



An MPEG-4-Based High-Definition Videotape Recorder

By Richard Lewis

The subject of this paper is an advanced tape format designed especially for digital cinema production and post-production, ultra-high-quality shooting, and blue-and-green-screen shooting. It is also suitable for final on-air program delivery and interchange of HDTV. This paper will concentrate on the coding techniques used, the multiformat capability demanded by users, and the implementation of a format converter to allow easy integration into a multistandard world. A key development is the implementation of an MPEG-4 Studio Profile chipset. With 10-bit recording at HD resolutions in either 4:2:2 (YCbCr) or full chroma bandwidth 4:4:4 (RGB) at up to 60 frames/sec, the videotape recorder (VTR) is also backwards-compatible with the two most widely used tape formats in the HD and SD world. The packing density on tape is well over 300 Mbits/sq. in., nearly five times that of the Type D6 format, and a portable version, which can record at double the net video data rate. That is 880 Mbits/sec, allowing a full 444 (RGB) recording at only 2:1 compression. This format has been proposed to SMPTE for standardization.

Two types of machine have been developed using MPEG-4 coding. First, a studio machine with full editing capabilities including pre-read, individual editing of the 12 x 24-bit audio channels, and so on; second, a portable (battery operated) machine is available and uniquely can support double tape speed operation, allowing extended recording capability.

Multiformat

The internal structure of the videotape recorder (VTR) has been designed to allow recording and playback at the frame rates specified in SMPTE 274M as indicated in Table 1.

Normal Tape Speed

The first thing to notice is that both 4:2:2 and 4:4:4 sampling are possible. When the tape speed is normal, 4:2:2 sampling up to 30PsF is compressed at 2.7:1, using a combination of MPEG-4 Studio Profile Discrete Cosine Transform (DCT) and Differential Pulse Code Modulation (DPCM) encoding. When recording 4:4:4 images up to 30PsF, the 4:4:4 standard quality (SQ) mode is used, at a compression rate of 4:1. Also, 720/50/ 59.94P according to SMPTE 296M can be recorded in this mode.

Double Tape Speed

A second mode of operation is possible, requiring double the normal tape speed. This functionality is incorporated only on the portable recorder and allows several variations of recording use. First, 4:4:4 recording can be made in high-quality (HQ) mode up to 30PsF, yielding a compression ratio of 2:1. Then 4:2:2 can be recorded at 60P with a compression of 2.7:1. Finally, a

Table 1—The Current and Future Multiformat Capabilities of the New VTR

Sampling	Resolution	Compression Ratio	Tape Speed	Scan Mode	Picture Rate
4:2:2	1920 x 1080	1/2.7	Normal	Interlace	50, 59.94, 60
				PsfF	23.98, 24, 25, 29.97, 30
				Progressive	50, 59.94
4:4:4 SQ	1920 x 1080	1/4		Interlace	50, 59.94, 60
				PsfF	23.98, 24, 25, 29.97, 30
				Double	
4:4:4 HQ	1920 x 1080	1/2	Interlace	50, 59.94, 60	
			PsfF	23.98, 24, 25, 29.97, 30	
			Interlace	50, 59.94, 60	
4:2:2 3D	1920 x 1080	1/2.7	PsfF	23.98, 24, 25, 29.97, 30	
			Progressive	59.94, 60	
4:2:2	1920 x 1080	1/2.7			

so-called 3-D mode is possible, where two entirely separate 4:2:2 sources at up to 30PsF can be recorded on one tape, ensuring perfect synchronism. Both channels can be played back separately by the studio machine by employing a 200% variable speed mode.

Format Converter

Because of the enormous variety of signal formats that are allowed within the various SMPTE documents, including standard definition, a format converter has been developed to allow easy interchange among them. The basic VTR itself can handle all of the nine different frame and field rates by changing the linear tape speed, thus avoiding the need for frame-based standards conversion. In addition, a 4:2:2 60P recorded tape can be played back at half-speed (30PsF mode) providing 50% slow down. In all cases, the tape footprint is the same.

This naturally changes the program length, but is a common feature with telecines, and so is well understood. The basic requirements are as follows:

- 2:3 pull-down for conversion from 23.98PsF to 59.94i (NTSC)
- 1080i and PsF conversion to 720P and vice versa
- 1080i and PsF conversion to 525i or 625i and vice versa
- 4:2:2 conversion to 4:4:4 and vice versa
- A combination of the above

A combination could be an original 4:4:4 tape at 23.98PsF converted to 4:2:2, adding 2:3 pull-down, and finally converting to 525i standard-definition. For example, transmission or offline

editing.

Note: The SD output marked “?” requires the format converter to be installed. Digital Betacam requires a processor board to be installed.

Table 2 shows the capabilities of the studio machine. Note that other popular 1/2-in. HD (SMPTE type D-11) and SD (Digital Betacam) tape format playback is also supported.

Codec

The codec is the key to the flexibility of the many different formats supported by this development. It has the

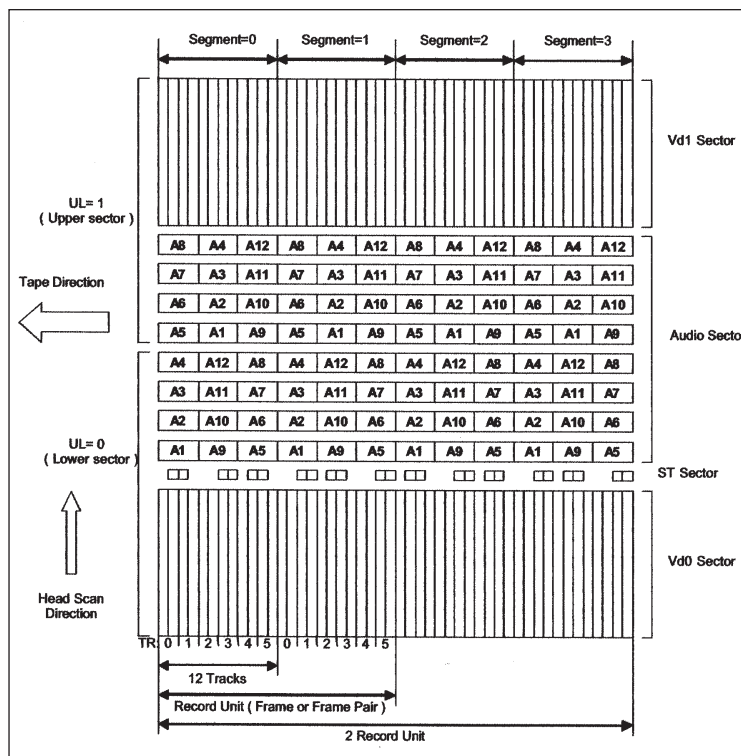


Figure 1. Record unit, segment, channel, and track pair counts.

Table 2—The Output Capabilities of the Studio Machine

Playback Tape Format			HD-SDI OUT		SD OUT	Format Converter OUT
MPEG-4 VTR or SMPTE Type D1	1080/422	23.98PsF	1080/422	23.98PsF	----- ? 525/59.94i	1080/444/23.98PsF 1080/422/59.94i
		24PsF		24PsF	-----	1080/444/59.94i 720/422/59.94P 1080/444/24PsF
		25PsF		25PsF	----	1080/422/60i 1080/444/60i
		29.97PsF		29.97PsF	625/50i	1080/444/25PsF 720/422/59.94P
		30PsF 50i		30PsF 50i	525/59.94i	1080/444/29.97PsF 1080/444/30PsF 1080/444/50i 720/422/59.94P
		59.94i		59.94i	525/59.94i	
		60i		60i	-----	1080/444/59.94i 1080/444/60i
MPEG-4 VTR	1080/444	23.98PsF	1080/444	23.98PsF	----- ? 525/59.94i	1080/422/23.98PsF 1080/422/59.94i 720/422/59.94P 1080/422/24PsF
		24PsF		24PsF	-----	1080/422/60i 1080/422/25PsF 1080/422/29.97PsF
		25PsF		25PsF	625/50i	
		29.97PsF		29.97PsF	525/59.94i	720/422/59.94P 1080/422/30PsF 1080/422/50i 720/422/59.94P
		30PsF 50i		30PsF 50i	525/59.94i	1080/422/59.94i 1080/422/60i
		59.94i		59.94i	-----	
		60i		60i	-----	
MPEG-4 VTR	720/422	59.94P	720/422	59.94P	525/59.94i	1080/422/59.94i 1080/444/59.94i
Digital Betacam	525	59.94i	1080/422	59.94i	525/59.94i	720/422/59.94P
	625	50i		50i	625/50i	-----

following attributes:

- DCT, DPCM, VLC are compliant with the MPEG-4 Simple Studio Profile coding tools.
- Shuffling and rate control are unique for this VTR.
- Multichip combination is possible in 1929 x 1080 as follows:
 - 1 chip can do 4:2:2 30PsF
 - 1 chip can do 1280 x 720/ 50/59.94P
 - 2 chips can do 4:4:4 30PsF, 4:2:2 30PsF 3-D mode and 4:2:2 60P
 - 4 chips can do 4:4:4 60P

- The sync block structure is compatible between 4:4:4 compression at 2.7:1 and 4:1.
- 3 lines per field are assigned to uncompressed 10-bit data words (metadata).
- Playback compatibility: 4:2:2 60P tape can play back at 4:2:2 24, 25, and 30PsF.

Physical Layout and Segments

Figure 1 shows the basic footprint as it is laid down on tape. Each segment is composed of 6 track pairs, so a frame at 4:2:2 30PsF equates to 24 tracks.

Picture Segmentation

Each 4:2:2 PsF 1920 x 1080 picture is first reconstituted into a progressive 1920 x 1080 frame, then each frame is divided into 8160 16 x 16 shuffle blocks for luminance and two cosited 8160 8 x 16 blocks for chrominance. In the case of 4:4:4 PsF, there are three 8160 16 x 16 blocks for each RGB. In the case of interlace signals, each field is treated as an independent 1920 x 540 field and is divided into 4080 16 x 16 blocks for luminance and two 4080 8 x 16 blocks for chrominance. An example for 4:2:2 PsF is shown in Fig. 2.

Each 120 x 68 shuffle block is then divided into 4 shuffle sets, each set containing 2040 shuffle blocks. Finally, each shuffle block in a shuffle set is allocated to a unique macroblock in one of 40 macroblock units. For a 1920 x 1080 picture, there are 204 macroblocks within a macroblock unit (204 macroblocks x 40 macroblock units = 8160 original sync blocks), as shown in Fig. 3. The actual data assigned to a macroblock and then to a macroblock unit is assigned by a pseudo-random equation, depending on the block number and allocation size. Thus, there are 40 macroblock units of shuffled picture data. These now pass to the DCT and DPCM processes for MPEG-4 encoding.

Segments

Segments are used to define how the MBU compressed data stream and twelve AES3 audio data streams are mapped to the helical tracks.

Figure 4 shows the segment mapping for normal 4:2:2 or 4:4:4 SQ recordings, where two segments comprise a frame. Figure 5 shows the mapping for a 4:4:4 HQ mode recording, where the frame is extended to four segments. The overall frame timing remains the same (1/30th sec), but the segment duration is reduced.

In the same way, it can be seen that in 4:2:2 60P mode, the segment timing would be reduced to 1/120th of a second, so that two segments equal 1/60th sec.

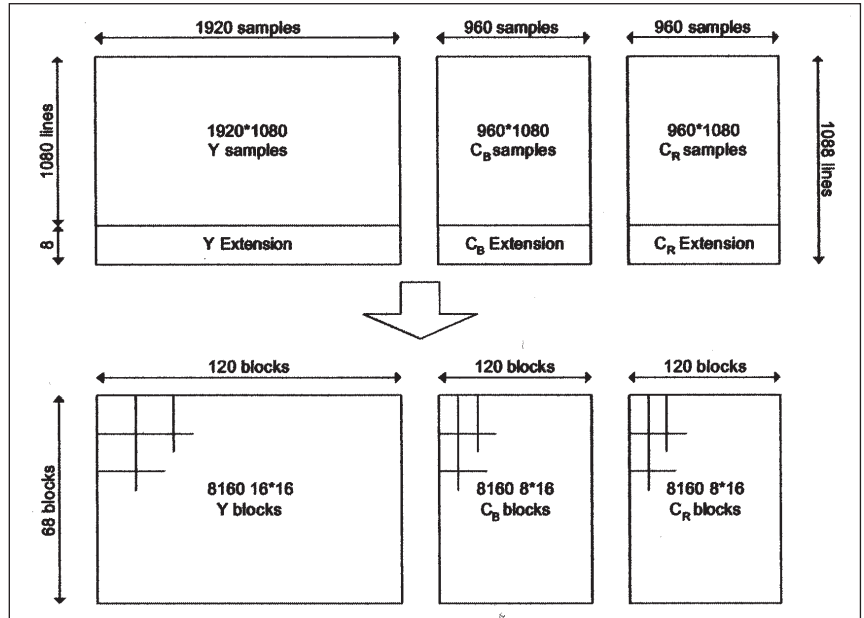


Figure 2. 1920 x 1080/PsF 4:2:2 YCbCr shuffle blocks.

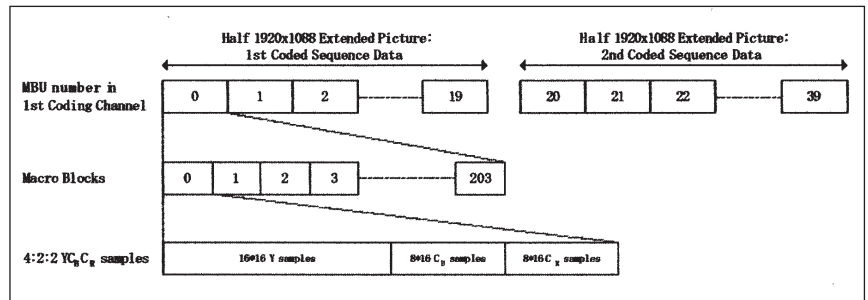


Figure 3. 1920 x 1080 4:2:2 YCbCr macroblock unit number allocation.

Thus, each segment can process one 4:2:2 or 4:4:4 SQ signal at 440 Mbits/sec as a net video data rate. Each segment also has its own error correction. So various formats in the future can be achieved simply by parallel processing of this one segment unit. Figure 6 shows the unique 4:2:2 3-D mode segment mapping; the left and right channels are recorded alternately on the tape. The portable machine can playback both the left and right channels simultaneously, while the studio machine can playback either channel separately, by employing a 200% variable speed mode. The left and right channels both record the same audio data to ensure normal speed audio playback, even if the tape is played back using the 200% variable mode on the studio machine.

As previously mentioned, each segment has its own error correction, the outer correction being extremely

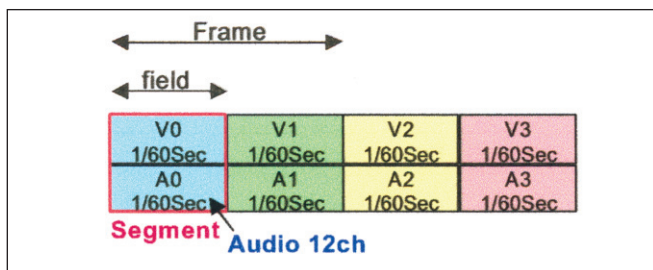


Figure 4. 4:2:2 60i or 4:4:4 SQ 60i segment mapping.

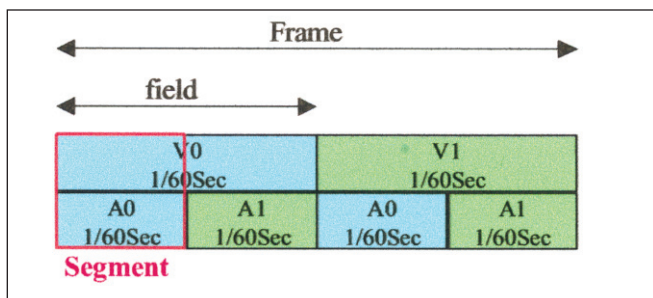


Figure 5. 4:4:4 HQ 2:1 compression segment mapping.

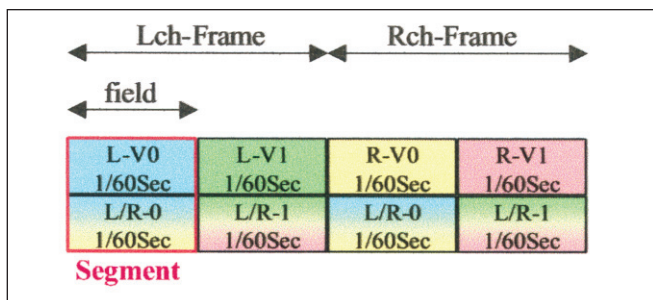


Figure 6. 4:2:2 3-D mode segment mapping.

powerful with 12 bytes of correction for every 114 byte of data (Fig. 7). In addition, 3 outer tables are recorded on each physical track, and these outer tables are shuffled among the 12 tracks comprising a record unit. The net result is that 1 whole track can be missed, and the data can be perfectly recreated.

The audio data is corrected in exactly the same manner except that there are 8 bytes of outer correction for every 8 bytes of data, in addition to the 24 audio tables recorded per field (2 tables per channel), which offers further physical 100% redundancy.

Figure 5 also shows that the audio is written twice; that is, A0 is written twice, once associated with V0 and once with V1. This means that in any condition that the video is skipped, for example, single-channel playback of a tape recorded in 3-D mode or a 60P tape played back at 24, 25, or 30PsF, there will still be continuous audio.

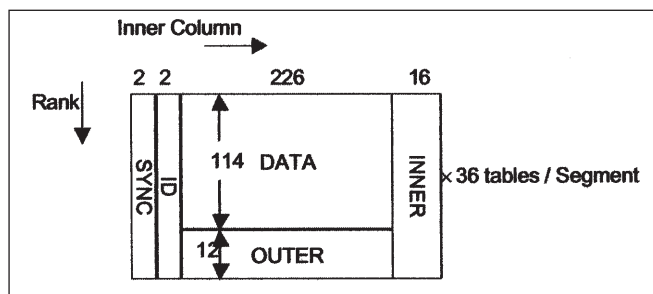


Figure 7. Video segment error-correction table.

Conclusion

The development of this equipment has required many new developments; however, one of the key ideas is the use of segments, as this allows the flexibility to multiply simple processing units to accommodate both current and future derivatives of existing formats. In particular, this will allow the portable unit, in the very near future, to record a variable frame rate from 1 frame/sec to 60 frames/sec with no external processing required to recover the data by deleting frames.

The quality of the full bandwidth 10-bit interface combined with the efficiency of MPEG-4 coding realizes a machine suitable for intensive post-production techniques and tasks such as matting, which have previously only been possible with nonportable hard-disk-based recorders.

Acknowledgments

Thanks to my development colleagues at Sony Atsugi, Japan, and the Sony U.K. Research Laboratories, who developed the Codec Chipset.

First published in the IBC 2004 Conference Proceedings, Amsterdam, The Netherlands, Sept. 9-13, 2004. Copyright © International Broadcasting Convention.

The Author

Richard Lewis is currently chief engineer, marketing, for Sony Europe, where he manages Sony digital cinema style products in acquisition, recording, and display products for Europe, the Middle East, and Africa. These include CineAlta, HDCAM SR, and the new SXR 4K D-Cinema projector.

Lewis has been with Sony for more than 23 years in product marketing roles, initially introducing U-Matic products and then Betacam into Europe. Subsequent moves were into camera, VTR, editors and post-production equipment. For the past 20 years he has been involved in high-definition.