

Integrated solutions for embedded Dolby E and AC-3

Romolo Magarelli and David Strachan, Evertz Microsystems Ltd.

Synopsis

Dolby E offers the television broadcaster the ability to carry up to 8 channels of audio, complete with metadata, on a single AES-3 digital audio cable. Many broadcasters are adopting this convenient technology as a way to reduce the number of cables needed to transport multiple audio channels around a facility, ensure phase coherent treatment of multiple related channels (e.g. 5.1 audio) and bind AC-3 encoded information (metadata) together with the audio. We need to be aware that, if all audio and video paths are not designed carefully, the advantages can be marred by lip sync problems, audio errors, and loss of information.

This paper describes the Dolby E process and ways to avoid common pitfalls. We show how to de-embed and decode Dolby E material and to re-embed the audio back onto the SDI or HDSDI video as PCM audio. The metadata is preserved throughout the process and can be re-embedded along with the PCM audio. Lastly we describe the process of converting embedded PCM audio back to embedded Dolby E or AC-3.

Receiving Dolby E or Dolby Digital encoded audio at the TV station

We will assume that the signal being received at a broadcast affiliate station, has been correctly encoded using a Dolby E or Dolby Digital (AC-3) encoder. There is a good chance that the broadcaster will need to decode the audio, listen to it, modify it and maybe transport it to other parts of the facility before re-encoding it for onward distribution. Figure 1 shows some of the devices which would be needed to achieve this.

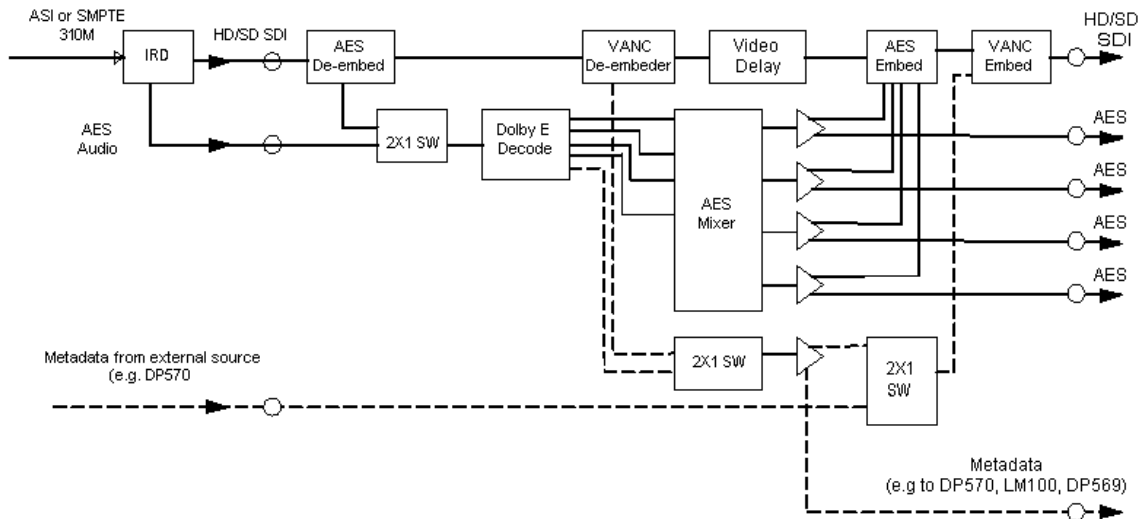


Figure 1 - Discrete devices to handle Dolby E content

Engineers are often content to learn the hard way - I guess because we like a challenge. But it can be time consuming and an expensive exercise for our companies, especially if an installation has to be re-worked to correct errors. The following are examples of the problems that broadcasters regularly find, as soon as they start playing with compressed audio.

- Audio clicks, pops and discontinuities when switching discrete or embedded audio
- Complete loss of audio
- Loss of audio channels
- Loss of audio metadata
- Lip sync problems between the program video and the audio

Everyone who has watched television has witnessed one or more of these annoying discrepancies. Dealing with compressed audio is a complex subject and a brief summary of Dolby audio signal structure may help us understand how things can so easily go wrong if we are not careful.

Dolby E and Dolby Digital

Modern home theatre systems use 5.1 or 7.1 audio channels, to provide a surround sound experience. The 5.1 systems have six channels of audio with Left (L), Right (R), Center (C), Left surround (Ls) and Right surround (Rs) together with a Low Frequency Effects (LFE) channel covering the lower 10% of the audio spectrum. More advanced home theater systems feature eight channels of audio, adding Left Back (Lb) and Right Back (Rb) loud speakers, to provide sounds which completely surround the audience.

Dolby Digital (AC-3) is the encoding scheme used by broadcasters to transmit all of the audio channels to the home. The audio data is carried in the ATSC or DVB encoded bit stream. Dolby E encoding, on the other hand, is used within the TV station. The main difference between the two systems, is that Dolby E is frame locked to the video, allowing clean switching between video channels, whereas AC-3 has no relationship to video. Dolby E is also more robust and can better survive multiple generations of transcoding. Both systems compress the audio to accommodate eight channels (Dolby E) or 6 channels (AC-3) of audio into one AES stream.

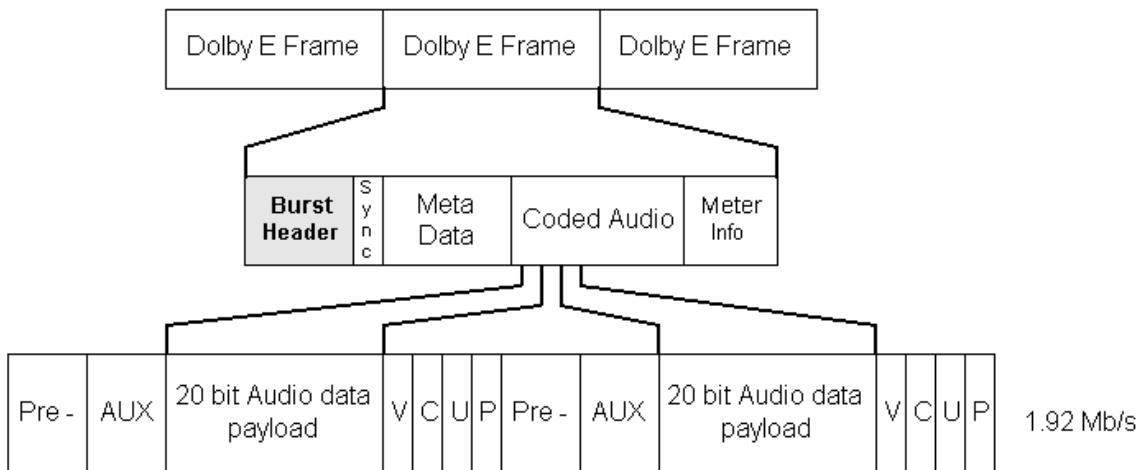


Figure 2 AES-3/Dolby E Packet Structure

The Dolby E encoder, bundles the audio data into packets corresponding to each video frame. Each frame starts with a burst header (guard band) consisting of a minimum of four zeros. This is followed by a sync signal, the Dolby encoded audio and additional data, which may be used for metering purposes. The sync data is defined in SMPTE 337M and 338M. Among other things, it contains information about the type of encoding which follows (i.e. Dolby E or AC-3). Any necessary switching between two audio streams must be made in the guard band at the start of each frame. The requirement is the same whether the audio data is carried over a discrete AES line or whether the data is embedded into an SDI (SMPTE 259M) or HDSDI (SMPTE 292M) bit-stream. In either case, the two audio streams must be frame synchronized, so Dolby E encoders must be supplied with a genlock video black burst signal.

When it becomes necessary for television engineers to process video and audio separately, Dolby E or AC-3 signals must be decoded back to the 3 or 4 separate AES streams, as shown in Figure 1.

Passing Dolby E material through the Television Facility

Dolby E is designed to pass through standard AES devices, but care must be taken because Dolby E is no longer audio - it is data. Some AES audio devices use Sample Rate Converters (SRC) to change the standard from one sample rate to another. It is important to turn off all SRCs through which the Dolby E data might pass, as the data will otherwise be destroyed.

Metadata

In addition to audio, the Dolby E stream transports information about how the 5.1 channels of audio are to be de-multiplexed by the receiver. This information is the *Metadata* and it precedes each frame of coded audio in the AES bit stream. The term metadata covers a very extensive range of audio parameters. The most significant ones are the average audio dialog level (Dialnorm), the receiver's control of the full dynamic range (Dynamic Range Control) and information about how the signal should be decoded for mono or stereo listening (Down-mixing). Whether or not the metadata is needed by the broadcaster, the information must be preserved during the decoding process, as it will be required later by the home entertainment system, to provide the intended listening experience.

Latency

As with many of today's processing devices, Dolby E introduces latency into the signal path. There is a 1 frame delay during the encoding process and a further 1 frame of delay for the decoding. Although Dolby E is the encoding process of choice for the professional television engineer, AC-3 is also quite common within the broadcast facility, because many incoming feeds contain Dolby Digital (AC-3) material. It is important to note that the AC-3 encoding process, introduces approximately 5 frames of audio delay. Consequently there is a need to introduce an equivalent video delay in order to maintain lip sync. Decoding AC-3 is easier and in this case the audio delay is about 1 video frame. (Television engineers are familiar with the need to delay audio to keep it synchronized with the video, because video frame synchronizers have to delay the video to line it up with the house reference. It is important to consider the overall audio delay, together with the overall video delay. It is usually necessary to add additional video delay in systems using Dolby encoding or decoding processes).

Embedded audio

SMPTE 272M defines the audio embedding for SDI at 270Mb/s. SMPTE 299M defines the audio embedding for HDSDI at 1.5Gb/s. In both cases there are 4 *groups* of audio, each containing 2 AES channels, which in turn contain 2 mono audio channels. Table 1 shows the SMPTE embedded audio structure.

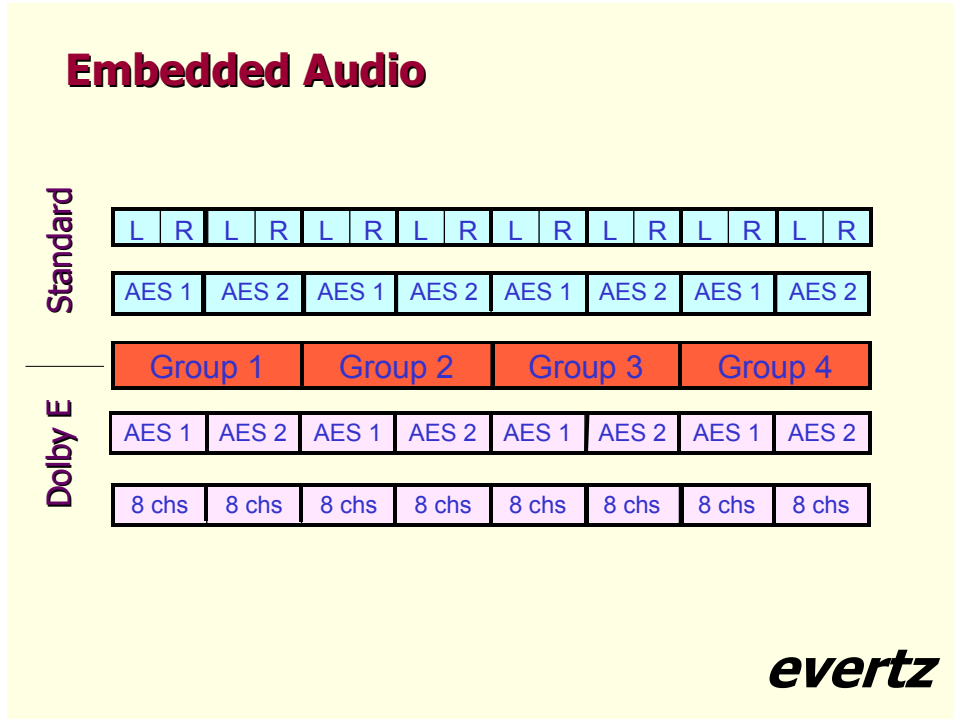


Table 1 Embedded Audio

Each uncompressed AES channel can carry 2 mono audio channels, or 8 compressed Dolby E audio channels (arranged as 4 Dolby E encoded AES channels). Embedded audio can carry up to 8 AES channels, so it is theoretically possible using Dolby E, to transport as many as 64 mono audio channels. Many broadcasters who carry multiple audio channels, will use Dolby E for the 5.1 and will carry separate stereo audio in another AES stream, either in the same group or in a second group. (It is also possible to carry the 5.1 audio together with an unrelated stereo audio, in a single Dolby E stream as 5.1 + 2).

System Example

Dolby E technology is great for the economical transportation of multiple audio channels, but PCM is needed for audio production. For this purpose, most broadcasters will need access to decoded audio. A simplified diagram is shown in Figure 3. Video with embedded Dolby E (or with AC-3) enters the facility and is converted to video with 4 channels of embedded PCM audio and the metadata. This video is transported throughout the facility in the usual way, using audio de-embedders and embedders as required. At the other end of the chain, the video with PCM and metadata, is converted back to video plus Dolby E (or Dolby AC-3) audio data, using a Dolby encoder. (The two 520AD4-xx devices will be described later).

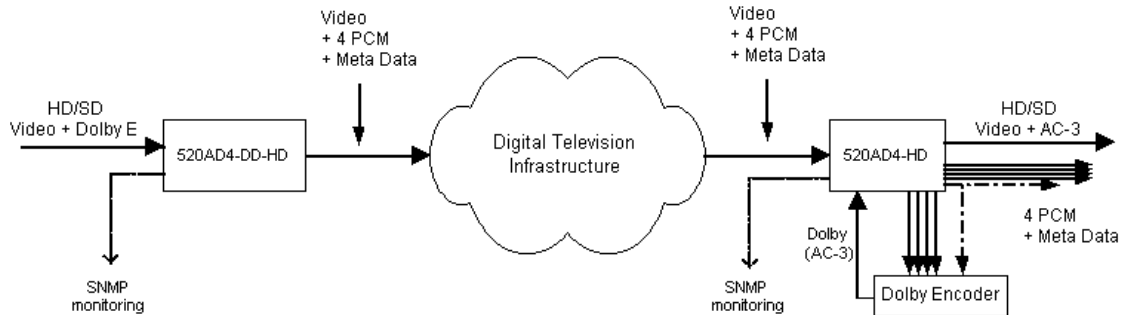


Figure 3 System Example

Integrated Solutions

As we saw earlier that discrete de-embedders, video delay cards and other devices, can be used with Dolby E decoders, to provide an end to end solution. But this can be expensive and far from convenient. To simplify these tasks, Evertz has developed integrated solutions in two 500 series plug-in modules. These devices can handle up to 8 channels of audio (which covers the 7.1 channels). The block diagram of the 520AD4-DD-HD module is shown in Figure 4.

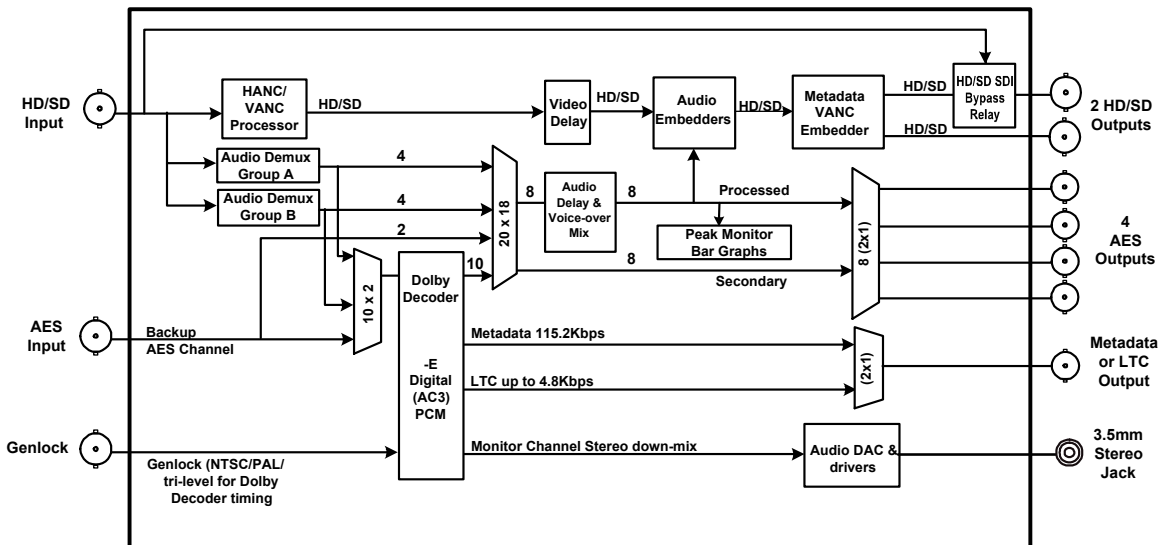


Figure 4 Evertz 520AD4-DD-HD de-embedder, Dolby E decoder and re-embedder (the numbers on each path are mono channels and provide for Dolby or PCM inputs)

The audio, contained in 2 specified groups, as defined by SMPTE 299M for 1.5Gb/s serial HDTV, or from SMPTE 272M for 270Mb/s serial SDTV video, is first de-embedded into 4 AES streams (4+4 mono channels). At this point, it is necessary to determine if either Dolby E or Dolby Digital is present and if so, what we want to do with it. If Dolby encoding is not present, the system will automatically default to any available PCM audio. The output from the Group A or Group B de-embedders, or from the separate AES input, can be routed through a Dolby E/Dolby Digital decoder, or may by-pass the decoder altogether. The decoding process recovers the 4 discrete stereo pairs and creates a 5th stereo down-mix of the 5.1 audio. These 5 decoded AES channels (10 mono) are fed to the 20x18 mono router, where they are joined by the other 5 stereo pairs (the ones that did not pass through the decoder). We refer to this router as a 20x18, because each of the 20 mono audio channels can be handled separately and routed to any of the 18 mono outputs (4+4+1 AES) outputs. The first 4 stereo pairs may be processed, by adding delay and stereo voice over, while the remaining 4 stereo pairs by-pass the processing and are fed directly to the last multi-channel router. It is here that the final 4 AES outputs are selected. (It is important to note that the genlock reference is used for the audio only, not the video. There is no video frame synchronizer capability on this module).

The video path at the top of Figure 4, shows a HANC/VANC processor, where unwanted embedded packets may be removed, a video delay function to compensate for Dolby decoding delay and an audio re-embedder. At this stage, any 4 AES channels, selected by the 20x18 router, can be embedded into the SDI or HD-SDI stream. For example Dolby E or Dolby Digital (AC-3) material presented at the input to the 520AD4-DD-HD, can be re-embedded in 2 groups as PCM audio. This is one of the main functions of the device.

As well as PCM audio, the Dolby decoder also provides some additional outputs. These are Dolby metadata, LTC (if present) and a copy of the stereo down-mix. The latter passes through a digital to analog converter to a stereo jack socket for headset listening. The metadata is brought out to the rear of the card on a BNC connector and is also routed to a metadata embedder where it can be carried throughout the broadcast facility along with the PCM audio. (A simple adaptor may be used to access the data via a common 'D' connector). The metadata can be embedded on any user defined line, but the default is line 10. For interoperability purposes, the packet is labeled with the data ID hex number of 45.

It can be seen by comparing Figure 4 with Figure 1, that this new plug-in circuit card can replace up to ten discrete devices.

Audio De-Embedder, Metadata Embedder

A second circuit card, available from Evertz completes the system. This device, the 520AD4-HD, takes the video signal with embedded PCM audio and metadata and, with the help of a Dolby E encoder, re-embeds the Dolby E into the same video stream. Figure 5 shows how it works.

As before, the audio is de-embedded to recover the four PCM AES channels and the metadata. These signals can then be fed to a Dolby E encoder, the output of which is returned to the AES input of the 520AD4-HD card, where it is re-embedded into the video stream. The card has been designed in such a way as to maximize the number of functions that might be required of it. Hence the separate metadata input and output, voice over capability, video and audio delay functions and audio channel swapping.

