

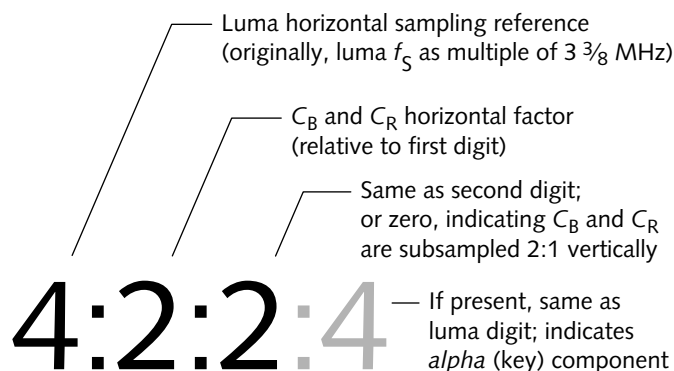
Chroma subsampling notation

Video systems convey image data in the form of one component that represents lightness (luma), and two components that represent color, disregarding lightness (chroma). This scheme exploits the poor color acuity of vision: As long as luma is conveyed with full detail, detail in the chroma components can be reduced by subsampling (filtering, or averaging). In digital video, the components are properly denoted $Y' C_B C_R$. ($Y'UV$, $Y'IQ$, and $Y'R_B P_R$ are incorrect.)

Subsampling is designated by a string of three (or sometimes four) small integers separated by colons. The relationship among the integers denotes the degree of vertical and horizontal subsampling. At the outset of digital video, subsampling notation was logical; unfortunately, technology outgrew the notation. In Figure 1 below, I strive to clarify today's notation.

The first digit originally specified luma sample rate relative to $3\frac{3}{8}$ MHz. (The commonly used leading digit of 4 is a historical reference to a sample rate roughly four times the NTSC or PALcolor subcarrier frequency, when subcarrier-locked sampling was under discussion for component video.) HDTV was once supposed to be described as 22:11:11! The leading digit has, thankfully, come to be relative to the sample rate in use. Until recently, the initial digit was always 4, since all chroma ratios have been powers of two – 4, 2, or 1. However, 3:1:1 subsampling has recently been commercialized in an HDTV production system (Sony's HDCAM), and in the SDL mode of consumer DV, and 3:1:0 has been commercialized in the SDL mode of consumer DV, so 3 may now appear as the leading digit. A leading digit of 2 is never used.

Figure 1 **Chroma subsampling notation** indicates, in the first digit, the luma horizontal sampling reference. The second digit specifies the horizontal subsampling of C_B and C_R with respect to luma. The third digit originally specified the horizontal subsampling of C_R . The notation developed without anticipating vertical subsampling; a third digit of zero now denotes 2:1 vertical subsampling of both C_B and C_R .



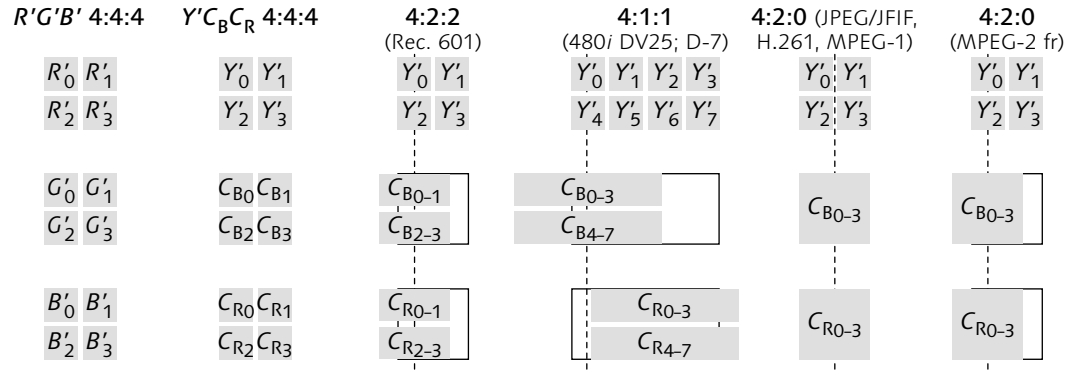


Figure 2 **Chroma subsampling.** A 2×2 array of $R'G'B'$ pixels is matrixed into a luma component Y' and two color difference components C_B and C_R . Color detail is reduced by subsampling C_B and C_R ; providing full luma detail is maintained, no degradation is perceptible. In this sketch, samples are shaded to indicate their spatial position and extent. In 4:2:2, in 4:1:1, and in 4:2:0 used in MPEG-2, C_B and C_R are cosited (positioned horizontally coincident with a luma sample). In 4:2:0 used in JPEG/JFIF, H.261, and MPEG-1, C_B and C_R are sited interstitially (midway between luma samples). In 4:2:0 used in MPEG-2 frame pictures, C_B and C_R are sited vertically midway between luma samples. In MPEG-2 field pictures, the situation is more complicated, and not sketched here.

Originally, the second digit specified the horizontal subsampling of C_B and the third digit specified the horizontal subsampling of C_R . That scheme failed to anticipate vertical subsampling, and in any event, all practical systems have the same subsampling ratios for both C_B and C_R . So, in today's notation, the second digit specifies the horizontal subsampling of both C_B and C_R with respect to luma.

The third digit now has two possibilities. If the third digit is the same as the second digit, there is no vertical subsampling. If the third digit is zero, this indicates 2:1 vertical subsampling of both C_B and C_R .

If a fourth digit is present, it must be identical to the first digit, and indicates the presence of a fourth signal channel containing transparency (key, or alpha) information, sampled identically to luma.

Several different subsampling schemes have been commercially deployed. Some of these are shown schematically in Figure 2. The left-hand column depicts a 2×2 array of $R'G'B'$ pixels. Prior to subsampling, with $R'G'B'$ each having 8 bits per sample, this 2×2 array would consume data capacity of 12 bytes.

4:4:4

Prior to subsampling, $R'G'B'$ video is denoted 4:4:4 $R'G'B'$. Each $R'G'B'$ triplet (pixel) can be transformed ("matrixed") into $Y'C_B C_R$, as shown in the second column; this is denoted 4:4:4 $Y'C_B C_R$. (Strictly speaking, subsampling has not yet taken place; however, the notation 4:4:4, akin to subsampling notation, is commonly used.)

In component digital video, data capacity is reduced by subsampling C_B and C_R using one of the schemes that I will now describe.

- 4:2:2 $Y'C_B C_R$ studio digital video according to Rec. 601, including professional DV50 systems and the 422 profile of MPEG-2 (sometimes denoted 422P) use 4:2:2 sampling. In the 4:2:2 scheme, C_B and C_R components are each subsampled by a factor of 2 horizontally; their effective positions are coincident (*cosited*) with alternate luma samples. (When sample numbers in an active line start at zero, as is standard, chroma samples are cosited with even-numbered luma samples.) The 12 bytes of $R'G'B'$ are reduced to 8, effecting 1.5:1 lossy compression.
- Progressive 483p59.94 ("dual-link") systems, sometimes denoted 4:2:2p, use 4:2:2 subsampling but that scheme is somewhat different than 4:2:2 studio subsampling: Strangely, C_B and C_R samples in 4:2:2p are centered in line-alternate fashion, and are not coincident in the image array.
- 4:1:1 Certain digital video systems, such as consumer 480i29.97 DV25, professional 480i29.97 DV25, and professional 576i25 DV25, use 4:1:1 sampling. In this scheme, C_B and C_R components are each subsampled by a factor of 4 horizontally, and cosited with every fourth luma sample. The 12 bytes of $R'G'B'$ are reduced to 6, effecting 2:1 compression.
- 4:2:0 This scheme is used in JPEG/JFIF stillframes in computing, in H.261 (for videoconferencing), in MPEG-1, in consumer 576i25 DV25, and in most variants of MPEG-2. C_B and C_R are each subsampled by a factor of 2 horizontally and a factor of 2 vertically. The 12 bytes of $R'G'B'$ are reduced to 6, yielding 2:1 lossy compression. C_B and C_R are effectively centered vertically halfway between scan lines.
- There are two variants of 4:2:0, having different horizontal siting. In JPEG/JFIF, H.261, and MPEG-1, C_B and C_R are sited interstitially, halfway between alternate luma samples. In MPEG-2, C_B and C_R are cosited horizontally.
- As I mentioned a moment ago, the 422 profile (422P) of MPEG-2 accommodates studio-style 4:2:2 subsampling.
- In this document I describe and show MPEG-2's 4:2:0 subsampling for frame-coded (progressive) pictures. For field-coded (top and bottom) pictures, the situation is more complicated; a description of chroma subsampling for field-coded pictures is outside the scope of this document.
- 3:1:1 Sony's HDCAM system, and DV HD mode, use 3:1:1 sampling, where C_B and C_R components are each subsampled by a factor of 3 horizontally. Chroma samples are cosited with every third luma sample. 36 bytes of $R'G'B'$ are reduced to 20, effecting approximately 2:1 compression.
- 3:1:0 Consumer DV25 in "long-play" (SDL) mode uses 3:1:0 sampling. In this scheme, C_B and C_R components are each subsampled by a factor of 2 vertically, and subsampled by a factor of 3 horizontally (cosited with every third luma sample). 36 bytes of $R'G'B'$ are reduced to 16, effecting 2.25:1 compression.