DOLBY E

Manufacturer Guidelines: Handling of Dolby® E in Environments Using Progressive Frame Rates (50/60p)

Note:

Dolby® E natively supports frame rates up to 30 fps, and all references in this document to 50p/50i/25 fps systems can be substituted for 59.94p/59.94i/29.97 fps and 60p/60i/30 fps systems.

All future references in this document to the SMPTE RP 168 reference point shall be determined as the SMPTE RP 168 vertical reference point—the point in time where line 1 for 720 59.94p, line 1 for 1080 59.94i, and line 4 for 480 29.97 signals align, and the point in time where line 1 for 720 50p, line 1 for 1080 50i, and line 1 for 576 25 signals align.

Introduction

In Europe, the initial uptake of high-definition (HD) broadcasting involved mainly interlaced video formats. As many European broadcasters and their equipment suppliers already had experience handling Dolby E surround audio in interlaced SD systems, good practices were easily transferred to HD. Considerations for carriage of Dolby in interlaced HD systems are similar to those for SD, as frame lengths are equal (see figure 1).

For 50i, there is one Dolby E frame (25 fps) that starts on the odd field, continues and crosses from the odd to the even field, and ends in the even field. This does not normally cause a problem because equipment switches at the SMPTE RP 168 switch area based on the odd (first) video field (frame-based switching), where the correctly aligned Dolby E guard band is located. Embedders, de-embedders, frame synchronizers, and switchers work fine in this mode—the output Dolby E is error-free.

With the popularity of progressive HD video formats increasing, it is apparent that a little more thought has to be devoted to ensuring that no problems occur with video frame lengths of half the duration of interlaced video frame lengths. The guard band of Dolby E (which protects against errors during a switch point) in this mode does not cover all SMPTE RP 168 switch areas, only every other switch area. This means that without care, a video switch point has a 50 percent chance of corrupting one Dolby E frame (figure 1).

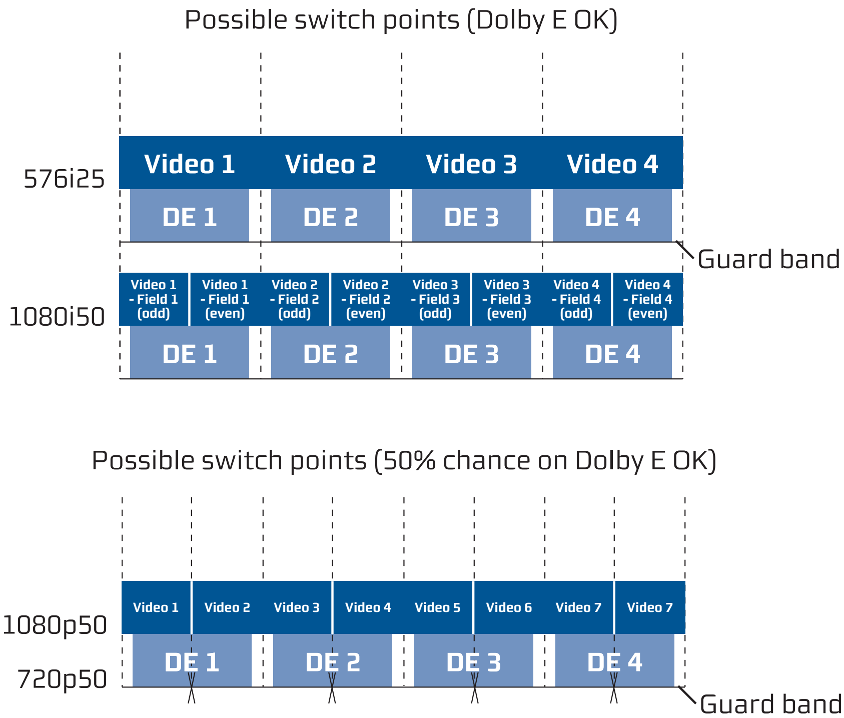

Figure 1. *Interlaced Video and Dolby E (top), Progressive Video and Dolby E (bottom)*

Where Dolby E is used in 50p environments, Dolby E frames have a length of two video frames (40 ms), as the Dolby E frames are still being generated at 25 Hz. This means, as in the previous example, the Dolby E frame starts on one video frame, continues and crosses from this frame into the next, then ends on the second frame. As current 50p equipment doesn't always switch on the first frame—it may switch on the second frame—a single frame of Dolby E may be corrupted at this switch point. This typically results in a momentary (one-frame) mute in suitably designed broadcast systems, which is considered acceptable (although not ideal) by several existing broadcast users. It should be noted that HD-SDI embedders and de-embedders work fine in this mode; it is only switching and frame synchronising that causes an issue.

720p 50/60 Dolby E Synchronization

As frame synchronizers are used to "firewall" incoming live-feed signals, it is essential to have the frame synchronizer operate in a "Dolby E safe" mode, allowing clean downstream switching. Two points must be considered in designing such a frame synchroniser:

1) Any frame synchronisers should drop two progressive frames, instead of one. Keeping Dolby E aligned by this method is considered to be worthwhile given the potential small motion artifact observed when dropping two progressive frames.

2) The dropping of frames should be aligned to both a 25 fps reference (bilevel or trilevel) and the embedded Dolby E header.

   Standard interlaced frame synchronizers (that drop an entire frame's worth of audio samples during a video frame drop) handle Dolby E well (figure 2).

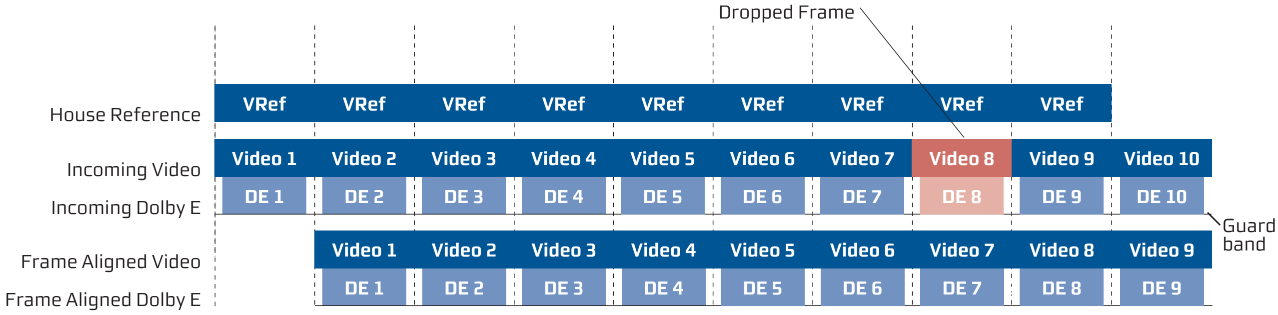

Figure 2. *Standard Interlaced Frame Synchronizer Locking to Interlaced Reference*

However, with a standard progressive frame synchronizer dropping single progressive frames, Dolby E frames would be corrupted (figure 3).

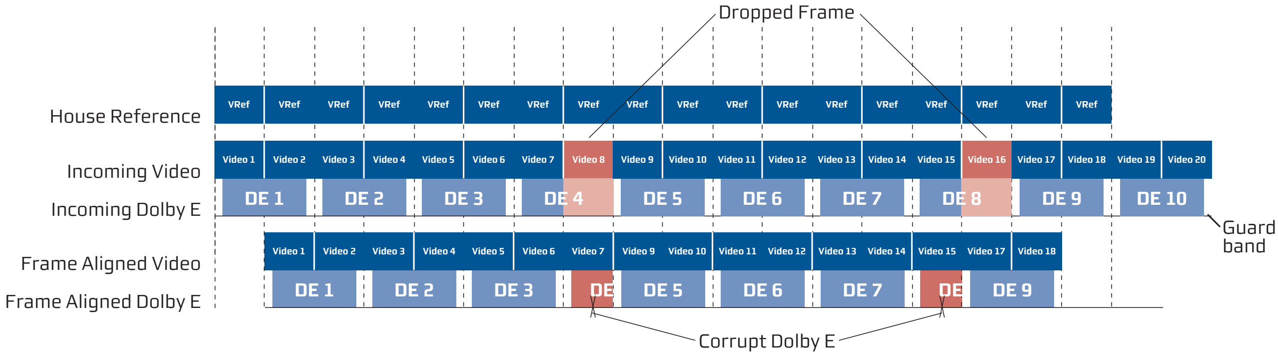

Figure 3. *Progressive Frame Synchronizer Locking to Progressive Reference*

Just having a progressive frame synchronizer lock to a 25 fps signal, and dropping two frames instead of one, is not enough on its own to ensure clean Dolby E. There is a 50 percent risk of starting the two-frame drop sequence incorrectly, by an offset of one frame (figure 4). This would also mean that downstream equipment relying on the same 25 fps reference to correctly switch Dolby E cannot guarantee error-free switching. (The 25 fps reference and embedded Dolby E are not aligned.)

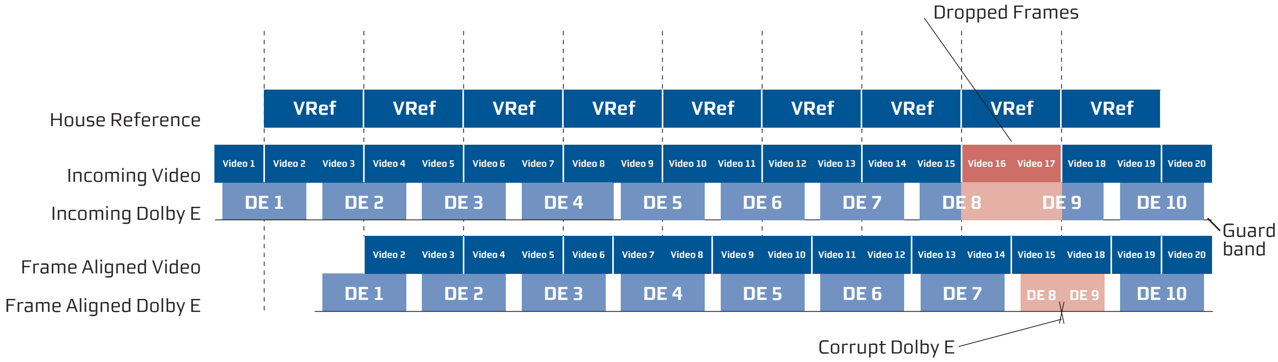

Figure 4. *Progressive Frame Synchronizer Locking to Interlaced Reference*

The answer is to have the frame synchronizer locked to a 25 fps reference signal and detect (from the Dolby E header) where to start the two-frame drop sequence (figure 5).

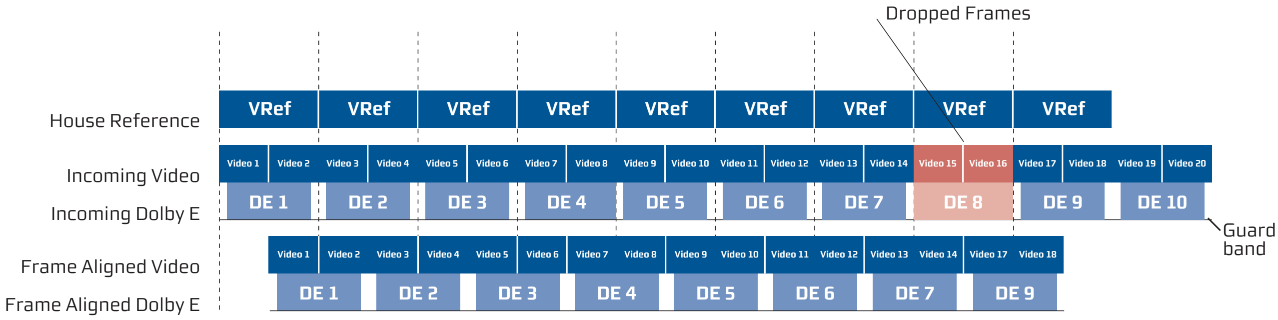

Figure 5. *"Dolby E Safe" Progressive Frame Synchronizer Locking to Interlaced Reference*

The detection of the Dolby E header (and hence the Dolby E alignment) can be done in several ways (see the "Dolby E Detection" section for details).

Optionally, frame synchronizers could perform realignment of the Dolby E on the incoming HD-SDI signal to the appropriate position. This would add the benefit of ensuring all embedded Dolby E output is not only error-free during frame drops but also correctly aligned for further processing, switching, and so on. The re-alignment should result in a 0.5 to 1.5 Dolby E frame delay, associated with a two progressive video frame delay. This allows for Dolby E to be realigned to the closest interlaced frame boundary, rather than realignment to the next complete interlaced frame boundary.

Consideration of previous upstream switches that were not "Dolby E safe" should be taken. In this case, the input stream may be as in figure 3. A frame synchronizer would deal with this type of stream only by parsing the entire Dolby E stream to detect the end of a complete Dolby E frame, rather than just detecting the start by the presence of SMPTE 337M headers. This would also require pairs of progressive video frames to be selected to be "dropped" based on the presence of full Dolby E frames only, and the latency adjusted so that the output Dolby E is always aligned to the local 25 Hz reference.

720p50/60 Dolby E Switching

Several methods of producing "Dolby E safe" switching equipment at progressive video frame rates are described in the following section.

Modes of Locking

1) Accommodate a 25 fps bilevel B&B signal to lock to, and ensure all switches, processes, and so on are done according to the 25 fps start of frame. The 25 fps reference and 50p video signal should be correctly aligned by the broadcast system's design, and the equipment requires a suitable reference input. A compromise of this approach is that it does not offer the gain in accuracy that would be achieved using a 50p trilevel signal. Ideally the B&B is the same as used to lock the Dolby E encoder, although this may not be practical with remote live feeds. Due to the switch points being different for 576i systems and 720/1080 systems, either the 25 fps reference input needs to be offset (retarded) by the broadcast systems design, or the switching unit needs to calculate an offset from the 576i switch point. This will ensure switch points at the correct time in the 720/1080 signal.

2) Accommodate a 25 fps 720p25 or 1080p25 trilevel reference signal to lock to. This would give a similar situation to the previous example but include the benefits of lower jitter from a trilevel reference signal. Again, this would require the broadcast systems design to provide suitably aligned reference signals.

If applicable, these modes of "field identification" could be used in addition to a 720p50 or 1080p50 trilevel reference signal, so both correct frame switch alignment and accurate switching points could be made. This would require two reference inputs.

If a 25 fps reference signal is used in this way, it would not seem unreasonable to assume (or make clear that) it is the responsibility of the broadcaster systems design to guarantee aligned 25 fps reference and Dolby E. This should not be an issue for a broadcaster that controls the Dolby E encoding equipment (which requires a 25 fps bilevel B&B signal anyway to operate).

For sources outside of the control of the broadcaster (for example, incoming live feeds that are frame synchronized), the frame synchronizer would have to operate in a "Dolby E safe" mode (see the preceding section on frame synchronizers), whereby two frames are dropped (instead of one), and these frame drops are aligned correctly with both the incoming Dolby E headers and the 25 fps reference signal. After this point, it would be assumed a similar situation to the broadcaster owned Dolby E encoders.

This method safely allows a sequential mixture of Dolby E and PCM to be presented on the same embedded pair (as is the situation in switched-out modes for standard PCM source).

Dolby E Detection

If locking to 25 fps bilevel or 720p25 or 1080p25 trilevel reference signal is not an option, then lock to either incoming HD-SDI or a 720p50 or 1080p50 trilevel reference signal, and analyze/parse the incoming Dolby E, and align (or "gate") any switches to be on the frame that contains the Dolby E header. It is proposed that the Dolby E header does not give specific timing information for when to make a switch, but does give an indication on whether a switch (aligned to the correct, odd-frame SMPTE RP 168 switch point) should be performed in any given frame.

In completely embedded systems, this requires analyzing/parsing the correct pair (containing the Dolby E) from the eight 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 considered necessary to go as far as parsing every embedded pair in a "catch-all" scheme.

There are potentially three methods of feeding a parsing function:

a) Internally de-embed the nominated audio pair completely, and parse the resultant AES words. The de-embedding delay needs to be known, so the "actual embedded" position of the Dolby E header can be calculated, and used to base the switching gate on.

b) Internally parse the HD-SDI, looking for the appropriate HANC words of the nominated audio pair. These can be calculated from SMPTE 299M mappings. Consideration needs to be given to the embedder behavior—that is, how samples are distributed around the switch line (which does not allow any embedded data).

c) Accommodate an external Dolby E input (on AES carrier) and parse the received AES words. The difference in timing between the Dolby E on AES and Dolby E embedded in HD-SDI needs to be known; if not, it must be of a reasonable, controlled, and low value.

There are potentially two methods of parsing audio to gate switches:

1) To detect the SMPTE 337M header (indicating the start of the Dolby E data burst), and base a future frame's switching on this detection. A valid SMPTE 337M detection cannot instigate a switch in the same frame because the switch point occurs earlier in time than the SMPTE 337M header in a correctly aligned system. The header detect parser would have had to receive a valid Dolby E header at least two 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 two frames later). The SMPTE 337M extended header detection method is recommended to avoid false detection of a header. This method would deal with upstream transitions to PCM (in the same embedded pair) as well, so switches could occur when no

Dolby E was present (that is, if a SMPTE 337M header has not been detected in the last two frames, assume the audio is PCM, and allow switching on any frame). The validity bit of an AES stream should not be relied upon for determining presence of Dolby E (data) or PCM (audio).

2) 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 frames contain two consecutive 0x00 stereo samples around halfway through the data burst, so it is recommended to detect at least three consecutive stereo 0x00 value sample words (a period of 62.5 us). This method also relies on typical audio content being encoded in the Dolby E (this contains less than 1 percent of stereo audio samples in the data burst with a value of 0x00). With each full bandwidth channel encoded with silence, there is approximately a 12 percent increase in stereo 0x00 samples within the data burst. Dolby E containing eight channels of silence would contain nearly 90 percent of silent stereo 0x00 samples within the data burst, 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 switches to PCM, as the detection method would never allow a switch if the PCM audio samples contained typical audio content (that is, nonsilent values). Because of this, it is recommended to only use this method in systems that can guarantee a constant Dolby E signal on the parsing function input. The validity bit of an AES stream should not be relied upon for determining presence of Dolby E (data) or PCM (audio).

Optionally, switching units could perform realignment of the Dolby E on the incoming HD-SDI signal to the appropriate position. This would add the benefit of ensuring all embedded Dolby E output is not only error-free during frame drops but also correctly aligned for further processing, switching, and so on.

The following assumptions are reasonable, and should be made clear to broadcast systems designers:

1) That all inputs to a unit have similar or identical Dolby E alignment. This removes the possibility of a switcher creating a period of corrupt data or silence consisting of one progressive video frame. It also removes the need to parse every input to a switching unit.

2) That if the video system is interlaced, the Dolby E alignment (on a systems level) has been designed to be in the correct position (see table 1).

	625 25 PAL SD	1080i 50 PAL HD	720p50** PAL HD	525 29.97 PAL HD	1080i 59.94 PAL HD	720p 59.94** PAL HD
Potentially earliest valid DE position TV line us*	8 450	13 450	17 450	12 510	18 510	23 510
Ideal DE line position - 80 us TV line us*	11 650	19 650	25 650	13 610	21 610	28 610
Ideal DE line position ± 80 us TV line ± 80 us us ± 80 us*	**12** **730**	**21** **730**	**28** **730**	**14** **690**	**24** **690**	**32** **690**
Ideal DE line position + 80 us TV line us*	13 810	23 810	31 810	16 770	26 770	35 770
Potentially latest valid DE position TV line us*	30 1860	53 1860	70 1860	26 1400	48 1400	63 1400

* In relation to SMPTE RP 168 reference point, and approximate values.
** In relation to the first (odd) frame.

Table 1. *Dolby E Positions*

Related to the proposals in this document, SMPTE 2020-1 defines progressive frame 1 of a pair of progressive frames as the frame that begins closest to the beginning of field 1 of an interlaced (PAL or NTSC) color black reference signal. The Dolby E burst has to be lined up with progressive frame 1 and any switching must occur on progressive frame 1 in order to secure the integrity of Dolby E.

Long-term Foresight

It is expected that the IEEE 1588 Precision Time Protocol will replace trilevel sync and black burst in the broadcast industry over the long term. The SMPTE/EBU task force on timing and synchronization 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 frames will continue to have a length of 33.3667 ms or 40 ms 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 also respect the interlace frame relationship that defines the position of the Dolby E frames.

SMPTE 12M-2 is a new document that specifies the transmission of timecode in the ancillary data space of an SD or HD signal, and explains how to set flags that can mark the first and second frames of a P frame signal. Here, the ANC data is part of the image, so the binding problem is mostly solved (some recorders may still drop the ANC data, and production switchers will replace the H and V intervals), but IF the flags in the timecode are properly used, they could help.

SMPTE 12M-2 also mentions that since timecode runs at 25 or 30 Hz, any 50 or 60 Hz video shall carry the same timecode values in each pair of P frames. This provides another 25 or 29.97 Hz reference signal to align Dolby E with.

720p MPEG Encoding and Dolby E Pass-Through

The following approach shall be recommended:

1. The video source must be locked to a dual-standard sync generator that is generating NTSC or PAL black burst along with 720p trilevel sync. This must be correctly phased per SMPTE RP 168-2002 Annex A, Section A.1.

2. Assuming that the audio and video are properly locked and timed at the inputs to a Dolby E encoder and a video frame sync placed before the MPEG encoder, the video must be delayed two frame times (2 × 16.6833 ms = 33.3667 ms for NTSC or 2 x 20 ms = 40 ms for PAL). The Dolby E encoder already has a single frame delay (33.3667 ms or 40 ms), so the additional video frame delay time will bring the audio back into sync with the video before encoding. An Evertz 7746FS-EAES4-HD or Harris/Leitch HFS6801+D, VFS-3901H, or X75HD may also be suitable for delaying the video. If this delay is done properly, the MPEG encoder will receive the resultant signals with proper audio/video timing.

3. The MPEG encoder will be provisioned for Dolby E at 29.97 Hz or 25 Hz operation.

4. The MPEG encoder will violate the provisions for SMPTE 302M, Section 5.9 in the following ways:

- It will treat the incoming Dolby E stream as if the encoder were operating at 29.97 Hz or 25 Hz.

- It will form one audio data PES packet per two video frames.

- The PES packet will have a PTS value nearly equal to that of the corresponding video PES header as required by SMPTE 302M, Section 5.10. The second video frame will not have an audio PES packet associated with it. The relative rate of audio samples to video frames is 8008/10 for 59.94 Hz video. Therefore, the audio PES packet will line up exactly only every ten video frames.

- The audio PES packet will be transmitted along with the corresponding video PES data in the resulting TS during the same frame time.

5. The MPEG decoder (PIRD) will bring the video back to HD-SDI and the Dolby E audio back to the AES/EBU interface.

6. A frame synchronizer must be used at the output of the PIRD to lock the HD-SDI and Dolby E audio to the receiving facility. If the device can handle both HD-SDI and Dolby E audio, then it will keep the video/audio synchronized properly. Care may need to be taken to ensure that it will handle two different frame rates simultaneously. The user must keep in mind that for each frame of audio (33.3667 ms or 40 ms), the video must be delayed two frames (2 × 16.6833 ms = 33.3667 ms or 2 x 20 ms = 40 ms) in order to keep proper audio/video timing. If the video frame synchronizer cannot handle nonaudio AES data (such as Dolby E), then a DP583 audio frame synchronizer should be used to synchronize the audio.

DOLBY E

Note: The Dolby DP583 Frame Synchronizer will delay the Dolby E bitstream between 33.3667 (40 ms) and 66.7334 ms (80 ms; similar to the behavior of a video frame synchronizer).

7. The frame synchroniser(s) must be locked to the same sync generator as is used to feed the Dolby E decoder with black burst. As with the originating facility, these two generators must have the vertical phase relationship specified in RP 168-2002 Annex A, Section A.1.

Conclusion

To summarize, the problem of Dolby E carriage in 50/60p environments arises from the lack of information in a 50/60p signal at less than 50/60 Hz, and information on 25 or 29.97 Hz references must be found elsewhere.

Using timecode as a reference to determine the correct switching point would be a good place to start in trying to resolve this issue if it was used consistently and in the manner SMPTE has described in trying to find a solution for this problem.

As it appears unlikely to expect a revision to SMPTE RP 168 that introduces markers for odd and even frames, either the proposal in this document, or SMPTE 12M-2 seems to offer the most straightforward short-term solution to deriving a 25 or 29.97 Hz reference until we can expect to see wide adoption of IEEE 1588.

Dolby has detailed information on how to implement Dolby E carriage and switching, which can be made available under NDA to interested equipment manufacturers.

Dolby Laboratories, Inc.
Wootton Bassett, Wiltshire SN4 8QJ England T 44-17093-842100 F 44-1793-842101
100 Potrero Avenue, San Francisco, CA 94103-4813 USA T 415-558-0200 F 415-863-1373
dolby.com

W08/20466