WHITE PAPER

The Benefits of using SMPTE 2020 to Transport Dolby Metadata

By Stewart McGhie
Head of Sales
Crystal Vision Ltd.

This paper is intended to give an explanation of SMPTE 2020, including why it was created and how it can be used in a broadcast environment.

SMPTE 2020 is a way of carrying easily accessible audio metadata in the same data stream as the video and embedded audio.

Crystal Vision

Why is Crystal Vision interested in SMPTE 2020?

Crystal Vision's interest in SMPTE 2020 stems from the fact that the company provides products to embed and de-embed audio into the SDI digital video stream, used within a television studio or transmission system. These products include Dolby E encoders and decoders, available as an add-on option for the TANDEM 310 audio embedder/de-embedder and the SYNNER 310 synchroniser. These Dolby options allow users to carry 5.1 multi-channel surround sound around their installations within the SDI video stream – in the same way that they would carry normal audio.

What is Dolby E?

Dolby E is a proprietary form of audio compression created by Dolby Laboratories. It allows surround sound such as 5.1 (or even 7.1) to be encoded into a single AES data stream. This in turn can then be embedded into SDI and carried as if it were just a stereo audio signal.

This was a great advantage when it was first created as a lot of audio equipment could not support the passing or storing of more than four channels of audio. It has been in use by broadcasters for over ten years now and it has become the main way that many big broadcasters distribute surround sound within their own installations or between different sites.

Dolby E should not be confused with Dolby Digital (also known as AC3) which is the format used for transmission to the consumer's set top box. Each format has its own place in the broadcast system. What does, however, link the two together is metadata. Metadata from the Dolby E system can be passed into the AC3 stream, and in fact *should* be passed for the surround sound to be correctly presented to the consumer at home.

The purpose of metadata

A Dolby E stream contains metadata. The metadata needs to tell the receiving equipment how to decode the audio data because Dolby E can carry different configurations of audio. There is obviously 5.1 surround sound, but a Dolby E stream could alternatively carry 7.1 surround sound, 5.1 surround sound with an additional stereo audio track, or even eight separate and unrelated audio tracks – such as for multi-language applications. The metadata will indicate what configuration of audio is in the stream.

The metadata also contains a key piece of information to manage audio levels. Audio encoded in the Dolby E stream may be at different reference levels, depending on the production requirements or set up. This reference level is called "Dialnorm" and is carried in the metadata.

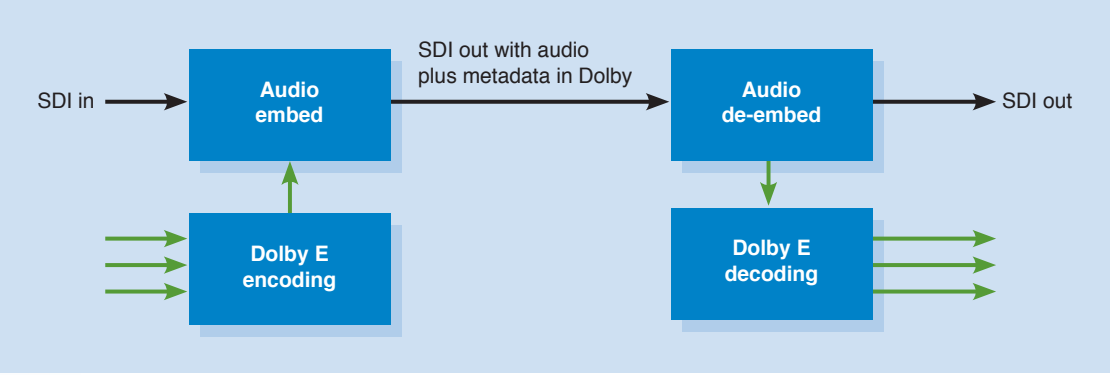
It defines the average audio level of the sound that has been encoded. For programmes with lots of dialogue, such as news, this could be -20dB but for other programmes – like films – it may be different, for example -27dB. This is because films require a greater dynamic range – with the loud parts really loud and the quiet parts very quiet. The dialnorm value allows the decoder to control its output levels based upon what the user, the consumer or the broadcaster wants to achieve in terms of their listening environment or their transmission system. Ensuring the dianorm value is passed through production chain can help keep the audio levels consistent across different programmes.

The audio encoded as Dolby E might have had pre-emphasis noise reduction or other filtering options applied – these again can be signalled within the metadata. There is also information in the metadata which indicates how a surround sound signal might be downmixed to stereo and so on. The Dolby website (www.dolby.com) provides full information about all of the different things which can be in the metadata.

The key point is that metadata is essential if the Dolby E signal chain is to function properly – which is equally applicable to a Dolby D/AC3 system as well.

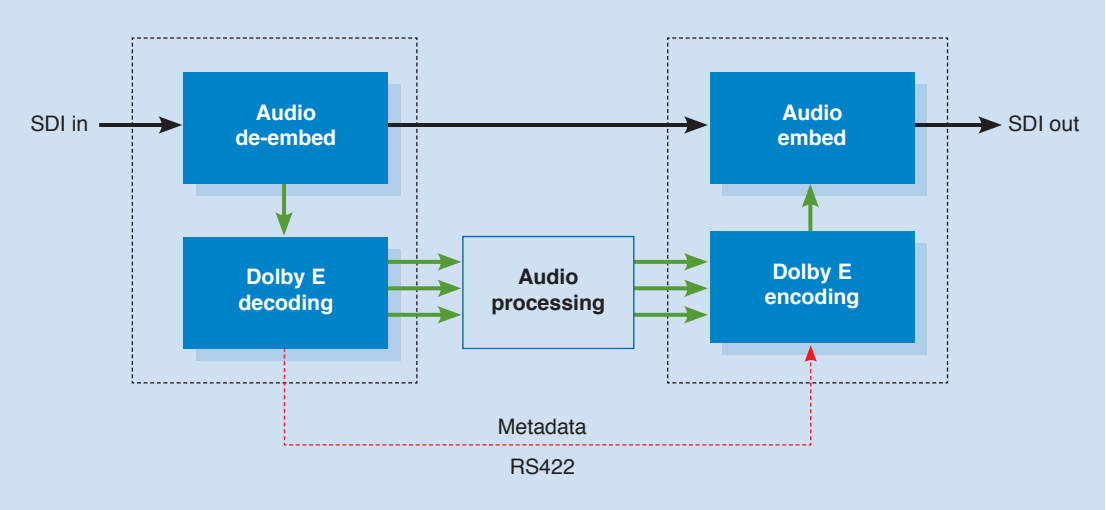
Different ways of transporting metadata

In a simple system the metadata is carried within the Dolby E bit stream. The user sets the metadata at the encoder. It travels within the AES/Dolby E stream and the decoder then uses this metadata to decode the audio properly. This is quite simple really and so you would think that users would not need to worry about metadata very much.



Metadata carried in Dolby E

However, if the Dolby encoded audio needs to processed – to mix in a voiceover or commentary or to have an audio level adjustment – then it needs to be decoded. Although you could set the metadata again, it would be easier and more dynamic if the metadata from the decoder was passed through to the encoder. This can be achieved by an RS422 link as shown on the diagram below.



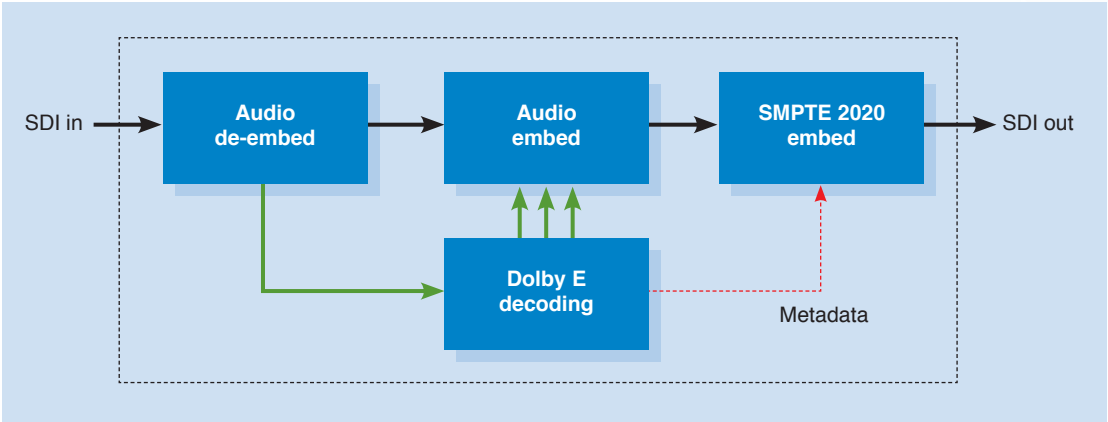
Metadata passed by RS422

But in a big system this is not easy, as the decoder and re-encoder are not necessarily physically close to one another and having a separate path for RS422 would be complicated. This is when you may want to have the metadata embedded into the SDI.

It may also be the case that a broadcaster does not wish to distribute the surround sound within their installation as Dolby E. Why? Doing so has the disadvantage that every time they want to process the audio to perform functions such as audio mixing, inserting voiceovers and so on, they need to decode and re-encode it. This increases the cost of the system and possibly introduces timing issues, as the decode or encode process introduces a 40 millisecond delay on the audio.

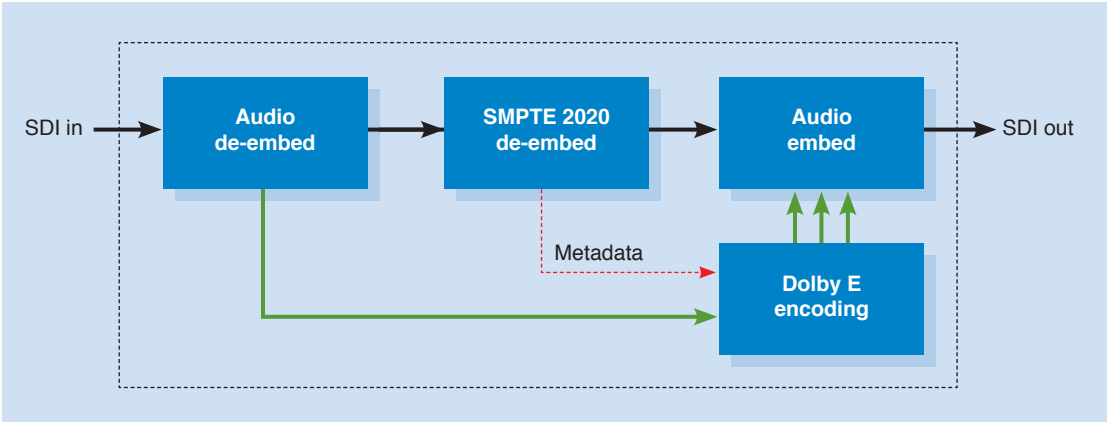
The broadcaster may therefore decode the Dolby E at the input to their system and then only re-encode it at the final output. The problem, then, is how to carry the metadata through their installation. The answer is SMPTE 2020.

SMPTE 2020 is the standard which defines how the metadata can be carried as an Ancillary Data Packet in the SDI stream, accompanying the video stream wherever it goes and making the metadata easily accessible. The SMPTE 2020 data packet is usually carried on line 9 in the Vertical Ancillary (VANC) area, but the standard does allow other lines to be used.



Metadata put into SDI (SMPTE 2020 embedding)

When arriving at the final output encoder it can then be de-embedded and applied to the encoder.



Metadata taken out of SDI (SMPTE 2020 de-embedding)

As SMPTE 2020 is an agreed standard for formatting the metadata in the SDI video stream, it is possible for different manufacturers' equipment to work together.

Crystal Vision can provide solutions for Dolby encoding and decoding which include the capability to embed and de-embed the SMPTE 2020 data and use this data as required.

In conclusion

SMPTE 2020 provides broadcasters with more flexibility when designing their systems. They can choose to pass the surround sound through their system as either encoded Dolby E or as discrete multi-channel audio – and still maintain the necessary metadata path.