but we will mention it here to re-enforce that while various compression and file formats use a number of formats such as 4:1:1 and 4:2:0, baseband digital video, which is all this section is about only uses 4:2:2. That means that active video samples are comprised of ½ luminance samples, and ½ chroma samples. Of the 3840 video samples per line of 1080i, or 2880 for 720p, half (1920 or 1440) are luminance and half are chroma of which ½ the chroma are for CB (B-Y) and the other half CR (R-Y).

The CB and CR samples are co-sited with the odd (1st, 3rd, 5th,etc.) Y samples. In the case of 1080, this creates 1920 multiplexed Y and CB \CR active video samples. Thus sample output data sequence for each horizontal line is CB,Y,CR,CB,Y,CR,Y…… The last sample in the line is a Y. Increasing data values for each luminance data means a brighter sample. Black is 40h and 3ADh is peak white. This allows for 877 discrete levels for luminance. Chroma is centered at 200h because originally analog chroma signals had positive and negative values. The highest chroma value is 3C0h and the lowest value is 40h. Chroma has 896 discrete levels.

Sources, such as cameras or graphics devices, use the following formulas for luminance and chroma signals:

$$Y = 0.21R + 0.72G + 0.07B$$

$$Cr = 0.63\,(R\text{-}Y) = 0.63((1R + 0G + 0B) – (0.21R + 0.72G + 0.07B)$$

$$= 0.63((1R – 0.21R) + (0G – 0.72G) + (0B – 0.07B))$$

$$= 0.63\,(0.79R – 0.72G – 0.07B)$$

$$= 0.50R – 0.45G – 0.04B$$

$$Cb = 0.54\,(B\text{-}Y) = 0.50B – 0.39G – 0.11R$$

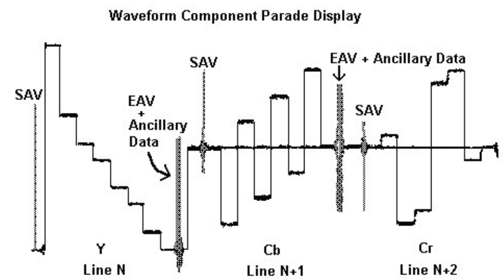

Figure 17 Analog representation of serial component digital video.

In place of sync pulses, groups of words called Timing Reference Signals (TRS) are inserted. There are 2 types of TRS strings, SAV (Start of Active Video) and EAV (End of Active Video). As their names imply they signify where the active video begins and ends. Active video and any ancillary data are not allowed to have values 000h – 003h, or 3FCh – 3FFh. Only TRS and Ancillary data headers can have values in those ranges. The first three TRS words are 3FFh, 000h, 000h. The 4th TRS word is a set of 3 flags. This word, referred to in the standard as the XYZ word uses 3 bits to indicate whether this TRS signal is:

1) An SAV or an EAV

2) In the vertical interval or not

3) Field 1 or Field 2

Most of the rest of the bits in the XYZ word are protection bits to prevent it from having illegal values below 003h and above 3FCh.

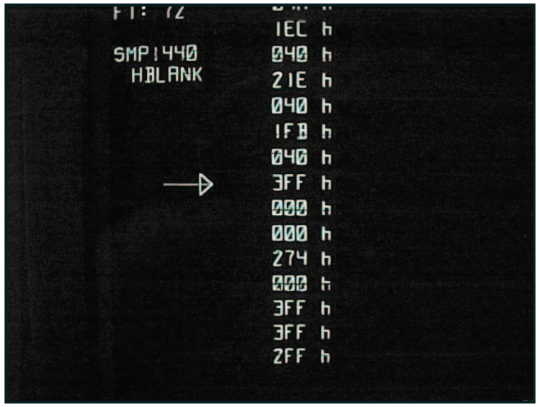

Figure 18 Hex decimal representation of EAV sequence in a SD signal

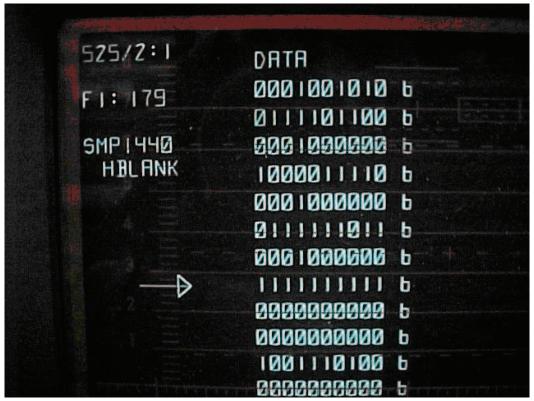

Figure 19 Same EAV sequence in Binary

The space between the EAV and SAV represents traditional horizontal blanking. 268 words are available during this time for use by ancillary data. The header for this data is 000h, 3FFh, 3FFh. Like the video Horizontal ANCillary (HANC) data consists of 10 bit words. The horizontal lines during the vertical interval can be used for ancillary data (VANC). Over the years, there has been some confusion about the use of VANC because early D1 VTRs, which where only 8 bit machines, used this data space. Originally to stay consistent digital standards specified 8 bit VANC. Therefore VANC was not used much.

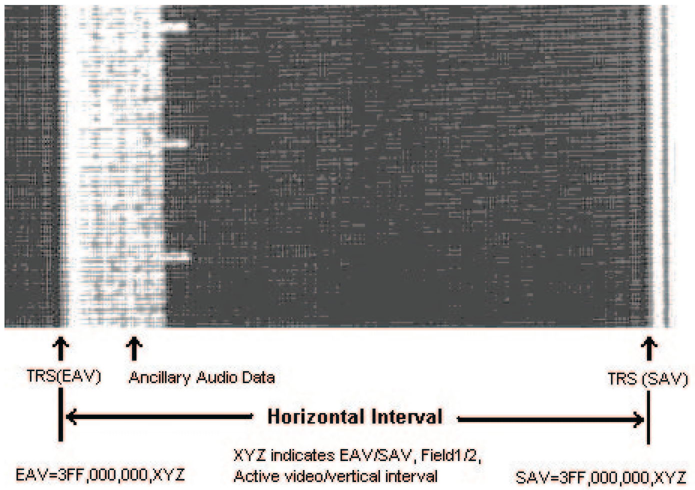

Figure 20 Video monitor in "Pulse-Cross" mode displaying embedded audio in the horizontal blanking interval. A common practice today is to embed the audio within the video—Figure 20 shows where the audio resides.

Since horizontal ancillary data in digital video is 10 bit words, and an AES frame consists of 64 bits containing up to two 24-bit audio words, some re-formatting must occur. This reformatted data will be inserted immediately after the EAV (End of Active Video) data words. They consist of three 10 bit words with the values 000, 3FF, 3FF (as opposed to 3FF,000,000 for EAV/SAV timing reference signals). Again only the EAV/SAV and ancillary data headers are allowed to have these values.

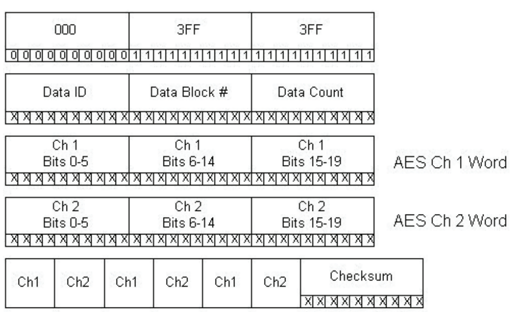

Figure 21 This chart shows how four samples of 16 – 24 bit AES audio can be mapped into 10-bit HANC space.

1440-1443	3FF	000	000	26C
1444-1447	000	3FF	3FF	2FF
1448-1451	200	212	1A0	2D3
1452-1455	21E	1A2	2D3	11E
1456-1459	170	2E8	21E	172
1460-1463	2D3	11E	198	2CC
1464-1467	21E	19A	2CC	11E
1468-1471	129	040	200	040
1472-1475	200	040	200	040
1476-1479	200	040	200	040

TRS (EAV)
Ancillary Data Preamble
Data ID
Data Block #
Data Count

Figure 22 This chart shows actual samples of AES audio being mapped into SD HANC space.

In Figure 22 you can see that the Ancillary Data Preamble immediately follows EAV. Next is the Data ID word. Its value indicates whether the audio data packets that follow are audio group 1 (channels 1-4), group 2 (channels 5-8), group 3 (channels 9-12), or group 4 (channels 13-16), or one of 4 extended audio data packets.

Next is a word that is the data block number. This number was meant to be incremented with each packet. Only 8 of the 10 bits are used to count to prevent illegal values, such as values reserved for TRS, and ancillary data preambles. The count value rolls back to 0 after reaching FF hex.

Some equipment uses a non-sequential value here from the previous packet to indicate that a switch has occurred in the audio data to a new source. A problem here early on, was that some products, such as some test equipment, did not increment this number with each packet. This caused intermittent muting in some receive equipment. If this value is not incremented, it is now kept at a value of 0h to eliminate any chance for confusion.

The next word indicates the number of data words to follow. A maximum of 255 bytes are possible because as with the data block number, only 8 of the 10 bits are used. The number of data words will depend on what horizontal line is being embedded with this data. Here is where it gets slightly confusing.

AES 20 bit samples are mapped into 3 successive 10 bit SMPTE259 words. The extra 10 bits left over ((3 * 10 bits) - (20 bit audio sample) = 10 bits) have AES audio channel status, and user bits mapped into four of them. Three bits indicate whether this is channel 1, 2, or a Z frame, which indicates the start of 192 sample AES block, and the final 3 are used to protect against illegal values, which again are reserved for EAV/SAV, or ancillary data preambles. This means that a single sample of channel 1, and 2 (1 AES frame) requires 6 SDI words. One way to implement this is to have most lines carry 6 AES samples. That's 36 SDI data words.

But the most common sampling frequency used in professional digital audio for television is 48 KHz (and the only one discussed here). This is not an exact multiple of the 29.94/59.94 field rates used commonly. This means that 8008 samples are required in the SDI stream every 10 fields. This works out to the necessary 48,000 necessary each second.

To spread these through the horizontal blanking, or horizontal ancillary data space, if most lines have 6 samples, every 12th line will have 8.

What about 24 bit audio? The first 20 bits are sent as just described, but the additional 4 bits of both channels in one AES frame are combined into 1 word and assembled into what is called an extended data packet. This extended data packet is sent immediately following the normal audio data packet. The extended data packet starts with the same preamble:

> (0000,3FFF,3FFF),
> data ID,
> Data Block Number,
> and Data Count words as the normal audio data packet.

Analog video is rapidly disappearing but many of the analog attributes still found at the digital physical layer lives on. Even the way digital video is organized is heavily reliant on methods pioneered during the analog era. Next we will look how all this data is organized into a system.