Preprint 2351 (N5)

Audio in the DVTR

J.H. Wilkinson and G.A. Walker

Sony Broadcast Limited

Presented at the 80th Convention 1986 March 4-7 Montreux, Switzerland

AN AUDIO ENGINEERING SOCIETY PREPRINT

AUDIO IN THE DVTR

J.H. Wilkinson and G.A. Walker

Sony Broadcast Limited

1. Introduction

After ten years research and many years of discussion between the standardisation committees representing both the manufacturers and broadcasters, the digital video tape recorder (DVTR) is now close to the point of production. This DVTR is based on a worldwide standard shortly to be ratified by the CCIR. It represents a major step forward in basic quality for both video and audio signals and will allow a high level of multi-generation use for programme assembly. The video part of the standard is concerned with component video and as a result, the standard is applicable, with only minor changes, to both major line standards 625/50 and 525/60. The component strategy removes all the differences created by the composite signals, PAL, NTSC, SECAM and their derivatives. The objective of this paper is to explain the areas of the format that directly relate to the audio signal processing and will include the topics of data formatting, error control and editing.

Firstly though, the basic specifications of the audio part of the machine are given below:

(i) Sampling frequency : 48kHz
(ii) Number of bits per sample : user selectable from 16 to 20
(iii) Number of channels : 4

The resulting performance will be at least as good as any machine yet available. Further, if the 'super hi-fi' option of 20 bits resolution is selected, then the potential analogue performance will be unmatched, leaving headroom for maladjusted signal levels and representing a considerable improvement over analogue video recorders. The four channels may be specified as single channels or stereo pairs and in the latter case DVTR will maintain sample accuracy between the two channels. For all audio channels, accurate lip sync performance will be maintained even after many generations of edit operations.

The DVTR format is designed to facilitate easy interfacing of audio and video signals at a digital level. The digital audio is accepted in the EBU/AES format, allowing for the sampling frequency of 48kHz $\pm$ 3 parts in 10^6 synchronous with the video. There are many variations possible within the format and these will be described in section 2. The coding in all cases is two's complement linear PCM.

In order to achieve the necessary head to tape writing speed for the video information a helical scan recorder presents the only practical choice. The tape is helically

wrapped around a drum containing the recording, erase and replay heads. Rotating the drum at a frequency of 150Hz produces the track pattern shown in figure 1. This figure also includes a number of key dimensions for reference purposes. The position of the audio bursts is an important feature. During early development of the DVTR the audio was located in two bursts at each end of the video track. As a result, the audio was prone to tape edge damage and tracking skew resulting from tape stretch. Considering the mechanics of recording a total data rate in excess of 220Mb/s, there were only two practical options for processor design; either a two channel design, or four channel. The latter option was adopted for its greater tolerance to likely (but extreme) fault conditions (ref 1). As a consequence, an edit gap had to be introduced into the centre of the track pattern for 525/60 recordings. Although not necessary, these edit gaps were also introduced into the 625/50 standard for the reason for consistency. Having gaps in the middle of each track means that it is easy to locate the four audio bursts in this area of the tape where it benefits from both a more reliable data pick-up and greater resistance to tracking skew errors.

Further, to aid cueing in tape shuttle a longitudinal analogue audio edit track is provided (fig 1). This is provided mainly to aid cueing when spooling the tape, however it does not preclude the use of digital audio recovered from the helical tracks when, and if, technology permits.

Independent insert editing is provided on all four audio channels through the use of data duplication. Figure 2 highlights the arrangement of audio bursts in the helical track format. Each audio segment corresponding to a time window of 6.67msecs for all four channels and is composed of sixteen sectors each containing either odd or even samples along with control words and error check words. It is evident from figure 2 that each sector is repeated. This is an essential part of the error protection strategy and is used at edit points to store the downstream audio data in the duplicate sectors. The processor can, during the edit period, cross fade between the original upstream audio and the newly inserted downstream audio on replay. The error protection is reduced during this period and this factor will be explained further in the section dealing with error control. In each sector there is a maximum of 161 audio samples arranged as shown in figure 3.

The next section will describe the function of the additional words contained within each sector.

2. Input Formatting

The digital audio input is presented in the EBU/AES format (ref 2). The DVTR format makes provision for recording a maximum of twenty bits per audio sample and a small key subset of the information provided by the user and status bits. However it is possible to extend the amount

user, channel status and validity information at the expense of audio bit resolution and all the possible options are shown in figure 4. There are three types of additional control words recorded within each sector; interface control words, processor control words and user control words.

The INTERFACE CONTROL words are decoded from the digital interface to the DVTR. PREF and CHAN are derived from bytes '0' and '1' of the twenty four channel status bytes within the EBU/AES format. LNGH indicates the number of bits within a word that represent an audio sample and which bits within a word are derived from the EBU/AES format (fig 4). SMARK 0 and SMARK 1 are pointers to the first and last block sync markers (from the EBU/AES format) respectively. As the block sync can generally only occur within either an odd sector or an even sector then AA_{16} is stored in each SMARK for the sector not containing the block sync. (Note: AA_{16} cannot naturally occur as a valid SMARK value). After editing there may be multiple markers, however only the first and last ones will be stored.

It is also necessary to record PROCESSOR CONTROL words which are used to convey information from the record side to the replay side of the processor. It will also enable exchange of programme material between different machines. ELAP is used to indicate that an edit has occurred during this segment and that for this segment the duplicate (upper) sectors contain downstream audio and the lower sectors retain the original upstream audio data. SEQN is an arbitrary count (0-15) that is used to aid data recovery at shuttle speeds. The final processor control word is BCNT which indicates the number of samples within the current block that are valid audio samples. This is normally 160 but can vary from 159 to 161. This is a facility which allows a synchroniser to be built in to the DVTR enabling the processor to be fed with asynchronous video sources. This flexibility in sector size is also required for 525/60 operation where the number of samples in a segment is non-integer over a segment period even for synchronously sampled inputs. This word also contains a flag to indicate the recorded line standard; 525/60 or 625/50.

Certain of the control words mentioned above are critical to the correct operation of the processor and hence they are recorded more than once in each sector to increase the probability of the replay processor decoding the correct value. The words affected are LNGH, BCNT and SEQN.

Finally there is provision to record eight USER CONTROL words within each sector; the use of these is not defined by the standard.

3. Error Control

Error control is one of the key aspects of any high density magnetic recording format. Even in a simple record-play machine, the errors created by the tape medium require a comprehensive error correction strategy. The already identified user requirements add further levels of complexity mainly as a result of the editing needs but also the desire for audio cueing in high speed shuttle. For the latter point, the digital VTR format provides a conventional analogue audio cue track but, as mentioned, there is the possibility, at this point unproven, of recovering and regenerating some semblance of recognisable audio in shuttle.

The prime objectives of the error control scheme can be itemised as follows;

(i) correction of all normal tape errors in play,
(ii) error correction performance still available during the period of an insert edit,
(iii) resistance to likely fault conditions such as:
loss of data from any one of the four tape heads,
a long dropout, up to 5mm longitudinal length,
a wide physical or magnetic tape scratch up to 0.5mm width and,
tracking errors due to tape stretch.

The selected scheme allows for all these requirements without resorting to concealment; and extrapolation of our measurements, on equipment demonstrated at last years' International Television Symposium in Montreux, suggests that the agreed format will result in concealment free audio regeneration. Computations in a later section of this paper will confirm this view.

The objectives of the error control scheme are met by a combination of coding processes, which at the recording side, are implemented in the following sequence:

(i) Windowing the audio data contained within a 6.67msec period (normally 320 samples if the audio inputs are synchronous to the video),
(ii) dividing the data in to two channels containing ODD and EVEN samples, each forming the basis of an audio sector,
(iii) applying a Reed-Solomon (R-S) outer code,
(iv) redistributing the audio samples in a shuffle store,
(v) applying a R-S inner code,
(vi) assembling inner codes into sync blocks with a sync code header and block identification,
(vii) applying a channel code and,
(viii) recording each audio sector twice.

Steps (i) and (ii) are relatively straightforward.

The 161 maximum allowable samples in each sector are renumbered 0 to 320 in steps of 2 for even sectors and 1 to 321 in steps of 2 for odd sectors. Normally, where the audio and video sampling frequencies are synchronous, and in 625/50, exactly 160 sample sites are occupied in each sector. However, in the special non-synchronous mode of operation the number of sample sites occupied can vary from 159 to 161.

The next three steps, (iii) to (v), are best described in combination by the product block details of figure 3. The diagram details not just the location of the audio samples but also the control words and user words. The core of the error correction process is the creation of an array with error correcting codes added to rows and columns of the array. The principle behind array codes is further described in the reference (3), but in essence the columns of the array form the outer error correcting code and the rows form the inner correcting code. Figures 5 and 6 highlight respectively the concept of the error correcting array and the block schematic which indicates the manner of writing to and reading from the array stores.

Now errors generated by the tape medium consist essentially of random bit errors caused by head noise and long burst errors created by physical tape defects such as dropouts and scratches. The decoder sequentially decodes the inner code then, after de-shuffling, the outer code. The inner code is only able to correct random bit-errors and very short burst errors. However if inner correction fails due to a dropout, (say), then that associated row is indicated as having an error in every location. After de-shuffling, the outer correction operates on the columns of the array and uses a mechanism called 'erasure correction'. R-S codes normally need to decode both error locations and magnitudes, i.e. for each error, two variables must be computed. However, erasure decoding is possible if the location variable is known. Then only the error magnitude needs to be computed. This greatly increases the outer code power and is achieved through the provision of error location indicators from the inner decoder.

The details of the inner and outer codes are as follows;

Inner Code, 60 + 4 bytes formed from each row of the array. 60 bytes of audio data to which are added 4 bytes of error check data. This total code size of 64 bytes is capable of correcting one byte in any one of the 64 locations and providing reliable indication of two or more errors. If two or more bytes are in error, then the entire row is indicated in error.

Outer Code, 7 + 3 nibbles (the term nibble represents a 4 bit word) formed from each column of the array. 7 nibbles of audio data to which are added 3 nibbles of error check data. The total code size of 10 nibbles is capable of correcting any 3 error nibbles by erasure correction. Since entire rows of the array will be either indicated in error or not, the overall effect of the outer code is to be able to correct any three rows of the array.

It is this last point which defines the power to correct dropouts and this will be used in later calculations.

The shuffling structure is defined in a particular manner to achieve not just optimum distribution for array correction but also to allow both better concealment when error correction does fail and optimum sample distribution for the potential recovery of audio in shuttle.

Each row of the array contains 23 samples of audio (each sample being 20 bits), and equals one seventh of the audio sector recorded. The 23 samples in each row are taken from every even or odd sample for optimum distribution. For example, row 0 contains audio samples 0, 14, 28, 42 ..... 308 for the even sector. Columns are not so simply distributed. Outer correction is added to the array when the sample sequence falls in the natural ascending order. Then the rows are shuffled, as whole rows, to the sequence of figure 3. Figure 7 illustrates this point. The incoming sequence on the left hand side is in natural order with the 3 error correction words located at the end of each (vertical) block, and the sequence on the right hand side is the distributed row sequence.

This now leads naturally to the sixth coding process, that of assembling inner codes in to sync blocks. The concept of a sync block is that is the minimum self sufficient unit. Inner blocks by themselves contain no indication of the block start position or block identification. The sync block performs both these functions and details of its specification are given in references 1 and 4. In essence, each sync block encompasses two inner blocks as indicated in figure 7 . Therefore, in severe dropout conditions or during fast tape shuttle, only one or two sync blocks may be recovered from any sector. The row distribution has been chosen to complement this factor by ensuring that sequentially rows of each array contain samples optimally spaced for concealment operation.

The seventh coding process, channel coding, is used to optimise the characteristics of the signal to the tape channel. Many different codes have been considered, and the one finally selected is called 'synchronised scrambled NRZ' or SSNRZ. This code randomises the data pattern and breaks up unfavourable bit patterns (refs 1 and 5). It is applied equally to both video and audio data.

The eighth and last coding process is data duplication as outlined in the introduction. This is used to improve the final concealment rate since although the protection assigned to each sector array is sufficient for the transient insert edit requirements, it is not the level of performance required for continuous operation.

Each factor in the overall error protection strategy has now been described albeit in some cases very briefly. How do these factors integrate to produce a reliable end result? The next section will identify key measures of performance.

4. Replay Performance

The error correction power is such as to be able to provide error free outputs under the following conditions;

(i) head loss (this can be created by an oxide short or by circuit failure)

or

(ii) dropouts and scratches up to 6.8mm of the helical track (≒ 0.7mm tape width).

Both conditions can be verified by inspection of figures 1 and 2. A head loss removes four sectors along one track, however all the data can be successfully recovered from the other three tracks. A scratch of 0.7mm tape width will delete eight sectors in two sequential rows (and the associated edit gaps). The remaining two rows are sufficient to recover all the data. Note, however, that data within an insert edit is not duplicated and as a result this level of robust performance is not applicable during this transient operation.

Under the condition of normal play, the error control scheme will be presented with normal tape error conditions within the limits of a tape error model agreed between manufacturers as representing the worst likely field conditions. This model is given in figure 8 and shows the error length distribution against probability of occurance. From this, the concealment rate can be readily computed (ref 3);

(i) Calculate the probability of inner block failure;

Failure rate due to random noise, P_{IR}, with a bit error rate of 10^{-4}; the inner code will fail where two or more errors occur;

$$P_{IR} \doteq 64 * 63/2 * (8 * 10^{-4}) = 1.29 * 10^{-3} \quad (1)$$

Failure rate due to dropouts > 1 byte; where the probability of a dropout > 1 byte is;

$2 * 10^{-6}$ per byte, $P_{ID} \doteq 64 * 2 * 10^{-6} = 1.28 * 10^{-3}$ (2)

Then, the inner block failure rate, P_I is given by;

$$P_I \doteq P_{IR} + P_{ID} = 2.57 * 10^{-3} \quad (3)$$

Now sector correction will fail when four or more inner blocks fail. The probability of sector failure due to multiple short dropouts, P_{SR}, is;

$$P_{SR} \doteq \frac{10 * 9 * 8 * 7}{2 * 3 * 4} * (P_I)^4 = 9 * 10^{-9} \quad (4)$$

However, a single long dropout of >192 bytes can also cause failure and with the probability of such a dropout at 10^{-7} per byte, then the probability of sector failure, P_{SD}, is;

$$P_{SD} \doteq 192 * 10^{-7} = 1.92 * 10^{-5} \quad (5)$$

Now each block is duplicated. The probability of both blocks being corrupt, P_{S2R}, is;

$$P_{S2R} = 2 * (P_{SD})^2 = 7.4 * 10^{-10} \quad (6)$$

This figure is added to the probability of a single dropout affecting both sectors. Since the distance of separation is;

$$((138 * 7) * 2) + (64 * 3) = 2124 \text{ bytes}, \quad (7)$$

the probability of such a dropout is approximately $1 * 10^{-10}$ per byte, then the probability of dual block failure, P_{S2D}, is;

$$P_{S2D} \doteq 2124 * 10^{-10} = 2 * 10^{-7} \quad (8)$$

The total probability is the sum of P_{S2R} and P_{S2D}, and equals $2 * 10^{-7}$ (9)

Finally, the probability of either odd or even sectors failing is then the probability of segment failure and is given by;

$$P_{S2D} * 2 = 4 * 10^{-7} \quad (10)$$

When this occurs, the entire contents of the sector fail since the dominant mode of failure is the long dropout.

The segment rate is 150/second, hence the segment failure rate, P_S, is;

$$P_S = 4 * 10^{-7} * 150 = 6 * 10^{-5}/s.$$

When duplicate data fails, then concealment will occur since only alternative samples will be available. However, this will occur at a rate of only once in 4.6 hours. Concealment caused by shorter dropout errors will be at a far lower rate and can be considered as negligable (from equation 6) since the dominant modes of failure are tape dropout and scratches.

The concealment rate must be further qualified. Firstly, the SMPTE tape model represents results which are ten times worse than those measured in the labs, and, secondly, audio in the centre portion of the tape is subjected to less head to tape contact (and consequently dropout) problems.

The conclusion is, therefore, that this worst case result represents an acceptable end result and that for even the most critical applications, the performance will be virtually perfect. This view is confirmed by the results obtained from the experimental equipment demonstrated in Montreux, May 1985.

5. Insert Editing

The principles behind sector duplication and insert editing have already been described. Figure 9 highlights that although the audio is grouped in to segment bursts, the crossfade start and complete points do not need to coincide with segment boundaries. Thus edit accuracy down to lmsec (and theoretically to one sample) can be achieved. Note that the BCNT control signal must be obeyed for the downstream edit data during the crossfade window period in order to avoid misalignment of samples on replay.

During the crossfade window, the duplication process is removed and each sector becomes more prone to errors. The equation for sector failure (5) gives a failure probability of $1.9 * 10^{-5}$. The probability for either sector failing is $3.8 * 10^{-5}$. At a segment rate of 150 per second, the rate of segment failure is then;

$$P_S = 150 * 3.8 * 10^{-5}/s, \qquad (11)$$

or once in 3 minutes (for each channel), when the contents of the affected segment will contain only alternate valid samples and heavy concealment will occur.

Editing and crossfading operations have a further effect on audio data channels where they are not AES block aligned. This problem is recognised and presents a choice of either pre-aligning AES blocks and accepting the resultant time shift (up to +2msec), or accepting misalignment and the consequent effects of break up in the continuity of the AES block format. This problem is relevant to other areas and is currently under study pending recommendations.

6. Conclusions

This paper has explained the basic outline of the agreed format encompassed in the 4:2:2 component digital VTR standard. Necessarily the descriptions have been brief, and in a number of areas the reasons for the choice of particular parameters has been defined by aspects of both the video and audio requirements. The format certainly provides a number of features in excess of those required for a basic machine. An instance of this is the area of timecode. Each audio channel has contained within the user bits, the option for two timecodes, one for origination and one for regular use. A further pair of timecodes

are available in the longitudinal track (fig 1). This totals ten timecodes in all. The operational and organisational aspects of this must be approached with care in order to create usable and friendly timecode environment if the system is to be successful. A further area of concern is where the recording format is in 16 + 4 or 18 + 2 mode, where channel status data may be recorded twice, once in the 'V' bit and duplicated in the 'CHAN' and 'PREF' control words.

These points are, however, operational problems and do not detract from the fundamental point which is that agreement on a worldwide digital component VTR standard has been achieved. This format is the result of many years of extensive research in to the many engineering challenges allied to the requirements of equipment users and should allow the dream of multigeneration dubbing without perceptible loss.

7. Acknowledgements

The authors would like to take this opportunity to thank colleagues at Sony Corporation for their considerable part in the digital VTR development program and to their contributions to ideas presented in this paper.

8. References

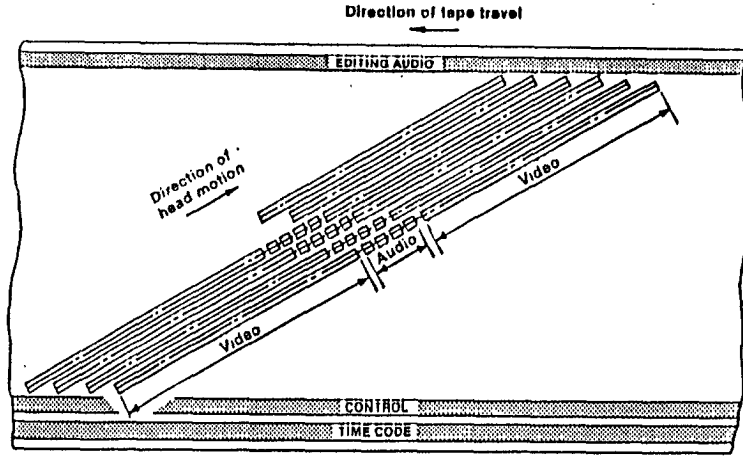
1. "A review of the signal format specification for the 4:2:2 component digital VTR", J.H. Wilkinson, paper to be presented at the 'Video and Data Recording Conference', Brighton, UK, 17 - 21 March 1986.

2. 'AES recommended practice for digital audio engineering - serial transmission format for linearly represented digital audio data', ANSI S4.40-1985.

3. "The D-1 digital television recorder -error control", J.H. Wilkinson, 20th Annual SMPTE Television Conference, Chicago,USA, 8th February 1986.

4. "Standard for recording digital television signals on magnetic tape in cassettes", EBU Tech 3252.

5. "An experimental channel coding for the digital video tape recorder", K. Yokoyama, S. Nakagawa, 12th International Television Symposium, 1981, pp 251-260.



PRINCIPLE TRACK DIMENSIONS

SECTOR LENGTH	VIDEO (2 OFF)	:	77.78mm	INCLUDING
	AUDIO (4 OFF)	:	2.56mm	PREAMBLE & POSTAMBLE
EDIT GAP LENGTH	(5 OFF)	:	0.84mm	
TOTAL TRACK LENGTH		:	170mm	
TRACK WIDTH	HELICAL	:	16 mm	
	CUE AUDIO	:	0.7mm	
	CONTROL	:	0.5mm	
	TIME CODE	:	0.5mm	
TOTAL TRACK WIDTH		:	19.01mm	

FIGURE 1 - Track Layout

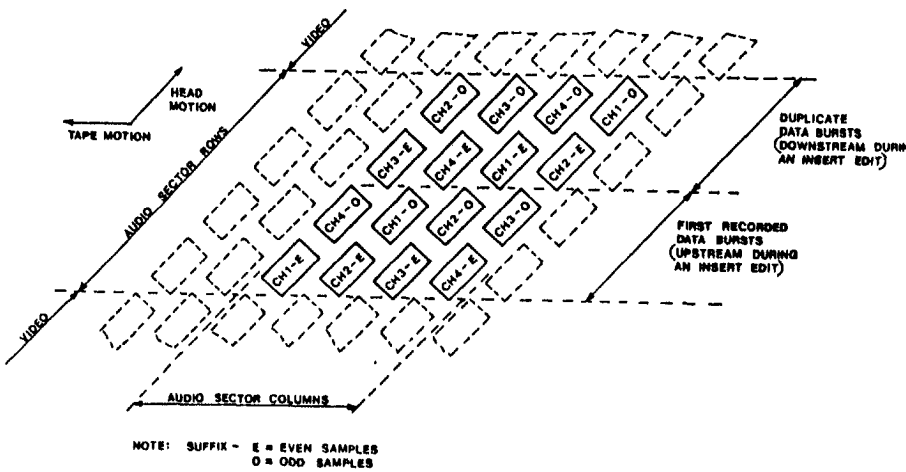
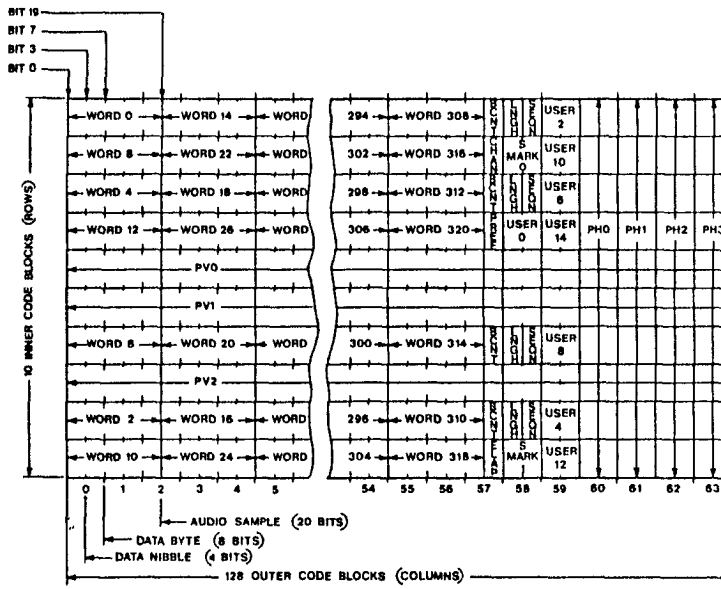


FIGURE 2 - Audio Sector Arrangement



NOTES: 1. EVEN SECTOR NUMBERS SHOWN. FOR ODD SECTOR NUMBERS, ADD 1 TO EACH NUMBER ABOVE (EXCEPT S MARK 0 AND S MARK 1)
2. WORD 320 ABOVE, AND WORDS 319 AND 321 IN ODD SECTORS CONTAIN DATA AS DEFINED BY CONTROL WORD 'BCNT.'

FIGURE 3 - Layout for an Audio Product Block

WORD MODE LNGH		BIT ASSIGNMENTS 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
* 0	(000)	C U V R 0 A 15
1	(001)	C U V 0 A 16
2	(010)	C V 0 A 17
* 3	(011)	C U 0 A 17
4	(100)	C 0 A 18
5	(101)	V 0 A 18
6	(110)	U 0 A 18
* 7	(111)	0 A 19

NOMENCLATURE: A = AUDIO DATA
C = CHANNEL STATUS BIT
U = USER BIT
V = VALIDITY BIT
R = UNDEFINED BIT RESERVED FOR INTERNAL USE
* = RECOMMENDED MODE FOR GENERAL USE

FIGURE 4 - Assignment of Word Options

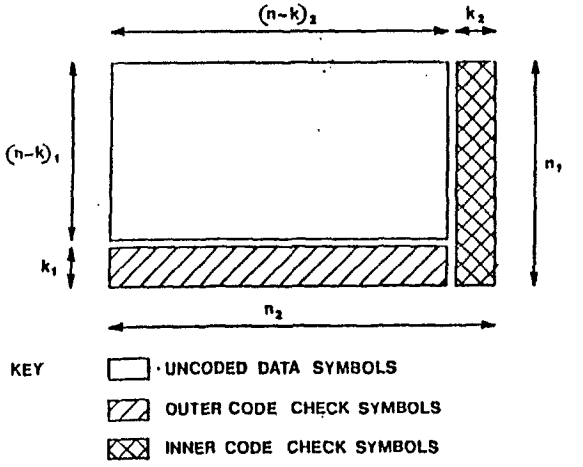


FIGURE 5

Formatting the error control code in an array

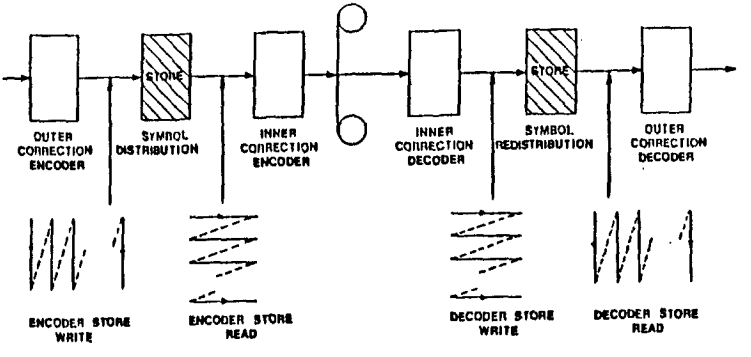


FIGURE 6

Inner and outer correction block schematic with associated array scanning sequences

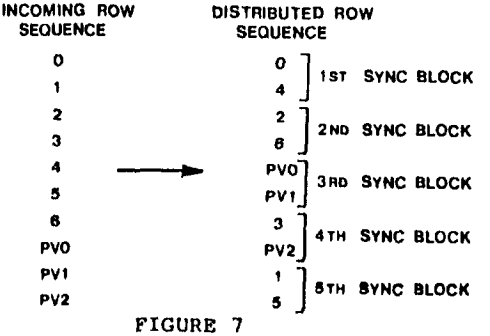


FIGURE 7

Distribution of row sequence

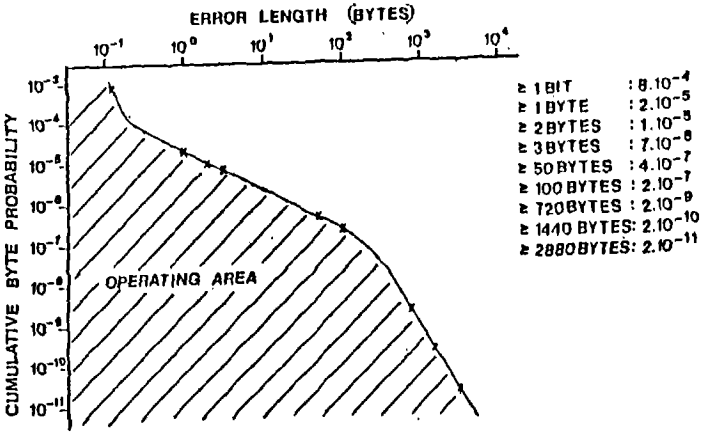


FIGURE 8

Cummulative error probability as a function of error length

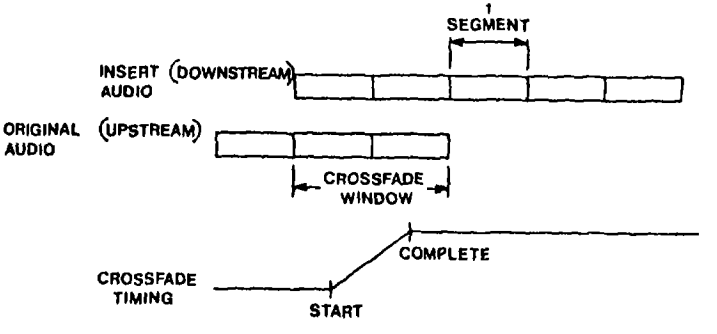


FIGURE 9

Cross Fading at an Insert Edit