Ancillary Data and the Serial Digital Interface

By Randy Conrod

Illustrated by Michel Proulx and Randy Conrod

The previous tutorial in this series discussed EDH (error detection and handling). The EDH information is inserted into the serial data stream as ancillary data. Other signals, such as digital audio and time code, have also found their place as an ancillary data type. Signal source identification and closed captioning would be the next two obvious candidates for standardization. Now that many signals can be carried along a single coaxial, there are benefits to the broadcaster. As digital audio signals can be embedded into the serial video signal, a review of digital audio is in order.

The digital audio standard AES3-1992 is commonly referred to as AES/EBU. It was developed by the Audio Engineering Society and the European Broadcasting Union. This is another example of a worldwide standard being developed for the broadcast industry.

There are other standards used in the industry (Table 1). SDIF (Sony Digital Interface Format) is a digital standard that uses three connections: one for each channel in a stereo pair and one for a synchronization signal. The analog signal is sampled with either a 16 or 20-bit word at 44.1 or 48 kHz. This format is not used in any new equipment manufactured currently but may be encountered in some older systems. This digital format is not used to embed audio information in the serial video signal.

SPDIF is the Sony Philips digital interface format. This standard is similar to the AES/EBU format and is used quite extensively in the industry, although this signal is not embedded into the serial video signal. A single-ended transmission through RCA connectors is used as well, and there is an optical interface between equipment. This interface is used for CD (44.1 kHz sampling) and DAT (48-kHz sampling) type devices.

MADI (Multichannel Audio Digital Interface) is a standard for carrying several channels of AES/EBU. Up to 56 channels of AES/EBU in a serial format can be transmitted along coaxial cable at a bit rate of 125 Mbits/sec using the NRZI (nonreturn to zero inverted) encoding method. BNC-type connectors with a 75-Ω characteristic impedance are used and typically provide a connection for many audio signals between a recorder and a console.

AES3-1992 or AES/EBU is the serial digital audio standard that is used widely in studio applications. An AES/EBU channel supports two audio channels. The samples are represented in a linear 2's complement form (20 or 24 bits). Although the standard supports a range of sampling frequencies, 48 kHz is most common for studio applications. The choice of 48 kHz is advantageous to the broadcast engineer in that there is a relatively simple relationship between the audio sample rate and the video frame rate, which allows for simpler audio synchronization equipment and embedding equipment. The bit rate is 3.072 Mbits/sec for 48-kHz sampling, and the channel coding utilizes a bi-phase code. Manchester coding ensures a DC (direct current) free transmission,

Table 1 — Summary of Digital Audio Formats

Type	Form	Interface	Notes
SDIF	16 or 20 bits @ 44.1 to 48 kHz	3 wire 2 channel, 1 clock	Out of date
SPDIF	Like AES	Single-ended (RCA) or optical	CD (44.21 kHz DAT (48 kHz)
MADI	125 Mbits	75-Ωcoaxial	Carries 56 channels AES
AES	20 or 24 bits @ 30 to 50 kHz	Twisted pair (XLR) 75 Ω under review	Embedded in serial video signal

Table 2 — Summary of Ancillary Data in Composite Digital ($4f_{sc}$)

Pulse	No. Words/ Occurrence	No. Occurences/ Frame	Total Words/Frame
Horizontal	55	507	27,885
Broad	376	12	504
Equalizing	21	24	4,512
		Total 10-bit words/frame	32,901
		Overall data rate:	9.87 Mbits/sec

This is the seventh in a series of tutorial articles by these authors. Randy Conrod is with Leitch Technology International, Inc., North York, Ont., Canada M3B 1V7; Michel Proulx is with Leitch Inc., Chesapeake, VA 23320.

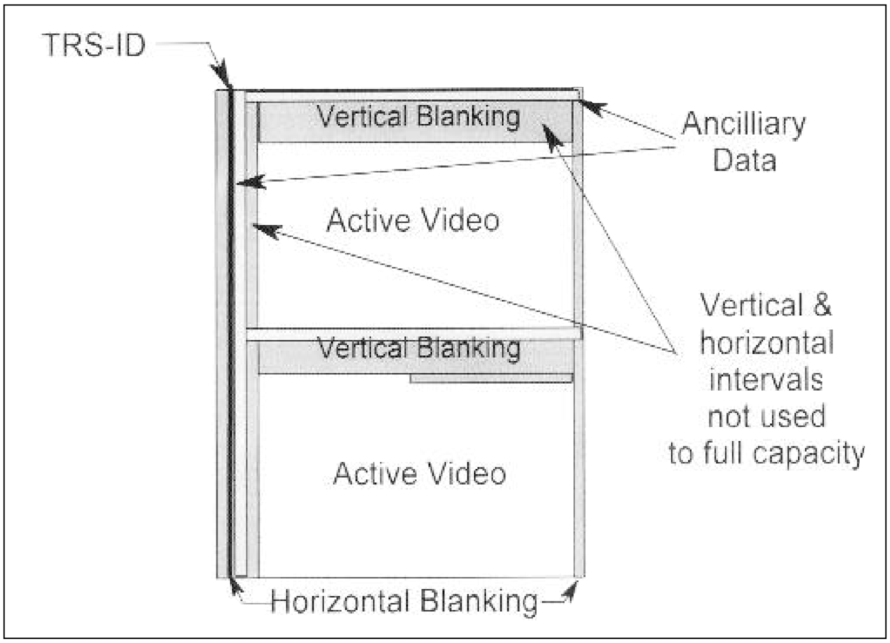

Figure 1. Ancillary data in composite digital ($4f_{sc}$).

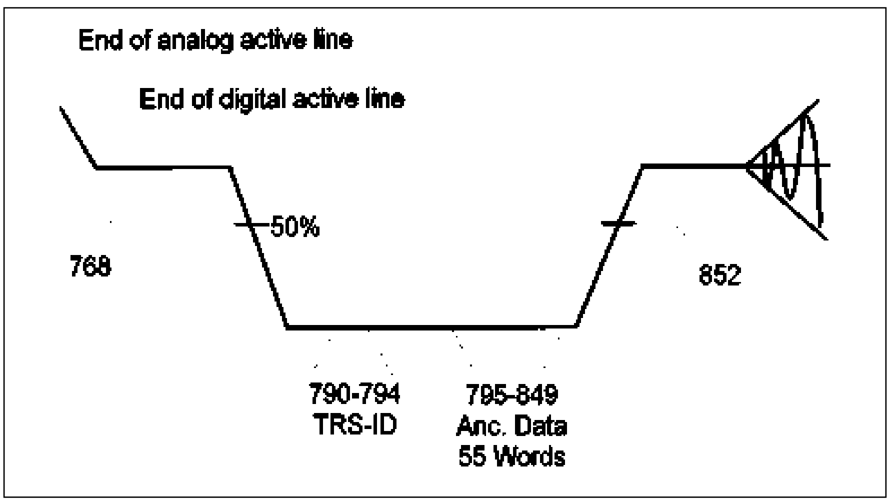

Figure 2. Data in tips of horizontal sync pulses.

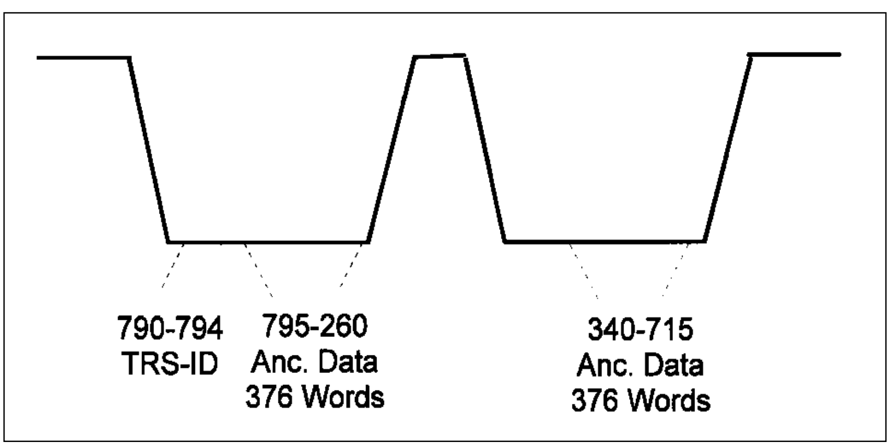

Figure 3. Data in tips of vertical sync pulses.

thereby facilitating simpler clock recovery circuitry on the receiver. The physical interface specifies a balanced transmission over twisted pair with XLR connectors. Many manufacturers have strayed from the use of the bulky XLR connectors for reasons of convenience. SMPTE standards groups have recently been busy pulling together an unbalanced 75Ω coaxial interface. For those broadcast engineers concerned with what they will do with those analog video distribution amplifiers (DAs) that were cast aside for the shiny new digital edit suite, there is no cause to fret. The coaxial interface was designed so that the standard analog DA can be used as single-ended digital audio DA.

AES/EBU digital audio has the distinction of being the first widely used ancillary data type.

Ancillary Data

Ancillary data is defined as information added to the serial data stream that is not active picture information. At this time there is a standard for embedding audio, EDH information, and time code. There are future possibilities for standards for the embedding of source identification and closed-captioning information.

Let's talk about the ancillary data region and how it corresponds to the composite ($4f_{sc}$) and component (4:2:2) serial digital standards.

$4\,f_{sc}$

Ancillary data is embedded into the tips of synchronizing pulses in the composite video signal. One important thing to note is that ancillary data in the composite digital domain is only available in serial interfaces — not parallel. Figure 1 gives a pictorial representation of where the ancillary data can be placed. The TRS-ID is the timing reference signal identification, which represents sync in the digital domain and takes up some room in the horizontal and vertical blanking intervals. Note that the entire blanking intervals are not used up. The horizonal sync pulses can hold the TRS-ID and 55 ancillary data words (Fig. 2). The vertical sync pulses hold the

TRS-ID and a total of 752 ancillary data words in two adjacent vertical pulses (Fig. 3). The pre- and post-equalizing pulses hold a TRS-ID and a total of 42 ancillary data words in two adjacent pulses (Fig. 4). Table 2 gives a summary of ancillary data in the $4f_{sc}$ signal. When one does the math, the overall data rate for ancillary data is 9.87 Mbits/sec. If one considers $4f_{sc}$ sampled NTSC signals (143 Mbits/sec), the percentage of room available for ancillary data is approximately 7%.

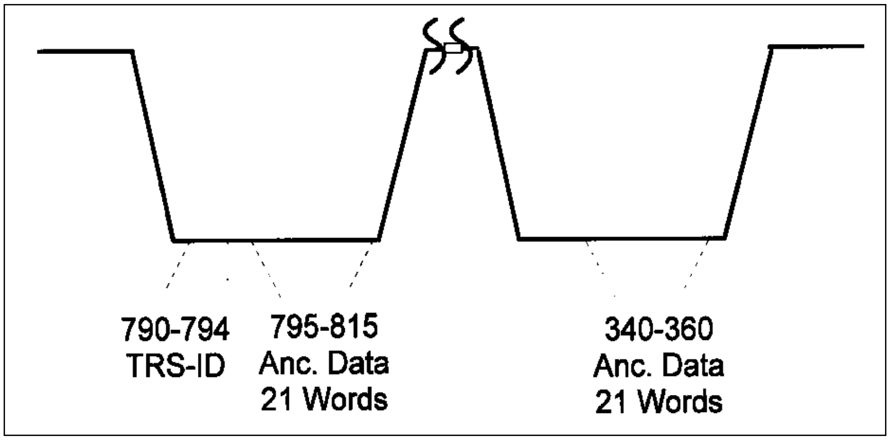

Figure 4. Data in tips of equalizing pulses.

4:2:2

Ancillary data in the 4:2:2 signal is placed in the vertical and horizontal blanking intervals in between the SAV (start active video) and EAV (end active video) data sync reference words. Remember that the blanking intervals are not sampled into the 4:2:2 domain (Fig. 5). VANC is the ancillary data space in the vertical blanking interval and HANC is the ancillary data space in the horizontal blanking interval.

Table 3 shows a summary of the number of data words that can be placed into the VANC and HANC. After one does the math, a data rate of 55.7 Mbits/sec is revealed and the percentage of ancillary data to video is approximately 21%, or one-fifth. Obviously, there is more room in the 4:2:2 signal for extra data than in $4f_{sc}$ signals.

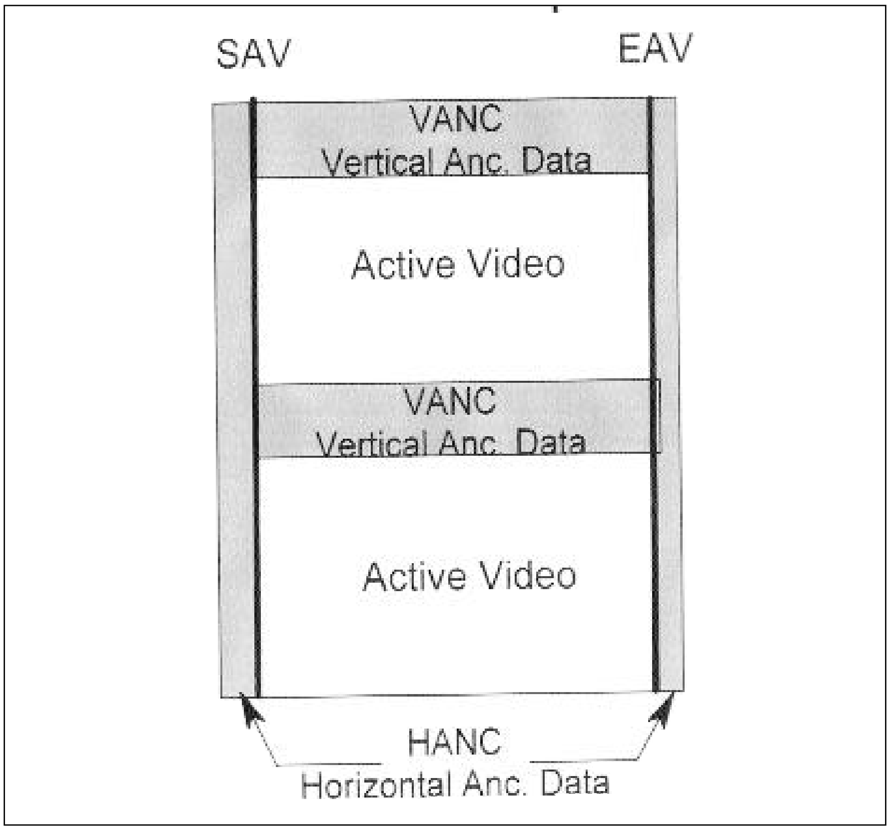

Figure 5. Ancillary data in composite digital.

Table 3 — Summary of Ancillary Data in 525-Line Component Digital (4:2:2)

Type	No. Words/ Occurrence	No. Occurrences/ Frame	Total Words/Frame
HANC	268	525	140,700
VANC	1440	39	56,160
		Total 10-bit words/frame:	198,860
		Overall data rate:	55.7 Mbits/sec

What About VITS (Vertical Interval Test Signals) and Closed Captioning?

Information that is encoded into the vertical blanking interval (VBI) of the composite analog signal (e.g., VITS or closed captioning) is not considered ancillary data. This information is digitized and survives the analog-to-digital (A/D) and digital-to-analog (D/A) conversions. Because of the hybrid nature of systems in use today and for the forseeable future, the traditional use of the VBI will continue for some time.

Embedded Audio

There is good reason to send the audio and video signals down a single coaxial cable. There is a savings in costs for distribution and router

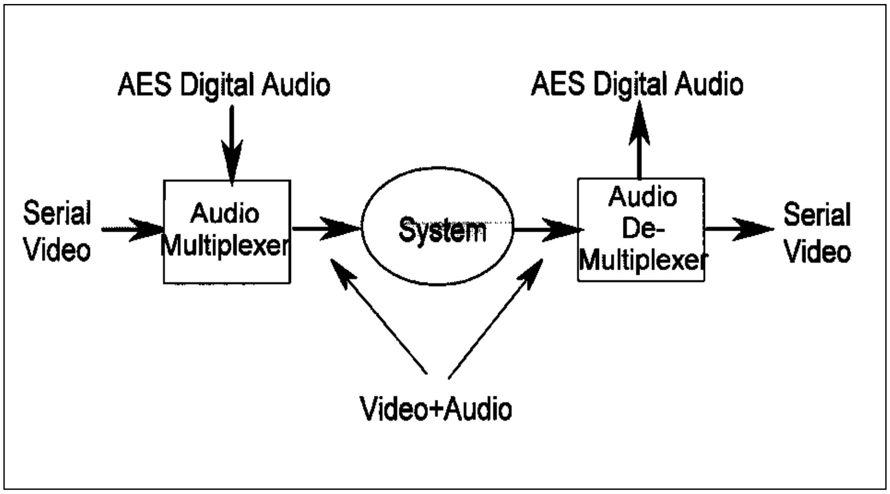

Figure 6. Embedded audio.

Table 4 — Number of Embedded Audio Channels

Channels	$4f_{sc}$	4:2:2
No. of groups	1	4
No. of AES	2	8
No. of “mono”	4	16

switcher equipment, as well as costs associated with cable. When audio signals are embedded into the serial data stream, a multiplexer is used, and when audio signals are separated from the video signal, a demultiplexer is used. Typically, these processing devices are called muxes and demuxes (Fig. 6).

As mentioned earlier, the AES/EBU standard of digital audio is the signal that is embedded into the serial video signal. There is sufficient room in the serial data stream for several channels of digital audio information. Remember that AES/EBU carries two channels, typically a stereo pair (left and right). Audio data is embedded in a group format. A group is defined as two AES/EBU signals or four monaural channels.

When embedding the digital audio information into the ancillary data region, there is no strict template as to where the audio information is placed. The digital audio information is mapped into data packets and placed in the ancillary data region. At the receive end a 64-sample buffer is used to reconstruct the digital audio signal before it is output.

An analogy might help explain how this works. Imagine that the serial signal is a train. The train engine is pulling many boxcars, which contain the serial video information. Every few boxcars an empty one can be found, and this would be the ancillary data region. This would occur during pulses in the blanking intervals of $4f_{sc}$ sampled signals and between the EAV and SAV of 4:2:2 sampled signals. The audio information is broken into small packages or data packets, and a data header or label is added to the data packet. This package is placed into the empty boxcar and sent on its way. At the receive end the 64-sample buffer fills with incoming data packets and is output as the AES/EBU signal. This would be like a loading dock onto which the data packets are placed after being moved from the boxcars. The trick is never to let the sample buffer or loading dock get to the point of being overloaded or completely empty.

All ancillary data (AES/EBU, EDH, time code) is multiplexed and demultiplexed in a similar manner. The header on the data packet identifies the type of data in the packet. In other words, the empty boxcar could be filled up with audio, EDH, and/or time-code information.

In $4f_{sc}$ sampled signals, there is room for one group of AES/EBU signals (four monaural channels). In 4:2:2 sampled signals, it is suggested that the audio information be placed between the EAV and SAV during the horizonal blanking interval, and there is room for four groups of AES/EBU (Table 4). This is 16 channels of monaural audio information!

Conclusion

To Embed or Not to Embed

Now that it is possible to carry audio and video information on the same coaxial cable, one must decide whether this is an advantage for each system. Obviously, if the video and audio signals do not have to be processed separately, as in a network switching facility, embedding audio saves levels of routing switchers and distribution amplifiers.

In addition, the audio and video always pass through the same amount of equipment, which eliminates lip sync problems. For instance, if the video signal passes through a frame synchronizer, the audio will now lead the video signal and create lip sync problems. But if one must process the audio separately from the video, for instance, moving the signals around on separate cables may make more sense. Also, remember that if the video is processed through a production switcher, the audio must be dumuxed before the processing and muxed after the processing.

The question left to the systems engineer is whether to embed or not to embed. Thank goodness for ASICs (application specific intergrated circuits) and VLSI (very large scale integration), which will drive the costs down as these new standards gain market acceptance.

We’ll have to leave you with your calculator to figure out if embedded audio is for you!