SMPTE REGISTERED DISCLOSURE DOCUMENT

for Television — Guidelines on the Use of Dolby® E with Video Signals at Frame Rates Greater than 30 Hz

The attached document is a Registered Disclosure Document prepared by the proponent identified below. It has been examined by the appropriate SMPTE Technology Committee and is believed to contain adequate information to satisfy the objectives defined in the Scope, and to be technically consistent.

This document is NOT a Standard, Recommended Practice or Engineering Guideline, and does NOT imply a finding or representation of the Society.

Errors in this document should be reported to the proponent identified below, with a copy to eng@smpte.org.

All other inquiries in respect of this document, including inquiries as to intellectual property requirements that may be attached to use of the disclosed technology, should be addressed to the proponent identified below.

Every attempt has been made to ensure that the information contained in this document is accurate. Dolby Laboratories is, however, not responsible for the action, effect, consequences, etc. of any non-Dolby implementation of any device(s) based on this document. If any errors, inconsistencies, or any points that require clarification are found, contact Dolby Laboratories at the address below as soon as possible.

Proponent contact information:

Dolby Laboratories Inc.
100 Potrero Avenue
San Francisco, CA 94103

Attention: Broadcast Services and Support

broadcastsupport@dolby.com

Approved
February 4, 2011

Table of Contents

Page

1 Scope

This document describes how Dolby® E audio codecs can be used in conjunction with Television signals operating at frame rates higher than 30 Hz. Because the Dolby E encoded audio data is normally carried in an AES3 stream, wrapped per SMPTE ST 337, it may be transported separately, or embedded in a video signal. It also provides advice to equipment designers and manufacturers on the requirements of associated equipment and to systems integrators engaged in building complete systems.

2 Related Documents

Note: All references in this document to other SMPTE documents use the current numbering style (e.g. SMPTE ST 318:1999) although, during a transitional phase, the document as published (printed or PDF) may bear an older designation (such as SMPTE 318M-1999). Documents with the same root number (e.g. 318) and publication year (e.g. 1999) are functionally identical.

SMPTE ST 318:1999, Television and Audio — Synchronization of 59.94- or 50-Hz Related Video and Audio Systems in Analog and Digital Areas — Reference Signals

SMPTE RP 168:2009, Definition of Vertical Interval Switching Point for Synchronous Video Switching.

SMPTE ST 12-2:2008, Transmission of Time Code in the Ancillary Data Space

SMPTE ST 302:2007, Mapping of AES3 Data into an MPEG-2 Transport Stream

SMPTE ST 337:2008, Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface

SMPTE ST 2051:2010, Two-Frame Marker for 50-Hz and 60(/1.001)-Hz Progressive Digital Video Signals on 1.5 Gb/s and 3 Gb/s Interfaces

3 Introduction

Dolby Laboratories has developed Dolby E, a method of reducing the data rate of PCM (baseband) audio signals while preserving the subjective quality of the audio through up to ten encode – decode cycles. The data rate reduction ratio allows up to eight audio signals to be carried in a single AES3 digital audio stream. Further, the Dolby E encoded audio data is transmitted as a sequence of blocks of data, separated by guard band intervals containing null data.

Dolby E encoders generate data blocks that are synchronized with video frame rates of up to 30 frames per second. The guard bands between blocks are co-located with the Vertical Interval Switch points of the associated video signal, so that the Dolby E data stream and the video may be switched without causing any disruption of the encoded data and hence no disruption of the decoded audio signals.

The Dolby E data blocks cannot be generated at higher than a 30-Hz rate, thus cannot be switched at higher rates without corrupting the data. If a Dolby E stream is to be switched simultaneously with a progressive video signal operating at 50 or 60 frames per second, then the Dolby E data stream may only be switched on a Progressive frame (or P Frame – See Terms and Definitions) boundary that occurs during the Dolby E guard band intervals, or in other words, on every second P Frame boundary.

SMPTE RP 168 states that “with both interlaced video at a specific frame rate and progressive video at double that frame rate, devices handling the progressive video should be referenced to a signal derived from an interlaced format at the interlaced frame rate. Having established this referencing relationship, progressive video devices should switch using the recommended line during Field 1 of the reference signal”. As can be seen in Figure 1, since the Dolby E guard band intervals are aligned with Field 1, the Dolby E data will not be harmed by an audio-follow-video switch of a progressive signal operating at greater than 30-Hz rate as long as the switch happens during P Frame one.

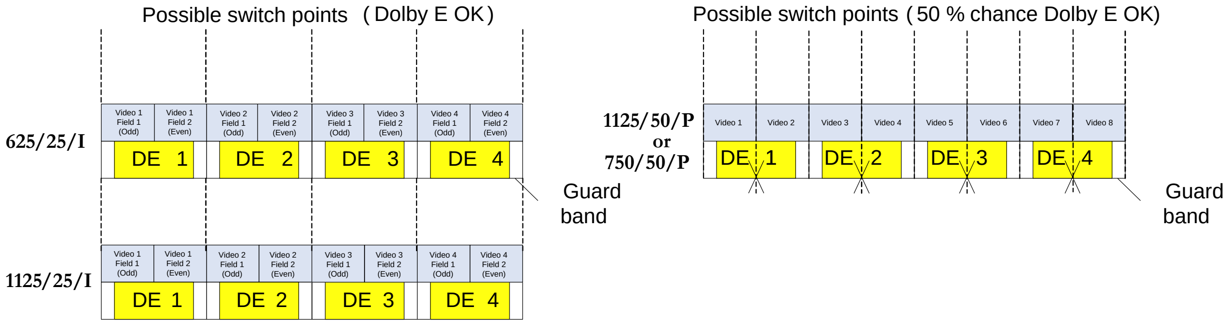

Figure 1 - Interlaced video with co-timed Dolby E (left) and Progressive video with Dolby E (right). Unless the Audio follow Video switch occurs at the switch point of the first P Frame, the first part of the Dolby E data stream will be taken from the audio associated with the first source, and the second part from the audio associated with the second source, making the checksum incorrect. The decoder will mute for one frame, or until it receives a Dolby E data block with a correct checksum.

4 Terms and Definitions

Television System notation: References to television systems in this document will follow the conventions used by documents such as SMPTE RP 168, which state the total number of lines, the total number of complete images per second and whether the serial interface structure is Interlaced or Progressive. Examples are 525/59.94/I, 1125/50/I, 750/60/P. Note that other notation systems state the number of active lines, the serial interface structure and the number of complete pictures per second. Examples are 1080I/25 and 1080P/50.

An **Interlaced raster structure** consists of a pair of Fields, called Fields 1 and 2, or the Odd and Even fields. The complete image is called a Frame (or an I Frame) and consists of Fields 1 and 2.

A **Progressive raster structure** consists of a sequence of complete images, called **P Frames**. Notionally, each P Frame can be temporally related to one Field of an Interlaced television system operating at half the frame rate of the progressive system. The P Frame temporally related to Field 1 is called P Frame 1 or the Odd P Frame, and the P Frame temporally related to Field 2 is called P Frame 2, or the Even P Frame.

DE is an abbreviation of Dolby E. DE 1, DE 2 seen in the Figures refer to Dolby E data block 1, Dolby E data block 2, etc.

5 Synchronization of Dolby E and Progressive Video, Operating at 50, 59.94 and 60 Frames Per Second

As frame synchronizers are used to synchronize incoming signals with the “house sync”, it is essential that the frame synchronizer to operate in a “Dolby E safe” mode, allowing clean downstream switching. Two points must be considered in designing such a frame synchronizer:

1) Any frame synchronizer must drop (or repeat) two P Frames, instead of one. The benefit of keeping Dolby E aligned by this method is considered to be worthwhile given the potential small motion artifact observed when dropping 2 progressive frames.

2) The P Frames to be dropped (or repeated) must be a P Frame 1 and P Frame 2 pair (in that order) and must be aligned with the start of the Dolby E block.

Standard Interlaced frame synchronizers that drop or repeat an entire I Frame's worth of audio samples during a video frame drop or repeat handle Dolby E well (Figure 2).

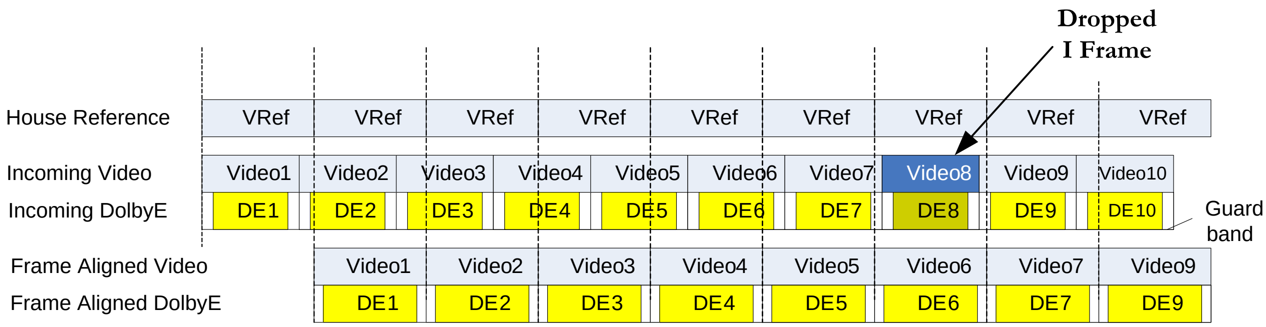

Figure 2 – An Interlaced Frame Synchronizer locking incoming video and Dolby E to an Interlaced house reference

Dolby E data blocks will be corrupted if only a single P Frame is dropped or repeated by a progressive frame synchronizer.

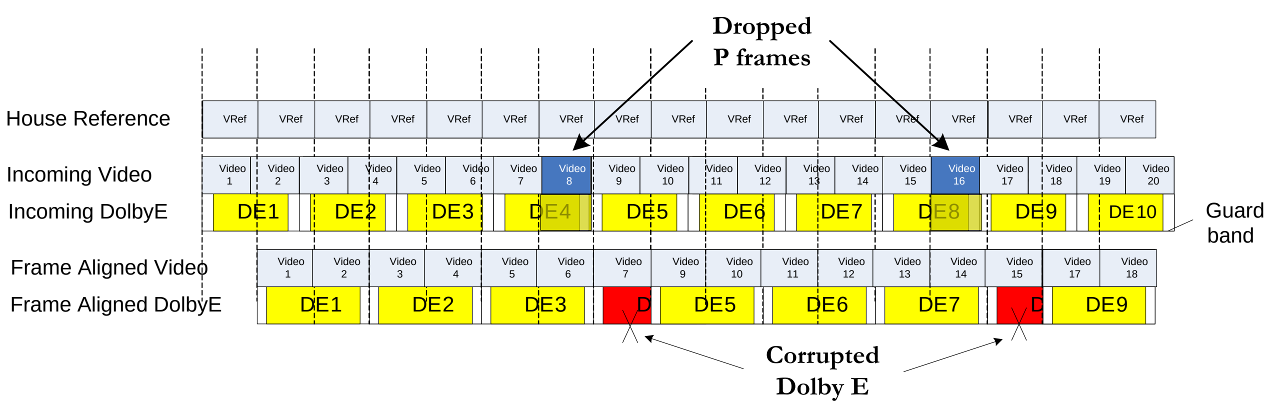

Figure 3 – A Progressive Frame Synchronizer locked to a progressive house reference signal Note that the single P Frames dropped correspond to half of a Dolby E data block.

Just having a progressive frame synchronizer lock to a 25 or 30 Frames per second (fps) signal, and dropping two P Frames instead of one is not enough to ensure undamaged Dolby E. The synchronizer must drop (or repeat) a P Frame 1, P Frame 2 pair (in that order) to avoid the 50% risk of starting the two P Frame drop sequence incorrectly, with a P Frame 2. This would also mean that downstream equipment relying on the same 25 or 30 fps reference to correctly switch Dolby E cannot guarantee error-free switching.

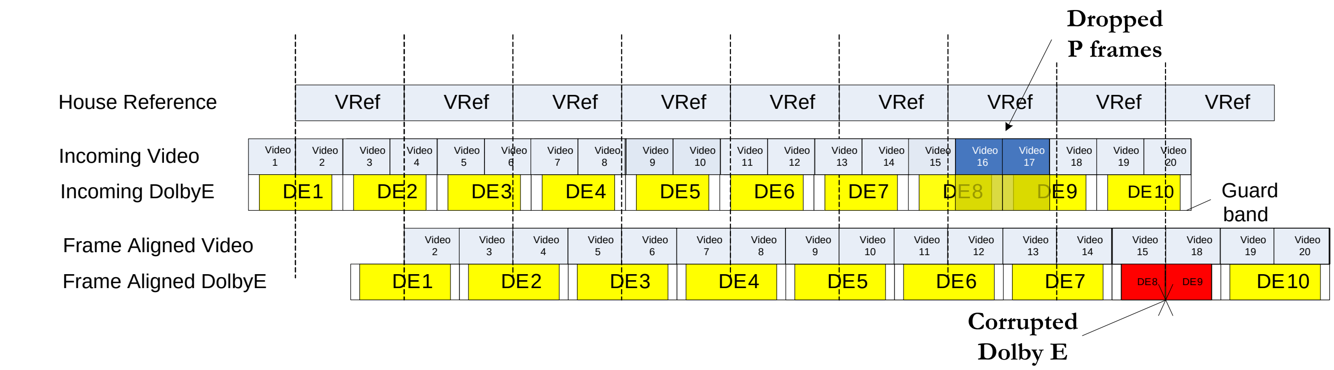

Figure 4 – A frame synchronizer correctly dropping a progressive frame pair, but incorrectly aligned with the interlaced reference and the P Frame One, P Frame Two cadence

The solution is to lock the frame synchronizer to a 25 or 30 fps reference signal, and to detect the Dolby E header to determine where the P Frame One, P Frame Two cadence begins, and hence where to start the two P Frame drop sequence, as shown in Figure 5. See Section 7 - Dolby E. Detection.

Frame synchronizers could also re-align the Dolby E on the incoming HD-SDI signal to the appropriate position. This would add the benefit of ensuring all embedded Dolby E outputs are not only error-free during frame drops, but also correctly aligned for further processing / switching, etc.

The realignment results in a two P Frame delay with a 0.5 to 1.5 Dolby E data block delay.

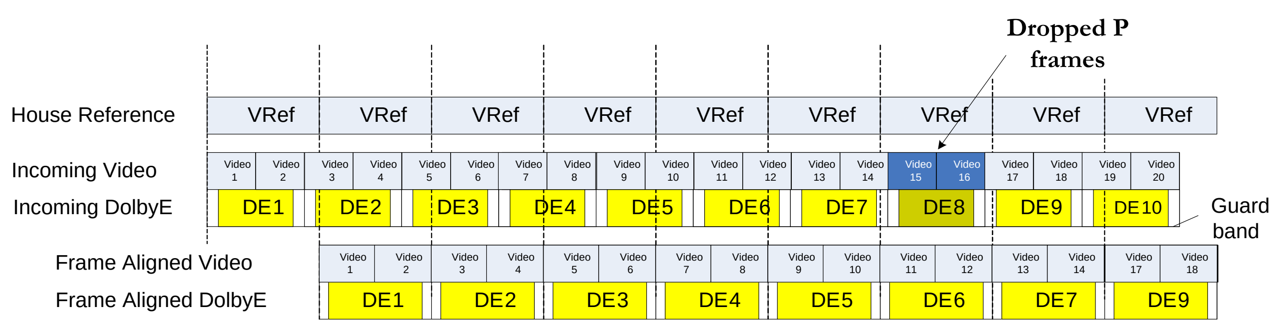

Figure 5 – "Dolby E safe" P Frame synchronizer locked to an interlaced reference, and the correct P Frame cadence.

By searching for the end of a Dolby E data block as well as the beginning, a frame synchronizer could eliminate the kind of errors shown in Figure 3, caused by dropping single P Frame, or a switch during the P Frame 2, P Frame 1 transition. Pairs of P Frames would be dropped (or repeated) only when aligned with complete Dolby E data blocks, and the latency adjusted to align the Dolby E data blocks to the interlaced house reference signal.

6 Switching of Dolby E with Progressive Video, Operating at 50, 59.94 and 60 Frames Per Second

SMPTE RP 168, (Definition of Vertical Interval Switching Point for Synchronous Video Switching) states, in Section 6, that “for a system with both interlaced video at a specific frame rate and progressive video at double that frame rate, devices handling the progressive video should be referenced to a signal derived from an interlaced format at the interlaced frame rate. Having established this referencing relationship, progressive video devices should switch using the recommended line during Field 1 of the reference signal”.

The same interlaced reference signal should be used the synchronize the Dolby E encoder, thus ensuring that the vertical interval switching area at the beginning of P Frame 1 (the Odd P Frame) of the video signal will fall within the guard band of the Dolby E signal.

7 Dolby E Detection

If a reference signal at 25 or 30 fps is not available, then lock to the incoming program signal, or to a 50 or 60 fps reference signal, and parse the Dolby E stream to locate the data blocks. The guard band preceding the beginning of a Dolby E data block can be used to enable the audio follow video switching command, which will ensure that the video will be switched during the vertical interval (per SMPTE RP 168) and that the Dolby E stream will be safely switched during the guard band interval, preventing corruption of the data.

In embedded audio systems, this requires parsing the pair containing the Dolby E data from among the 8 pairs available in HD-SDI. The choice of pair is important, and would have to be manually set based on the broadcaster systems design. As a broadcaster would need to know this pair number for a functioning system anyway, it is not thought that it is necessary to go as far as parsing every embedded pair, in a “catch-all” scheme.

7.1 The parser may receive its input from any of three sources:

a) Disembed the audio pair carrying the Dolby E data, and parse the resultant AES words. The disembedding delay needs to be known, so the actual embedded position of the Dolby E header can be calculated, and used to enable the switching command.

b) The HD-SDI signal itself can be parsed to locate the appropriate HANC words of the audio pair carrying the Dolby E data. The location of the HANC words can be calculated from SMPTE 299M mappings. The way the embedder distributes audio samples around the switch line (which does not allow any embedded data) also needs to be known.

c) An external AES3 source carrying Dolby E data, derived from the embedded signal. The relative timing between the embedded signal and the AES3 data must be known.

7.2 There are two possible methods of parsing the audio data to create the switch enabling signal:

a) Detect the SMPTE ST 337 header (indicating the start of the Dolby E data block), and base a future frame’s switching on this detection. A valid SMPTE ST 337 detection cannot instigate a switch in the same P Frame because the switch point occurs earlier in time than the SMPTE ST 337 header in a correctly aligned system. The header detection parser would have had to receive a valid Dolby E header at least 2 progressive video frames earlier than the required switch point. As the headers are different for 16 bit and 20 bit Dolby E, both need to be searched for. This method assumes that the Dolby E alignment does not alter drastically between the detection of the header, to the next switch point (just less than 2 P Frames later). The SMPTE ST 337 extended header detection method is recommended to avoid false detection of a header. This method would deal with switches to PCM as well, so switches could occur when no Dolby E was present (i.e., if a SMPTE ST 337 header has not been detected in the last 2 P Frames, assume the audio is PCM). The Validity bit of an AES stream should not be relied upon for determining presence of Dolby E (data), or PCM (audio).

b) To gate switching upon silent (0x00) stereo audio samples, indicating the Dolby E guard band. This method would allow a switch to occur in the same frame as the detection method, because the guard band (on switchable frames) would be present immediately before, and during, the switch point. However, most Dolby E data blocks contain 2 consecutive 0x00 stereo samples around half way through the databurst, so it is recommended to detect at least 3 consecutive stereo 0x00 value sample words. This method also relies on typical audio content being encoded in the Dolby E (this contains <1% of stereo audio samples in the databurst with a value of 0x00). With each full bandwidth channel carrying encoded silence, there is approximately a 12% increase in stereo 0x00 samples within the databurst. Dolby E containing 8 channels of silence would contain nearly 90% of silent stereo 0x00 samples within the databurst, hence increasing the risk of false guard band detection. This risk is reduced given the assumption of a Dolby E alignment equivalent to a well designed SD / interlaced broadcast system. This method would not deal well with transitions to PCM, as the detection method would never allow a switch if the PCM audio samples contained typical audio content (i.e. non silent values). Because of this, it is recommended only to use this method in systems which can guarantee that the parser receives a constant Dolby E signal. The Validity bit of an AES stream should not be relied upon for determining presence of Dolby E (data), or PCM (audio).

7.3 Broadcast systems designers should ensure that:

a) That all inputs to a unit have similar or identical Dolby E alignment. This removes the possibility of a switcher creating a one P Frame period of corrupt data or silence.

b) That if the video system is interlaced, the Dolby E block position relative to the video is as required by Table 1.

Table 1 – Start of Dolby E data block timing

Signal Format	625/25/I	1125/25/I	750/50/P	1125/50/P	525/29.97/I	750/ 59.94/P	1125/ 59.94/I	1125/ 59.94/P	1125/60/P
Earliest valid Dolby E position									
TV line	7	11	15	22	11	21	16	31	31
uS *	390	390	390	390	450	450	450	450	450
Ideal Dolby E line position - 80uS									
TV line	11	19	25	37	13	28	21	42	42
uS *	650	650	650	650	610	610	610	610	610
Ideal Dolby E line position ± 80uS									
TV line ± 80uS	12	21	28	41	14	32	24	47	47
uS ± 80uS *	730	730	730	730	690	690	690	690	690
Ideal Dolby E line position + 80uS									
TV line	13	23	31	46	15	35	26	52	52
uS *	810	810	810	810	770	770	770	770	770
Latest valid Dolby E position									
TV line	31	55	73	110	27	67	50	100	100
uS *	1950	1950	1950	1950	1470	1470	1470	1470	1470
* - in relation to SMPTE RP 168 reference point and approximate values									

8 The Medium and Long Term View

The recent SMPTE ST 2051 Standard (Two-Frame Marker for 50-Hz and 60(/1.001)-Hz Progressive Digital Video Signals on 1.5 Gb/s and 3 Gb/s Interfaces) defines a SMPTE ST 291 Type 2 ancillary data packet that is used to identify P Frame 1 and P Frame 2 of the progressive signal carrying it, so can be used to align a Dolby E signal properly. The packet is identified by a DID value of 46h and an SDID value of 01h. There is a one byte user data word, with bit b0 = 0 indicating P Frame 1 or with bit b0 = 1 indicating P Frame 2. The rest of the bits are reserved and set to 0. Refer to the Standard for complete details.

SMPTE ST 12-2 is a standard that specifies the transmission of time code in the ANC data space of an SD or HD video signal, and explains how to set flags that can identify the odd and even P Frames. Since the ANC data is carried by the video signal, the binding problems is simplified, but not completely solved, as some recorders may drop the timecode data, as does video processing equipment. SMPTE ST 12-2 also specifies that since timecode runs at 25 or 30 Hz, video operating at 50 or 60 fps shall carry the same timecode value in each P Frame of a pair. This provides an odd and even P Frame marker that can be used to align a Dolby E signal properly.

It is expected that the IEEE 1588 Precision Time Protocol will replace tri-level sync and black burst in the broadcast industry over the long-term. The SMPTE Technology committee on timing and synchronization (33TS) is working on solutions that will make the transition as smooth as possible. For Dolby E in particular it is important to note that Dolby E data blocks will continue to have a length or 33.3667ms or 40ms even though the timing signal will have a higher granularity. This calls for devices that not only translate the IEEE 1588 protocol back into a black and burst signal, but these devices will also have to respect the interlaced frame relationship that define the position of the Dolby E data blocks.

9 Guidelines for MPEG Encoding

1. The video source MUST be locked to a dual-standard sync generator which is generating 625/25/I or 525/29.97/I black burst along with tri-level sync for progressive video systems, operating at twice the interlaced rates, and correctly phased per SMPTE RP 168.

2. Assume that the audio and video are both locked to the same sync generator and correctly timed at the inputs to a Dolby E encoder and a video delay unit[1] placed before the MPEG encoder. The video must be delayed by a total of two P Frame durations to match the Dolby E encoding delay, which is one Interlaced frame duration. The audio in the Dolby E stream and the video will thus be in sync at the input to the MPEG encoder

3. The MPEG encoder must be provisioned for Dolby E at 25 fps or 29.97 fps operation.

4. The MPEG encoder must violate the provisions for SMPTE ST 302 in the following ways:

    1. It must treat the incoming Dolby E stream as if it was operating at 25 or 29.97 fps.

    2. It must form one audio data PES packets for every two video P Frames.

    3. The audio PES packet shall have a PTS value nearly equal to that of the corresponding video PES header as required by SMPTE ST 302. The second video P Frame must not have an audio PES packet associated with it. The relative rate of audio samples to video P Frames is 8008/10 for 59.94 Hz video. Audio sample boundaries and video frame boundaries line up exactly only every ten video P Frames.

    4. The audio PES packet shall be transmitted along with the corresponding video PES data in the resulting Transport Stream.

[1] An Evertz frame synchronizer from the 7746FS family may be suitable for either delaying or re-synchronizing the video, or for both tasks.

10 Dolby E Pass-Through

A frame synchronizer must be used at the output of the professional integrated receiver decoder (PIRD) to lock the incoming HD-SDI and Dolby E audio stream to the local reference signal of the pass through (receiving) facility. If the device can handle both HD-SDI and Dolby E data in the “Dolby E safe” mode, per Section 5, then it will re-lock and keep the video/audio synchronized properly. The same considerations apply to a frame synchronizer that may be built into the PIRD.

If the internal or external frame synchronizer cannot operate in the Dolby E safe mode, then the Dolby E stream from the PIRD must be re-synchronized using an external Dolby DP583 Frame Synchronizer, or equivalent. An additional delay of one interlaced frame or two P Frames must be added to the video signal path to compensate for the delay introduced by the Dolby E decoder. Any additional static delay introduced by either the audio or video frame synchronizers must also be compensated for. After that, the dynamic delays introduced by the audio and video frame synchronizers must track each other to preserve audio to video synchronization (lip sync).