Radiant exitance from a nonemissive surface is simply its irradiance times its reflectance.

In photography, the symbol I is often used for irradiance or illuminance, instead of intensity. Beware that *sound* intensity is conceptually unlike light intensity: Sound intensity has dimensions of power per unit area, comparable to irradiance of light.

Radiant flux in a specified direction – formally, radiant flux per unit solid angle – is *radiant intensity* (I); its SI unit is watts per steradian [$\mathrm{W \cdot sr^{-1}}$ or W/sr]. Intensity must be specified in a particular direction. Intensity is a property of a point-like source; it is independent of distance to the observer or measurement device. Intensity does not follow the inverse square law! Intensity has many other meanings in physics, with the most common being power per unit area, but the unambiguous term for that quantity is *irradiance.*

Palmer, James M. (1993), "Getting intense on intensity," in *Metrologia* **30** (4): 371–372.

The late James M. Palmer pointed out that the term *intensity* is widely misused, for no good reason, because it is one of the seven base units of the SI system!

Beware so-called intensity expressed in units other than watts per steradian, W·sr^{-1}. Some authors in thermal engineering, and some authors in the computer graphics field of radiosity, use the term *radiant intensity* for what I call *radiance*, which I will now describe.

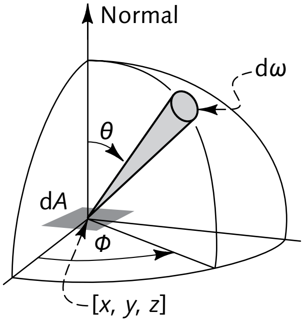


Figure B.1 **Geometry associated with the definition of radiance** The quantity dA represents unit area; the quantity dω represents unit solid angle. Projected area falls off as cos θ.

Radiant flux density in a specified direction is *radiance* (symbol L). Formally, radiance is radiant flux differentiated with respect to both solid angle and projected area; the geometry of this definition is depicted in Figure B.1. Radiance is expressed in units of watts per steradian per meter squared [W/sr/m^2, or preferably, W·sr^{-1}·m^{-2}]. For a large, nonpoint source, radiance is independent of distance.

Confusingly, radiance is often called *intensity* in some domains, especially heat transfer, astrophysics, and astronomy. Also, what is today called radiance was once called *radiometric brightness,* or just *brightness.* The term *brightness* remains in use in astronomy, but it is deprecated for image and colour science.

Radiance can be considered to be the fundamental quantity of radiometry: All other radiometric quantities can be computed from it. You might find it intuitive to start with radiance, and then consider the following:

- Radiant intensity is radiance integrated across an area.
- Irradiance is radiance integrated through solid angle, that is, integrated across all directions in a hemisphere.
- Flux is irradiance integrated across area, or equivalently, radiant intensity integrated through solid angle.

All of these radiometric terms relate to a broad spectrum of wavelengths. Any of these terms may be limited to a narrow spectrum by prepending *spectral* to the term, subscripting the letter symbol with λ, and appending *per nanometer* (·nm^{-1}) to the units.

Photometry

So far, I have discussed the physical quantities of radiometry. Photometry is entirely analogous, except that the spectral composition of each quantity is weighted by the spectral sensitivity of human vision, standardized as the luminous efficiency of the CIE Standard Observer (graphed in Figure 20.1 on page 205).

A lumen is produced by about 1.5 mW of monochromatic power at 555.5 nm. That wavelength, a frequency of 540 THz, corresponds to the peak luminous efficiency of vision.

Radiometry and photometry are linked by the definition of the candela: One candela [cd] is the luminous intensity of a monochromatic 540 THz source having a radiant intensity of $\frac{1}{683}$ $W \cdot sr^{-1}$. Once this definition is established, the remaining photometric quantities and units parallel those of radiometry. The relationships are sketched in Figure B.2 opposite.

The photometric analog of radiant flux is luminous flux (Φ_V). To quote James Palmer, luminous flux is what you want when you buy a light bulb. Its brightness is lumens, the photometric analog of watts; its efficacy is measured in lumens per watt. One lumen appears equally bright regardless of its spectral composition.

The quantity *illuminance* was formerly called *illumination.* However, use of *illumination* for this quantity is deprecated, owing to the more general meaning of the word as the act of illuminating, or the state of being illuminated.

The old *footcandle* and *metercandle* units for luminous exitance were misleading, because the old *candle* was a unit of intensity, not flux.

Luminous flux per unit area – that is, luminous flux density – arriving at point is *illuminance* (E_V), having SI units of lux [lx]. One lux is defined as 1 $lm \cdot m^{-2}$. Illuminance is the photometric analog of irradiance; it is the quantity measured by an incident light meter. Luminous flux per unit area leaving a point is *luminous exitance* (M_V). One lux equals 1 $lm \cdot m^{-2}$, whether the light is coming or going; however, traditionalists often state luminous exitance in units of $lm \cdot m^{-2}$. Luminous exitance from a nonemissive surface is its illuminance times its reflectance.

Luminous flux in a particular, specified direction is *luminous intensity* (I_V), expressed in units of $lm \cdot sr^{-1}$, or candela [cd]. The candela is the modern equivalent of the old *candle* (colloquially, candlepower). Intensity is independent of distance: A candle has a luminous intensity of about 1 cd, at any viewing distance.

The unit of luminous energy is the *talbot* [T]. A talbot is a lumen-second.

Nit isn't an official SI unit but it is synonymous with $cd \cdot m^{-2}$. It has been used by the CIE since 1947 and is in common use today. *Nit* is derived from the Latin word *nitere,* "to shine."

It is an error, sadly common among home theatre calibrators, to abbreviate candela per meter squared as candela. Candela is a different unit. To abbreviate $cd \cdot m^{-2}$, use *nit* [nt].

Luminous flux density in a particular direction is *luminance* (L_V). Formally, luminance is luminous flux differentiated with respect to both solid angle and projected area. Luminance is the photometric analog of radiance; it is expressed in units of $cd \cdot m^{-2}$, or colloqially, *nit* [nt]. Luminance is an important and useful measure, because it is invariant under transformation by a lens. Luminance of 1 $cd \cdot m^{-2}$ corresponds to roughly a million photons at 560 nm, per square degree, per second. Brightness is the perceptual correlate of luminance; however, brightness perception is very complex

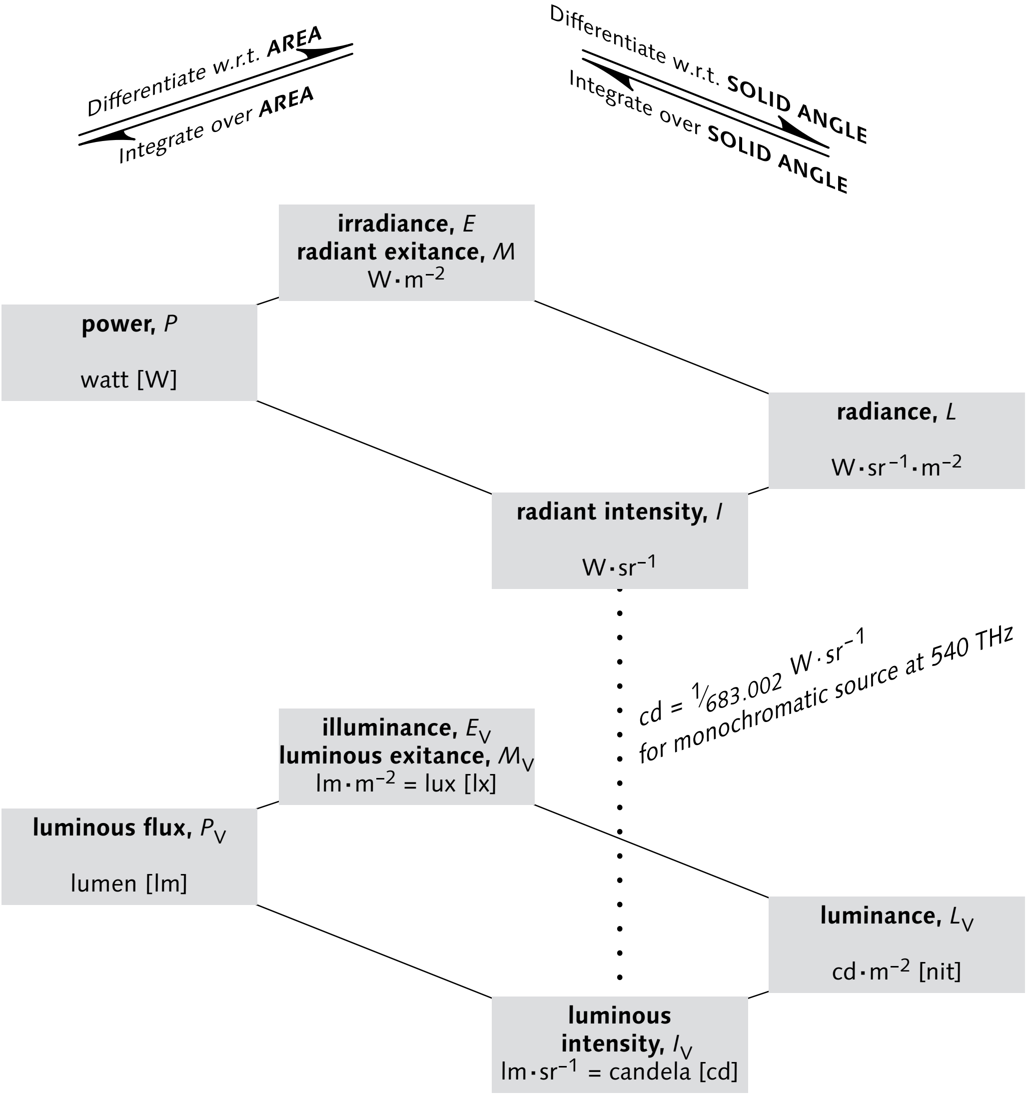


Figure B.2 **Radiometric and photometric quantities** and units are related in this diagram. The top quad shows radiometric quantities – radiant flux, irradiance, radiant intensity, and radiance. At the bottom are photometric quantities – luminous flux, illuminance, luminous intensity, and luminance. The systems are linked by the definition of the candela in terms of radiant intensity.

and highly nonlinear. By CIE's definition, brightness is a subjective quantity that can't be measured.

A perfect Lambertian (diffuse) reflector exhibits luminance [in $\mathrm{cd \cdot m^{-2}}$, nt] of $^1/_\pi$ times its illuminance [in lux].

Light level examples

The following table gives examples of light levels that are encountered in everyday life.

Situation	*Illuminance [lx]*	*Luminance of 90% diffuse reflector [nt]*	*Luminance of 18% diffuse reflector [nt]*
Clear, bright sunlight at noon	100 000	30 000	6 000
Typical daylight	10 000	3 000	600
Overcast daylight sky; TV studio	1 000	300	60
Very dark overcast sky; living room	100	30	6
Twilight; candle at 33 cm	10	3	0.6
Deep twilight; candle at 1 m	1	0.3	0.06
Full moon overhead, clear sky	0.1	0.03	0.006
Half moon overhead, clear sky	0.01	0.003	0.000 6
Starlight + airglow	0.002		
Total starlight, overcast night	0.000 1		
Sirius (m_V = –1.47)	0.000 01		
Vega (m_V = 0)	0.000 003		

Table B.2 **Light level examples**

Image science

Absolute luminance has units of $cd \cdot m^{-2}$, colloquially called nits [nt]. In image science, luminance is usually normalized to a range of 100 (or as I prefer, 1) with respect to a specified or implied white reference, and expressed without units. So normalized, its symbol is *Y*. The term *luminance* is often used as shorthand to refer to this pure quantity; however, it is properly called *relative luminance*. The term *luminance factor* should be avoided for this quantity, since the latter term refers to a property of a surface or material: *Luminance factor* is the ratio of luminance of a surface, under specified conditions of light source, incidence, and observation, to the luminance of a perfectly diffusing ("Lambertian") surface, under the same conditions.

Relative luminance (*Y*) is one of three distinguished *tristimulus values* standardized by the CIE; the other two distinguished tristimulus values are *X* and *Z*. Other tristimulus values such as [*R*, *G*, *B*] are related to CIE [*X*, *Y*, *Z*] values by a 3×3 linear matrix product. Relative luminance (*Y*) and other tristimulus values such as *X*, *Z*, *R*, *G*, or *B* are pure numbers.

Michael Brill and Bob Hunt agree that *R*, *G*, and *B* tristimulus values have no units. See HUNT, R.W.G. (1997), "The heights of the CIE colour-matching functions," in *Color Research and Application*, **22** (5): 337 (Oct.).

Units

Many bizarre units have been used for illuminance and luminance. I urge you to abandon these, and to adopt the standard SI units. Radiometry and photometry are sufficiently difficult without having to deal with a plethora of arcane units. If radiometry and photometry are new to you, I believe that your understanding will come more rapidly if you ignore the traditional units – which were deprecated by the scientific community 40 years ago – and adopt the SI units. If you are a practitioner who learned the science and the craft using the traditional Imperial units, please don't stubbornly stick to them: According to the *CIA World Factbook,* only three countries – Burma, Liberia, and the United States – have not adopted International System of Units (SI, or metric system) as their official system of weights and measures. I urge you to use SI units. Simply learn to multiply footlamberts by 3.4 to get candelas per meter squared. A movie screen has a typical white luminance of about 14 fL; call this 48 nits. A studio reference display in North America typically has a reference white luminance of about 100 nits (about 33 fL).

It was inconsistent mixture of US customary (*Imperial!*) units and SI units that led to the 1998 crash of NASA's Mars Climate Orbiter. American cinema experts often express luminance in footlamberts; why then do they refer to cinema film as 35 mm instead of 1 $^3/_8$-inch?

To convert illuminance into lux [lx], use Table B.3:

$$\frac{10^6}{25.4^2 \cdot 12^2} = \frac{1}{0.3048^2} \approx 10.764$$

To obtain $\text{lm}\cdot\text{m}^{-2}$ [lx], *multiply unit below*	*by*	*numerically*
$\text{lm}\cdot\text{ft}^{-2}$, footcandle, fc	10.764	10.764
metercandle	1	1

Table B.3 **Conversion of illuminance into lux**

To convert luminance into candelas per meter squared, use Table B.4:

To obtain $\text{cd}\cdot\text{m}^{-2}$ [nit, nt], *multiply unit below*	*by*	*numerically*
lambert, L	$^{10\,000}/_{\pi}$	3183.1
millilambert, mL	$^{10}/_{\pi}$	3.1831
$\text{cd}\cdot\text{ft}^{-2}$	10.764	10.764
footlambert, fL	$^{10.764}/_{\pi}$	3.4263

Table B.4 **Conversion of luminance into** $\mathbf{cd\cdot m^{-2}}$

Further reading

Chapter 1 of Ian Ashdown's book presents a very approachable introduction to measuring light. (The remainder of the book details the computer graphics technique called *radiosity*.) A version of that chapter is available on the web.

D. Allan Roberts offers a terse summary of the basic quantities of radiometry and photometry, and describes the confusing units. (It is worth seeking out the 1994 edition; sadly, the description of radiometry in the 2006 edition of this handbook is not nearly as lucid.)

ASHDOWN, IAN (1994), *Radiosity: A Programmer's Perspective* (New York: Wiley).

ASHDOWN, IAN (2002), *Photometry and Radiometry – A Tour Guide for Computer Graphics Enthusiasts*, <http://www.helios32.com/Measuring%20Light.pdf>.

ROBERTS, D. ALLAN (1994), "A Guide to Speaking the Language of Radiometry and Photometry," in *Photonics Design and Applications Handbook*, 1994 edition, vol. **3**, pages H-70 to H-73 (Pittsfield, Mass.: Laurin Publications).

Glossary

α–ω 0–9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

This glossary supplements the main text of *Digital Video and HD Algorithms and Interfaces.* This glossary is self-contained; it contains no references to the main text, and it is not indexed. For discussion in the main text concerning any term herein, consult the index. I spell colour with a *u.*

Terms specific to legacy analog NTSC and PAL are not included here. For those terms, see the *Glossary of legacy video signal terms* in the book *Composite NTSC and PAL: Legacy Video Systems.*

For the purpose of collating, we place Greek letters first, then numbers, then alphabetic entries. Adjacent letters form units separated by spaces, hyphens, and other symbols; numbers (including decimal point) are also units. Sorting is by unit, so *A/V C* and *AVC* lie several pages apart, and C_B lies with C/PH, not next to CC. Sorting of numbers is by increasing numeric value (considering any decimal fraction). Entries for values between 0 and 9 come first, then 10 to 99, then three-digit numbers, then four – for example, 0_H, 1-bit, 2-2 pulldown, 3:1:1, 3.58 MHz, $4f_{SC}$, 7.5-percent setup, 8 VSB, 10-bit, 18%, 24*p*, 109%, 525/59.94, 601, 1080*i*, 9300 K.

α (alpha)	**1**	In optics, absorptance, the fraction of light absorbed – that is, neither reflected nor transmitted.
	2	For computer graphics usage, see *alpha, ·* , on page 591.
γ (gamma)		See several entries for *gamma,* starting on page 621.
0_H datum		In a video stream, the reference point of horizontal (line) sync. In bilevel sync, the 50%-point of the leading edge of the transition to synctip level. In trilevel sync, the zero-crossing between the negative and positive pulses.
0_V datum		In a video stream, the reference point of vertical sync; the start of line 1.
0–255		The standard range of 8-bit pixel values (R', G', B', or Y') from reference black to reference white according to the sRGB standard IEC 61966-2-1 and common computer graphics practice.

1:1 pixel mapping	A system or subsystem that maps video samples to display pixels without any image resampling.
1-chip	A camera having a single image sensor. In the usual case of a colour camera, a 1-chip sensor incorporates mosaic colour separation filters. Compare to *3-CCD, 3-CMOS, 3-chip,* below.
1-D LUT	A LUT (see page 637) that enumerates the scalar results of a function over a range of whole numbers, typically 0 to $2^N - 1$. Useful in greyscale imaging, and useful component-wise in colour imaging exhibiting additive colour mixture.
1 GbE	Informal designation for 1 Gb/s Ethernet; formally *1000BASE-T* and informally also *GigE.*
2-2 pulldown	A process whereby a picture sequence originated at 24 frames per second (cinema or 24*p* digital material) is transferred to video at a field (or, in progressive scanning, frame) rate of 50 Hz. Each 24*p* frame is scanned twice; once to produce the first field (or frame), and once again to produce the second field (or frame). Synonymous with 24 @ 25, but different from 24 @ 25 pulldown! See *24 @ 25 pulldown* and *2-3 pulldown*.
2-2-2-4 pulldown	A deprecated process whereby a picture sequence originated at 23.976 fps [or 24 fps] (motion picture film, or 23.976*p* or 24*p* digital material) is transferred to video at a field (or, in progressive scanning, frame) rate of 59.94 Hz [or 60 Hz]. Groups of four film frames are scanned two, two, two, then four times to form successive video pictures. When such material is viewed directly, the process yields severe stutter.
2-3 pulldown	A process whereby a picture sequence originated at 23.976 fps [or 24 fps] (motion picture film, or 23.976*p* or 24*p* digital material) is transferred to video at a field (or, in progressive scanning, frame) rate of 59.94 Hz [or 60 Hz]. Alternate film frames are scanned first two then three times to form successive video pictures.
2-3-3-2 pulldown	A process whereby a picture sequence originated at 23.976 fps [or 24 fps] (motion picture film, or 23.976*p* or 24*p* digital material) is transferred to video at a field (or, in progressive scanning, frame) rate of 59.94 Hz or 60 Hz. Alternate film frames are scanned two, three, three, then two times to form successive video pictures. The advantage over classic 2-3 pulldown is that when interlaced video is produced, every video frame contains two fields from a single film frame; the disadvantage is that such material viewed directly exhibits somewhat more stutter than 2:3 pulldown.
2.0	Two-channel (stereophonic) sound, contrasted with 5.1.
2:1 interlace	See *interlace,* on page 628.
2 K	Relating to image representations having 2048 image columns (samples per picture width), such as 2048×1080; particularly

	the system for digital cinema devised by the DCI and standardized by SMPTE. Use of 2048 image columns instead of 1920 reflects an industry differentiation rather than a technical advantage; however, 2 K image data is normally represented in a colour space that mimics cinema, not the BT.709 colour space that characterizes HD.
2-pop	A 1 kHz audio tone burst at –20 dB_{FS} occupying exactly one frame interval ending exactly 2 seconds (or in 25 fps systems, 1.92 s) prior to the first frame of action of a program. In cinema, the 2-pop is synchronized with the leader frame that displays the digit “2.”
2.4:1	The “CinemaScope” aspect ratio commonly used for movies.
3-CCD, 3-CMOS, 3-chip	A camera having three optically aligned image sensors, achieving colour separation through an optical beamsplitter (“prism”). Compare to *1-chip,* above.
3:1:1	Component digital video wherein each C_B and C_R component is horizontally subsampled by a factor of 3 with respect to luma, and not subsampled vertically. Not widely used.
3-2 pulldown	See *2-3 pulldown,* above. The notation *3-2 pulldown* is inaccurate because SMPTE standards denote as the *A-frame* the first 24*p* frame in the four-frame sequence; that frame is associated with two fields (the first and second) of one picture, which is scanned twice (not three times).
3 dB	Three decibels: half power, equivalent to 0.707 signal amplitude. Relative amplitude at a filter's corner frequency.
3-D	**1** Relating to imaging in general, stereoscopic: A system that acquires, processes, records, transmits, and/or displays two views of the same scene from slightly different positions, one destined for the viewer's left eye and the other for the right. **2** Relating to colour processing, techniques – ordinarily involving lookup tables – that process three components in conjunction using nonlinear operations instead of processing components separately or combining them in linear manner.
3-D LUT	A LUT (see page 637) that enumerates the three-dimensional vector results of a function over three-dimensional vector arguments. Arguments and results are typically whole numbers 0 to 2^N-1. Useful in colour imaging exhibiting nonadditive colour mixture (that is, exhibiting nonlinear colour crosstalk).
3-D LUT interpolation	A process using a 3-D LUT (see above) that includes an algorithm that rather coarsely quantizes argument values (perhaps to 17 or 33 levels) and interpolates between result values.
3G-SDI	A serial digital interface (SDI) having a data rate of either $^{2.97}/_{1.001}$ Gb/s or 2.97 Gb/s, commonly used to transport 1080*p*30, 4:4:4 video and also capable of 1080*p*60, 4:2:2.

3 K	**1** Relating to image representations having roughly 3000 image columns (samples per picture width), such as 2880×2160 (i.e., 1.5 times 1920×1080) or 3072×1728 (i.e., $3\cdot 1024\times {}^{9}/_{16}\cdot 3\cdot 1024$); particularly systems intended for digital cinema acquisition. 3 K image data is normally represented in a colour space that mimics cinema. The notation 3 K is ambiguous with respect to whether a pixel comprises one or three colour components; see *4 K,* on page 584. **2** The approximate temperature, in units of kelvin, of the cosmic background radiation, the discovery of which earned the Nobel prize in physics for Arno Penzias and Robert Wilson.
3-way colour correction	Colour correction of video-coded image data whereby the colour balance of shadow tones, midtones, and light tones is adjusted separately. The "thresholds" defining the transition points between these regions (in terms of luma) can typically be adjusted. Often presented as a set of three colour wheels. Not to be confused with *lift-gamma-gain.*
3.58 MHz	More precisely, 3.579545 MHz, or exactly, $5\times{}^{63}/_{88}$ MHz: the colour subcarrier frequency of 480*i* NTSC video.
$4f_{SC}$	Obsolete composite digital video using a sampling frequency of 4 times the colour subcarrier frequency. There were 480*i* NTSC and 576*i* PAL versions of $4f_{SC}$.
4 K	Relating to image formats having 4096 image columns (samples per picture width), such as 4096×2160; particularly the system for digital cinema devised by the DCI and standardized by SMPTE. 4 K imagery is normally represented in a colour space that mimics cinema. The notation *4 K* is ambiguous. When used to characterize a digital cinema camera, the each of the 4096×2160 "pixels" typically comprise a single colour component (red, green, *or* blue, or perhaps another green). When used to characterize a digital cinema film scanner, or a projector, each of the 4096×2160 pixels typically comprises three colour components – red, green, *and* blue.
4:0:0	Greyscale video (sometimes confusingly called *monochrome*).
4:1:1	Chroma subsampling wherein C_B and C_R components are horizontally subsampled by a factor of 4 with respect to luma, and not subsampled vertically.
4:2:0	This confusing notation denotes chroma subsampling wherein C_B and C_R chroma components are subsampled both vertically and horizontally by a factor of 2, with respect to luma. There are two variants of 4:2:0 chroma: *interstitial* 4:2:0, used in JPEG/JFIF, H.261, and MPEG-1; and *cosited* 4:2:0, used in MPEG-2.

4:2:2	**1** Chroma subsampling wherein each C_B and C_R component is horizontally subsampled by a factor of 2 with respect to luma, and not subsampled vertically.
	2 An SD component digital video coding or interface standard, based upon BT.601, using 4:2:2 chroma subsampling, having versions for both 480*i* or 576*i* scanning. The corresponding 19 mm videotape format is denoted D-1.
4:2:2:4	A 4:2:2 system, as in 4:2:2 above (2), augmented by an opacity (or *alpha,* or *key*) component sampled at the same rate as the luma component. See *alpha,* α, on page 591.
4:3	The standard aspect ratio of SD.
4:4:4	Component digital video where $R'G'B'$ (or rarely, $Y'C_BC_R$) components are conveyed with equal data rate.
4:4:4:4	A 4:4:4 system, as above, augmented by a transparency (also known as *alpha*, or *key*) component sampled at the same rate as the *R' or Y'* component. See *alpha,* α, on page 591.
4.43 MHz	Expressed exactly, 4.433618750 MHz: the colour subcarrier frequency of 576*i* PAL-B/G/H/I video.
5-5-5	A 15 bit per pixel image format, used in low-end computer graphics, having 5 bits for each component of a pixel (i.e., red, green, and blue). Historically called *Hi-color* [sic].
5-6-5	A 16 bit per pixel image format, used in low-end computer graphics, having 5, 6, and 5 bits per component of red, green, and blue, respectively. Historically called *Hi-color* [sic].
5.1	Notation invented by Tom Holman for surround sound having five channels (front left, centre, and right; back left and right; and *low frequency effects* (LFE), the .1 in the notation.
6-way	**1** Video colour adjustment whereby hue and chroma (and possibly luminance or luma) of red, green, blue, cyan, magenta, and yellow are individually controlled. Additive colour mixing properties are easily broken. Also called *6-axis.*
	2 Video colour adjustment whereby the off-diagonal elements of a linear-light 3×3 matrix in *RGB* are altered in a linear manner.
7.5% setup	See *setup,* on page 657.
8-VSB	See *VSB,* on page 664.
10-bit Cineon/DPX	A data value, or a file containing image data values, where transmittance of exposed and developed camera negative film is transformed logarithmically to a quantity proportional to optical density, then digitized and encoded into integers ranging 0…1023. (For transport across HD-SDI, codes 0…3

	and 1020…1023 are excluded.) Data encoding is usually in accordance with SMPTE ST 268 ("DPX") *printing density*.
10-bit log	**1** 10-bit log *RGB:* Estimated *RGB* tristimulus values, transformed approximately logarithmically, then digitized and encoded into integers ranging 0…1023. See *log RGB,* on page 635. **2** 10-bit Cineon/DPX, see above.
10 GbE	10 Gb/s Ethernet; formally, 10GBASE-T.
13.5 MHz	The standard luma sampling rate for SD.
14:9	A compromise aspect ratio used in Europe during the transition to HD. Programming at 16:9 was created "14:9-safe" to allow cropping to 14:9 without harm. On 4:3 receivers, narrow top and bottom bars (less objectionable than wide bars) are displayed.
16:9	The standard aspect ratio of HD.
16–235	The standard range of 8-bit video code interface values (R', G', B', or Y') from reference black to reference white according to BT.601 and BT.709. Code values 0 and 255 are reserved for sync; 1–15 and 236–254 are permitted for transient elements.
16-way	Video colour adjustment whereby hue and chroma (and possibly luminance or luma) at 16 different hue angles are individually controlled. Additive colour mixing properties are easily broken.
18%	Reflectance factor (in the linear-light domain, relative to a perfect diffuse Lambertian reflector) of a "mid grey" test card.
23.976 Hz	A common frame rate for cinema production involving interface to SD or HD video. Expressed exactly, $^{24}/_{1.001}$.
24 Hz	The standard frame rate for cinema production and exhibition.
24 @ 25	A process whereby pictures (motion picture film, or 24*p* digital material) originated at 24 frames per second is run 4% fast and transferred to progressive video at 50 Hz frame rate (or interlaced video at 50 Hz field rate). Each film frame is scanned twice. (In interlace, one scan produces the first field and the next produces the second field.) Synonymous with *2-2 pulldown*.
24 @ 25 pulldown	The process of converting motion picture film or 24*p* digital material at 24 frames per second into a 50 picture per second video representation (e.g., 576*i*50), wherein exactly 24 frames per second are conveyed, but pictures having original 24*p* timecodes :11 and :23 are replicated. The result could be described as 2:2:2:2:2:2:2:2:2:2:2:3:2:2:2:2:2:2:2:2:2:2:2:3 pulldown; such material is unsuitable for direct viewing but is fairly easily processed.

24*a*	24 frames per second, progressive "advanced," whereby 24 fps progressive frames are subject to 2:3 pulldown in the camera for recording in 59.94 Hz or 60 Hz interlaced form, and each recording starts on an A-frame.
24*p*	24 frames per second, progressive. Preferably written *p*24 to correspond to notations such as 720*p*24 and 1080*p*24.
24*p*A	24 frames per second, progressive "advanced,"carried across a 29.97 Hz video stream. See *2-3-3-2 pulldown*, on page 582.
24PsF	24 frames per second, progressive segmented-frame. The image format is typically 1920×1080. See *PsF*, on page 648.
25 Hz	The usual frame rate for interlaced video production in Europe and Oceana.
29.94 Hz	The usual frame rate for interlaced video production in North America and Japan. Expressed exactly, $^{30}/_{1.001}$.
50 Hz	The usual frame rate for interlaced video production in Europe and Oceana.
59.94 Hz	The usual frame rate for interlaced video production in North America and Japan. Expressed exactly, $^{60}/_{1.001}$.
64–940	The standard range of 10-bit video code interface values (R', G', B', or Y') from reference black to reference white according to BT.601 and BT.709. Code values 0–4 and 1020–1023 are reserved for sync; 4–63 and 941–1019 are permitted for transient values outside the reference range.
74.176 MHz	The standard luma sampling rate (rounded to three decimal places) for HD at picture rates altered by the ratio $^{1000}/_{1001}$.
74.25 MHz	The standard luma sampling rate for HD at integer picture rates.
90%	Reflectance factor – in the linear-light domain, and relative to a perfect diffuse Lambertian reflector – of a typical white test card. (Some people prefer 89.1%, representing an optical density of 0.05 relative to a perfect diffuse reflector.)
109%	Encoded Y', R', G', or B' signal level – in the gamma-corrected, perceptual domain – at the peak white level defined in BT.601 or BT.709. Expressed exactly, $^{238}/_{2.19}$.
360*p*	A progressive video standard having image structure of 480×360; suitable for square-sampled SD content at 16:9 aspect ratio.
422P	The 4:2:2 profile of MPEG-2. (The colons are elided; the P is written in Roman uppercase.)

480*i*, 480*i*29.97	An interlaced scanning standard or image format for SD used primarily in North America, Japan, Korea, and Taiwan, having 525 total lines per frame, approximately 480 image rows (usually in an image structure of 704×480 or 720×480), and 29.97 frames per second. The notation 480*i*29.97 does not specify colour coding; colour in 480*i*29.97 systems is conveyed in the studio using $R'G'B'$, $Y'C_BC_R$, or $Y'P_BP_R$ components, and was historically encoded for transmission using composite NTSC. Historically denoted 525/59.94 (see below), or *ITU-R System M*. Often loosely referred to as *525/60*. Often incorrectly called *NTSC,* which properly refers to a colour-encoding standard, not a scanning standard.
525/59.94	Archaic notation for what is now called 480*i*29.97; see above.
540*p*	A progressive image format having an image structure of 960×540. A variant deployed by Apple is denoted *iFrame;* its frame rate (written after the *p*) is not publicly documented, but is reported to be 29.97 Hz. (I-frame compression is apparently used for this image format.)
555 format	See *5-5-5,* on page 585.
565 format	See *5-6-5,* on page 585.
576*i*, 576*i*25	An interlaced scanning standard or image format for SD used primarily in Europe, Australia, and parts of Asia, having 625 total lines per frame, 576 image rows (usually in an image structure of 720×576), and 25 frames per second. A raster notation such as 576*i*25 does not specify colour coding; colour in 576*i*25 systems is commonly conveyed in the studio using $R'G'B'$, $Y'C_BC_R$, or $Y'P_BP_R$ components, and typically conveyed through analog distribution by composite PAL. Historically denoted 625/50. Often incorrectly called *PAL,* which properly refers to a colour-encoding standard, not a scanning standard. Sometimes inaccurately called *CCIR*.
601	See *BT.601,* on page 597.
625/50	Archaic notation for what is now called 576*i*25; see above.
656	See *BT.656,* on page 598.
709	See *BT.709,* on page 598.
720*p*	A progressive image format for HD, having an image structure of 1280×720, and any of several frame rates including 23.976, 24, 29.97, 30, 59.94, or 60 Hz (or rarely, 50 Hz).
1035*i*	A developmental interlaced image format for HD, now obsolete, having an image structure of 1920×1035, a frame rate (written after the *i*) of 30.00 Hz, and nonsquare sampling.

1080*i*	An interlaced image format for HD having an image structure of 1920×1080 and a frame rate (written after the *i*) of 29.97 Hz or 30.00 Hz (or rarely, 25 Hz).
1080*p*	A progressive image format for HD, having an image structure of 1920×1080, and any of several frame rates (written after the *p*) including 23.976, 24, 25, 29.97, or 30.00 Hz, and potentially 50, 59.94, or 60.00 Hz.
1125/59.94/2:1	An interlaced scanning standard for HD, having a field rate of 59.94 Hz, and 1125 total lines per frame (of which formerly 1035, and now 1080, contain picture). The standard system with 1080 lines is now denoted 1080*i*29.97.
1125/60/2:1	An interlaced scanning standard for HD, having a field rate of 60 Hz, and 1125 total lines per frame (of which formerly 1035, and now 1080, contain picture). The standard system with 1080 lines is now denoted 1080*i*30.
1280×720	A standard image array for HD. (Aspect ratio is 16:9, 1.78:1.)
1556*p*24	A progressive image format compatible with film having an image structure of 2048×1556 and frame rate of 24 Hz or 23.976 Hz (though the latter should be written 1556*p*23.976). The count of 1556 image rows is suitable for “open gate”; 1536 image rows suffices for exact 4:3 aspect ratio (“Academy aperture”).
1920×1080	A standard image array for HD. (Aspect ratio is 16:9, 1.78:1.)
1998×1080	An image array, having the 1.85:1 aspect ratio of typical movies, commonly used for 2 K digital cinema.
2048×1080	An image array, having aspect ratio of about 1.9:1, commonly used for 2 K digital cinema.
2048×858	An image array, having aspect ratio 2.4:1 (*CinemaScope*), commonly used for 2 K digital cinema.
4096×2160	An image array, having aspect ratio of about 1.9:1, commonly used for 4 K digital cinema.
6500 K	In colour science generally, and in video and computer graphics, a white reference corresponding to CIE Illuminant D_{65}, whose correlated colour temperature is approximately 6504 K.
9300 K	In computer graphics, and in studio video in Asia, a white reference whose correlated colour temperature is 9300 K.
A-frame	In 2-3 pulldown, the first 24*p* frame in a sequence of four frames A, B, C, and D, or the corresponding video frame. The A-frame is scanned twice, to produce a first field then a second field. If nondropframe timecode HH:MM:SS:F*u* is coherent with 24*p* scanning, the A-frame produces first and second fields

	where $u=0$; the B-frame produces first and second fields where $u=1$, then a first field where $u=2$; the C-frame produces a second field with $u=2$ and a first field with $u=3$; and finally, the D-frame produces a second field where $u=3$, then first and second fields where $u=4$. The sequence continues with A (5), B (6, 7), C (7, 8), then D (8, 9).
A/VC	In IEEE 1394, audio/video control. A mechanism to control audio and video devices across an IEEE 1394 interface. Sometimes written AV/C. Not to be confused with AVC (see below).
AAC	Advanced audio coding: A lossy audio compression system defined in MPEG-2 Part 7.
ABL	Automatic brightness limiter: In a PDP display, signal processing circuitry that reduces display *R*, *G*, and *B* tristimulus values uniformly for any image where the sum of the input signal tristimuli exceeds about 25% of the peak relative luminance of the panel. For example, the luminance of a full white signal is reduced when it occupies more than about 25% of the image area. The limitation is imposed to avoid excessive power dissipation in the panel. See *power loading,* on page 648.
ABR, adaptive bit rate	Schemes capable of varying transmission bit rate depending upon available data capacity. Contrast *CBR, constant bit rate,* on page 599, and *VBR, variable bit rate,* on page 663. VBR ordinarily applies to storage media, and ABR to network delivery.
AC	**1** Alternating current: Historically, an electrical current or voltage that reverses in polarity periodically – that is, whose sign alternates periodically between positive and negative. **2** In modern usage, a signal whose value varies periodically. Distinguished from *DC, direct current;* see page 611. **3** In JPEG and MPEG, any or all DCT coefficients in an 8×8 block apart from the DC coefficient.
AC-3	Originally, *Audio Coding 3:* Dolby Labs' designation of a digital audio compression standard incorporated into the ATSC HDTV system. AC-3 is now called *Dolby Digital*.
Academy	Academy of Motion Picture Arts and Sciences (AMPAS).
Academy aperture	1.375:1 aspect ratio, standardized by the Academy in 1932.
accuracy	The degree of closeness of a measurement to a true value. Distinguished from *precision* (page 648).
ACES	Academy Color Encoding Specification.
achromatic	Without hue: In digital imagery, lightness-related information (or component) only.

active	**1** Historically, a signal element (sample, or image row, or in analog systems, scan line) defined by an image representation or scanning standard to contain part of the picture or its associated blanking transition. In a digital representation (such as 1920×1080), all of the indicated image matrix elements are "active." The term *visible picture element* is an oxymoron: If an element is in the picture, it's intended to be visible! (Exceptionally, in 480*i*29.97, line 21 closed caption data is considered to be active, despite the fact that it does not contain picture.) **2** In the context of *active format description* (AFD, see below), *active* refers to portion of the pixel array that contains the image intended to be viewed by the consumer, not including any fixed content surrounding the picture that is necessary to meet the aspect ratio requirements of the container.
Adobe RGB (1998)	A colour exchange standard promulgated by Adobe and used in graphic arts and professional digital photography, having wider gamut than BT.709.
ADU	Analog-to-digital unit: The digital number produced by analog-to-digital conversion of an sensor signal. Sometimes denoted DN (digital number) or DCV (digital code value). For a conventional linear sensor such as a CCD, ADU and DN are linear-light measures; DCV may incorporate nonlinear coding.
AES3	An interface, for professional use, for uncompressed digital audio, standardized for several physical interfaces including balanced twisted-pair, unbalanced coax, and optical fibre. Standardized in IEC 60958 Part 4.
AFD, active format description	Metadata (defined in SMPTE ST 2016) accompanying video to indicate the portion of the raster that contains the image intended to be viewed by the consumer.
albedo	In physics, average diffuse reflectance across the visible spectrum (and perhaps extending into the infrared and ultraviolet).
alpha, α	In computer graphics, a component of a pixel indicating the opacity – conventionally between black (0, fully transparent) and white (1, fully opaque) – of the pixel's colour components. Colour component values ($R'G'B'$, $Y'P_BP_R$, or $Y'C_BC_R$) may have been premultiplied by the value of a corresponding alpha value; this is sometimes called *shaped video*. Colour component values that have not been so premultiplied are *unassociated* (or *unshaped*, or *nonpremultiplied*). See also *key,* on page 631.
alychne	"Absence of lightness line": In a chromaticity diagram, the locus of coordinates having zero luminance. In the CIE [x, y] diagram, the x-axis.
anamorphic	A subsidiary format (or its associated lens) standing in relation to a base format having relatively narrow aspect ratio, wherein the horizontal dimension of a widescreen image is squeezed by

	some factor with respect to the horizontal dimension of the base format. In cinema, the widescreen (anamorphic) image conventionally has 2.4:1 aspect ratio and the squeeze is by a factor of 2. In video, the widescreen (anamorphic) image has 16:9 aspect ratio, and the squeeze is typically by a factor 4⁄3 .
ANC	In SDI and HD-SDI, ancillary (nonessence) information typically conveyed during vertical and horizontal blanking intervals.
anchor picture	See *reference picture,* on page 651.
ANSI lumens	Total light output of a projector measured in accordance with ANSI/NAPM IT7.228 (withdrawn in 2003). To achieve 48 nt white on a perfect unity-gain screen, luminous flux of $48 \cdot \pi$ lumens is required for each square meter of screen area.
APC	Automatic power control: A signal processing mechanism common in plasma display panels whereby the video level is reduced in order to limit total power to the display surface. See *loading, luminance,* on page 635.
APL, average picture level	A historical term, now ambiguous: **1** Traditionally in video, APL is equivalent to average pixel level, see below. **2** Average relative luminance. This is a linear-light measure unlike average pixel level. Properly termed *average relative luminance* (ARL).
APL, average pixel level	The average of luma (Y') throughout the image area of a frame, sequence, scene, or program. Average *pixel* level is preferred to the historical term *average picture level* for disambiguation, to make clear that it is gamma-corrected pixel values (not their luminance or tristimulus equivalents) that are averaged.
ARL, average relative luminance	The average of luminance (Y) throughout the entire image area of a frame, sequence, scene, or program. ARL is a linear-light measure (unlike average pixel level).
ARS, adaptive rate streaming	Streaming (network) delivery wherein bit rate potentially varies depending upon network loading. See also *ABR, adaptive bit rate,* on page 590.
ASI	Asynchronous serial interface: An industry standard electrical interface, standardized by DVB, used to convey an MPEG-2 transport stream.
aspect ratio	The ratio of the width of an image to its height. (Some authors, such as MPEG, write this improperly as height:width.)
aspect ratio, sample (or pixel)	The ratio of horizontal sample pitch to vertical sample pitch.
ATSC	Advanced Television Systems Committee: A U.S.-based organization that standardizes and promotes digital SDTV and HDTV

	broadcasting. ATSC advocates MPEG-2 video compression and Dolby Digital (AC-3) audio compression, supplemented by ATSC terrestrial broadcasting transmission standards. Note that *Systems* is plural (contrary to the singular *System* in NTSC).
AV	Audiovisual: Electronic technology to capture, process, record, transmit, and/or present moving pictures along with associated sound.
AVC	Originally, advanced video compression: The effort within IEC, ISO, and ITU-T that produced the H.264 standard (see page 623). Now more clearly expressed as *H.264.*
AVC-Intra	A subset of H.264 video coding (see page 623) for professional applications, using I-frame only compression of 720*p*,1080*i*, or 1080*p* video, typically at data rates of 50 Mb/s (using the Hi10PIntra profile) or 100 Mb/s (using the Hi422PIntra profile).
AVCHD	A consumer HD system for 720*p*, 1080*i*, and 1080*p*24, adapted for professional use, typically using 12 cm DVD-R media, SDHC flash memory cards, or hard disk drive recording, using long-GoP H.264 video coding, Dolby Digital audio coding, and having a bit rate between about 6 Mb/s and 18 Mb/s.
AVCCAM	Panasonic's adaptation of AVCHD (see above) to professional markets.
average luminance	The mean (CIE, linear-light) luminance across the image area of a frame, sequence, scene, or program. The term is often used to refer to what is more properly called average *relative* luminance, that is, computed relative to the luminance of reference white. Not to be confused with *APL* (see above), which averages a gamma-corrected (perceptually uniform) signal.
average pixel level	See *APL*, *average pixel level,* on page 592.
AVS	Audio and Video coding Standard: A collection of standards of the People's Republic of China (PRC) relating to compression of audio and video. The standards specify a system comparable to H.264 but generally avoid its intellectual property.
AWGN, additive white Gaussian noise	Noise that is additive (i.e., has the same amplitude across the whole range of values of the associated signal), white in the frequency sense (that is, having power uniformly distributed across the frequency spectrum), and Gaussian (i.e., has a probability distribution function that follows the statistical "normal" curve). Some kinds of noise are well characterized as AWGN (e.g., sensor read noise); other kinds are not (e.g., photon shot noise).
b, bit	Binary digit: The elemental unit of information, valued either 0 or 1 (sometimes interpreted as false/true, no/yes, off/on, etc.).
B, byte	Byte: An ordered collection of eight bits, capable of representing whole numbers 0 through 255 (i.e., 0 through 2^8-1),

	integers from –128 to +127 in two's complement form, other number systems, or characters in a variety of encodings.
B-field	In MPEG, a field-coded B-picture. B-fields come in pairs (either top then bottom, or bottom then top). See *B-picture,* below.
B-frame	In MPEG and related standards, either a frame-coded B-picture, or a pair of B-fields (one top field and one bottom field, in either order). See *B-picture,* below.
B-picture	In MPEG, a bidirectionally predictive-coded picture: A picture, or coded picture information, in which one or more macroblocks involve prediction from a preceding or a following anchor picture. B-pictures exploit temporal coherence. They are computed and displayed, but do not form the basis for any subsequent predictions.
b/s	Bits per second; the *b* is preferably set in lowercase.
B/s	Bytes per second; the *B* is preferably set in uppercase.
$B'-Y'$, $R'-Y'$	A pair of colour difference components, B' minus luma and R' minus luma. Following decoding of Y', $B'-Y'$, and $R'-Y'$, the resulting red, green, and blue components ($R'G'B'$) are subject to an EOCF to produce tristimulus (linear-light). $B'-Y'$, and $R'-Y'$ colour differences may be scaled to form C_B and C_R for component digital systems, scaled to form P_B and P_R for component analog systems, or scaled to form U and V (or, in specialized forms of NTSC, I and Q) for composite encoding.
band-interleaved by line (BIL)	A method of storing pixels whereby like components of an image row occupy adjacent storage locations. A 3×2 8-bit *RGB* image could be stored in the order RRRGGGBBBRRRGGGBBB. See also *interleaved, component interleaved,* on page 628, and *planar,* on page 647.
bandwidth	**1** Technically, the frequency or frequency range where an analog or digital signal's power has fallen 3 dB – that is, to 0.707 – from its value at a reference frequency (usually zero frequency, *DC*). Equivalently, the frequency or frequency range where an analog or digital signal's amplitude has fallen 3 dB – that is, to 0.707 – from its value at a reference frequency. **2** In common language, *data rate;* see page 610.
Bayer pattern	The mosaic pattern (see page 639), named for Kodak researcher Bryce E.Bayer, comprising a 2×2 arrangement of photosites or pixel components representing red, green, green, and blue.
BD	Blu-ray disc.
BER	Bit error ratio: The probability that recording or transmission in an error-prone medium corrupts any single bit recorded or transmitted. Sometimes incorrectly expressed as bit error *rate,*

		which properly refers to the *rate* of occurrence of erroneous bits.
Betacam		Sony's trademarked term for a professional component analog videotape format for 480*i* or 576*i* on $\frac{1}{2}$-inch tape. The successor system, with higher bandwidth, is denoted Betacam SP. See also *Digital Betacam, Digital-β,* on page 612.
bias	**1**	In signal processing in general, an additive term contributing to a signal value; offset.
	2	In display systems, a low-level adjustment – traditionally necessary for CRT displays owing to analog drift, and now thoughtlessly replicated in many fixed-pixel displays – to set *RGB* biases individually on a component-by-component basis. (Black level or brightness sets bias on all components together.) Sometimes called cutoff or offset; in the home theatre community, sometimes called *RGB*-low. See also *drive* (**2**) on page 614.
	3	In the home theatre community, surround lighting.
bias light	**1**	An obsolete scheme used in certain tube-type video cameras and telecines to uniformly illuminate the sensor with a low level of illumination to ameliorate lag.
	2	In the home theatre community, surround lighting.
bilevel		A system where each pixel contains one bit representing either full black or full white; no shades of grey are possible. In computer graphics, such imagery or equipment is traditionally called *monochrome,* but that usage is misleading because it conflicts with the colour science definition of *monochrome.*
bilevel sync		Analog sync information conveyed through a single pulse having a transition from blanking level to a level more negative than blanking (synctip level), then a transition back to blanking level. In analog systems, synctip level is either $-285\frac{5}{7}$ mV or –300 mV. Bilevel sync is used in SD; distinguished from trilevel sync (see page 662), standardized for HD.
bit error rate		See *BER,* above.
bit splitting		In a pulse-duration modulated display such as a PDP or DLP, video codes are converted to tristimuli (linear-light) values, thence to pulses of light having durations that are small multiples (typically $^{1}/_{1000}$ or less) of the frame time. Bit splitting refers to the assignment of linear-light values to time intervals.
BITC		Burnt-in timecode: Timecode in visual form overlaid (keyed) onto picture content, so as to be human readable (but contaminating the picture, and not ordinarily machine readable).
black level		The level representing black: nominally 7.5 units for analog System *M* outside Japan and zero in other systems. See also *pedestal,* on page 645, and *reference black,* on page 651.

BLACK LEVEL	User-accessible means to adjust black level, traditionally by imposing the same additive offset to each of the gamma-corrected *R'G'B'* components. The term BLACK LEVEL is preferred to BRIGHTNESS.
black-to-white excursion	The excursion from reference black to reference white. Conventionally 92.5 units ($^{37}/_{56}$ V, approximately 660 mV) for System M, 100 units ($^{5}/_{7}$ V, approximately 714 mV) in NTSC-J, 700 mV in other analog systems, and codes 16 through 235 at an 8-bit component (BT.601 or BT.709) digital interface.
blanking interval	The time interval – in the vertical domain, the horizontal domain, or both – during which a video signal is defined by an interface standard not to contain picture. Ancillary signals such as VITC may be conveyed during blanking.
blanking level	Zero level; 0 units by definition. Identical to reference black level (see page 651) except in System M with setup.
block	In JPEG, M-JPEG, and MPEG, an 8×8 array of samples, or coded information representing them.
Blu-ray	A set of standards for data recording on and playback from optical media, typically HD, using shortwave (blue light) laser.
BNC	Bayonet Neill-Concelman (contrary to the entry in the *IEEE Standard Dictionary of Electrical and Electronics Terms*): A coaxial connector, now standardized in IEC 169-8, used in video. Paul Neill, working at Bell Telephone Laboratories, developed a threaded connector adopted by the U.S. Navy and named the N connector, after him. Carl Concelman, working at Amphenol, came up with a bayonet version (slide on and twist), called the C connector. The two collaborated on a miniature version, which became the BNC. A screw-on relative, the threaded Neill-Concelman connector, is the TNC. [Mark Schubin/*Videography*]
BOB	Break-out box: A panel of connectors remote from the associated equipment, ordinarily attached with a proprietary cable.
bob	A deinterlacing technique, common in PC video, where vertically adjacent samples in a single field are averaged to create synthetic intermediate image rows, effecting a crude form of interlace-to-progressive conversion. See also *weave* (sense **2**), on page 665.
bottom field	In MPEG, the field that contains the bottom coded image row of a frame; typically the first field in 480*i* and the second field in 576*i*.
bpc, bits per component (or bits per channel)	The number of bits allocated to each colour component (channel) of an image – that is, the "bit depth" per colour component (or sample). It is implicit that all colour components have the same number of bits. An image may comprise more than three components.

bpp, bits per pixel	The number of bits allocated to each pixel; in an uncompressed representation, the "bit depth" of the pixel. It is implicit that each pixel contains a complete set of colour components: The term is not directly applicable to chroma-subsampled representations, although it is used sometimes to give the effective (average) number of bits. A pixel may have more than three components.
BRCR	Bright room contrast ratio: Contrast ratio measured under relatively high ambient illuminance, higher than about 100 lx. Compare *DRCR,* on page 614.
brightness	*The attribute of* a *visual sensation according to which an area appears to emit more or less light* [CIE]. Brightness is, by definition, subjective: It cannot be measured or quantified, and so is inappropriate to describe pixel values. *Luminance* (see page 636) is a related objective quantity; loosely speaking, brightness is apparent luminance. Brightness is absolute, not relative; unlike *lightness* (see page 633), it is not expressed relative to any reference level.
BRIGHTNESS	User-accessible means to adjust *black level* (page 595). The term BLACK LEVEL is preferred to BRIGHTNESS. (See also *brightness,* above.)
brightness, photometric	Archaic, deprecated term for luminance (see page 636).
brightness, radiometric	Archaic, deprecated term for radiance (see page 650).
broad pulse	In analog video, a pulse – part of the vertical sync sequence – that remains at sync level for substantially longer than normal line sync and indicates vertical sync.
BT.601	**1** Formally, ITU-R Recommendation BT.601: The international standard for studio digital video sampling for SD. BT.601 specifies a sampling frequency of 13.5 MHz (for both 480*i*29.97 and 576*i*25 SD), $Y'C_BC_R$ coding, and this luma equation: $^{601}Y' = 0.299\,R' + 0.587\,G' + 0.114\,B'$ BT.601 is silent concerning *RGB* chromaticities. It is implicit that 480*i* systems use SMPTE RP 145 primaries, and that 576*i* systems use EBU Tech. 3213 primaries. BT.601 is silent concerning encoding gamma. BT.601 specifies "studio swing"; see page 659. **2** Loosely, in computer graphics, an interface for BT.601-style digital SD video stream where synchronization is accomplished using separate horizontal and vertical drive logic signals. The notations "ITU 601" and "Rec. 601" are ambiguous – and should be avoided – because ITU-T has an unrelated recommendation G.601 that could be called ITU 601 or Rec. 601.

BT.656	**1** Formally, *ITU-R Recommendation BT.656*. The international standard for parallel or serial interface of BT.601 digital video SD signals. **2** Loosely, in computer graphics, an interface for BT.601-style digital SD video stream where synchronization is accomplished with TRS (SAV and EAV) sequences embedded in video data.
BT.709	Formally, *ITU-R Recommendation BT.709:* The international standard for studio digital video sampling and colour encoding for HD. Chromaticity and transfer function parameters of BT.709 have been introduced into modern studio standards for 480*i* and 576*i*. BT.709 specifies this luma equation (whose coefficients are unfortunately different from the BT.601 coefficients of SD): $^{709}Y' = 0.2126\,R' + 0.7152\,G' + 0.0722\,B'$ The notations "ITU 709" and "Rec. 709" are ambiguous, and should be avoided, because ITU-T has an unrelated recommendation G.709 that could be called ITU 709 or Rec. 709.
BT.1886	Formally, *ITU-R Recommendation BT.1886*. The international standard defining the EOCF of studio reference displays for HD.
BTB	Blacker than black. See *superblack,* on page 659.
BTW	Black-to-white: A nonstandard measure of the transition time ("response time") of a display (often an LCD) from black level to white level. See also, *GTG,* on page 623.
burn, burn-in	**1** To impose permanently upon image data – for example, by keying – auxiliary information such as timecode. **2** To manipulate image data values so as to permanently impose a colour interpretation. Also known as *bake.* **3** Permanent aging (evident as loss of brightness) in certain types of display (e.g., PDP) when stationary bright or colourful elements are displayed for a long time (days or weeks). **4** Significant alteration of performance observed in the first few hours, days, or weeks of operation of a new device, or the act of operating a new device for such a period of time.
burst	A brief sample of eight to ten cycles of unmodulated colour subcarrier inserted by an NTSC or PAL encoder onto the back porch of a composite video signal. Burst enables a decoder to regenerate the continuous-wave colour subcarrier.
C_B, C_R	**1** Versions of colour difference components B'–Y' and R'–Y', scaled and offset with "studio swing" for digital component transmission. At an 8-bit interface, C_B and C_R have excursion 16 through 240. See also [B'–Y', R'–Y'], [P_B, P_R], [U, V], and [I, Q]. In systems using BT.601 luma, such as 480*i* and 576*i*, it is

	standard to apply these scale factors (and interface offsets, shown below in grey) to B'–Y' and R'–Y': $C_B = 128 + 112\frac{1}{0.886}(B' - Y'); \quad C_R = 128 + 112\frac{1}{0.701}(R' - Y')$ In HD systems using BT.709 luma, such as 1280×720 and 1920×1080, it is standard to apply the following scale factors (and offsets) to B'–Y' and R'–Y': $C_B = 128 + 112\frac{1}{0.9278}(B' - Y'); \; C_R = 128 + 112\frac{1}{0.7874}(R' - Y')$ **2** Versions of colour difference components B'–Y' and R'–Y', scaled to “full-swing” or “full-range” (±128, with code +128 clipped) for use in stillframe JPEG/JFIF. It is usual to apply the following scale factors (and offsets) to B'–Y' and R'–Y': $C_B = 128 + 128\frac{1}{0.886}(B' - Y'); \quad C_R = 128 + 128\frac{1}{0.701}(R' - Y')$
C/PH	Cycles per picture height: A unit of resolution corresponding to a cycle (a black and a white element, comparable to a line pair) across the height of the picture.
C/PW	Cycles per picture width: A unit of resolution corresponding to a cycle (a black and a white element, comparable to a line pair) across the width of the picture.
cadence	In a motion image sequence having 2-3 pulldown (or some variant thereof), the property of having strictly periodic A-frames. Careful editing preserves continuous cadence; careless editing disrupts it.
calibration (colour)	Modification of the colour reproduction parameters of a particular device – effected either within the device itself, or within associated equipment – to bring the device into conformance with an absolute reference associated with a standard or exemplified by a reference device. Distinguished from *characterization (colour)* (see page 601), which is passive with respect to the device.
candela [cd]	The SI unit for luminous intensity; one of the seven base SI units. See *intensity* (**1**, on page 628).
CBR, constant bit rate	Any compression format in which the bit (or byte) count of compressed data is the same from one second to the next.
CC	**1** Colour correction; see page 604. **2** Closed caption; see page 603.
CCA	Means for a display user (or technician) to alter the effective chromaticity of red, green, and blue by building the appropriate 3×3 matrix on the fly. The term *CCA* refers to various approaches offered by various manufacturers. *Comprehensive*

	Color Adjustment (CCA), is trademarked by Christie Digital, but other vendors use terms such as *Color Coordinate Adjustment.*
CCD	Charge-coupled device. In modern use of the term, a microelectronic image sensor constructed using MOS technology. A CCD converts incident photons to electrons; the electrons are collected and then transported to an on-chip converter and amplifier then output from the device.
CCIR	**1** *Comité Consultatif Internationale des Radiocommunications* (International Radio Consultative Committee): A treaty organization, as of 1993 renamed ITU-R.
	2 Sometimes incorrectly used to denote 576*i* scanning.
CCIR Rec. 601, Rec. 709	Obsolete designations, now properly referred to as ITU-R Rec. BT.601 (colloquially, BT.601), or ITU-R Rec. BT.709 (colloquially, BT.709). See *BT.601* and *BT.709,* on page 598.
CCO	Centre cut-out, see below.
CCT	Correlated colour temperature: The temperature (in units of kelvin, K) of the point on the blackbody radiator's locus where a line drawn on a CIE 1960 [*u, v*] diagram from a colour stimulus intersects the blackbody curve perpendicularly. If that sounds obscure, it is: If a source's chromaticity lies near the blackbody curve, then CCT gives a reasonable expression of its chromaticity, but ordinarily chromaticity coordinates should be used instead.
CCTV	**1** Closed circuit television. An archaic term denoting video systems used in nonbroadcast applications such as security.
	2 China Central Television (in the Peoples' Republic of China).
CCU	Camera control unit: A device that enables remote control of the iris of a camera lens and basic camera settings, most importantly gain, white balance, and pedestal.
$\mathrm{cd}\cdot\mathrm{m}^{-2}$	Candela per meter squared: the SI unit for absolute luminance.
CE	Consumer electronics.
centre-cut, centre cut-out	A widescreen image that has been cropped to 4:3 aspect ratio.
CFA	Colour filter array. A regular arrangement of coloured filters placed over neighbouring photosites in an image sensor. The Bayer structure (see page 594) is commonly used, but other structures (such as interleaved red, green, and blue vertical stripes placed over successive photosite columns) are also used.
CGI	Computer-generated imagery: Synthetic image data, generated by computation (as opposed to being acquired from a physical scene by a camera).

characterization (colour)	Modelling, measurement, and/or estimation of colour parameters of a particular device, class of devices, or subsystem. Distinguished from *calibration (colour)* (see page 599).
chroma	**1** In colour science, *colourfulness of an area judged in proportion to the brightness of* a *similarly illuminated area that appears to be white or highly transmitting* [CIE].
	2 Generally, a component or set of components such as [C_B, C_R] that conveys colour independent of luma or luminance.
	3 In component video, colour independent of (or accompanied by) luma, conveyed as a pair of colour difference signals such as [C_B, C_R], or [P_B, P_R].
	4 In composite video, colour subcarrier modulated using the NTSC or PAL technique by two colour difference components [*U, V*] to form a *modulated chroma* signal, *C.*
	5 In video, the polar-coordinate modulus (radius) of a colour difference pair in C_B, C_R coordinates for component digital video or P_B, P_R coordinates for component analog video.
	6 User-accessible means to adjust colour *saturation* (**5,** on page 655), sometimes called CHROMA GAIN. This adjustment is often called COLOUR; however, that term fails to make clear whether *which* colour (i.e., hue) or the *amount* of the colour (i.e., chroma) is affected. To avoid ambiguity, use HUE for the former and CHROMA for the latter.
CHROMA GAIN	User-accessible means to adjust colour *saturation* (**5**); preferably called CHROMA.
CHROMA PHASE	**1** Phase of modulated NTSC or PAL chroma.
	2 User adjustment of *chroma phase* (**1**); preferably called HUE, see *hue* (**4**).
chromakey	An archaic technique of matte extraction based upon *hue* (**3,** on page 626) – and possibly *chroma* (**5,** on page 601) – of foreground video, followed by keying.
chromaticity	**1** Specification of colour in terms of CIE [*x, y*] or [*u′, v′*] coordinates – that is, in terms of a projective transform of tristimuli.
	2 Loosely, the *chromaticity* (**1**) of the red, green, and blue primaries, and the chromaticity of the white reference, of a video system.
chrominance	**1** Formally, the colour of a scene element or image element independent of its luminance; usually expressed in the form of CIE [*x, y*] chromaticity.
	2 Loosely, *chroma;* see above.

CIE	*Commission Internationale de L'Éclairage* (International Commission on Illumination): The international standards organization that sets colorimetry standards.
CIE D_{65}	The standard spectral radiance (SPD) or chromaticity of white, representative of northern daylight and having a colour temperature of approximately 6504 K. See *reference white,* on page 651.
CIE luminance, CIE Y	See *luminance (page 636)*. A qualifier *CIE* or *linear-light* is sometimes used to emphasize that the associated quantity is representative of tristimulus value proportional to intensity, as opposed to the nonlinear, gamma-corrected quantity *luma.*
CIF, common image format	**1** Historically, the elusive goal of a single (common) worldwide standard pixel array for digital video, perhaps at different frame rates, not achieved in SD standardization, but achieved for HD as 1920×1080. Confusingly, the acronym collides with *CIF, common intermediate format (see below).* **2** In MPEG-2 or other video compression systems excepting those for videoconferencing, an image format of either 720×480 or 720×576. The term is an oxymoron, since for SD there is not a single ("common") format, but rather two different formats. (In the deliberations that led to digital SD studio standards, agreement was not reached upon a common image format!)
CIF, common intermediate format	In ITU-T Rec. H.261 and related standards, a progressive 352×288 image format with 4:2:0 chroma subsampling, a frame rate of 29.97 Hz, and a sample aspect ratio of 12:11 (width:height). CIF image data is ordinarily subsampled from SD. The format is a compromise derived from the image structure of 576*i*25 and the frame rate of 480*i*29.97. Distinguished from *CIF, common image format*, above. See also *QCIF,* on page 649.
CinemaScope	Material intended for display at 2.4:1 aspect ratio.
Cineon	A project and set of products from Kodak, discontinued in 1994, that started the digital cinema revolution. The legacy of the project lives on in *Cineon printing density, CPD* (see below).
Cineon printing density, CPD	Colour data metric representing optical density values measured from (or approximating) exposed and developed colour-negative photographic film as 10-bit log data.
clamp (*v.*)	**1** Imposition (by addition or subtraction) of a DC offset (bias) onto a signal, so as to place a certain signal feature (such as back porch) at a specific level (such as blanking level, 0 IRE). **2** Commonly but incorrectly used to indicate *clip;* see below.

clean aperture	The specified or standardized rectangular portion of the pixel array that remains subjectively free from intrusion of artifacts resulting from filtering of the picture edges.
clip (*v.*)	A process of forcing a signal not to exceed a certain maximum level (or not to fall below a certain minimum level).
clipped (*adj.*)	A signal that has been limited to a certain maximum (or minimum) level.
clone (*n*. or *v.*)	Digital copy with no degradation from a master. See also *dub.*
closed caption (CC)	Digital data conveying textual information that can be decoded and displayed for the benefit of hearing-impaired viewers. In 480*i* NTSC analog video, closed caption data is inserted into line 21. (Unlike other vertical interval signals, NTSC line 21 is classified as active picture video.)
CLUT	Colour lookup table: A LUT (see page 637) that maps from a set of integer pixel values (often 0 through 255) to $R'G'B'$ triplets (encoded red, green, and blue component values). Each $R'G'B'$ value (typically 8 bits) is proportional to the $^{1}/_{2.2}$-power of the associated display tristimulus value. See *pseudocolour.*
CM	In DV compression, a coded (compressed) macroblock.
CMF	See *colour matching functions*, on page 605.
CMOS	Complementary-symmetry metal-oxide-semiconductor (pronounced *see-moss*): The most common type of analog or digital microelectronic semiconductor technology.
CMOS image sensor	An image sensor constructed using CMOS technology. Each photosite has a photodiode that converts incident photons to electrons, and a small number of transistors (typically 3 or 4) that amplify and gate the corresponding voltage to an amplifier and/or analog-to-digital converter for subsequent transmission off-chip.
CMY[K]	Cyan, magenta, yellow: The traditional primaries in subtractive colour reproduction. For halftone printing, CMY are usually augmented by K (historically, *key;* in modern terms, black).
codec	Coder/decoder: Hardware circuitry, firmware, software, or equipment to encode or decode data between two formats (perhaps between analog and digital, or between two digital formats), often including signal compression or decompression.
coherence, frequency	The property whereby two or more periodic signals are phase-locked to a common reference frequency. The unmodulated colour subcarrier of a studio-quality NTSC or PAL composite video signal is coherent with its sync.

coherence, spatial	In a single image, the property whereby adjacent samples have values that are correlated.
coherence, temporal	In a motion image sequence, the property whereby corresponding samples in successive images, perhaps subject to spatial displacement of moving image elements, are correlated.
COFDM	Coded orthogonal frequency-division multiplexing: An RF modulation system using a large number of subcarriers to spread the information content of a signal evenly across a transmission channel. The subcarriers of COFDM are individually modulated, typically using QPSK or QAM. COFDM is used in DVB-T.
COLOR/COLOUR	User-accessible means to adjust *chroma* (**5**). The term COLOUR fails to make clear whether *which* colour (i.e., hue) or the *amount* of the colour (i.e., chroma) is affected. To avoid ambiguity, use HUE for the former and CHROMA for the latter.
colorimetry	**1** Formally, the science of measuring colour, especially as standardized by the CIE. Even Canadians spell this word without *u.* **2** In video, *colorimetry* (**1**) as above, augmented by concerns outside the domain of classical colorimetry, including some or all of the following: the parameters of the opto-electronic conversion function (OECF) applied to the linear-light (tristimulus) *RGB* components of classical colorimetry to form *R'G'B'*; the parameters of the 3×3 matrix transform applied to *R'G'B'* to form luma and two colour difference components; the inverse of that matrix; and the parameters of the electro-optical conversion function (EOCF) applied to *R'G'B'* signals to form linear-light *RGB* (tristimulus) values at the display.
colourburst	See *burst,* on page 598.
colour correction	This term can refer to either of two processes which are, in a well-designed imaging system, separate. If the first process is omitted (owing to being absent from the camera design, or having been disabled by the camera operator), done poorly by the camera manufacturer, or done poorly owing to misadjustment by the camera operator, then the second will be difficult: **1** The process – ordinarily included in an SD or HD camera where it is called *linear matrixing,* but potentially implemented in postproduction or DI – by which linear-light sensor signals are processed by a 3×3 linear matrix having coefficients carefully chosen such that the resulting signals have a specific colorimetric interpretation (for example as additive *RGB* with a specific set of primaries and a specific reference white chromaticity). **2** The process, ordinarily carried out in postproduction by a skilled person called a *colourist,* of altering image colour component values in order to implement the visual æsthetic goals of a program intended to be coded into a standard colour

	image interchange space and subsequently viewed on a standardized display (ordinarily characterized as additive *RGB* with a specific set of primaries, a specific reference white chromaticity, and a standard EOCF) in a known environment.
colour correction, primary	Colour correction (**2**), where colour alterations are effected uniformly across all pixels in the image ("globally"), and where the alteration is limited to adjustment – either to all components at once ("master"), or to individual *RGB* components – of black level, power function exponent (gamma), and gain. (These three controls are often called "lift, gamma, and gain.")
colour correction, secondary	Colour correction (**2**), where colours are altered selectively by luminance or luma range, by colour range, by spatial extent, or by any combination of these.
colour difference	**1** A numerical measure of the perceptual distance between two colours; for example, CIE ΔE^*_{uv} ("delta-*E*"). **2** A signal that vanishes – that is, becomes identically zero – for pure luma without colour. A video system conveys a colour image using a set of three signals: a luma signal (Y') and a pair of colour difference signals. Spatial filtering may be applied to reduce the information rate of the colour difference components without perceptible degradation. The usual colour-difference pairs are $[B'-Y', R'-Y']$, $[C_B, C_R]$, $[P_B, P_R]$, and $[U, V]$.
colourfulness	*The attribute of* a *visual sensation according to which an area appears to exhibit more or less of its own hue* [CIE].
colour management	**1** Generally, and specifically in graphics arts and desktop publishing, techniques that use colour profiles to quantify the colour reproduction characteristics of colour devices, and facilities to impose colour transforms based upon the profiles with the goal of achieving predictable colour. **2** In home theatre displays, facilities to adjust chromaticity, luma, and/or luminance of primary colours (red, green, and blue) and secondary colours (cyan, magenta, and yellow) independently. If a display correctly implements additive colour reproduction, so-called colour management is antagonistic to accurate display. If a display does not correctly implement additive colour reproduction, there is no assurance that colour management facilities can be used to bring it into conformance.
colour matching functions	A set of three functions across the wavelength interval 380 nm to 780 nm, closely associated with the CIE Standard Observer, that weights a spectral power distribution (see page 658) and yields a 3-vector comprising tristimulus values (see page 662).
colour saturation	*The colourfulness of an area judged in proportion to its brightness* [CIE]. Subjective, by definition. Saturation runs from neutral grey through pastel to saturated colours. Roughly speaking, the more an SPD is concentrated at one wavelength, the more saturated the associated colour becomes. A colour

	can be desaturated by adding light that contains power distributed across a wide range of wavelengths.
COLOUR SATURATION	User-accessible means to adjust colour saturation. Preferably called CHROMA; see *chroma* (**6,** on page 601).
colour standard	The parameters associated with encoding of colour information – for example, $R'G'B'$ or $Y'C_BC_R$ *component* video standards, or historical NTSC or PAL *composite* video standards. Distinguished from *scanning standard* (see page 656).
colour subcarrier	**1** A continuous sinewave signal at about 3.58 MHz or 4.43 MHz used as the basis for quadrature modulation or demodulation of *chroma* (**2,** on page 601) in an NTSC or PAL composite video system. See also *burst,* on page 598. **2** Colour subcarrier (see above), onto which two colour difference signals have been imposed by quadrature modulation. Properly, modulated chroma.
colour temperature	**1** Characterization of an illuminant or a white reference in terms of the absolute temperature (in units of kelvin) of a blackbody radiator having the same chromaticity (in 1960 [*u*, *v*] coordinates). **2** User- or technician-accessible adjustment of *colour temperature* (**1**), by which *RGB* gains are adjusted to achieve the intended chromaticity of reference white.
comb filter	**1** Generally, a filter having magnitude frequency response with periodic equal-magnitude maxima and equal-magnitude minima. **2** In video, a *comb filter* (**1**) incorporating delay elements with line, field, or frame time duration. **3** In a composite NTSC video decoder, circuitry incorporating one or more line delay elements (linestores) to exploit the frequency interleaving of modulated chroma to separate chroma from luma. A comb filter provides better separation than a notch filter, owing to its suppression of *cross-colour* and *cross-luma* artifacts. A *3-D comb filter* incorporates at least one fieldstore. **4** In a composite PAL video decoder, circuitry incorporating a line delay element (linestore) to separate the modulated *U* chroma component from the modulated *V* chroma component.
component (*adj.*)	In video, a system that conveys three colour values or signals independently, free from mutual interference. Examples are $R'G'B'$ and $Y'C_BC_R$. Distinguished from *composite (adj.)*, below.
component (*n.*)	**1** Generally, a device or a piece of equipment. **2** In mathematics or signal processing, one element of a vector.

	3 One value or signal from the set of three necessary to completely specify a colour. **4** A value, channel, or signal – such as transparency or depth – that is spatially associated with image data, or temporally associated with an image in a sequence, that does not contribute to the specification of colour.
component analog	An analog video system (as opposed to digital) using $R'G'B'$ or $Y'P_BP_R$ component colour coding (as opposed to using composite colour coding such as NTSC or PAL).
component digital	A digital video system (as opposed to analog), using $R'G'B'$ or $Y'C_BC_R$ component colour coding (as opposed to using composite colour coding such as NTSC or PAL). Component digital SD systems are sometimes called "4:2:2," though the latter notation strictly refers to just the colour subsampling, not any of the other encoding parameters.
composite (*adj.*)	Combined, as in combined vertical and horizontal sync elements [see *composite sync*, below]; combined luma and chroma [see *composite video (**1**)*, below]; or combined video and sync [see *composite video* (**2**), below].
composite (*v.*)	To combine images by layering, keying, matting, or a similar process usually performed as $R = \alpha \cdot FG + (1-\alpha) \cdot BG$, where *FG* represents foreground (*fill*) image or video data and *BG* represents background. Foreground image data that has been premultiplied by the key is called *shaped* in video (or *associated,* or, in computer graphics, *premultiplied*). Foreground image data that has not been premultiplied by the key is called *unshaped* in video (or *unassociated,* or, in computer graphics, *nonpremultiplied*). Image data may be represented in linear-light form (typical of D-cinema DI and postproduction) or in gamma-corrected form (typical of studio video).
composite digital	A digital video system (as opposed to analog), using composite colour coding such as NTSC or PAL (as opposed to using component colour coding such as $Y'C_BC_R$). All standard composite systems sample at four times the colour subcarrier frequency, so composite digital video is also known as $4f_{SC}$.
composite sync	A deprecated term meaning *sync*. The word *sync* alone implies both horizontal and vertical elements, so *composite* is redundant. The adjective *composite* more meaningfully applies to *video* or *colour*, so its use with *sync* is confusing.
composite video	**1** A video system in which three colour components are simultaneously present in a single signal. Examples are NTSC and PAL, which use the *frequency-interleaving* principle to encode (combine) luma and chroma. SECAM is another form of composite video. Distinguished from *component (adj.)*, above. **2** A *composite video (**1**)* signal, including luma, sync, chroma, and burst components; called *CVBS* in Europe.

concatenated	In compression, two or more compression systems in series. Also known as *tandem codecs.*
constant luminance	In a colour video system that dedicates one component to greyscale-related information, the property that true (CIE) relative luminance reproduced at the display is unaffected by the values of the other two components. All standard video systems, including NTSC, PAL, $Y'C_BC_R$, HD, JPEG, MPEG, and H.264 approximate constant luminance operation; however, because luma in these systems represents a weighted sum of nonlinear primary components ($R'G'B'$), "true" constant luminance operation is not achieved: A certain amount of (CIE) luminance "leaks" into the colour difference components and induces second-order artifacts (*Livingston errors*).
contrast, ANSI	Contrast ratio (see below) of a display (typically a projector) measured in accordance with ANSI/NAPM IT7.228 (withdrawn in 2003): Contrast ratio derived from an illuminance measurement taken by averaging black and white rectangles of a 4×4 "checkerboard" pattern without considering any effect of the screen or the viewing environment.
contrast	**1** Generally, a large or small difference in luminance or colour. **2** *contrast ratio;* see below.
CONTRAST	User-accessible means to adjust the luminance of reference white (sometimes incidentally changing the luminance of reference black as well). In video processing equipment, preferably called VIDEO GAIN; in consumer display equipment, preferably called PICTURE or WHITE LEVEL.
contrast ratio	The ratio between specified light and dark luminances, typically the luminance associated with the peak white or reference white of a display system and the luminance associated with reference black. *Inter-image* (or *on/off,* or *sequential*) contrast ratio is measured between separate full-screen white and black images. *Intra-image* (or *simultaneous*) contrast ratio is specified with respect to white and black measurements from a single test image such as that specified by ANSI ("checkerboard") or ITU-R BT.815. Different results are obtained if ambient light is present or absent in the measurement. Display manufacturers routinely exclude ambient light from contrast ratio measurements; however, estimating visual performance requires that ambient light be included.
corner frequency	The frequency at which the output power of a lowpass or highpass filter or subsystem has fallen to 0.707 of its value at a reference frequency, typically DC. (For a digital or constant-impedance system, this is equivalent to the frequency at which the output magnitude has been attenuated 3 dB.)
cosited	Chroma subsampling in which each subsampled chroma sample is located at the same horizontal position as a luma sample. BT.601, BT.709, and MPEG-2 standards specify cosited chroma

	subsampling. (MPEG-2, 4:2:0 chroma subsampling places chroma samples interstitially in the vertical domain.)
CSC	Colour space conversion, typically in gamma-corrected colour space, typically $Y'C_BC_R$ to $R'G'B'$ or $R'G'B'$ to $Y'C_BC_R$, but potentially including 3×3 processing in the linear-light domain.
CRC	Cyclic redundancy check (code). Information inserted prior to recording or transmission that allows playback or receiver equipment to determine whether errors were introduced. A CRC with a small number of bits provides error-detection capability; a CRC with a large number of bits provides error-correction capability. CRC codes involve multiplying and dividing polynomials whose coefficients are chosen from the set {0, 1}. Similar capability can be achieved using codes based upon mathematical principles other than CRC; see *ECC,* on page 616.
CRF	Camera response function. Synonym used in the computer vision community for OECF (see page 643). See also *RSR*, on page 653.
cross-colour	An artifact of composite (NTSC or PAL) video encoding and/or decoding that involves the erroneous interpretation of luma information as colour. The cross-colour artifact appears frequently when luma information having a frequency near that of the colour subcarrier appears as a swirling colour rainbow pattern.
cross-luma	An artifact of composite video encoding and/or decoding involving erroneous interpretation of colour signals as luma. Cross-luma frequently appears as *dot crawl* or *hanging dots*.
crossover (bar/chip)	In an optical greyscale step (chip) chart having an odd number of chips (often 9 or 11), the chip in the middle. (That chip conventionally has relative luminance of about 18%.)
CRT	Cathode-ray tube.
CRU	A dockable hard disk drive carrier, commonly used in digital cinema, manufactured by the company CRU-DataPort.
CUE	Chroma upsampling error: An implementation error in tens of millions of DVD players whereby subsampled chroma was reconstructed with incorrect position.
cutoff (*n*.)	**1** The phenomenon (particularly in a vacuum tube such as a CRT) where for a sufficiently low input signal the electron current (and in a CRT, emitted light) drops to essentially zero. **2** See *bias* (**2**) on page 595.
cutoff frequency	See *corner frequency,* on page 608.

cuts	In home theatre terminology, individual red, green, and blue cutoff adjustments; see *bias* (**2**) on page 595.
CVBS	Composite video with burst and syncs: A European term for *composite video* (**2**).
D-5 HD (*HD-D5*)	A component HD digital videotape format utilizing $^1/_2$-inch tape cassettes and recording BT.709 Y'CBCR signals, based upon either 720*p*60 or 1080*i*30, mildly compressed to about 270 Mb/s using motion-JPEG. Also known as *HD-D5.*
D-6	A SMPTE-standard component HD digital videotape format utilizing $^1/_2$-inch tape in cassettes, recording uncompressed BT.709 Y'CBCR signals, subsampled 4:2:2, at about 1.5 Gb/s.
D-7 (*DVCPRO/DVCPRO50*)	The SMPTE-standard SD compression and recording scheme introduced by Panasonic as *DVCPRO* (at 25 Mb/s) or *DVCPRO50* (at 50 Mb/s). See *DVCPRO,* on page 615.
D-9 (*Digital-S*)	SMPTE designation for JVC's obsolete Digital-S; see page 613.
D-10 (*MPEG IMX*)	SMPTE standard designation of Sony's *MPEG IMX;* see page 641.
D-11 (*HDCAM*)	SMPTE standard designation for Sony's *HDCAM;* see page 624.
D-12 (*DVCPRO HD*)	SMPTE designation for Panasonic's *DVCPRO HD;* see page 615.
D_{65}	See *reference white,* on page 651; and *CIE* D_{65}, on page 602.
D-cinema	Digital cinema: relating to the production, postproduction, distribution, and exhibition of movies using digital technology instead of photochemical film. See also *E-cinema,* on page 616. (E-cinema refers to distribution of movie-like material using HD technology.)
D-SLR	Digital single-lens reflex. A type of digital camera incorporating a swinging mirror that normally admits light to an optical viewfinder, but can be repositioned to admit light to a sensor.
dailies	Image sequences delivered to a set or location to enable production staff to quickly decide whether shots have been adequately captured from that set or that location. Dailies were historically film prints from developed camera negative film; now SD or HD video media or QuickTime files are typical. The term *dailies* originated when film processing imposed a latency of $^1/_2$ a day or so, perhaps overnight. The term *rushes* is synonymous.
data rate	Information rate of digital transmission, in bits per second (b/s) or bytes per second (B/s). Colloquially (and incorrectly) called *bandwidth* (see page 594).
datum	See 0_H *datum* and 0_V *datum,* on page 581.

dB, decibel	**1** Twenty times the logarithm (to base 10) of the ratio of a signal level (e.g., voltage, or digital code value) to a reference signal level.
	2 Ten times the logarithm (to base 10) of the ratio of a power to a reference power. Exceptionally, in video, HD, digital cinema, and still cameras, when optical power is being characterized, the $20\cdot\log_{10}$ convention (of sense **1**) is used even though the $10\cdot\log_{10}$ convention is strictly correct.
DC, direct current	**1** Historically, an electrical current or voltage having no periodic reversal in polarity.
	2 In modern usage, having zero frequency.
	3 In video, a signal having frequency substantially lower than the frame rate.
	4 In JPEG and MPEG, that spatial frequency component having uniform response over an 8×8 block. Distinguished from *AC* (**3**), on page 590.
DC, dynamic contrast	Mechanisms separate from the main light modulator – such as backlight control in an LCD, or mechanical iris in a projector – that reduce light output for dark images or scenes.
DCI P3	A set of *RGB* chromaticities and a white reference, standardized by the Digital Cinema Initiative (DCI) and in SMPTE ST 431, for use in digital cinema projectors.
DC restoration	In analog video, *clamp (v.)* (**1**) at blanking level (or in low-quality systems, at synctip level).
DCT	Discrete cosine transform: In image and video compression, the mathematics at the heart of the JPEG and MPEG algorithms.
DCV	Digital code value: A pixel component value, ordinarily a whole number representable in 8, 10, or 12 bits. The term is applicable to any image coding scheme (linear-light, gamma-corrected, pseudolog, etc.). See *ADU,* on page 591.
DCT	Discrete cosine transform.
DDC	Display data channel: A scheme standardized by VESA to convey display parameters upstream to a graphics subsystem.
DDL	Digital drive level: A pixel component data value that crosses an interface – typically DVI, HDMI, or DisplayPort – and drives display equipment. For "full-range" data, each component value is interpreted as an integer 0 … 2^K-1 (where K is the bit depth at the interface); DDL 0 produces minimum tristimulus value, and DDL 2^K-1 produces maximum.
DDWG	Digital Display Working Group, the now-defunct organization that developed and promulgated the DVI interface.

decade	Factor of 10: one $\log_{10}$ unit; almost exactly (or some would say, exactly) $3\,{}^1/_3$ stops.
decimation	Producing fewer output samples per unit time than input samples. Decimation could involve just dropping samples; if proper filtering is used, the term *downsampling* is preferred.
decoding	**1** Generally, converting one or more coded signals into uncompressed form, reversing a previous encoding operation that was applied to reduce data rate for transmission or recording. **2** In traditional video usage, taking NTSC or PAL composite video, performing luma/chroma separation and chroma demodulation, then producing component video output such as $Y'C_BC_R$ or $R'G'B'$. **3** In modern video usage, taking coded picture information (such as a JPEG, M-JPEG, or MPEG compressed bitstream) and recovering uncompressed $Y'C_BC_R$ or $R'G'B'$ picture data.
deep colour	Pixels having more than 8 bits per colour component, particularly when conveyed across a display interface.
degamma	See *inverse gamma correction,* on page 629.
deinterlace	A process of spatiotemporal upsampling that produces estimated progressive image data from interlaced image data. Also known as *I-P* (interlace-to-progressive) conversion.
demosaic	The process of processing spatially multiplexed colour samples, as produced by a CFA sensor (see page 600) to create three spatially coincident colour samples (typically three samples per photosite). Also called *deBayer,* which strictly applies only to Bayer-pattern sensors (see *Bayer pattern,* on page 594).
DF	dropframe; see page 614.
DFC, dynamic false contouring	An artifact of pulse-width modulated displays (such as PDP and DLP) whereby the time course of pulse modulation interacts with the viewers' eye-tracking of image elements in motion to create spatiotemporal artifacts.
DI	Digital intermediate. Historically, digital image data comprising the entirety of a movie, where the data originated with scanning cinema camera negative film, and was destined to be recorded onto cinema print film. In modern usage, image data typically originates from a digital cinema or HD camera, and data is typically destined for digital cinema release.
DIF	Digital interface standardized for DV bitstreams.
Digital Betacam, Digital-*β*	Sony's trademarked term for a component SD digital videotape format for professional use, utilizing $\frac{1}{2}$-inch tape and recording BT.601 $Y'C_BC_R$ signals, based upon either 480*i* or

	576*i* scanning, mildly compressed using M-JPEG to about 90 Mb/s.
Digital-S	An obsolete SD digital videotape format for professional use, utilizing ½-inch tape in VHS-type cassettes and recording BT.601 $Y'C_BC_R$ signals based upon either 480*i* or 576*i* scanning, mildly compressed to about 50 Mb/s using the DV motion-JPEG technique (DV50). Standardized as SMPTE D-9.
Dirac, Basic	A long-GoP motion-compensated wavelet codec developed by the BBC and released as open source.
Dirac PRO	A standard initially developed by BBC and subsequently standardized as SMPTE ST 2047-1 (VC-2), for the lossy mezzanine-level compression of 1080*p* HD video using intraframe wavelet compression at high bit rates. An open-source implementation (*Schrödinger*) is available.
disc	Rotating optical media, for example, Blu-ray disc.
disk	Rotating magnetic media, for example, hard disk.
display-referred	Image signal values (e.g., *R'G'B'*) having a well-defined mapping to tristimulus values and absolute luminance at an intended display in a specified viewing condition.
DisplayPort	A display interface promulgated by VESA.
DivX5	An implementation of MPEG-4 ASP.
DLP	Digital light processing: A trademark of Texas Instruments referring to projection displays based upon an array of digital micromirrors fabricated on silicon using MEMS techniques.
DN	Digital number. See *ADU,* on page 591.
DNxHD	A standard, developed and deployed by Avid for use in post-production, subsequently standardized as SMPTE ST 2042-1 (VC-3), for the lossy compression of digital motion images at bit rates between about 60 Mb/s and 220 Mb/s. DNxHD data is typically conveyed in MXF files.
Dolby Digital	Trademark term designating a digital audio compression standard incorporated into ATSC's DTV system; formerly, *AC-3*.
dominant field	See *field dominance,* on page 619.
dot crawl	A cross-luma artifact that results from a notch filter decoder, appearing as fine luma detail crawling up a vertical edge in a picture that contains a saturated colour transition.
dot-by-dot (dot-for-dot)	**1** In video, a system or subsystem that maps video samples directly to display pixels without image resampling.

	2 In printing technology, mapping halftone dots without descreening.
downconversion	In video, conversion to a scanning standard at the same frame rate having substantially lower pixel count (e.g., HD to SD).
downsampling	Resampling that produces fewer output samples than the number of input samples provided. (*Downsampling* does *not* properly describe rounding or truncating to shallower bit depth.)
digital picture exchange, DPX	An image file format standardized in SMPTE ST 268 that accommodates image data with a variety of bit depths and coding metrics. Commonly used to convey three 10-bit components according to the Cineon printing density (CPD) metric.
DRCR	Dark room contrast ratio: Contrast ratio measured under relatively high ambient illuminance, exceeding about 100 lx. Compare *BRCR*, on page 597.
drive (*n.*)	**1** A periodic pulse signal, now rarely used, that conveys either horizontal or vertical synchronization information.
	2 A low-level adjustment (DRIVE) – traditionally necessary for CRT displays owing to analog drift, and now thoughtlessly replicated in many fixed-pixel displays – to set *RGB* gain individually on a component-by-component basis. See *gain* (**2**).
dropframe	A timecode stream associated with scanning at a field or frame rate of 59.94 Hz, wherein timecodes of the form *HH*:*Tu*:00:00 and *HH*:*Tu*:00:01 are omitted from the count sequence whenever *u* (the units digit of minutes) is nonzero; counting frames in this manner obtains a very close approximation to clock time. This adjustment almost exactly compensates for the field or frame rate being a factor of exactly $^{1000}/_{1001}$ slower than 60 Hz.
DTIM	See *DCI P3, on page 611*.
DTV	Digital television: A generic term including digital SDTV and digital HDTV. Generally, broadcast is implied.
dual-link HD-SDI	Dual-link high-definition serial digital interface: A SMPTE-standard interface using a pair of HD-SDI links to transmit 4:4:4:4 HD.
dub (*n.* or *v.*)	Copy, typically degraded slightly from a master. (A digital copy with no degradation is a *clone.*)
DV	**1** Generally, digital video.
	2 A specific motion-JPEG compression technique, videotape recording format, and/or digital interface (DIF) bitstream, for $Y'C_BC_R$ digital video. See *DV25*, *DV50*, and *DV100*, below.

DV25	DV (see above), coded at 25 Mb/s. DV25 is widely implemented in consumer DVC and Digital8 equipment, and in professional D-7 (DVCPRO) and DVCAM equipment.
DV50	DV (see above), coded at 50 Mb/s. DV50 is used in professional D-7 (DVCPRO50), D-9 (Digital-S), DVCAM, and DVCPRO50 equipment.
DV100	DV (see above), coded at 100 Mb/s. DV100 is used in DVCPRO HD equipment.
DVB	Digital Video Broadcasting: An organization that standardizes and promotes DTV broadcasting. or the suite of standards promulgated by this organization. DVB advocates MPEG-2 video and audio compression, supplemented by DVB transmission standards for cable (DVB-C), satellite (DVB-S), and terrestrial (DVB-T) broadcasting (for which DVB-T specifies COFDM transmission).
DVC	Digital video cassette: A component SD digital videotape format for consumer use, taking BT.601 $Y'C_BC_R$ video having either 480*i* or 576*i* scanning, compressing to about 25 Mb/s using the DV motion-JPEG technique (DV25), and recording on 6.35 mm tape encased in a cassette of one of two sizes – small ("MiniDV") or large.
DVCAM	Sony's trademarked term for a component SD digital videotape format for professional use, utilizing 6.35 mm tape cassettes and recording BT.601 $Y'C_BC_R$ video having either 480*i* or 576*i* scanning, compressed to about 25 Mb/s using the DV motion-JPEG technique.
DVCPRO	Digital video cassette, professional version: A component SD digital videotape format for professional use, utilizing 6.35 mm tape encased in a cassette of one of three sizes (small, medium, or large), recording BT.601 $Y'C_BC_R$ signals, based upon either 480*i* or 576*i* scanning, compressed using the DV motion-JPEG technique to about 25 Mb/s. Standardized as SMPTE D-7.
DVCPRO50	Digital video cassette, professional version, 50 megabits per second: A component SD digital videotape format for professional use, utilizing 6.35 mm tape encased in a cassette of one of three sizes (small, medium, or large), recording BT.601 $Y'C_BC_R$ signals, based upon either 480*i* or 576*i* scanning, compressed using the DV motion-JPEG technique to about 50 Mb/s. Standardized as SMPTE D-7.
DVCPRO HD (*DVCPRO100*)	Digital video cassette, professional version, high-definition: A component HD digital videotape format utilizing 6.35 mm tape encased in a cassette of one of three sizes (small, medium, or large), recording 720*p*, 1080*i*, or 1080*p*, 4:2:2 signals, compressed to about 100 Mb/s using the DV motion-JPEG technique (DV100). Standardized as SMPTE D-12.

DVI	Digital Visual Interface: An interface established in 1999 and promulgated by the now-defunct Digital Display Working Group (DDWG) for the connection from a PC to a display.
DVTR	Digital videotape recorder. See *VTR,* on page 665.
DWT	Discrete wavelet transform: In image compression, the mathematical technique at the heart of the JPEG 2000 algorithm.
dynamic range	The ratio of physical quantities associated with the maximum and minimum meaningful or usable values, usually expressed as decibels (dB) using the $20 \cdot \log_{10}$ convention. For an image sensor, dynamic range is the ratio between the saturation-equivalent exposure and the noise-equivalent exposure.
EAV	End of active video: A sequence of four words inserted into a component digital video data stream, marking the end of picture samples on a line. See also *SAV, on page 655*, and *TRS, on page 662*.
EBU	**1** European Broadcasting Union: An organization mainly comprising European state broadcasters.
	2 In the context of colour standards: EBU Tech. 3213, which defines the chromaticities historically used in 576*i* video.
ECC	Error checking and correction: A method of inserting redundant information prior to digital storage, recording or transmission, and processing that information upon subsequent playback or reception, so that recording or transmission errors can be detected (and in some cases, perfectly corrected). ECC systems can perfectly correct errors having certain statistical properties. Synonymous with EDC.
E-cinema	Electronic cinema: relating to the distribution and exhibition of movies using HD technology instead of photochemical film. See also *D-cinema,* on page 610. (D-cinema applies to the entire cinema chain starting at production, and is not limited to HD technology. The term *E-cinema* generally implies use of video equipment and lower quality than cinema film or D-cinema.)
EDC	Error detection and correction: A synonym for ECC; see above.
EDH	Error detection and handling: A system standardized by SMPTE for encoding transmission error status into data conveyed across a series of SDI interfaces.
EDID	Extended Display Identification Data: A VESA standard specifying data to be conveyed from a display to a host graphics subsystem (across a display interface such as DVI, HDMI, or DisplayPort), giving display parameters such as raster timing and colour space.

encoding	**1** Generally, the process of converting one or more signals into a more complex representation, with the goal of reducing data rate for transmission or recording. **2** In traditional video usage, the process of taking component video input (e.g., $Y'C_BC_R$ or $R'G'B'$), performing chroma modulation and luma/chroma summation, and producing composite video (e.g., NTSC or PAL). **3** In modern video usage, the processing of uncompressed image data to produce a compressed bitstream (such as in JPEG, M-JPEG, MPEG, or H.264 compression).
ENOB	Effective number of bits (of an analog-to-digital converter). See IEEE Standard 1241.
EOCF	Electro-optical conversion function. The function that maps an $R'G'B'$ video component signal value to a tristimulus value at the display. Typical EOCFs approximate power functions having exponents between 2.2 and 2.6. EOCFs are standardized for graphics arts and for digital cinema; as I write, there is no standard EOCF for studio video. See also *gamma, decoding* (γ_D), on page 621.
EOTF	Electro-optical transfer function. See *EOCF.* The acronym EOCF is preferred owing to its use in ISO standards such as ISO 22028.
equalization	**1** The correction of undesired frequency or phase response. Coaxial cable introduces a high-frequency rolloff that is dependent upon cable length and proportional to $1/\sqrt{f}$ (pronounced *one over root f*); this is corrected by a subsystem called an *equalizer.* A naively designed analog lowpass filter, or a simple digital IIR filter, has poor phase response; this can be corrected by an *equalizer* filter section. **2** equalization pulse; see below.
equalization pulse	In analog SD sync, a pulse, part of vertical sync, that has approximately half the duration of a normal sync and occurs at 0_H or halfway between two 0_H instants.
error concealment	Masking, by playback or receiver circuits, of errors introduced in recording or transmission. Concealment is enabled by the playback or receiver circuits' detection of errors by using ECC codes. Concealment is accomplished by replacing errored samples by interpolated (estimated) signal information.
error correction	Perfect correction, by playback or receiver circuits, of errors introduced in recording or transmission. Correction is effected by the decoder using the redundant ECC information inserted by the recorder or transmitter to perfectly reconstruct the errored bits.

ETTR, expose to the right	Methodology for setting camera exposure that chooses exposure as high as possible just short of saturation – that is, without clipping highlights. The term derives from placement of image data on a histogram display. The technique maximizes photon shot noise performance.
essence	Data representing the sampled, quantized, and possibly compressed representations of continuous sound and/or image signals. Distinguished from metadata.
even field	Historically, in 480*i* (interlaced) scanning, the field whose first broad pulse starts halfway between two line syncs. Compare *odd field,* on page 643. The terms *odd* and *even* should be avoided, and *first* and *second* used instead.
excursion	The amplitude difference between two widely separated levels. Unless otherwise noted, reference excursion.
exponential function	A function of the form $y = b^x$, where b (the *base*) is constant. Exponential functions are rarely used in video. Gamma correction, and the EOCFs of typical displays, are approximate *power* functions, not exponential functions (as is frequently claimed).
exposure latitude	The range of camera exposures, typically expressed in stops with respect to ideal exposure, over which image quality is not significantly degraded. The term latitude is sometimes used to describe the maximum and minimum exposures for which the recorded image exhibits measurable contrast variation; however, *dynamic range* is preferred for the latter usage. The term *latitude* is sometimes used to describe the ratio (typically in number of stops) between the exposure associated with sensor saturation and the exposure associated with a PDR in the scene; however, *headroom* is preferred for that usage.
extended range	See *full-range, full-swing,* on page 620.
f-stop	**1** Diameter of the optical aperture, assumed circular, of a lens relative to the focal length (denoted *f*) of the lens. **2** Particular *f*-stop (**1**) settings (see above) in a $2^{-0.5k}$ sequence (e.g., *f*/1.4, *f*/2, *f*/2.8,*f*/ 4, *f*/5.6, etc.), where each step is associated with approximately a factor of 2 (one $\log_2$ unit) of light power, radiance, luminance, or tristimulus value.
FEC	Forward error correction: A synonym for ECC (see page 616), particularly when used in transmission or video recording systems.
FHA	Full-height anamorphic: Imagery occupying the full height of the format used for transport. (The image carried is spatially distorted with respect to the transport standard.)
field	In interlaced scanning, the smallest time interval that contains a set of scanning lines covering the height of the entire picture, along with all associated preceding sync elements. Fields were

	once denoted *odd* and *even;* these terms should be avoided, and *first* and *second* (as appropriate) used instead.
field dominance	In an interlaced motion image sequence, the field (first or second) immediately prior to which temporal coherence is susceptible to interruption due to editing. In principle, video edits can be made at any field, but good practice calls for edits in the vertical interval prior to the first field – that is, first field dominant. Good practice also calls for editing on frame *pairs,* in the vertical interval prior to even timecode numbers.
field merging	A deinterlacing technique, also knows as *weave,* that merges fields together, without interpolation, to form a frame. The technique is suitable for deinterlacing field pairs (or regions) that exhibit minimal motion; however, in the presence of significant motion, the zipper or sawtooth artifact will result.
field sync	In interlaced scanning, the analog sync pulse pattern that defines the start of a field. Field sync contains the 0_V datum. In 480*i* and 576*i* systems, field sync is a sequence comprising preequalization pulses, broad pulses, and postequalization pulses. In 480*i* systems there are six of each; in 576*i* systems there are five.
first field	In interlaced scanning, the first field of the pair of fields comprising a frame. In analog 480*i*, the field whose first equalization pulse starts coincident with 0_H. In analog 576*i*, the field whose first broad pulse starts coincident with 0_H. (See also *field dominance*, above, and *second field,* on page 657.)
first frame of action, FFoA	The first picture (frame) of a program, excluding leader.
fL, footlambert	Deprecated unit for luminance. Multiply by $^{10.764}/_{\pi}$, about 3.4263, to obtain $\mathrm{cd{\cdot}m^{-2}}$ [nit, or nt]. See *cd·m-2,* on page 600.
flare	Light, associated with a region of an image or an entire image, that is reflected and diffused in a manner that interferes with another part (or the entirety) of the image, either at a camera or a display. Flare is image-related: The unwanted light is modulated by image content. (See also *glare,* which is unmodulated. Both terms are sometimes used in a manner that fails to distinguish image-related and non-image-related sources.)
footprint, NTSC; footprint, PAL	The first time that the NTSC or PAL luma and modulated chroma components of an image are added together into a single composite signal, cross-luma and cross-colour artifacts become permanently embedded: Subsequent decoding and reencoding cannot remove them. The permanence of these artifacts is referred to as the *NTSC footprint* or the *PAL footprint*.
format conversion	An ambiguous term. See *transcoding,* on page 661, *scan conversion,* on page 656, *downconversion,* on page 614, *upconversion,* on page 663, and *standards conversion,* on page 658.

fourcc	Four-character code: A scheme to select among a potentially large number of options or formats – for example, to choose a video data structure – by indexing with a 32-bit field that contains four bytes coded in ASCII, for example, “YUV2”, “YV12”, or “r408”.
FPD	Fixed-pixel display (or, flat-panel display): A direct-view display (not a projector) having a flat display surface and discrete individually addressable light-emitting or light-modulating elements.
fps	Frames per second. Alternatively, express frame rate in hertz.
frame	The time interval or set of data that contains all of the elements of one picture, complete with all of the associated preceding sync elements. In analog systems, the frame occupies the interval between 0_V instants; in digital systems, the frame occupies the interval between the EAVs preceding line 1. In an interlaced system, a frame comprises two fields, *first* and *second*, which normally exhibit temporal coherence; each field contains half the image rows of the frame.
frame A	Deprecated. See *A-frame,* on page 589.
frequency interleaving	Modulation of chroma, and summation with luma, such that the modulated chroma signal occupies frequencies disjoint from the integer multiples of the line rate (at and near which the luma signal is concentrated).
FTA	Free-to-air: Terrestrial broadcasting, implicitly including content that has been or will be relayed by cable or satellite TV.
full HD	Generally, 1920×1080 HD, as opposed to subsampled variants such as 1366×768 or 1440×1080*i*.
full-range, full-swing	**1** In computing, an R', G', B', or Y' signal where reference black is at code 0, and reference white is at the largest possible code value in the code range available (e.g., 255 in an 8-bit system). In a full-swing $Y'C_BC_R$ system, C_BC_R are typically scaled to have $^{256}/_{255}$ the excursion of Y'. Full-swing $Y'C_BC_R$ coding is rarely used in video systems having more than 8 bits per component, though it is sometimes used in digital cinema. Preferably called *full-swing.* Contrast with *studio-swing,* on page 659. **2** In a 10-bit studio HD or digital cinema interface, coding where reference black is well below 10-bit interface code 64 (and typically code 0 or code 4), or reference white is well above interface code 940 (typically code 1019 or 1023), or both. Contrast with *studio-swing,* on page 659. Sometimes misleadingly called *extended range.*
full width half maximum, FWHM	The duration or spatial extent of a pulse, assumed to be symmetric, measured between the 50% points of amplitude.

$G'B'R'$	Green, blue, and red. An alternate notation for $R'G'B'$, with the components reordered to associate with [Y', C_B, C_R], respectively (or with [Y', P_B, P_R], respectively). G' associates with Y' because green dominates luma. Properly written with primes, but sometimes sloppily written *GBR*.
gain	**1** In signal processing, multiplying signal amplitude by a given factor (sometimes adjustable). In video, gain control ordinarily applies to Y' or to $R'G'B'$ components; the zero reference for gain adjustment is ordinarily taken at reference black level. **2** In displays, individual R', G', and B' gain control; sometimes called *drive.* (CONTRAST sets gain on all components together.) A closely related adjustment is *colour temperature* (**2**). In home theatre terminology, individual $R'G'B'$ gain adjustments are sometimes called *RGB-HIGH*. Distinguished from *bias* (**2**) on page 595.
gamma	A numerical parameter giving the exponent of a power function assumed to approximate the relationship between a signal quantity (such as video signal code) and light power. See *gamma, decoding* (γ_D), *gamma, encoding* (γ_E), and *gamma, system* below. The term *gamma* is so overloaded and so confused that it is best avoided. Use OECF (page 643) and EOCF (page 617) where possible. Certain relationships between applied signal and light power are not well approximated by a single-parameter power function; for example, the native electro-optical response of an LCD often takes an S-curve shape, which is not a power function. In such cases, *gamma* is not appropriate.
gamma = 1.0	A subsystem whose output values are linear with respect to its input values – that is, the subsystem has no net nonlinear transform (such as an OECF or EOCF). Sometimes used to indicate blending in the linear-light domain.
gamma, decoding (γ_D)	The exponent (greater than unity) of a power function taking the form V^{γ_D} that characterizes an EOCF (see page 617). The parameter γ_D is the exponent to which a video component signal R', G', *or* B' is raised to obtain a linear-light (luminance or tristimulus) value. In practical displays for electronic imaging, the value of decoding gamma is typically between 2.2 and 2.6. (For HD studio video, BT.1886 standardizes the value at 2.4.) Gamma characterizes a *display*, and is critical to achieving perceptually uniform coding. See *EOCF,* on page 617.
gamma, encoding (γ_E)	The exponent, less than 1, of a power function of the form L^{γ_E} that approximates, in a single parameter, an OECF (see page 643). A linear-light signal (for example, a relative luminance or tristimulus estimate) is transformed through the OECF to obtain a video signal R', G', or B'. (Subsequently, luma, Y', is computed.) Encoding gamma today typically has a value between approximately 0.4 and 0.5. Historical NTSC and FCC standards for SD mention 2.2, but do not state whether this value relates to encoding or decoding. Standards for *576i*

	systems mention display (decoding) gamma of 2.8; however, such a high value is never found in practice. For television, encoding gamma is typically about 1.2 times the reciprocal of the decoding (display) gamma; for modern systems, it is about 0.5. See *picture rendering,* on page 646. Encoding gamma properly has a value less than unity, though sometimes its reciprocal is quoted. See also *OECF,* on page 643.
gamma, system	The product of all of the power function exponents to which image data is subjected as it traverses a set of subsystems, starting from linear-light components captured from a scene by a camera (or from linear-light components captured from a previously reproduced image by a scanner), and ending with linear-light components reproduced at an image display. The term is best avoided owing to the difficulty of identifying exactly what constitutes the "system," and because it is used so widely without any consideration of picture rendering.
gamma-corrected	Image data to which gamma correction has been applied – that is, a signal that is intended to be converted to tristimulus through a power function (*OECF;* see page 643) having an exponent between about 2.0 and 2.6. Because gamma correction produces a video signal that mimics the lightness sensitivity of human vision, a gamma-corrected signal exhibits good perceptual uniformity: Noise or quantization error introduced into the signal is approximately equally perceptible across the tone range of the system from black to white.
gamma correction	The process by which a quantity proportional to intensity, such as an estimate of CIE luminance or some other tristimulus signal, is transformed into a signal intended for display through a power function having an exponent in the range roughly 2.0 to 2.6. See *OECF,* on page 643. In video, gamma correction is ordinarily performed at a video camera or its control unit.
gamma shift	Undesired alteration of effective gamma that results from inadvertent application of Macintosh-related gamma upon import or export of video involving a Macintosh computer. Gamma shift usually involves inadvertent application of a 1.45-power function or its inverse, a $^{1}/_{1.45}$ (i.e., 0.69) power function.
gamut	Generally, the largest possible set of colours of a particular device or circumstance. **1** Of a display device, the set of colours that can be produced in a particular viewing condition. **2** Of a colour interchange space, the set of colours that can be represented across all permitted codeword combinations – for example, *R'G'B'* video signals each ranging from reference black to peak white – when displayed as intended and analysed colorimetrically in a particular viewing condition. **3** Of a camera whose output is characterized colorimetrically, the set of colours represented across all possible output codeword

	combinations, displayed as intended and analysed colorimetrically in a particular viewing condition. (A camera whose output is not characterized colorimetrically, for example, a camera delivering "camera raw" data, has no defined gamut.)
gamut alarm	A studio device that warns of *R'G'B'* codewords outside the permitted set (for example, in narrow-gamut colour, codewords outside the *R'G'B'* unit cube. See *legal,* on page 632).
GbE	Gigabit [per second] Ethernet: A set of standards relating to conveying of data across shielded or unshielded twisted pair ("copper") wire, or optical fibre, at the rate indicated (typically either 1 Gb/s [implicit if no number is shown] or 10 Gb/s).
genlock (*v.*)	[Sync] generator lock: Synchronization, to a video source, of the scan timing of video origination or processing equipment.
glare	Light from the ambient environment (or potentially other sources) that interferes with an image. Glare sources are not image-related. (See also *flare,* which is modulated by image-related light. Both terms are sometimes used in a manner that fails to distinguish image-related and non-image-related light.)
global shutter	In an image sensor, operation (or resulting data) wherein exposure of all image rows starts at the same instant of time. (Opposed to *rolling shutter;* see page 653.)
GoP	Group of pictures: In MPEG or H.264, a set of consecutive pictures starting with a coded I-picture. A GoP typically extends to (but does not include) the following I-picture; however, a GoP may contain more than one I-picture. A short GoP has roughly 6 or fewer pictures; a long GoP, 15 or more.
grading	In the context of cinema, *colour correction* (**2**).
greyscale	**1** Generally, the range of achromatic values – that is, tones devoid of colour – from black to white.
	2 In video or computer graphics, the achromatic (or black and white, or lightness) component of image data.
GTG	Grey-to-grey: A nonstandard measure of the transition time ("response time") of a display (often an LCD) from one grey level to another. See also, *BTW,* on page 598.
H.264	Formally, ITU-T H.264, also published as ISO/IEC 14496-10 and known as MPEG-4 Part 10: A standard, jointly developed by ISO, IEC, and ITU-T, for the lossy compression of digital motion images and associated audio. The H.264 algorithm is based upon the basic principles of MPEG-2, but has many additional features that offer improved bit rate for the same performance (at the expense of additional encoder and/or decoder complexity). H.264 is sometimes loosely referred to as *MPEG-4;* however, that term does not uniquely identify H.264 because MPEG-4 Part 2 defines unrelated (SP/ASP and SStP) codecs.

HANC	Horizontal interval ancillary data: Ancillary data multiplexed into the horizontal blanking interval of an SDI or HD-SDI interface.
hanging dots	A cross-luma artifact appearing as a fine alternating pattern of dark and light dots along a horizontal edge in a picture having a saturated vertical colour transition, when decoded by a comb filter. Hanging dots are particularly evident when viewing the SMPTE colourbar test signal.
HD	**1** High-definition (video): There is no official definition; generally, a video system having aspect ratio of 16:9, frame rate of 23.976 Hz or higher, image data comprising 729 Kpixels (about ¾-million pixels) or more, and at least two channels of digital audio. Commonly either 720*p* (see page 588) or 1080*i* or 1080*p* (see page 589). Appending the letters TV (*HDTV*) implies entertainment programming. **2** Hard disk; to avoid ambiguity, use *hard disk drive, HDD.*
HD-CIF	ITU-R term referring to HD common image format – that is, 1920×1080.
HD-D5	See *D-5 HD,* on page 610.
HD-SDI	High-definition serial digital interface: A SMPTE-standard interface with a data rate of about 1.485 Gb/s or 2.97 Gb/s ("3G"), typically for uncompressed 4:2:2 studio-quality HD though other variants are accommodated.
HD-SLR	A D-SLR camera (see *D-SLR,* on page 610) adapted to record HD (often 1080*p*) video.
HDCAM	A component HD digital videotape format for professional use, utilizing ½-inch tape in Beta-type cassettes and recording $Y'C_BC_R$ signals based upon a 1440×1080 image structure, scanned progressive or interlaced at any of several frame rates, chroma subsampled 3:1:1, and mildly compressed to about 50 Mb/s using a motion-JPEG technique. Standardized by SMPTE as D-11.
HDCP	High-bandwidth Digital Content Protection: A scheme, promulgated by Digital Content Protection LLC, for encrypting uncompressed AV content conveyed across consumer digital interfaces (such as DVI, HDMI, and DisplayPort).
HDMI	High-Definition Multimedia Interface: An interface, promulgated by HDMI Licensing, LLC, to convey uncompressed audiovisual data.
HDTV	High-definition television: HD (see above); the *TV* suffix implies recording or transmission of entertainment programming.
HDV	A consumer HD system introduced by Canon, JVC, Sharp, and Sony, for 720*p* and 1080*i* (and in some systems,1080*p*), using

	long-GoP MPEG-2 MP@H-14 video compressed to a bit rate between about 19 Mb/s and 25 Mb/s and recording to DV tape, hard drive, or flash media.
HEVC	High Efficiency Video Coding: A compression scheme in development by the Joint Collaborative Team on Video Coding (JCTVC) of ITU-T VCEG and ISO/IEC MPEG, intended to compress video up to UHD to bit rate roughly half that of H.264.
HHR	Half horizontal resolution. An image format related to a base format (usually 704×480 or 704×576) by 2:1 horizontal downsampling.
Hi-color	A degenerate form of *truecolour* (page 662), now largely obsolete, wherein a pixel comprises 5 bits each of *R′*, *G′*, and *B′*. (Some Hi-color systems use 6 bits for green.) Not to be confused with *High Color* (see below).
Hi-Vision	NHK's term for 1080*i*29.97 HD.
hidef, high def	Slang term for *high-definition* (HD). I suggest that you avoid these terms: *def* suggests lack of hearing, but HD comes with sound.
High Color	Microsoft's term (in Windows 7) for pixel representations having 30 or 48 bits per pixel. Not to be confused with Hi-color (see above).
HLS	**1** Hue, lightness, saturation: Three basic perceptual attributes of human colour vision.
	2 http live streaming, HLS, see below.
hold-type	A display wherein a newly written pixel value is converted to light that is sustained for more than about half the frame time. Such a display is expected to exhibit motion blur induced by the viewers' eye-tracking of image elements in motion with respect to the frame.
horizontal blanking	The time interval – usually expressed in microseconds or sample counts, or sometimes the fraction of line time – between the end (or right edge) of picture information on one line and the start (or left edge) of picture information on the following picture line.
HomePNA, HPNA	HomePNA Alliance, formerly the Home Phoneline Networking Alliance: An organization promulgating standards for consumer IPTV. Its standards have been adopted by ITU-T (e.g., G.9954).
HSDL	High-speed data link. An adaptation of HD-SDI to convey 10-bit image data (typically *R′G′B′*) having image structure 2048×1556, typically scanned from film slower than realtime (at about 15 fps). Obsolete.

HTiB	Home theatre in a box. The notion that a consumer is sufficiently skillful to install a surround sound system.
http live streaming, HLS	Technology introduced by Apple to stream video across an IP network using just http (web) protocols. The stream is broken into components transferred across http as files.
HTVL	*Horizontal* resolution, expressed in units of *TVL*/PH: A unit of horizontal resolution used in CCTV and video security systems.
hue	**1** *The attribute of* a *visual sensation according to which an area appears to be similar to one of the perceived colours, red, yellow, green and blue, or* a *combination of two of them* [CIE]. Roughly speaking, if the dominant wavelength of an SPD shifts, the hue of the associated colour will shift. **2** In colour science, h^*_{uv} or h^*_{ab}, the polar-coordinate angle of a colour difference value in CIE $L^*u^*v^*$ or CIE $L^*a^*b^*$ components. **3** In video, the polar-coordinate angle of a colour difference value as displayed on a vectorscope, in C_B, C_R coordinates for component digital video or P_B, P_R coordinates for component analog video. **4** User-accessible means to adjust *hue* (**3**), sometimes confusingly called tint or phase.
HVS	**1** Human visual system. **2** Hue, value, and saturation: Three basic perceptual attributes of human colour vision.
I-field	In MPEG, a field-coded I-picture. I-fields come in pairs, either top then bottom, or bottom then top. See *I-picture,* below.
I-frame	In MPEG, either (i) a frame-coded I-picture, or (ii) a field-coded I-picture [either top or bottom] followed by a field-coded I-picture or P-picture [of opposite parity]. In the second case, the two fields form what is sometimes called an *IP-frame;* the P-field may involve prediction from the I-field. See *I-picture,* below.
I-frame-only	More generally, I-*picture*-only: A compression system that performs only *intra* compression – that is, compresses individual pictures (fields or frames) without exploiting coherence between frames.
I-P conversion	Interlace-to-progressive conversion: an alternate term for *deinterlace;* see page 612.
I-picture	In MPEG, an intraframe picture (field or frame): A picture, or coded picture information, that makes no reference to preceding or following pictures. An I-picture makes no use of temporal coherence.

I, Q	In-phase and Quadrature colour difference components of NTSC: U and V components rotated +33° and then axis-exchanged. When NTSC was established in 1953, modulated chroma was based on I and Q colour differences; I had considerably more bandwidth than Q. Since about 1970, NTSC colour modulation has been performed on equiband U and V components. When I and Q components are explained in a contemporary publication in any context other than purely historical, the presence of such an explanation signals that the author is uninformed about video practice during the last four decades.
IAR, image aspect ratio	See *aspect ratio,* on page 592.
IBP	See *long GoP,* on page 635.
IDCT	**1** In image compression, inverse discrete cosine transform.
	2 In scene-linear workflow, input device calibration transform.
ICT	In JPEG2000, irreversible colour transform: A mapping producing $Y'C_BC_R$ from $R'G'B'$ using BT.601 luma coefficients [0.299, 0.587, 0.114]. When inverted with fixed-point binary arithmetic, roundoff error results, so the transform is not perfectly reversible. Compare *RCT,* on page 651.
iFrame	Apple designation for a progressive video format having an image structure of 960×540 at aspect ratio 16:9, typically having a frame rate of 29.97 or 30 Hz – i.e., 540*p*30. (Apple apparently uses I-frame-only compression for this format.)
illegal	Not *legal* (page 632), each of the 3 senses.
image	A two-dimensional distribution of numerical values representing light levels associated with pictorial information.
image array	A two-dimensional array of sets of numerical values (picture elements, usually in a matrix) representing pictorial information.
image column	A column of the image matrix. (An image column corresponds to the “active” region of a historical analog scan line.)
image matrix	R', G', B', or Y' (luma) samples of a frame (or in an interlaced system, a field), considered as image rows and columns.
image row	A row of the image matrix. (An image row corresponds to an “active” historical scan line.)
image sequence	A set of images, conventionally uncompressed, stored in separate files (often TIFF or DPX) whose filenames incorporate sequence numbers or timecode that allows files in the set to be played out in order to present motion.
impulsive noise	**1** In analog transmission, noise where the signal is contaminated by very brief pulses. Limited channel bandwidth typically

	causes impulsive noise to be distributed across a few pixels once sampled. **2** In digital transmission, noise where individual, isolated digital sample values are corrupted (often to full black or full white, in which case the noise is sometimes referred to as “salt and pepper”). True impulsive noise is rare to nonexistent today. In uncompressed digital video, almost all useful channel coding methods implement error checking or correction. In compressed digital video, an individual pixel has no direct counterpart in the data stream, consequently pixels cannot be corrupted individually.
IMX	See *MPEG IMX,* on page 641.
intensity	**1** Of light, the amount of radiant or luminant flux (often radiated from or incident onto a surface), in a specified direction (that is, per unit solid angle). Intensity is a linear-light measure properly expressed in radiometric units such as watts per steradian $[\mathrm{W\cdot sr^{-1}}]$ or in photometric units such as candela [cd]. Image scientists and video engineers are usually interested in luminous intensity *per unit projected area* – that is, they are usually interested not in intensity but *luminance;* see page 636. **2** Of sound, power per unit area (in SI units, $\mathrm{W\cdot m^{-2}}$). Intensity of sound is comparable to irradiance or illuminance of light. **3** In computer graphics, video, and digital image processing, often used carelessly. The intensity produced by an image display subsystem is ordinarily *not* proportional to the applied signal, but is proportional to approximately the 2.4-power of the signal. Instead of *intensity* or *pixel intensity,* use *pixel value*.
intensity level	See *intensity* (**3**), above.
interlace	A scanning standard or image format in which alternate raster lines of a frame are displaced vertically by half the scan-line pitch and displaced temporally by half the frame time to form a *first field* and a *second field*. Examples are 480*i*29.97, 576*i*25, and 1080*i*30. Modern usage of the term *interlace* implies 2:1 interlace. See also *field,* on page 618.
interlace factor	The ratio between the number of picture lines in a reference progressive system and the number of picture lines necessary to defeat twitter in an interlaced system having equivalent spatial resolution. Distinguished from, but often mistakenly described as, *Kell factor, k*.
interleaved, component interleaved	A method of storing pixels whereby all components of a pixel occupy adjacent storage locations. A 3×2 image matrix of 8-bit *RGB* data could be stored as bytes in the order RGBRGBRGB RGBRGBRGB. Also known as *band-interleaved by pixel* (BIP), *chunky, packed pixel*, or *pixel interleaved.* See also *planar,* on page 647, and *band-interleaved by line (BIL),* on page 594.

interpolation	Resampling that produces more output samples than original samples (synonymous with *upsampling*), or that produces the same number of output samples as input samples (phase shifting).
interstitial	**1** Chroma subsampling wherein each subsampled chroma sample is effectively horizontally positioned halfway between adjacent luma samples. Interstitial 4:2:0 chroma subsampling is implicit in the JPEG/JFIF, H.261, and MPEG-1 standards.
	2 In television programming, a short element such as a commercial inserted into or between programs.
intra	A compression system that compresses individual pictures (fields or frames) without exploiting interpicture coherence.
inverse gamma correction	The process by which a video component signal value is transformed by an EOCF (resembling a power function having an exponent in the range roughly 2.0 to 2.6) to produce a tristimulus signal proportional to intended light power output at a display. See *EOCF,* on page 617. Inverse gamma correction is ordinarily performed as part of display processing.
inverse telecine	A process of taking interlaced video at 59.94 fps (or rarely, 60 fps) that has been subject to 2:3 pulldown, and recovering noninterlaced images corresponding to the original 23.976 Hz (or 24*p*) frames. In the studio, inverse telecine is trivially simple if the video is locked to an available timecode stream and cadence has been maintained. In consumer equipment, all bets are off, and inverse telecine usually involves comparison and analysis of several video fields.
IP	**1** Internet protocol: The foundation standards of the internet.
	2 Intellectual property: Copyright, patents, and trademarks.
	3 Intellectual property in the form of trade secrets incorporated into electronic subsystems (typically expressed in hardware description language such as Verilog or VHDL) known as *cores.*
IPR	Intellectual property rights: The right of a patent holder to control (and gain revenue from) deployment of a technology.
IPTV	Internet protocol television: Distribution of compressed digital video and associated audio over TCP/IP networks.
IRD	Integrated receiver-decoder: A device that receives an RF signal carrying digital television, and produces an uncompressed video signal. An IRD performs demodulation, demultiplexing, and MPEG-2 decoding. An IRD is typically a set-top device (*set-top box*, or STB).
IRE	*Institute of Radio Engineers,* the predecessor of the IEEE.

IRE unit	A historical unit of video signal amplitude adopted by the IRE (see above): In modern video systems having zero setup (including all forms of HD), one-hundredth of the excursion from reference black level (0 IRE) to reference white level (100 IRE). Today, *unit* or *video level* is preferred (see page 664).
IT	Information technology: Computer technology (or in one aspect, personal computing).
ITU-R	*International Telecommunications Union, Radiocommunications Sector;* successor to the *Comité Consultatif Internationale des Radiocommunications* (CCIR, International Radio Consultative Committee): A treaty organization that obtains international agreement on standards for radio and television broadcasting. The ITU-R BT (Broadcast Technology) series of Recommendations and Reports deals with radio and television. Although studio standards do not strictly involve radio frequency transmission, they are used in the international exchange of programs, so they fall under the jurisdiction of ITU-R.
ITU-R Rec. BT. 601, 709	Colloquially, *BT.601* and *BT.709* (or *Rec. 601* and *Rec. 709*); see *BT.601* and *BT.709,* on page 598.
ISO	International Standards Organization.
iso	Isolated. Referring to one camera of several that are recorded simultaneously to enable editing between shots in post-production; contrasted with *film style,* where a single camera is used to acquire several viewpoints of the same action shot multiple times.
ISP, image signal processor	A series of processing stages (a "pipeline") that receives digital signals (usually in mosaic colour form, commonly in a "Bayer" pattern) from an image sensor, and delivers processed image data with spatially coincident colour components.
JFIF	JPEG file interchange format. A file format, adopted in 1992 by an industry group led by C-Cube, that encapsulates a JPEG-compressed image, along with a small amount of supplementary data. If you are presented with an image data file described as JPEG, in all likelihood it is JFIF internally.
JND, just noticeable difference	The magnitude of change in a perceptually relevant physical stimulus that produces a 75:25 proportion of correct and incorrect responses in a two-alternative, forced-choice (2AFC) experiment. Observers are asked if they can detect a difference; the 75:25 proportion results from a stimulus that produces a 50% correct response rate; the remaining responses are guesses, half of which are correct by chance.
JPEG	**1** Joint Photographic Experts Group: A standards committee constituted jointly by the ISO and IEC and formally denoted ISO/IEC JTC1.

	2 A standard, formally denoted ISO/IEC 10918, adopted by *JPEG* (**1**) for the lossy compression of digital still images (either colour or greyscale).
judder	See *stutter,* on page 659.
JustScan	See *1:1 pixel mapping,* on page 582.
KB	Kilobyte: 2^{10} (or 1 024) bytes.
K	Unit of absolute temperature, kelvin. Properly written with no degree sign. In colour science, commonly used to quantify correlated colour temperature.
K-factor, *K*-rating	A numerical characterization of frequency response characteristics as evidenced by pulse fidelity, obtained by measuring the tightest fit of a specific time-domain envelope to a raised cosine test pulse. Distinguished from *Kell factor, k,* below.
Kell effect	In a video system – including sensor, signal processing, and display – *Kell effect* refers to the loss of vertical resolution, compared to the Nyquist limit, caused by the spatial dispersion of light power. Some dispersion is necessary to avoid aliasing upon capture, and to avoid objectionable scan line (or pixel) structure at display.
Kell factor, *k*	Historically, the ratio between effective vertical resolution and theoretically obtainable vertical resolution for a given number of picture lines. Generally between 0.7 and 0.9. Now deprecated. Distinguished from *interlace factor (page 628);* also distinguished from *K-factor, K-rating* (above).
kelvin	Unit of absolute temperature; its unit symbol is K, written with no degree sign. In colour science, commonly used to quantify colour temperature. See also *mirek,* MK^{-1}, on page 639.
key	**1** A component signal indicating opacity of the accompanying foreground image data, coded between black (0, fully transparent) and white (1, fully opaque). In computer graphics, called *alpha,* · (see page 591). See *composite (v.)* on page 607.
	2 The key component of KLV, see below.
KLV	Key, length, value: A data encoding defined in SMPTE ST 336 used to embed structured information in files (e.g., MXF).
knee	That portion of a video camera's OECF, or a comparable transfer function used in postproduction, that lies above a diffuse scene reflectance of about 90%. Video cameras are usually adjusted to compress (highlight) information beyond the knee.
L	Symbol for absolute luminance, having SI units $\mathrm{cd \cdot m^{-2}}$, “nit.”

L_A, active lines	The count of scan lines containing the picture. L_A for a frame is equivalent to the number of rows of image samples. In modern terminology, the count of image rows (N_R).
L_T, total lines	In raster scanning, the count of total scan lines in a frame.
L_V, luminance (visual)	Absolute (linear-light) luminance, preferably in SI units of $\mathrm{cd \cdot m^{-2}}$, or informally, *nit* [nt]. See *luminance, absolute* on page 636.
latitude	See *exposure latitude,* on page 618.
LB	Letterbox, see below.
LCD	Liquid crystal display. Most commonly direct view, but LCD technology (usually, LCoS) is used in projectors.
LCoS	Liquid crystal on silicon. A reflective type of LCD microdisplay structure used in projectors. (Pronounced *EL-koss.*)
leader	Film or tape media preceding and following program content, historically used for threading cameras, recorders, and projectors; by extension, digital imagery outside the bounds of a program. (Pronounced *LEED-er.*)
LED	Light-emitting diode, a solid-state semiconductor light source. LEDs are often used to backlight LCD displays.
legal	**1** In component video processing or transmission, the condition where each signal of a component set does not exceed its reference range: In an $R'G'B'$-legal combination, none of R', G', and B' exceed their reference ranges except perhaps for brief transients. (See also *valid,* on page 663.) **2** In NTSC and PAL processing or transmission, compliance of the composite NTSC or PAL signal (or its luma and chroma components) with broadcast standards. NTSC analog broadcast transmission has a 120 unit amplitude limit that requires limiting the chroma at hue values near yellow and cyan. **3** In JPEG, MPEG, or H.264, a bitstream that is compliant with standards, an encoder that produces only compliant bitstreams, or a decoder that correctly decodes any compliant bitstream.
legalizer	A studio device operating on $R'G'B'$ or $Y'C_BC_R$ video, warning of excursions outside the unit $R'G'B'$ cube and/or clipping to the $R'G'B'$ unit cube. A legalizer is incompatible with wide-gamut xyYCC.
letterbox	A widescreen image (such as 16:9 aspect ratio) conveyed or presented in a format having a narrower aspect ratio (such as 4:3), using the full width of the narrower format but not using the full height.

level	**1** In video, generally, the amplitude of a video signal, or one of its components, expressed on an abstract scale having reference values 0 and 1, or in volts, millivolts, IRE units, or digital code values.
	2 In JPEG, MPEG, and DV compression, the magnitude (i.e., absolute value) of a DCT coefficient.
level shifting	In JPEG, MPEG, and similar transform-based compression systems, the process effected prior to encoding of subtracting from each $Y'C_BC_R$ component half of its maximum excursion, so as to form a signed number that is subject to the transform mathematics (e.g., DCT). The process is reversed at decoding.
lift	**1** European term for *pedestal* (both senses); see page 645.
	2 A user control (LIFT) that introduces black-level offset, but changes gain correspondingly so as to maintain white level.
lift/gamma/gain	Three basic controls for colour correction of video-coded image data, often controlled by a set of three trackballs. See *lift* (above), and several entries for *gamma,* starting on page 621. *Gain* refers simply to a multiplicative scale factor applied to pixel values. (The scheme is generally inappropriate for image data in CPD or quasilog coding.)
lightness	**1** *The brightness of an area judged relative to the brightness of a similarly illuminated area that appears to be white or highly transmitting* [CIE]. Loosely, lightness is apparent reflectance. Lightness is expressed relative to some reference luminance or reflectance.
	2 CIE L^*: An objective quantity defined by the CIE, approximately the 0.42-power of relative luminance.
limited range	HDMI term for *studio-swing,* on page 659.
line, active; line, picture	A scanning line – or in modern terms, image row – that is specified by a scanning standard to contain picture. (Exceptionally, in 480*i* systems, closed caption lines are considered to be "active.") In modern terms, image row.
line doubling	A special case of image resampling that produces an image having exactly twice the number of image rows as the source.
line frequency	**1** In video, the frequency of horizontal scanning; about 15.7 kHz for SD, and 33.7 kHz or higher for HD.
	2 AC power line (mains) frequency, typically 50 Hz or 60 Hz, usually similar to the field rate of video.
	The term *line frequency* should be used with care in video because it may refer to the frequency of *horizontal* scanning, or to AC power (mains) frequency, which is usually similar or identical to the frequency of *vertical* scanning.

line-locked	**1** In digital video, having an integer number of samples per total line: If 0_H were digitized, it would take the same value every line. A line-locked system has coherent sampling and line frequencies, as in BT.601 or $4f_{SC}$ NTSC. Owing to the +25 Hz offset of the PAL subcarrier in PAL-B/G/H/I and PAL-N, PAL sampled at $4f_{SC}$ is not line-locked. **2** In CCTV, industrial, or security video, having vertical scan frequency locked to the AC power line (mains) frequency.
line sync	The sync signal pulse that defines the start of a scan line. In 480*i* and 576*i* systems, line sync may be the start of a normal sync or the start of certain equalization or broad pulses. See also *0_H datum,* on page 581.
line time	The time interval from the 0_H datum of one line to the 0_H datum of the next. In a digital system, from EAV to EAV.
linear (*adj.*)	A dangerously ambiguous term: **1** In mathematics, actual or presumed adherence to this equation: $g\left(a \cdot x + b \cdot y\right) \equiv a \cdot g\left(x\right) + b \cdot g\left(y\right) \quad \text{[for scalar } a, b\text{]}$ In digital image and video processing, the quantities *x* and *y* in this equation are the "signals"; *a* and *b* are "constants." The assertion that a process is linear offers no information concerning the nature of *x* and *y.* **2** *linear-light (adj.)*, see below. **3** Often confusingly used by digital imaging technicians (DITs) to denote video acquisition (e.g., BT.709), which is *not* linear-light, contradicting sense **2**. **4** Storage or processing of audio or video where the arrangement of data on the media is in direct correspondence to the timeline. Magnetic tape is sometimes referred to as "linear" media; hard drive media is sometimes called "nonlinear."
linear colour space, linear image data	Colour data wherein each component value, over nearly all of its range, is proportional to light power – that is, proportional to radiance, intensity, luminance, or tristimulus value.
linear-light (*adj.*)	Pertaining to a radiometrically linear signal – that is, proportional to radiance, intensity, luminance, or tristimulus value, or an estimate of one of these. Video signals are *not* encoded in linear-light form. Instead, individual *R'G'B'* values are *gamma-corrected,* that is, proportional to a power function (typically having an exponent between $^1/_{2.2}$ and $^1/_{2.6}$) of intended display luminance. Other forms of linear-light coding may be used to encode approximations of luminance in the original scene (*scene-referred* coding).

lines	**1** In raster scanning, the total number of lines per frame, L_T. **2** In an image format, active lines, L_A: image rows. **3** A unit of resolution; properly, *TV lines per picture height, TVL/PH* (see page 662).
loading, luminance	In PDP displays, sometimes used inaccurately to refer to *power loading;* see page 648.
log	**1** (*n.*) A recording of a series of events; (*v.*) to make a recording of a series of events. **2** Abbreviation of *logarithm.*
log colour space	Colour data wherein each component value, over much of its range, is proportional to the logarithm of light power – that is, proportional to the logarithm of radiance, intensity, luminance, or tristimulus value. See *log neg* and *log RGB*, below.
log neg	Image coding conveying integer codes representative of optical density values measured from (or approximating) exposed and developed colour-negative photographic film. *Cineon printing density* (CPD) is a widely used example of log neg coding.
log *RGB*	Image coding whereby scene-linear *RGB* values are mapped into integer codes through an approximately logarithmic OECF. It is usual to avoid the logarithm's singularity at zero by imposing an additive offset to a linear-light signal prior to taking the logarithm, so log *RGB* coding is usually more accurately called *quasilog.*
long GoP	Long group of pictures. Literally, a group of pictures (see *GoP,* on page 623) comprising more than about 10 pictures. Figuratively, an MPEG-, H.264-, or VC-1-class video compression system operating in non-intra mode using long GoPs.
LP/mm	Line pairs per millimeter: A unit of absolute resolution corresponding to a cycle (a black and a white element, comparable to a line pair) across one millimeter of physical distance.
LP/PH	Line pairs per picture height: A unit of resolution corresponding to a cycle (a black and a white element, comparable to a line pair) across the height of the picture.
LP/PW	Line pairs per picture width: A unit of resolution corresponding to a cycle (a black and a white element, comparable to a line pair) across the width of the picture.
luma	A video signal component representative of the greyscale component of an image, computed as a suitably weighted sum of nonlinear primary components (*R'G'B'*). At decoding, luma is combined with colour differences to produce *R'G'B'*, each of which is subjected to an EOCF in the conversion to tristimulus values at a display. A luma signal is approximately perceptually

uniform across the range from black to white. Luma is symbolized *Y'*. For SD, BT.601 standardizes these coefficients:

$$^{601}Y' = 0.299\,R' + 0.587\,G' + 0.114\,B'$$

For HD, BT.709 standardizes these coefficients:

$$^{709}Y' = 0.2126\,R' + 0.7152\,G' + 0.0722\,B'$$

Luma is nonlinearly related to relative luminance; luma alone is not sufficient to determine relative luminance.

luma coefficients

The coefficients of nonlinear (gamma-corrected) *R'G'B'* in the weighted sum that represents luma. Luma coefficients differ between SD and HD; see *luma,* above.

luminance

Properly, either *luminance, absolute* (see below) or *luminance, relative* (see below). The term *luminance* is often carelessly used in video engineering to refer to *luma;* see above.

In video, and computer graphics, much confusion surrounds the term *luminance*. In colour science and physics, *luminance* is proportional to intensity ("linear-light" light power). Absolute luminance has the symbol L or L_V and SI units of $\mathrm{cd \cdot m^{-2}}$ (or informally, *nit* [nt]). Relative luminance has the symbol *Y,* and is a dimensionless quantity typically between 0 and 1 (or 0 and 100); it can be computed as a properly weighted sum of *RGB* tristimulus values. In video, *luma* represents a weighted sum of nonlinear (gamma-corrected) *R'G'B'* components, and is properly denoted *Y'*. Confusion is heightened because luma coefficients are typically similar to or identical to the coefficients for relative luminance. The term *luma*, and the primed symbol *Y'*, denote the nonlinear quantity in a manner that avoids ambiguity. However, the term *luminance* is often sloppily used for this quantity, and the prime on the symbol is often omitted. Sloppy use of the word *luminance* and omission of the prime renders both the term *luminance* and the symbol *Y* ambiguous: Whether the associated quantity is CIE luminance (linear) or video luma (nonlinear) must then be determined from context.

luminance, absolute (*L*)

Luminous flux density (i.e., power per unit projected area) in a particular direction (i.e., per unit solid angle): The spectral radiance of a scene, in a particular direction (that is, per unit solid angle), weighted by the luminous efficiency function $\bar{y}(\lambda)$ of the CIE Standard Observer. Denoted L_V; properly expressed in SI units of $\mathrm{cd \cdot m^{-2}}$ (or informally, *nit* [nt]). Luminance is the photometric analog of radiance; it is the objective correlate of brightness. If that sounds complex, it is. Briefly and informally, absolute luminance is the physical linear-light quantity that is associated with the lightness sensation of human vision.

luminance, relative (*Y*)

1 Absolute luminance (see above), relative to a certain absolute white luminance (often a perfect diffuse reflector, or its depiction on a display). Relative luminance is symbolized *Y;* it is a dimensionless quantity having no units, and typically has a reference value of 1 or 100. In video, absolute reproduction

	of luminance is unnecessary; video normally involves an estimate of relative luminance.
	2 An estimate of relative luminance (**1**) obtained from a camera or scanner whose spectral sensitivities cannot be linearly combined to exactly match the CIE $\bar{y}(\lambda)$ function.
luminance coefficients	The coefficients of linear-light (tristimulus) *RGB* in the weighted sum that forms luminance. For the primaries of BT.709 – the primary set of HD and sRGB, also representative of modern SD – the luminance coefficients are 0.2126, 0.7152, and 0.0722.
luminance loading	In PDP displays, sometimes used inaccurately to refer to *power loading;* see page 648.
luminosity	Historically, the ratio of luminous flux to total radiant flux (over a wide range of wavelengths). Now deprecated. Used widely and incorrectly by Adobe to refer to (BT.601) luma.
LUT	Lookup table: A table enumerating the scalar or low-dimensional vector results of a function over a scalar or low-dimensional vector argument. The argument and result components are typically whole numbers in the range 0 to 2^K-1.
M-JPEG	A technique or file format using *JPEG* (**2**), or a JPEG-like algorithm, to individually compress each field or frame in a motion image sequence, without exploiting interfield or interframe coherence. M-JPEG is not standardized.
M/NTSC	See *NTSC, National Television System Committee (**3**)*, on page 642. The *M* is an archaic reference to the nomenclature of ITU-R Rep. 470.
M-frame, midframe	In 2:3 pulldown to 29.97 Hz interlaced video, the video frame – unique in the 5-frame sequence – comprising a first field from one 24*p* frame (conventionally frame B) and a second field from another 24*p* frame (conventionally 24*p* frame C). An edit in the 24*p sequence* could intervene. See *2-2 pulldown,* on page 582 and *A-frame,* on page 589.
.m2ts	File extension commonly used for an MPEG-2 transport stream (including the Blu-ray container format).
macroblock	**1** In MPEG, image data comprising, or coded picture information representing, one of the 16×16 arrays of luma samples that tile the image, accompanied by the requisite number and arrangement of associated blocks of C_B and C_R. In the common MPEG-2 case of 4:2:0 chroma subsampling, four 8×8 luma blocks are accompanied by an 8×8 block of C_B and an 8×8 block of C_R; in this case, a macroblock comprises six blocks.
	2 In DV, image data comprising – or coded picture information representing – an 8×8 block of C_B, an 8×8 block of C_R, and the associated two (4:2:2), three (3:1:1), four (4:1:1 or 4:2:0), or six (3:1:0) 8×8 blocks of luma.

	3 In JPEG, an MCU (minimum coded unit; see page 638).
MADI	Motion-adaptive deinterlacing.
MAE	Mean absolute error: Sum of absolute differences (see *SAD,* on page 654), divided by the number of samples considered.
Matroska	An open-source container format used in Google's WebM project.
mb	See *macroblock*, above. It's sensible to write the abbreviation in lowercase, to disambiguate from megabyte; in compressed video there are both megabytes per second (MB/s) and macroblocks per second (mb/s).
Mb	Megabit: 2^{20} (or 1,048,576) bits; 131,072 bytes, or 128 KB.
MB	**1** Megabyte: 2^{20} (or 1,048,576) bytes, or in disk storage, $2^{10} \cdot 10^3$ (or 1,024,000) bytes.
	2 In MPEG and associated standards, macroblock.
MCU, minimum coded unit	In JPEG without subsampling, image data comprising, or coded picture information representing, a set of 8×8 blocks of image data, one per component. In $Y'C_BC_R$ JPEG, with 4:2:0 subsampling, image data comprising, or coded picture information representing, an 8×8 block of C_B, the corresponding 8×8 block of C_R, and the four associated 8×8 luma blocks.
median filter	A nonlinear "rank order" filter whose output is the middle value of the input samples, sorted by amplitude (in the case of image data, sorted by pixel value). (If an even number of input samples are provided, the output is the average of the two middle values.)
mesopic	Pertaining to vision at adaptation levels between about 0.1 nt and 3 nt, where the retina's rod cells and code cells are both active ("mez-*AH*-pick"). See mesopic, scotopic.
metadata	Generally, data about data. *Structural metadata* describes how essence is stored. *Descriptive metadata* describes the content of the essence.
metamerism	**1** In colour science, the condition that two different spectral power distributions, when weighted according to the CMFs (spectral response curves) of the CIE Standard Observer, produce identical tristimulus values – that is, appear to a colour-normal observer as the same colour. Metamerism frequently holds for spectra that are markedly different.
	2 In a camera or scanner, the condition that two objects that are metameric with respect to the CIE Standard Observer produce different sets of *RGB* components, owing to the spectral response of the device departing from the CIE curves. A camera

	or scanner has metamerism errors when it “sees” colour differently from a colour-normal human observer.
mezzanine	Pertaining to “light” compression for use in applications intermediate between acquisition/postproduction (where a high data rate is important to maintain image quality) and distribution (where a low data rate is important for economy).
midgrey, midscale, midtone	Reflectance in a scene, or the corresponding image data value carried in a film or video system, that represents diffuse scene reflectance of roughly 18% of a perfect diffuse reflector.
MiniDV	The small cassette size variant of DVC. (See page 615.)
mirek, MK^{-1}	Reciprocal megakelvin: a unit of correlated colour temperature, $^{10^6}/_{t}$, where t is CCT in kelvin [K]. Formerly, *mired.*
modulated chroma	In NTSC and PAL, a colour subcarrier (**1**) onto which two colour difference signals (typically *U* and *V*) have been imposed by quadrature modulation.
monochrome, monochromatic	**1** In colour science, shades and tones of a single hue: A colour stimulus having a single (usually narrowband) spectral peak.
	2 In video, the greyscale (or achromatic, or black and white, or lightness) component of image data. This usage of *monochrome* is confusing because it contradicts sense **1**; to avoid confusion, use *greyscale* (page 623) or *achromatic* (page 590).
	3 In computing, image data having one bit per pixel, representing either full black or full white. This usage is confusing because it contradicts sense **2**; also, the suffix *-chrome* suggests colour, but in computer terminology a so-called monochrome image has no colour. A colour scientist would describe a sodium vapour streetlamp as *monochrome,* but its light is coloured! Use *bilevel* (page 595) to describe image data with one bit per pixel.
MOS	**1** Metal-oxide-semiconductor (pronounced *moss*): The layers in a certain type of semiconductor structure widely used in integrated circuits (e.g., CMOS).
	2 In cinema acquisition, without sound; pronounced *em-oh-ess.* Some people believe the acronym to originate from *minus optical stripe*, referring to the optical sound track on a print. However, the term is associated with image capture where the sound is not recorded on film. The Hollywood legend is probably correct that the term arose from the famous German director Otto Preminger, who in frustration with the noisy cameras of his day is reported to have shouted, “Mit out sound!” indicating that shooting should commence immediately without concern for ambient noise, where the intended audio would be created later in postproduction.
mosaic	See *CFA,* on page 600.

mosquito noise	Visual noise having very high spatial frequency produced by motion-compensated transform coding (such as MPEG and H.264) using aggressive compression. Such noise is typically incoherent picture-to-picture, and can appear like buzzing mosquitoes surrounding pronounced edges in an image.
motion-JPEG	See *M-JPEG,* on page 637.
motion compensated interpolation, MCI	A technique of using one or more motion vectors to compute spatially displaced image information representing a certain time instant, derived from one or more pictures sampled at different time instants.
motion detection, MD	A technique of estimating, for a certain pixel or a certain block of pixels, whether an underlying scene is in motion (with respect to the picture boundary) at a speed greater than a certain number of pixels per picture time.
motion estimation, ME	A technique of estimating, for a certain pixel or a certain block of pixels, the speed and direction of motion (with respect to the picture boundary) of assumed underlying scene elements.
motion vector, MV	A vector, associated with a pixel or a block of pixels, that estimates the displacement (horizontally and vertically) of an assumed scene element from its position in a previous or future picture.
movie mode	In consumer television displays, a setting that typically disables proprietary signal processing features believed by the manufacturer to “enhance” or “improve” the picture. Such circuitry distorts material that is presented to the display in pristine form; *movie mode* defeats the “enhancement.”
MP3 (.mp3)	Formally, MPEG-1 Audio Layer III: An audio compression standard, defined in MPEG-1 and MPEG-2, that is widely used for music distribution. Sometimes incorrectly called MPEG-3.
MP4 (.mp4)	A container (file) format used in MPEG-4 Parts 1 and 14 (not Part 10/AVC) that is based upon the container format used in Apple's QuickTime system.
MPD	Motion-picture disturbance (or moving pixel distortion): See *DFC, dynamic false contouring,* on page 612.
MPEG	Moving (not Motion!) Picture Experts Group: A standards committee, jointly constituted by ISO/IEC and ITU-T, that develops standards for the lossy compression of digital motion images and associated audio. The MPEG algorithms exploit the temporal coherence found in video and audio data (and the spatial coherence found in video data). The MPEG-2 standard (see below) is of interest to digital video and HD. Its predecessor, now denoted MPEG-1, offers VHS-quality. Other MPEG standards, such as MPEG-7, and MPEG-21, are for applications other than broadcast television.

MPEG-1	A standard, adopted by MPEG (see above), formally denoted ISO/IEC 11172, optimized for SD video data rates of about 1.5 Mb/s and having approximately VHS quality.
MPEG-2	A standard, adopted by MPEG (see above), co-published as ISO/IEC 13818 and ITU-T standard Rec. H.262), optimized for SD and HD at data rates of 4 Mb/s and higher.
MPEG-3	There is no MPEG-3. The term is sometimes mistakenly applied to MPEG Audio Layer III; see *MP3 (.mp3),* above.
MPEG-4	A set of standards promulgated by ISO and ITU-T MPEG; however, as commonly used, MPEG-4 is an ambiguous term: **1** MPEG-4 Part 2, formally known as ISO/IEC 14496-2 ("Part 2"), and informally known as SP/ASP; intended for low bit rate applications such as mobile and handheld broadcasting, but largely superseded in commercial applications by H.264. **2** MPEG-4 Part 10, formally known as ISO/IEC 14496-10 ("Part 10") defines a video compression scheme used for broadcast. To avoid confusion with MPEG-4 Part 2 (SP/ASP), MPEG-4 Part 10 is better denoted by its ITU-T designation, H.264; see *H.264,* on page 623.
MPEG-4 Part 2 ASP	Formally, ISO/IEC 14496-2 ("Part 2"): A standard for video compression intended for low bit rate applications such as mobile and handheld broadcasting; it has been largely superseded commercially by H.264.
MPEG IMX	A component SD digital videotape format introduced by Sony for professional use, utilizing ½-inch tape in Beta-type cassettes and recording BT.601 $Y'C_BC_R$, 4:2:2 signals based upon either 480*i* or 576*i* scanning, mildly compressed to between about 30 Mb/s and 50 Mb/s using MPEG-2, I-frame-only video compression and conveying eight 24-bit uncompressed digital audio streams. Standardized as SMPTE D-10.
MVC	Multiview Video Coding: A mechanism standardized in Annex H of H.264 to efficiently code two or more pictures that are temporally coincident and highly spatially correlated, usually the left and right images of a stereo pair.
MXF	Material eXchange Format: A container or "wrapper" file format for exchange of professional digital video and audio media ("essence") and associated metadata, defined by SMPTE ST 377 and related standards.
ND	Neutral density: an optical filter having uniform transmittance across the visible spectrum.
NDF	Nondropframe. See *dropframe,* on page 614.
NG	No good. Common Japanese technical acronym; opposite of OK.

NHK 日本放送協会		Nippon Hoso Kyokai (Japan Broadcasting Corporation): The organization that invented HDTV.
nit [nt]		Colloquial term – derived from the Latin *nitere,* to shine – for the SI unit for luminance, $\mathrm{cd}\cdot\mathrm{m}^{-2}$. See $cd\cdot m^{-2}$, on page 600.
nonlinear	**1**	A process that does not exhibit mathematical linearity; see *linear (adj.),* on page 634.
	2	Storage or processing of audio or video where the arrangement of data on the media is *not* in direct correspondence to the timeline. Magnetic tape is sometimes referred to as “linear” media; hard drive media is sometimes called “nonlinear.”
normal line sync		In analog SD, a line sync pulse that remains at sync level for about 4.7 μs. In interlaced systems, the leading edge of equalization and broad pulses are utilized as line syncs.
notch filter		In a composite video decoder, circuitry that separates chroma from a composite signal using a simple bandpass filter centered at the colour subcarrier frequency. A notch filter introduces dot crawl artifacts into any picture that has luma detail at frequencies near the colour subcarrier.
NPM		Normalized primary matrix: A 3×3 matrix which, when matrix-multiplied by a linear-light *RGB* column vector on the right, produces an *XYZ* column vector. The middle row of an NPM gives the luminance coefficients for a set of display primaries. Owing to normalization, the middle row sums to unity.
nt		Abbreviation for *nit [nt]*, see above.
NTSC, National Television System Committee	**1**	The group, now referred to as *NTSC-I*, that in 1941 standardized 525-line, 60.00 Hz field rate, interlaced monochrome television in the United States.
	2	The group, formally referred to as *NTSC-II*, that in 1953 standardized 525-line, 59.94 Hz field rate, interlaced colour television in the United States. NTSC-II introduced the composite video technique.
	3	A method of composite video encoding based on quadrature modulation of *I* and *Q* (or *U* and *V*) colour difference components onto a colour subcarrier, then summing the resulting chroma signal with luma. Used only with 480*i* scanning, with a subcarrier frequency nominally $^{455}\!/_{2}$ times the horizontal line rate (i.e., a subcarrier frequency of about 3.579545 MHz).
	4	Often imprecisely used to denote 480*i*29.97 (525/59.94) scanning or 480*i*29.97 image format.
NTSC-J		NTSC as practiced in Japan: NTSC (**3**) with zero setup (and luma and chroma levels modified accordingly).

NTSC-legal	The condition where an NTSC signal is *R'G'B'*-legal and additionally has no chroma content that would cause the composite signal to exceed +120 units.
octave	Factor of 2, one $\log_2$ unit. Usually applied to frequency. [A factor of 2 in amplitude is a *stop (f-stop, T-stop)*; see page 659.]
odd field	In 480*i* (interlaced) scanning, the field whose first broad pulse is coincident with line sync. Compare *even field,* on page 618. The terms *odd* and *even* should be avoided, and *first* and *second* used instead as appropriate.
OECF	Opto-electronic conversion function. The function that maps estimated tristimulus value in a scene into an *R'G'B'* video component signal value. A typical OECF resembles a power function whose exponent lies between 0.4 and 0.5; however, OECF is routinely subject to various creative manipulations that cause it to diverge from a power function. Standards such as BT.709 and SMPTE 274M purport to standardize OECF, but those standards are useful only for engineering purposes and not for video production. See also *gamma, encoding* (γ_E), on page 621.
OETF	Opto-electronic transfer function. See *OECF* (above). The acronym OECF is preferred owing to its being firmly established in ISO standards such as ISO 22028.
OFDM	Orthogonal frequency-division multiplexing. In video transmission, OFDM is always applied to digital data, and referred to as *coded;* see *COFDM* on page 604.
offset sampling	A digital image format in which samples of one image row are offset horizontally by one-half the sample pitch from samples of the previous image row. Also known as *quincunx sampling*. Contrasted with *orthogonal sampling*, which is now ubiquitous.
OP-Atom	In MXF (see page 641), an operational pattern wherein video and audio tracks are stored in separate files. See SMPTE ST 390. Used in Panasonic's P2.
OP1a	In MXF (see page 641), a simple operational pattern in which video and associated audio are interleaved and stored in one file; typically, audio data is stored nearby the associated video. Used in Sony's XDCAM. OP1a is sometimes described as having "self-contained essence."
open (caption, subtitle, etc.)	*Open,* in the context of auxiliary information for the consumer that originates from nonimage sources, refers to information that is composited onto the image at some point in the production or distribution chain. (Contrasted with *closed,* referring to information embedded in metadata and only available to viewers having suitable, enabled decoding equipment.) Technical information that is composited onto the image but which is not intended for consumers is called *burnt-in.*

optical density	The base-10 logarithm of the reciprocal of transmission through partially absorbing optical material (in cinema, either negative or positive film). Various spectral weightings are used.
Opto-electronic conversion function	See *OECF,* above.
orthogonal sampling	A digital image format in which samples of one image row are vertically aligned with samples of the previous line of the field (or frame). Contrasted with offset sampling (see above).
OTA	Over-the-air: Terrestrial UHF broadcasting (implicitly DTV).
OTT	Over-the-top: Any system for video delivery that bypasses the traditional OTA, cable, and satellite providers.
overscan	The unfortunate practice, common in consumer electronics, of signal processing that crops some number of image rows at the top and bottom edges and some number of image columns at the left and right edges, then spatially expands by a factor slightly greater than 1. Overscan between about 3% and 5% is common in consumer HD equipment. If a source having 1920×1080 structure is displayed on a native 1920×1080 display, then overscan is unnecessary. If overscan is used, picture elements will be lost and scaling artifacts are likely.
P-field	In MPEG, a field-coded P-picture. P-fields come in pairs (either top then bottom, or bottom then top). See *P-picture,* below.
P-frame	In MPEG, either a frame-coded P-picture, or a pair of P-fields (one top field and one bottom field, in either order). See *P-picture,* below.
P-picture	In MPEG, a predictive-coded picture: A picture, or coded picture information, in which one or more macroblocks are predicted from a preceding anchor picture, and which may itself be used as the basis of subsequent predictions. P-pictures exploit temporal coherence.
P2	Panasonic flash-memory video storage media, based upon the 32-bit PCMCIA/Cardbus interface.
P3	See *DCI P3, on page 611.*
packed, packed pixel	In the context of storage of pixels in memory, see *interleaved, component interleaved,* on page 628.
paint, painting (*v.*)	With respect to camera signal processing, the act of adjusting gain, black level ("pedestal"), OECF ("gamma"), and other parameters in order to overcome technical deficiencies and/or to achieve a desired æsthetic effect. See also *shading,* on page 657.
PAL (phase alternate line, or phase alternation line)	**1** A composite video standard comparable to NTSC, except that the modulated *V* colour difference component inverts phase on

	alternate scan lines, and burst meander is applied. Usually used in 576*i* systems with a subcarrier frequency of 4.433618750 MHz, but also used with subcarriers of about 3.58 MHz in PAL-N and PAL-M analog broadcasting. **2** Often incorrectly used to denote 576*i* (625/50) scanning.
PAR, pixel aspect ratio	Properly, *SAR, sample aspect ratio*; see page 655.
parade	Presentation of three signal components (usually red, green, and blue) in sequential order on a waveform display.
pathological sequence	In SDI or HD-SDI, an encoded bit-serial sequence (or the pixel values causing such a sequence) that presents a long string of consecutive like-valued bits in the channel (many 0s or many 1s), potentially leading to PLL stress or equalizer stress.
P_B, P_R	Scaled colour difference components, blue and red, used in component analog video: Versions of B' minus luma (B'–Y') and R' minus luma (R'–Y') scaled for excursion nominally identical to luma for component analog transmission. P_B and P_R in the range ±0.5 according to the EBU N10 standard are equivalent to C_B and C_R scaled by the factor $^{1}/_{224}$; however, various different industry standards are in wide use for 480*i*. See also C_B, C_R, on page 598, and *U*, *V*, on page 662.
PDP	Plasma display panel.
PDR, perfect diffuse reflector	A perfectly Lambertian diffuse (nonspecular) 100% reflector. A PDR is well approximated by pressed titanium dioxide or magnesium oxide powder.
peak white	The maximum value of absolute luminance, relative luminance, or luma. Distinguished from *reference white:* Studio video systems typically allow signals to excurse to a peak somewhat above reference white – for BT.601 and BT.709, peak white luma, R', G', or B' lies at $^{238}/_{219}$ of the reference black to reference white excursion; with 2.4-gamma, peak white luminance is about 1.23 times reference white luminance.
pedestal	**1** black level (see page 595) expressed as an offset in voltage or units relative to blanking level. Conventionally 7.5 units (about 54 mV) in 480*i* SD and zero in all other systems (where blanking level and black level are identical). Pedestal is properly an offset or a level; it is incorrect to express pedestal as a percentage. See also *setup,* on page 657. **2** In camera or processing equipment, adjustment of *pedestal* (**1**). In a camera, provision may be made for operator adjustment of pedestal in the linear-light domain (i.e., prior to application of the opto-electronic transfer function, OECF).
percent	In general, one hundredth. In video and digital cinema, some uses (e.g., 18%) refer to relative luminance (in the linear-light domain). Other uses (e.g., 109%) refer to the encoded Y', R',

	G', or B' signal (in the gamma-corrected, perceptual domain); in that case, *IRE* and/or *unit* are synonymous.
PH	Picture height, the basis for a relative measure of viewing distance.
PHASE	Means accessible to the studio technician to adjust composite NTSC or PAL subcarrier phase. The associated consumer control is preferably called HUE – see *hue* (**4,** on page 626).
photometric (*adj.*)	Relating to a quantity of light, proportional to power or energy, that has been weighted spectrally by a function similar or identical to the CIE luminous efficiency function [denoted $V(\lambda)$ or $\bar{y}(\lambda)$]. See also *radiometric (adj.),* on page 651.
photopic	Pertaining to vision at adaptation levels above about 3 nt, where the retina's rod cells are inactive ("pho-*TAH*-pick").
picture	An ordered set of image rows (or historically, scan lines) covering the height of the picture. In a progressive system, the whole frame; in an interlaced system, either a top field or a bottom field.
picture excursion	The excursion from blanking to reference white. In 480*i*, 100 units by definition. In analog System M, $\frac{5}{7}$ V (about 714 mV); in other systems, particularly 576*i* and HD, 700 mV. Confusingly, in 480*i* systems having setup, *picture* in this term includes blanking (nonpicture information) occupying levels from 0 units to 7.5 units.
picture rendering	Encoding and subsequent decoding of estimated relative luminance – or, in a colour system, estimated tristimulus values – incorporating correction for effects owing to display characteristics and viewing environment characteristics so that colour appearance is correctly presented. In video, picture rendering is imposed by the combination of encoding to approximately the 0.5-power of scene tristimulus value, and decoding with a 2.4-power function; an end-to-end power function exponent of approximately 1.2 results.
pillarbox	An image, with an aspect ratio such as 4:3, conveyed or presented at correct aspect ratio in a format having a wider aspect ratio (such as 16:9), using the full height of the widescreen format but not using the full width. The term echoes *letterbox;* in common language in the U.K., a pillarbox is a tall postbox.
pixel	Picture element. Unfortunately, a deeply ambiguous term: **1** Historically, in greyscale digital imaging in general, and greyscale ("monochrome") video in particular, the quantized sample value specific to a single spatial sampling site in an image. **2** Historically, in colour digital imaging in general and colour video in particular, and in modern systems that accomplish

	colour separation or recombination using optical superposition, a set of three spatially coincident colour component samples; perhaps augmented by spatially coincident data such as opacity (*alpha* or *key*) data. Even in its historical interpretation, the term *pixel* is ambiguous when chroma subsampling is involved.
	3 In the terminology of digital still cameras – and, by extension, in mosaic-sensor-based digital cinema cameras – any single colour component sample.
	To resolve the ambiguity, I suggest using the term *triplet* to refer to a set of three spatially coincident colour samples, and the term *photosite* to refer to an image sensor element.
pixel intensity	See *intensity* (**3,** on page 628). Because imaging systems are rarely designed to have pixel values proportional to (physical) intensity, the term *pixel value* (see below) is preferred.
pixel-by-pixel, pixel-mapped,	A system or subsystem that maps optical samples to pixels (or pixels to optical samples) without the necessity for image data resampling. Sometimes called *dot-by-dot;* sometimes known by trade names such as *Just Scan.*
pixel value, PV	The numerical value of a component of a pixel, ordinarily a whole number representable in 8, 10, 12, 16, or some other number of bits, in certain cases represented in floating point form (e.g., OpenEXR).
PJ	Abbreviation, commonly used in Japanese, for projector.
PLUGE	*Picture line-up generator:* Originally, equipment, but now a test signal, that produces video signal elements slightly below reference black, exactly reference black, and slightly above reference black, to aid a technician in setting display equipment BLACK LEVEL. Modern standards for PLUGE call for 8-bit interface codes 12, 16, 20, and 24 (about –2%, 0%, +2%, and +4%).
planar	A method of storing pixels whereby like components of an image occupy adjacent storage locations. A 3×2 8-bit *RGB* image could be stored in the order RRRRRRGGGGGGBBBBBB. Also known as *band sequential* (BSQ), *plane interleaved,* or (confusingly, in Photoshop) *per channel.* Distinguished from *interleaved, component interleaved* (page 628), and *band-interleaved by line (BIL)* (page 594).
PNM	Pulse number modulation: A process that produces apparently continuous sample values by altering the number of pulses present per frame time. Each pulse is typically much shorter than a frame time. Sometimes loosely called PWM; see page 649.
postage stamp	When an image at 16:9 aspect ratio is padded with bars top and bottom for conveyance in a 4:3 container, then subsequently the 4:3 container is padded out to 16:9 format by the

	insertion of bars on both sides, the result is a “postage stamp” image having correct aspect ratio but shrunk to 56.25% of the intended area. Distinguished from *letterbox* (page 632) and *pillarbox* (page 646).
power function	A function of the form $y = x^a$ (where a is constant). Distinguished from *exponential function*, which has the form $y = a^x$ (where a is constant). Gamma correction in video is often approximated as a power function $V = L^{\gamma_E}$, where γ_E (the *encoding gamma*) symbolizes a numerical parameter typically having a value between about 0.4 and 0.5.
power loading	A phenomenon of PDP signal processing whereby above a certain requested tristimulus value (R, G, or B) the displayed value is reduced from that requested. The power limit is imposed to prevent excessive heating of the panel. See *ABL*, on page 590. Sometimes inaccurately called *luminance loading.*
precision	The degree to which repeated measurements under unchanged conditions show the same results. Distinguished from *accuracy* (page 590).
proc amp	Processing amplifier: A unit of processing equipment that has at least GAIN and OFFSET (BLACK LEVEL) controls, and if intended for composite video, CHROMA GAIN and CHROMA PHASE controls.
production aperture	The active samples of a video format: The pixel array, comprising S_{AL} image columns and L_A image rows.
progressive	A scanning standard or image format in which spatially adjacent picture lines are associated with consecutive periodic (or identical) instants in time. Examples are 1080*p*24 and 720*p*60. Distinguished from *interlace*. See also *PsF*, below.
ProRes	A family of proprietary intraframe codecs developed by Apple, compressing 10-bit video to data rates between about 82 Mb/s and 264 Mb/s.
proxy	A subsampled (relatively low pixel-count) image, or sequence of images, used to represent an image or a sequence that would require substantial storage, network, or compute resources if represented directly.
pseudocolour	Image data wherein each pixel (typically 8 bits) represents a scalar colour index value. Pseudocolour image data is accompanied by a CLUT (see page 603) that maps each possible index value to an *R'G'B'* triplet. Each encoded data value is a whole number (typically 0 … 255) proportional to the $^1/_{2.2}$-power of the associated additive *RGB* display tristimulus value.
PsF	Progressive segmented frame: A transport scheme for progressive imagery whereby image rows are rearranged to resemble an interlaced scheme for transmission or recording. Unlike interlace, the first and second fields of the transmission format are temporally coincident and are properly displayed during the

	same time interval. No vertical filtering is necessary at origination. Correct reconstruction is achieved by simply weaving the transmission fields back into a frame. The technique is ordinarily used at 24 fps; see *24PsF,* on page 587.
PSF	Point spread function.
PSNR	Peak (to peak) signal to noise ratio. The ratio – ordinarily expressed in decibels, computed as 20 times the base-10 log of the arithmetic ratio – of peak-to-peak signal to RMS noise. PSNR can apply to a physically related quantity such as luminance, but is usually applied to a nonlinearly coded quantity such as luma or *R'G'B'*.
pulldown	The term is ambiguous, potentially referring to two different processes used individually or in combination: **1** 2-3 pulldown: The process of converting 24 fps material to 60 fps, or 23.976 fps material to 59.94 fps, by repeating first 2 pictures then 3 pictures. See *2-3 pulldown,* on page 582. **2** Alteration of the picture rate of video or the sampling rate of digital audio by the factor $^{1000}/_{1001}$ (about –0.1%), by reclocking without altering any sample values; for example, to represent 30.00 fps video at 29.97 fps, or audio recorded at 48.048 kHz at 48 kHz.
pullup	Alteration of the picture rate of video or the sampling rate of digital audio by the factor $^{1001}/_{1000}$ (about +0.1%), by simply reclocking without altering any sample values; for example, to represent video recorded at 29.97 fps at 30.00 fps.
PW, peak white	*See peak white,* on page 645.
PWM	Historically, pulse width modulation: A process that produces apparently continuous sample values by altering the width of each pulse in a train of binary (off/on) pulses occurring at a fixed, high rate. Strictly speaking, modern displays such as PDP and DLP do not use PWM; see *PNM,* on page 647.
QAM	Quadrature amplitude modulation: A modulation system wherein two information signals independently modulate two subcarriers that are in *quadrature* (that is, offset in phase by 90°), which are then summed to form the modulated subcarrier. An analog version of QAM, usually called just *quadrature modulation*, combines two colour difference components onto a colour subcarrier in NTSC and PAL composite video. A digital version of QAM is used for RF modulation in some digital television transmission systems (e.g., 16-QAM, 64-QAM), particularly for cable television.
QCIF	Quarter common intermediate format: In ITU-T standards for videoconferencing (e.g., H.261), a progressively scanned raster with 4:2:0 chroma subsampling having 176×144 luma samples at 29.97 frames per second. QCIF image data is ordinarily

	subsampled from SD. See *CIF, common intermediate format,* on page 602.
QPSK	Quadrature phase-shift keying: A modulation system wherein the modulating signal alters the phase of a carrier (or subcarrier). In video, digital QPSK is used for RF modulation.
Quad-HD	Video systems having 3840×2160 image format, comparable to 4 K.
quantization	The process of assigning a discrete, numbered level to each of two or more intervals of amplitude of a data value. (In video or audio, there are typically hundreds or thousands of intervals.) In the usual *uniform quantization*, the *steps* between levels have equal amplitude. Quantization can be performed upon sample values in the time domain (e.g., audio), upon sample values in the spatial domain (e.g., images or video), or upon transform coefficients (e.g., in DCT or wavelet-based compression algorithms).
quasi-interlace	Term in consumer electronics denoting progressive segmented frame; see *PsF,* on page 648.
quasilog	Partially log. Typical “log *RGB*” image coding is usually more accurately called quasilog because a small offset (often about $^{1}/_{60}$ or $^{1}/_{90}$ of the reference black-to-white range in the linear-light scale) is introduced prior to taking the logarithm.
QuickTime	**1** Apple's trademark identifying a system for encoding, recording, decoding, and playing back realtime media on computers. **2** A proxy represented in a QuickTime format.
quincunx sampling	Synonymous with *offset sampling;* see page 643. The term became obsolete in the early 1990s, but was resurrected around 2008 for stereoscopic video. The term stems from the arrangement of club, diamond, heart, or spade symbols on a playing card of value five. The term is misleading: There is nothing special about the eponymous geometrical arrangement of five samples, and it would be nonsensical to tile an image array with that pattern.
R'G'B'	Red, green, and blue nonlinear primary components. The prime symbol makes gamma correction explicit: *R'*, *G'*, and *B'* denote *RGB* tristimulus signals that are intended to be converted to tristimulus through a power function *EOCF* (page 617) having an exponent between about 2.0 and 2.6. The precise colour interpretation of *RGB* values depends on the characteristics of the *RGB* primaries; see *RGB*, below.
R'–Y'	See *B'-Y', R'-Y',* on page 594.
radiance	Radiant flux per unit projected area.

raw	Image data encoding wherein no picture rendering and no chroma subsampling has taken place. Usually, image data is in scene-referred, linear-light form (though some systems use nonlinear conversion functions); usually, no compression has been applied (though some systems use wavelet or other compression schemes).
radiometric (*adj.*)	Relating to a quantity of light, proportional to power or energy, that has been weighted uniformly across some (wide) region of the spectrum. Distinguished from *photometric (adj.)* (page 646).
RCT	In JPEG2000, reversible colour transform: A mapping producing $Y'C_BC_R$ from $R'G'B'$ using simple binary integer luma coefficients $[{}^1/_4, {}^1/_2, {}^1/_4]$. Called “reversible” because it is perfectly invertible in fixed-point binary arithmetic without roundoff error. Compare *ICT,* on page 627.
Rec. 601	See *BT.601,* on page 597.
Rec. 709	See *BT.709,* on page 598.
raster	The pattern of parallel horizontal scan lines that paints out a picture in a system that uses scanning. The raster is the static spatial pattern that is refreshed with successive frames of video. Historically relates to an entire scanning pattern (including blanking intervals); in modern usage, may relate to image alone.
reference black	The reference level for abstract video signal value 0. In an 8-bit system, code 16; in a 10-bit system, code 64; historically, in 480*i*, 7.5 units; in all other systems, 0 units. The absolute luminance displayed for reference black in the studio depends upon the studio technician's setting of BLACK LEVEL (using PLUGE); for 100 nt white, it is typically between 0.01 nt and 0.1 nt.
reference picture	In interframe compression, a picture, or coded picture information that is available as the basis for prediction of a subsequent picture in transmission order. (It is misleading to refer to an reference *frame*, because when MPEG-2 or H.264 is used with interlaced video, a picture may comprise a single field.) Also called *anchor picture.*
reference white	The video signal level corresponding to white (100 units by definition), or the corresponding relative or absolute luminance. In video, it is standard for reference white to have the colorimetric properties of CIE Illuminant D_{65} (except in Japan and in certain other Asian regions where the standard white reference is 9300 K). As I write, there is no effective standard for the luminance of white, but most studios use about 100 nt.
relative luminance	See *luminance, relative,* on page 636.
rendering	**1** Depiction of a scene, usually involving an æsthetic treatment.

2 A software process to convert a synthetic scene described in geometric primitives into a raster image, or a sequence of raster images. The process is usually slower than realtime.

3 (picture rendering) Application of algorithmic processes to raster image data to impose the appearance associated with a particular display and/or viewing condition.

reordering

In compression systems such as MPEG-2 and H.264, when B-pictures are used they are reconstructed at the decoder from reference pictures that must already be present at the decoder. Transmission order is therefore different from display order.

resampling

The process of estimating, from a given set of samples, the samples that would have been produced had sampling taken place at different instants or at different positions.

resolution

A heavily overloaded term that, strictly speaking, refers to spatial or temporal properties of a bandlimited continuous analog signal or its sampled digital representation:

1 Generally, a measure of the ability of an imaging system or component, or of human vision, to delineate picture detail.

2 In image science, horizontal resolution in cycles per picture width [C/PW] is the maximum number of line pairs (where each “pair” comprises a black line and a white line) that can be visually discriminated from a test chart containing vertically disposed alternating black and white lines (square wave).

3 In image science, vertical resolution in cycles per picture height [C/PH] is the maximum number of cycles that can be visually discriminated per picture height from a test chart containing horizontally disposed alternating black and white lines.

4 Traditionally, in analog video, if unqualified by *horizontal* or *vertical*, horizontal resolution: twice the number of vertical black and white pairs (cycles) that can be visually discerned across a horizontal distance equal to the picture height, expressed in TVL/PH or colloquially, “*TV lines*” (see page 662). Horizontal resolution in video is sometimes expressed in units of megahertz. Also known casually as *limiting resolution;* see IEEE Std. 208.

5 In computing, *resolution* usually refers to pixel count, the count of image columns and image rows of a device or an image (that is, the number or columns and rows in the pixel array), without regard to the amount of picture detail carried or displayed.

6 Often improperly used to refer to the number of quantization levels (or bits per sample).

7 Often improperly used to express what is properly called sample density, for example in “dots per inch” [dpi].

reverse telecine	See *inverse telecine,* on page 629.
RF modulation	In video, a composite video signal that has been modulated onto a *radio frequency* (VHF or UHF) carrier in the range 50 MHz to 1 GHz. RF-modulated video in electrical form is usually conveyed with coaxial cable using Type-F (cable TV) connectors. Historically, NTSC consumer video signals conveyed from a video source to a receiver are often RF modulated onto channel 3 or channel 4 (both VHF).
RFF, repeat first field	In a compressed video bitstream (such as MPEG-2 or H.264) conveying an interlaced picture, a bit that asserts that the first field should be repeated to reconstruct a display at 2.5 times the frame rate represented in the compressed sequence. RFF enables the decoder to reconstruct the intended 2:3 pulldown.
RGB	**1** Strictly, red, green, and blue tristimulus components (linear-light). The precise colour interpretation of *RGB* values depends on the *chromaticity coordinates* of the primaries and the chromaticity coordinates of reference white. Different primary chromaticities are specified by the FCC 1953 NTSC standard (obsolete), SMPTE RP 145, EBU Tech. 3213, and BT.709; however, BT.709 is ubiquitous today in the consumer domain. **2** Loosely, red, green, and blue nonlinear primary components, properly denoted *R'G'B'* (page 650).
RGBA	**1** In video and computing, red, green, blue, and alpha (key). **2** In LEDs, red, green, blue, and amber.
RMS	Root-mean-square: A set of numbers, individually squared, then arithmetically averaged (to form the mean value), then square-rooted. (The order of operations is right-to-left: *square,* then *mean,* then *root.*) Equivalent to Euclidean distance.
rolling shutter	In an image sensor, operation (or resulting data) wherein exposure of successive image rows is delayed by a certain fixed time interval from the previous row. The last image row is exposed a significant fraction of the frame time later than the first row. (Opposed to *global shutter;* see page 623.)
RSR	Relative spectral response: The response of one colour channel of a sensor or camera to an elemental band of wavelength, in arbitrary units having dimensions of flux per unit wavelength.
RTP	Realtime Transport Protocol: A set of IP protocols for realtime streaming (typically, of video) defined by RFC 3550.
rushes	Synonym for *dailies* (see page 610).
S-HD, S-HDTV	Super high-definition [television]: Experimental video systems with a 3840×2160 image format (frame rate is undecided).

S-video	An analog interface conveying luma (Y') and quadrature-modulated chroma (C) separately on a specific four-pin mini-DIN connector. S-video is not exactly component video, and not exactly composite video. There are three types of S-video: *S-video-525*, *S-video-525-J* (used in Japan), and *S-video-625*. S-video uses quadrature modulation. If a VCR has an S-video interface, its S-video signal almost certainly does not exhibit frequency interleaving – the colour subcarrier is likely incoherent.
S_{AL}, samples per active line	The count of luma samples in a scan line that are permitted by a scanning standard to convey picture (and the associated blanking transitions). Digital video systems typically store just active samples. In modern terms, the number of columns of image samples (N_C).
S_{PW}, samples per picture width	The number of samples in a scan line corresponding to the width of the picture, measured at the 50% points of a white flatfield. In modern terminology, the count of image columns.
S_{TL}, samples per total line	The number of sample intervals between consecutive 0_{H} instants in a scanning standard.
S/PDIF	Sony/Philips digital interface: An interface specified in IEC 60958 (formerly IEC 958) for uncompressed consumer digital audio.
S3D	Stereoscopic, 3-dimensional: Imagery created, acquired, processed, or displayed with a pair of views, one intended for the viewer's left eye and the other for the right. The term *S3D* distinguishes stereoscopic imagery from planar raster images created from 3D models (e.g., computer-generated imagery).
SAD	Sum of absolute differences. A particular similarity metric commonly used in algorithms for motion estimation whereby pixel values of a reference macroblock are subtracted from pixel values from a displaced target macroblock, absolute values of the differences are formed, then the sum is computed. The task is computationally intensive. See also *MAE,* on page 638.
sample	**1** The value of a bandlimited, continuous signal at an instant of time and/or space. Usually, but not necessarily, quantized. **2** Component; see page 606. See also S_{PW}*, samples per picture width* and S_{TL}*, samples per total line,* on page 654.
sampling, 1-D	The process of forming, from a continuous bandlimited one-dimensional function of time, a series of discrete values, each of which is a function of the distribution of intensity across a small time interval. *Uniform sampling*, where the time intervals are of equal duration, is nearly always used.

sampling, 2-D	The process of assigning, to each element of a sampling grid (or lattice), a value that is a function of the distribution of intensity over the corresponding small area of the image plane. In digital video and in conventional image processing, the samples lie on a regular, rectangular grid.
SAR, sample aspect ratio	The ratio of horizontal distance between luma (or RGB) samples to vertical distance between samples. HD standards have square sampling (where SAR is unity).
saturation	**1** The condition that a signal has reached the maximum value that can be carried on the circuit or channel on which it is being carried. More accurately called *clipping* (see *clip,* on page 603).
	2 The condition that exposure to a high light level in a scene has caused an image sensor (or channel) to reach *saturation* (**1**).
	3 In colour science, colour saturation: *The colourfulness of an area judged in proportion to its brightness* [CIE]. Subjective, by definition. Saturation runs from neutral grey through pastel to saturated colours. Roughly speaking, the more an SPD is concentrated at one wavelength, the more saturated the associated colour becomes. A colour can be desaturated by adding light that contains power distributed across a wide range of wavelengths.
	4 The radius of a colour difference value as displayed on a vectorscope (in polar coordinates), ordinarily in C_B, C_R coordinates for component digital video (or historically in P_B, P_R coordinates for component analog video or U, V coordinates for composite video). Loosely, chroma.
	5 SATURATION: User-accessible means to adjust colour *saturation* (**4**), typically implemented by altering chroma (C_B, C_R) gain; preferably labelled CHROMA.
SAV	Start of active video: A sequence of four words inserted into a 4:2:2 component digital video data stream, marking the start of active samples on a line. See also *EAV*, *TRS*.
sawtooth artifact	See *zipper artifact.*
SbS	Side-by-side: A frame-packing scheme for stereo 3-D whereby left and right images are subsampled 2:1 horizontally then assembled horizontally into one picture for transmission.
scaling (*v.*)	**1** Generally, and in mathematics, multiplying by a constant factor.
	2 With respect to image data samples, the process of multiplying by a constant factor (for example, to effect video gain).
	3 Term used in desktop computing and consumer electronics referring to the process of resizing – by the same factor, and without temporal filtering – every image in an image sequence.

	Scaling converts from one image format to another having different spatial structure (image row and column counts), interpolating to increase pixel count or decimating to decrease pixel count. Also called *upconversion* or *downconversion;* historically, *scan conversion*.
scaler	A subsystem to accomplish *scaling* (3). Not to be confused with *scalar.*
scan conversion	Conversion, without temporal filtering, between video signals having different sampling structures.
scanning standard	The set of parameters of raster scanning of video equipment or a video signal. Historically, a scanning standard was denoted by its total line count and its field rate (in hertz), separated by a virgule (slash); for example, 525/59.94, 625/50, or 1125/60. Interlace was implicit. Modern preference is to provide image format notation, comprising the count of picture lines, *p* for progressive or *i* for interlace, and the frame rate; for example, 480*i*29.97, 576*i*25, 720*p*60, 1080*i*30, or 1080*p*60.
scene-linear	An image signal whose pixel component values are proportional to tristimulus values (or estimates of them) in the scene.
scene-referred	An image signal that has a well-defined mapping (such as a quasi-power function, or a quasilog function) to tristimulus values (or estimates of them) in the scene.
scotopic	Pertaining to vision adapted to very low light levels, below about 0.1 nt, where only the retina's rod cells are active.
screen (*n.*)	**1** Generally, a surface upon which a display image is formed. **2** An archaic term; see *bias* (**2**).
SD	Standard definition (video). There is no official definition, but generally, a video system having frame rate 23.976 Hz or greater whose digital image comprises fewer than about $^{3}/_{4}$ million pixels. The most widely deployed SD studio and broadcasting systems are 480*i* (see page 588) and 576*i* (see page 588). See also SDTV, below. The term *SDTV* implies transmission or recording of entertainment programming.
SDI	Serial digital interface: A SMPTE-standard studio video interface having data rate between 143 Mb/s and 360 Mb/s (usually 270 Mb/s). Usually, uncompressed SD video is conveyed, though the SDTI variant may be used to wrap compressed data (such as from D-7). See also *HD-SDI,* on page 624.
SDI-safe	A signal that excludes codes 0 and 255 (in 8-bit code) or 0–3 and 1020–1023 (in 10-bit code), avoiding the TRS codes required for synchronization across an SDI or HD-SDI interface.

SDTI	Serial data transmission interface: A SMPTE-standard variant of SDI used to convey arbitrary data (or compressed video data) instead of uncompressed digital video.
SDTV	Standard-definition television. See SD, above. *SDTV* implies transmission or recording of entertainment programming.
SECAM	*Séquential couleur avec mémoire:* An obsolete *composite video (**1**)* system based on line-alternate *B′–Y′* and *R′–Y′* colour difference signals, frequency modulated onto a subcarrier, then summed with luma. Neither quadrature modulation nor frequency interleaving was used. SECAM was used for broadcast in certain countries with 576*i* scanning (e.g., France and Russia). There was no SECAM production equipment: 576*i* component equipment or PAL composite equipment was used instead; signals were transcoded to SECAM at transmission.
second field	In interlaced scanning, the second field of the pair of fields comprising a frame. In analog 480*i*, the field containing the top image row (whose first equalization pulse starts midline). In analog 576*i*, the field whose first broad pulse starts at midline. (See also *field dominance*, and *first field, on page 619*.)
serration	**1** In bilevel analog sync, the interval between the end of a broad pulse and the start of the following sync pulse. This term refers to the absence of a pulse rather than the presence of one, and is deprecated in favor of the terms *equalization*, *broad*, and *normal* sync pulses. **2** See *zipper artifact*.
setup	Reference black level expressed as a percentage of the blanking-to-reference-white excursion. In modern video systems, blanking level and black level are identical, so setup is zero. Setup is 7.5% in 480*i* SD in North America. Setup is properly expressed as a percentage: It is incorrect to express setup in voltage, level, or units. See also *pedestal,* on page 645. Confusion notwithstanding, PAL has always had zero setup, and setup has never been present in digital video or HD.
shading (*v.*)	With respect to camera signal processing, the act of imposing position-dependent *R*, *G*, and *B* gain alteration in order to overcome technical deficiencies of a camera optical system. See also *paint*, *painting (v.),* on page 644.
SHV	Super Hi-Vision; see *Super Hi-Vision* (page 659).
shoulder	That portion of the response of photochemical film at relatively high exposure where the slope of the transfer function decreases significantly compared to the slope at midscale. In camera negative film, scene highlights are exposed on the shoulder. (Video uses the term *knee.*)
sidebar format	An image with an aspect ratio such as 4:3 conveyed or presented in a format having a wider aspect ratio (such as

	16:9), using the full height of the widescreen format but not using the full width. Synonymous with *pillarbox* (see page 646).
sinc	*Sinus cardinalis*, the function $^{(\sin \pi x)}/_{\pi x}$. An argument of 1 corresponds to the sampling frequency. Loosely known as “sine *x* over *x*,” see below.
sine *x* over *x*	Colloquial expression for *sinc;* see above (but here π is absent).
SMPTE	**1** Society of Motion Picture and Television Engineers: A professional society and ANSI-accredited standards-writing organization.
	2 In the context of colour standards: SMPTE RP 145, which defines the chromaticities historically used in 480*i* video.
SMPTE 170M	The defining standard for 480*i*29.97 (525/59.94) NTSC studio video.
SNR	Signal to noise ratio. The ratio – ordinarily expressed in decibels, computed as 20 times the base-10 log of the arithmetic ratio – of peak-to-peak signal to RMS noise. SNR can apply to a physically related quantity such as luminance, or to a nonlinearly coded quantity such as luma. See also *PSNR* (page 649).
SPD	Spectral power distribution: The amount of power (or more usually, radiance) per unit wavelength (in SI units, per nanometer).
SPTS	Single program transport stream: In MPEG-2, a transport stream that contains a single program.
square sampling (“square pixel”)	Image sampling wherein horizontal sample pitch is identical to vertical sample pitch. Use *square sampling* instead of “square pixel”: The latter term offers the possibility of round pixels.
sRGB	Formally, IEC 61966-2-1: An international standard for *R’G’B’* colour image coding; sRGB incorporates BT.709 primary chromaticities, is display-referred, and has a 2.2-power EOCF.
SStP	Simple Studio Profile: A profile defined in MPEG-4 Part 2. SStP is used by Sony in HDCAM SR.
standards conversion	Conversion, including temporal filtering, of a video input signal having one scanning standard into an output signal having a different image format and a different frame rate. Historically, the output signal had similar pixel count to the input, for example, a 480*i*-to-576*i* standards converter (loosely known as an NTSC-to-PAL standards converter), though today “standards conversion” may incorporate upconversion or downconversion. See *scanning standard*, on page 656. See also *transcoding*, *scan conversion*, *downconversion*, and *upconversion*.

stereoscopic	A system that acquires, processes, records, transmits, and/or displays two separate views of a (possibly synthetic) scene, one destined for the viewer's left eye and another for the right.
stop (*f*-stop, *T*-stop)	Factor of 2 (or $10^{0.3}$) of light power, radiance, luminance, tristimulus value, or other amplitude value.
studio-swing	An R', G', B', or Y' signal having reference black-to-white excursion of $219 \cdot 2^{k-8}$, where k ($8 \le k$) is the number of bits in the representation. It is standard to add an offset of $+16 \cdot 2^{k-8}$ at a digital video interface, so the studio-swing range is 16 to 235 at an 8-bit interface and 64 to 940 at a 10-bit interface. In a $Y'C_BC_R$ system, C_BC_R are scaled to have $^{224}/_{219}$ the excursion of Y'. Distinguished from *full-swing* (see page 620).
stutter	A motion artifact whereby image elements in motion appear to move intermittently instead of smoothly, typically caused by temporal disturbances at frequencies of 10 Hz or lower.
subfield, subframe	In a pulse-duration modulated display such as a PDP or DLP, temporal repeating or upsampling of the video signal such that the information is dispersed temporally so as to be relatively uniformly distributed across the frame time. In 2011 it is typical to have between 8 and 12 subframes per frame at 60 Hz. (See also *bit splitting* on page 595.)
subpixel	**1** In a fixed-pixel sensor or display using spatial multiplexing of colour, a colour component of a pixel. A pixel typically comprises red, green, and blue subpixels. **2** In spatial resampling, the number of potential interpolated (synthetic) positions between two original samples.
Super Hi-Vision	NHK's term for Ultra-HDTV (U-HD, U-HDTV) experimental equipment with an image format of about 7680×4320 ("8 K").
superblack	The condition where an $R'G'B'$ component signal or a luma signal within a picture is intentionally sustained below reference black level for more than a few sample intervals. In studio video, the practice has historically been discouraged.
superwhite	The condition where an $R'G'B'$ component signal or a luma signal within a picture is intentionally sustained above reference white level for more than a few sample intervals. Specular highlights commonly excurse briefly into the headroom region; the term *superwhite* generally relates to content other than speculars. In some consumer gear (e.g., PS3), *SuperWhite* is an option that enables the interface to carry footroom and headroom codes; when not set, interface codes are clipped.
SVC	Scalable video coding: In H.264, A mechanism standardized in Annex G to convey information structured in a hierarchical manner to allow of portions of the bitstream at lower bit rate than the complete sequence to be extracted to enable decoding of pictures with multiple image structures (for

	sequences encoded with spatial scalability), pictures at multiple picture rates (for sequences encoded with temporal scalability), and/or pictures with multiple levels of image quality (for sequences encoded with quality scalability).
SxS	Sony flash-memory storage media (pronounced *ess-by-ess*); based upon ExpressCard/34 (not to be confused with CardBus, although both were developed by the PCMCIA organization).
sync (*n*.)	**1** A signal comprising solely the horizontal and vertical timing elements necessary to accomplish synchronization.
	2 The component of a video signal that conveys horizontal and vertical synchronization information.
	3 *sync level;* see below.
sync (*v*.)	Synchronization, to a video source, of the scan timing of receiving, processing, or display equipment. See also *genlock*, on page 623.
sync level	The analog level of synctip. Conventionally –40 units ($-285\frac{5}{7}$ mV) in System M, and –300 mV in other systems.
System M	Formerly CCIR System M; now properly referred to as ITU-R System M: An archaic designation specifying 480*i* scanning along with certain analog and RF transmission parameters.
TaB	Top-and-bottom: A frame-packing scheme for stereo 3-D whereby left and right images are subsampled 2:1 vertically, then assembled vertically into one picture for transmission.
tandem	Two or more systems or subsystems that are cascaded in series.
telecine	Equipment to scan motion picture film in realtime to produce video or HD. Distinguished from *film scanner,* which typically is slower than realtime. Pronounced *tell-e-SIN-eee.*
theatre black	Luminance (relative or absolute) of a cinema screen when the projector is commanded to produce no light.
timecode	A number of the form *HH:MM:SS:FF* (hours, minutes, seconds, frames) that designates a single frame in a video or cinema motion image sequence.
timing	See *grading,* on page 623. The term *timing,* now obsolete, originated with control of exposure time in photographic printing.
TINT	User-accessible means to adjust hue. Artists use the term "tint" to refer to adding white to a colour, thereby decreasing chroma, so it isn't clear whether TINT is related to hue or to chroma. To avoid ambiguity, this control should be called HUE. (In composite NTSC studio equipment, often called PHASE.)

toe	**1** In photochemical film, that portion of the response at relatively low exposure where the slope of the transfer function in log-log coordinates is significantly less than its slope at midtone exposure. See also *shoulder,* on page 657.
	2 In video, that portion of the camera's OECF that lies within about 2% of optical black. See also *knee,* on page 631.
top field	The field that contains the uppermost coded image row of a frame; typically the second field in 480*i* and 1080*i* and the first field in 576*i*.
TFF, top field first	In a compressed video bitstream (such as MPEG-2 or H.264) conveying an interlaced picture, a bit that asserts that the field containing the uppermost image row should be displayed first in temporal order. Typically negated for 480*i* and HD and asserted for other formats.
transcoding	**1** Traditionally, converting a video signal having one colour-encoding method into a signal having a different colour-encoding method, without altering the scanning standard; for example, 576*i* PAL to 576*i* SECAM.
	2 In compressed digital video distribution, various methods of recoding a compressed bitstream, or decompressing then recompressing.
transition sample	An image data value near the left or right edge of the picture whose amplitude is reduced or forced to blanking level so as to limit the high-frequency content of the video signal at the picture edges.
TRC	Tone reproduction curve, or tone response curve: A (generally nonlinear) function that relates a greyscale or colour signal value to a physical characteristic such as reflectance or relative luminance. The term is common in graphics arts.
triad	In image science, a set of three discrete colour emitting or modulating elements – almost always red, green, and blue – that form an elemental colour picture element. Triads regularly tile the display surface. In an *FPD* (page 620), triads are individually addressable.
trichromaticity	The property of human vision whereby additive mixtures of exactly three properly chosen primary components are necessary and sufficient to match a wide range of colours. That such matching is possible is surprising considering that physical spectra are infinitely variable; but not surprising considering that the human retina contains just three types of colour-sensitive photoreceptor (cone) cells.
trick mode	Achieving usable image display when modes such as pause, fast-forward, rewind, and slow-motion are used in video playback (particularly from videotape).

trilevel sync	Analog HD sync information conveyed by a pulse having a transition from blanking level to +300 mV, then a transition from +300 mV to –300 mV, then a final transition back to blanking level. Standard for HD. Distinguished from *bilevel sync*, used in SD.
tristimulus	One of a set of three component values that together represent relative spectral radiance weighted by a spectral sensitivity function having significance with respect to the trichromaticity of human vision (see *trichromaticity*, above). Tristimulus values (such as L, M, S; R, G, B; or X, Y, Z) are proportional to intensity. RGB signals are typically subject to *gamma correction,* forming $R'G'B'$, as part of their conversion into video signals.
TRS	**1** Timing reference signal: A sequence of four words across an SDI or HD-SDI interface that signals sync. See *EAV,* on page 616, and *SAV, on page 655*. **2** Tip, ring, sleeve: A connector – ordinarily 6.35 mm ($^{1}/_{4}$-inch) in diameter – typically used for unbalanced analog stereo audio.
truecolour	Image data containing encoded values of independent additive components red, green, and blue (symbolized $R'G'B'$); typically, each encoded data value is an 8-bit quantity proportional to the $^{1}/_{2.2}$-power of the associated display tristimulus value.
TrueHD, True 24*p*, etc.	“Truth is relative. It is only a matter of opinion.” [Pythagoras] Truth is probabilistic, dependent upon speaker and listener, and upon historical, cultural, social, and technical context. TrueHD, True 24*p* – or “true” anything else – must be qualified by context in order to be meaningful. Be suspicious when a provider of information finds it necessary to tag it with “truth” (or related terms like *fact* or *myth*). For further enlightenment on this subject, consult a Zen master.
TV lines per picture height, TVL/PH	A unit of resolution – in the horizontal, vertical, diagonal, or any other direction – equivalent to half of a cycle spanning the picture height. If a fine vertical pattern of 400 dark/light line pairs (cycles) can just barely be discriminated in a system having 4:3 aspect ratio, then resolution is $^{3}/_{4} \cdot 2 \cdot 400$ or 600 TVL/PH. One TV line corresponds to a pixel, or half a cycle; or in film, one line width or half a line pair. C/PH is preferred; see page 599; C/PW may also be used. See also *resolution,* on page 652.
U, *V*	**1** Historically, colour difference components, blue minus luma [$B'–Y'$] and red minus luma [$R'–Y'$], scaled by the factors 0.492111 and 0.877283, respectively, such that after quadrature modulation the reference excursion of the composite video signal is contained within the range $-\frac{1}{3}$ ($-33\frac{1}{3}$ units) to $+\frac{4}{3}$ ($+133\frac{1}{3}$ units). See also [C_B, C_R]; [I, Q]; [P_B, P_R]. **2** In modern usage, the symbols U and V refer to unscaled $B'–Y'$ and $R'–Y'$ components; to C_B and C_R components scaled for component digital transmission; or (most commonly) to P_B and

	P_R components scaled ±0.5 – that is, having the same reference excursion as luma (Y'). (There are many exceptions and errors in the scale factors.)
Ultra-HDTV (U-HD, U-HDTV)	Ultra high-definition [television]: Experimental video systems with a 7680×4320 image structure (frame rate is undecided). Also known as *Super Hi-Vision.*
uncompressed	In video, signal recording or transmission without using JPEG, M-JPEG, MPEG, or wavelet techniques. (Chroma subsampling effects lossy compression with a ratio of about 1.5:1 or 2:1; however, video with chroma subsampling is deemed *uncompressed;* in video, the term *compression* is reserved for transform techniques.)
unit	Loosely, video level expressed in percentage: one-hundredth of the excursion from reference black level (0 units) to reference white level (100 Iunits). Historically, *IRE unit*; see page 630.
upconversion	In video, conversion to an image format, usually at the same frame rate, having substantially higher pixel count (e.g., SD to HD).
upsampling	Resampling where more output samples are produced than the number of input samples provided.
V	See U, V, on page 662.
VANC	Vertical interval ancillary data: Ancillary data multiplexed into the vertical blanking interval of an SDI or HD-SDI interface.
valid	The condition where a video signal is $R'G'B'$-legal (see page 632) – that is, where none of the corresponding R', G', and B' signals exceeds its reference range except perhaps for brief transients.
value	In colour science, measures of lightness apart from CIE L^* (typically expressed in the range 0 to 10).
VBI	See *vertical blanking interval (VBI),* below.
VBR, variable bit rate	Pertaining to a compression format wherein data rate (bit or byte count per second) may vary from one frame to the next.
VC-1	A standard, developed and deployed by Microsoft as part of Windows Media 9, subsequently standardized as SMPTE 421M, for the lossy compression of digital motion images and associated audio. The VC-1 algorithm shares the basic principles of H.264, but differs in many details. VC-1 decoding is mandated in the Blu-ray disc specification.
VC-2	See *Dirac PRO,* on page 613.
VC-3	See *DNxHD,* on page 613.

VCR	Videocassette recorder. Implicitly, consumer-grade: In professional usage, VTR (with *T* for *tape*) is used even if the tape medium is encased in a cassette.
veiling glare	Light from the viewing environment reflected from a display surface, producing unwanted luminance.
vertical blanking interval (VBI)	Those scan lines of a field (or frame) that are precluded by an interface standard from containing picture. The vertical interval may contain nonpicture video, audio, or ancillary information.
vertical frequency	**1** The vertical component of spatial frequency. **2** In interlaced scanning, field rate; in progressive scanning, frame rate.
vertical interval	Vertical blanking interval (VBI); see above.
vertical sync	Those nonpicture elements of a video signal that delimit the boundary between fields or frames.
video level	$R'G'B'$ or Y' signal level, in digital video typically expressed as a percentage where 0 corresponds to reference black and 100 corresponds to reference white, or expressed as 8-bit or 10-bit digital interface code values (where the reference range is 16 through 235 or 64 through 940, respectively). Modest excursions outside the reference levels are permitted, generally just for transient content.
visual density	**1** For partially absorbing optical material (such as photochemical film), the base-10 logarithm of the reciprocal of transmission weighted by the CIE luminous efficiency function. **2** For a light emitting, transmitting, or reflecting element, the base-10 logarithm of the reciprocal of relative luminance.
VITC	Vertical interval timecode: Timecode data encoded in an analog representation and conveyed in the VBI.
VLI	Video line index; see SMPTE RP 186.
VOB	Video object: A container file format used in DVD comprising a restricted MPEG-2 program stream with the addition of private streams specifying elements such as menus and subtitles. (All VOB files are MPEG-2 program streams, but not all MPEG-2 program streams are DVD compliant.) A VOB file contains 2^{20} bytes (1 GiB) or less.
Vorbis (Ogg Vorbis)	An open-source audio compression system; part of WebM.
VP8	An open-source video compression system, originally designed and implemented by On2; part of WebM.
VSB	Vestigial sideband: An RF modulation system. Analog VSB is used in the NTSC and PAL standards for terrestrial television.

	A form of digital VSB (namely, *8-VSB*) is used in the ATSC standard for terrestrial digital television.
VTR	Videotape recorder. Implies professional: *T* for *tape* is used even if the tape medium is encased in a cassette.
VTVL	*Vertical* resolution, expressed in units of *TVL*/PH: A unit of vertical resolution used in CCTV and video security systems.
wave number	Reciprocal of wavelength, usually expressed in units of cm^{-1}.
weave	**1** In motion picture film, or video originated from film, erratic side-to-side motion of the image owing to imperfect registration of the sprocket holes to the film gate. **2** A deinterlacing technique, common in PC graphics, that merges two fields together, irrespective of interfield motion, to form a frame. Also known as *field merging.* See also *bob,* on page 596. (The terms "bob" and "weave" are said to originate from the sport of boxing.)
WebM	An open-source project sponsored by Google, or the associated video/audio files, based upon the Matroska container format, Vorbis audio compression, and VP8 video compression.
white	See *reference white,* on page 651.
window	**1** In general, a rectangular region in an image. **2** A rectangular region – often square and at the centre of an image – occupying about 10% of the total image area, filled with a uniform colour (often white).
working space	In colour management, the colour image encoding space in which the arithmetic of image manipulation is performed.
WTW	Whiter than white. See *superwhite,* on page 659.
WSS	Widescreen signalling: A mechanism implemented in analog transmission of 576*i*25 to signal widescreen image data.
x.v.Color, x.v.Colour	Sony's trademarked terms for xvYCC (see below).
XDCAM	Various systems commercialized by Sony to record and play SD and HD content from optical disc and SxS flash media using various image formats and various compression systems.
XLR3	A connector typically used for audio (either monophonic analog audio, typically about 770 mV RMS into 600 Ω impedance in balanced mode, or stereo digital according to AES3-4, in balanced mode).
xvYCC	A colour coding scheme defined in IEC 61966-2-4, purported to extend $Y'C_BC_R$ to allow chroma excursions outside the $R'G'B'$-legal range to represent wide-gamut colours. SD and HD

	versions, corresponding to BT.601 and BT.709, are defined. (The acronym is said to represent *extended video, $Y'C_BC_R$*.)
XYZ	A particular set of CIE tristimuli: linear-light quantities, where *Y* is luminance relative to a specified white reference.
X'Y'Z'	Nonlinear *XYZ* encoded for digital cinema: CIE *XYZ* tristimuli at the reference display, relative to reference white of 48 nt, each subject to a $^1/_{2.6}$-power function (inverse EOCF).
Y	**1** In physics and colour science, and when used carefully in video and computer graphics, the symbol for the CIE relative luminance tristimulus component, relative to an absolute reference white luminance. See *luminance, relative,* on page 636. **2** In video, in digital image processing, and in computer graphics, the symbol *Y* is often carelessly used to denote *luma* (properly symbolized *Y'*); see *Y'*, below.
Y'	In video, the symbol for luma: A quantity representing nonlinear *R'G'B'* primary components, each weighted by its luma coefficient. Luma may be associated with SD (BT.601) luma coefficients, or HD (BT.709) luma coefficients. *R'G'B'* components are intended to be converted to tristimulus through a power function (*OECF;* see page 643) having an exponent between about 2.0 and 2.6. *Y'* is distinguished from luminance, *Y*, which represents a weighted sum of linear-light (tristimulus) red, green, and blue primary components. Historically, the *Y* symbol in video was *primed* (*Y'*), but in modern times the prime is often carelessly elided, leading to widespread confusion with luminance.
Y'/C, *Y'/C* 3.58, *Y'/C* 4.43	Analog luma, *Y'* (not luminance), accompanied by a modulated chroma signal, *C*, quadrature modulated at approximately the subcarrier frequency indicated in megahertz. Preferably denoted S-video, S-video-525 (or in Japan, S-video-525-J), and S-video-625, respectively. May have stable or unstable timebase; may have coherent or incoherent colour subcarrier.
$Y'C_1C_2$	Luma, *Y'* (not luminance), accompanied by two colour difference signals, where the components C_1 and C_2 are specified (or evident from context) and may or may not be any of the common pairs [*B'–Y', R'–Y'*], [P_B, P_R], [C_B, C_R], or [*U, V*].
$Y'C_BC_R$	**1** In video, MPEG, and M-JPEG, luma, *Y'* (*not* luminance) accompanied by two colour difference components scaled independently to have a peak-to-peak excursion $^{224}/_{219}$ that of the luma excursion. In processing, reference black is at code 0 and reference white is at code 219, both referenced to 8-bit signals. At an 8-bit interface, luma has reference black code 16 and reference white code 235 ("studio-swing" or "normal-range"). C_B and C_R are scaled to a reference excursion of ±112; an offset of +128 is added. For processing or interfacing with more than 8 bits, additional bits provide additional precision but do not change the scaling. See *luma,* on page 635; and C_B, C_R (**1**) on

	page 598. Beware that BT.601 (SD) $Y'C_BC_R$ is coded differently than BT.709 (HD) $Y'C_BC_R$: The luma coefficients differ. **2** In JPEG/JFIF as used in computing, luma, Y' (*not* luminance) having “full-swing” or “full-range” excursion 0 through 255, accompanied by two colour difference components scaled independently to have 128±128 peak-to-peak excursion, $^{256}/_{255}$ that of the luma excursion; pure blue and pure red are clipped. See *luma,* on page 635; and *CB, CR* (**2**) on page 599. Full-swing $Y'C_BC_R$ is ordinarily coded according to BT.601 (SD) luma coefficients, independent of pixel count.
$Y'C_XC_Z$	A colour encoding space sometimes used for interfacing between a digital cinema server and a 48 Hz or stereo/3D projector, comprising $X'Y'Z'$ (see above) processed through the JPEG 2000 ICT matrix (a matrix identical to the BT.601 luma/chroma encoding); see *ICT,* on page 627.
$Y'P_BP_R$	**1** Historically, luma, Y' (not luminance) accompanied by analog [P_B, P_R] colour difference components. *P* historically stood for *parallel.* The EBU N10 standard specifies luma excursion of 700 mV and P_BP_R excursion of ±350 mV. This standard is used in 576*i* and in HD; however, various industry standard analog interfaces have been used for 480*i*. See *luma,* on page 635, and P_B, P_R, on page 645. **2** In modern usage, luma, Y' (not luminance) having reference excursion 0 to 1, accompanied by [P_B, P_R] colour difference components having reference excursion ±0.5 (that is, the same reference amplitude as luma).
$Y'UV$	**1** Historically, luma, Y' (not luminance) accompanied by two colour difference components [*U, V*] scaled for subsequent encoding into a composite video signal such as NTSC or PAL. For component analog video, [*U, V*] components are inappropriate, and [P_B, P_R] should be used. For component digital video, [*U, V*] components are inappropriate, and [C_B, C_R] should be used. See *luma,* on page 635; and *U, V,* on page 662. **2** In modern usage, the notation $Y'UV$ – or, carelessly written, *YUV* – is often used to denote any component system employing luma, Y' (not luminance), accompanied by two colour difference components derived from $B'–Y'$ and $R'–Y'$ such that *U* and *V* range ±0.5. (There are many exceptions and errors in the scale factors.)
YCM	The traditional term in the motion picture industry for a set of three black-and-white (“silver”) cinema films recording the yellow, cyan, and magenta transmittances of a movie release print, for archiving purposes.
zap delay	The delay (latency) associated with displaying newly acquired video upon changing channels.

zebra	A pattern of moving dashed diagonal lines overlaying an image indicating where image data occupies a certain range of code values, for example to indicate high exposure.
zipper artifact	An artifact associated with deinterlacing (particularly, field merging of field pairs) of source material containing rapid horizontal motion, where alternate image rows of the resulting frame contain horizontally displaced scene elements. Also called *sawtooth artifact* or *serration*.
ZOH	Zero-order hold: Retaining a sampled value for the whole duration of the sampling interval.

Index

α–ω 0–9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

About the author


sketch by Kevin Melia

Charles Poynton is an independent contractor specializing in the physics, mathematics, and engineering of digital colour imaging systems, including digital video, HD, and digital cinema. Apart from his professional work, he is a PhD candidate at Simon Fraser University.

In the early 1980s, Charles designed and built the digital video equipment used by NASA to convert video from the Space Shuttle into NTSC. In 1990, he initiated Sun Microsystems' HD research project, and introduced color management technology to Sun. He was Sun's founding member in what a few years later became the International Color Consortium (ICC).

Charles was a key contributor to current digital video and HD studio standards; he originated the number 1080 (as in 1080*p*60) in HD standards. A Fellow of the Society of Motion Picture and Television Engineers (SMPTE), he was awarded the Society's prestigious David Sarnoff Gold Medal for his work to integrate video technology with computing and communications.

Charles has taught many popular courses on video technology, HD, colour image coding, and colour science, including many SIGGRAPH courses.

Charles lives in Toronto with his wife Barbara – a psychotherapist. Their twenty-something daughter Quinn is back in Toronto after stints in Paris and New York. Their teenager Georgia is an undergrad at Dalhousie University in Halifax. It is owing to pressure from all three family members that he has reverted to spelling *colour* with a *u.* The sketch in the margin was made many years ago, just prior to Charles granting Georgia's ninth-birthday wish to shave off his beard.

This book is set in the Syntax typeface. Syntax was designed by Hans Eduard Meier, and first issued by Stempel in 1969. This book uses Linotype's revision issued in 2000 that includes bold italics, small caps, and old-style figures. The body type is 10.2 points, leaded to 12.8 points, set ragged-right.

The mathematical work underlying this book was accomplished using *Mathematica,* from Wolfram Research. The illustrations were executed in Adobe *Illustrator;* for raster (bitmap) images, Adobe *Photoshop* was used. The equations were set using Design Science *MathType.* Text editing, layout, and typesetting were accomplished using Adobe *FrameMaker.* Adobe *Acrobat* was employed for electronic distribution.

The work was accomplished using various Apple Macintosh computers.