On Several Standards for Converting PCM Signals into Video Signals*

TOSHITADA DOI, YOSHIKAZU TSUCHIYA, AND AKIRA IGA

Sony Corporation, Audio Development Laboratory, Tokyo, Japan

This paper discusses the engineering decisions in the design of a pulse-code-modulation audio recording system as an adaptor for video tape recorders (VTR). Standards are proposed for home use and professional VTRs which share sampling frequency and digital coding but differ in error detection and correction techniques.

0. INTRODUCTION: Many investigations concerning pulse-code-modulation (PCM) systems have been made in an attempt to break through the technical limitations of analog recording–playback systems of audio signals. One of the features of PCM is the necessity for wide-bandwidth recording media, and thus such video systems as video tape recorders or video disks are considered for use in audio PCM.

Among them, a PCM system as an adapter for video tape recorders, which converts PCM signals into video signals, has become of practical use [1], [2]. At this moment the problem of compatibility becomes important, and we must prevent confusion in the market caused by the emergence of many noncompatible systems.

This paper treats the proposal for the standards and their background for producing PCM recording and reproducing machines by utilizing home and professional video tape recorders (VTR) with rotary head systems.

The purpose of this standard is to secure complete compatibility where existing cassette tapes are used in the cassette VTR system, and also with the general use of VTRs.

Further, the proposal contained in this paper applies to VTRs for NTSC television standards (used in the United States, Japan, and some other countries), and the applicable VTR models can be roughly broken down into two types: VTRs for professional use with performance equal to or better than the U-matic® system, and VTRs for home use, such as Betamax® [3].

The reason for distinguishing between home-use VTRs and professional-use VTRs is the considerable difference in characteristics between these types. For convenience sake, we name the system for home-use VTR "standard B", and the system for professional use, such as U-matic, "standard A."

1. STANDARD B

In the PCM recording and reproducing unit for home-use VTRs we need to take economy into full consideration. At present the method of constructing PCM recording and reproducing units at the cheapest cost is by using PCM adapters without any modifications to the VTR. Even if the PCM recording and reproducing unit will be integrated with the VTR in the future, compatibility will be easily available.

The proposed standard and its background are described in the following.

1.1 Number of channels: 2

Considering such characteristics as frequency response, signal-to-noise ratio, and jitter of home-use VTRs, it is difficult to realize multichannel systems. Therefore, we propose limiting the system to two channels but, needless to say, the voice track of the VTR is open and is usable.

1.2 Form of Recorded Signal: Nonreturn to zero (NRZ).

If we make a PCM recording and reproducing machine by connecting a PCM unit to a VTR through the adapter method, it is necessary to convert an incoming signal so that it is of the same form as a standard video signal. Therefore it will convert PCM signals to video signals by

* Part of this paper was submitted to the Technical Group of Magnetic Recording, TG MR 77-24 (1977-11) IECE Japan, presented on November 28th, 1977, in Sendai City, Japan.

including vertical synchronizing and horizontal synchronizing pulses.

As the form of the PCM signal used here, we can consider NRZ, BIPHASE,[1] DM,[2] etc., and Fig. 1 shows the spectrum density in each of these modulations.

A characteristic of BIPHASE and DM is that the direct current portion is excluded, or is quite small. However, this has reduced importance because the video signal is not recorded directly on a tape, but is converted to FM recording at the final stage. Furthermore, 3 PM[3] encoding has recently been announced, but the circuits required are very complicated [4].

Modulation without self-clock (that is, NRZ) can easily be adopted by virtue of the horizontal synchronizing pulses between which the jitter is small enough. Additionally the modulators and demodulators for NRZ are the most simple and economic.

Therefore, we propose using NRZ.

1.3 Sampling Frequency: 44.05594 kHz.

The sampling frequency was determined by considering the following conditions.

1) The frequency band of the domestic signal capable of recording and reproducing is dc to 20 kHz. Therefore considering the problem in making filters to prevent aliasing noise, the minimum sampling frequency f_s is

$$f_s \geqq 42 \text{ kHz} \tag{1}$$

Note: Figs. 2 and 3 show typical characteristics of elliptic filters of 9 poles and 13 poles, respectively. Fig. 2**a** shows the circuit and the element values of a 9-pole filter designed for consumer-use PCM machines, where the equivalent resistances of coils measured at 20 kHz are added. Fig. 2**b** and **c** shows frequency-gain characteristics, and Fig. 2**d** and **e** shows phase and group-delay characteristics. Fig. 3**a**, **b**, and **c** shows the gain characteristics of a 13-pole filter, and Fig. 3**d** shows its phase characteristics.

2) If we record and reproduce PCM signals without redesigning the VTR, we need vertical synchronizing and horizontal synchronizing pulses, so we cannot input PCM data during these periods. Especially, the period of the vertical synchronizing pulse is equivalent to nine horizontal scan lines, and considering that the head switching point of helical scanning VTRs is about six horizontal lines maximum in advance of the front edge of the vertical synchronizing signals, there are 250 horizontal lines or less that we can put PCM data in one vertical (1V) field. Additionally, in view of the fact that we leave about two horizontal lines for system control, the number of horizontal lines, namely, H_{PCM}, when we can put PCM data in each field, is

$$H_{PCM} < 248. \tag{2}$$

3) The number of horizontal lines where PCM data can be put is subject to the condition shown in Eq. (2). Here we will determine the number of words to be put in each H line. We define one word as one sample including the data of two channels.

When considering the characteristics (mainly dynamic range) of the PCM recording and reproducing machine, the number of bits necessary for one word is 26 to 32, and further, additional bits are necessary to compensate for the dropout errors.

Therefore, considering the frequency response and jitter characteristics of VTRs, it is difficult to put four words in one horizontal (1H) line and three words are more appropriate. It is possible to put a nonintegral number of words in 1H (for instance, 3.5 words), but it is not advisable because the synchronizing and data-processing circuits become complicated.

In view of the above the number of words W_{PCM} of PCM data that can be put in 1V is

$$W_{PCM} < 3 \times H_{PCM} = 744. \tag{3}$$

4) The maximum number of words that can be put in 1V is given by Eq. (3), as a result of which the sampling frequency f_s is

$$f_s = W_{PCM} \cdot f_v < 44.5954 \text{ kHz} \tag{4}$$

where f_v is the frequency of the vertical synchronizing signal (NTSC)

$$(f_v = 59.94006 \text{ Hz}).$$

The master clock which generates horizontal synchronizing and vertical pulses, etc., must coincide with the master clock that makes the sampling frequency.

Therefore, the master clock must, at least, be an integer multiple of the sampling frequency, and $f_{eq} = 31.468528$ kHz, the frequency of the equalizing pulses, which is the highest frequency in relation to the synchronizing signal.

Now suitable sampling frequencies meeting the conditions shown in the Eqs. (1), (3), and (4) are indicated in Table I, which also shows the ratio of f_s to f_{eq}, namely, $f_s{:}f_{eq} = m{:}n$, where m and n are integers.

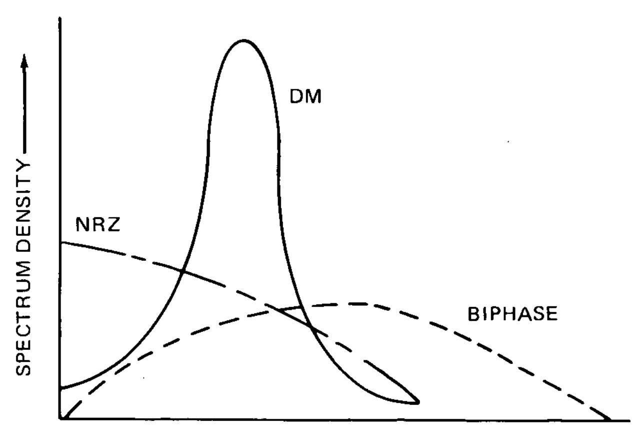

Fig. 1. Spectrum density for several modulations.

[1] A self-clocking code in which the logic level is indicated by transitions or lack of transitions instead of by absolute voltage.

[2] A delay modulation which is also called modified FM or Miller code.

[3] A three-position modulation which is a new kind of run length limited code.

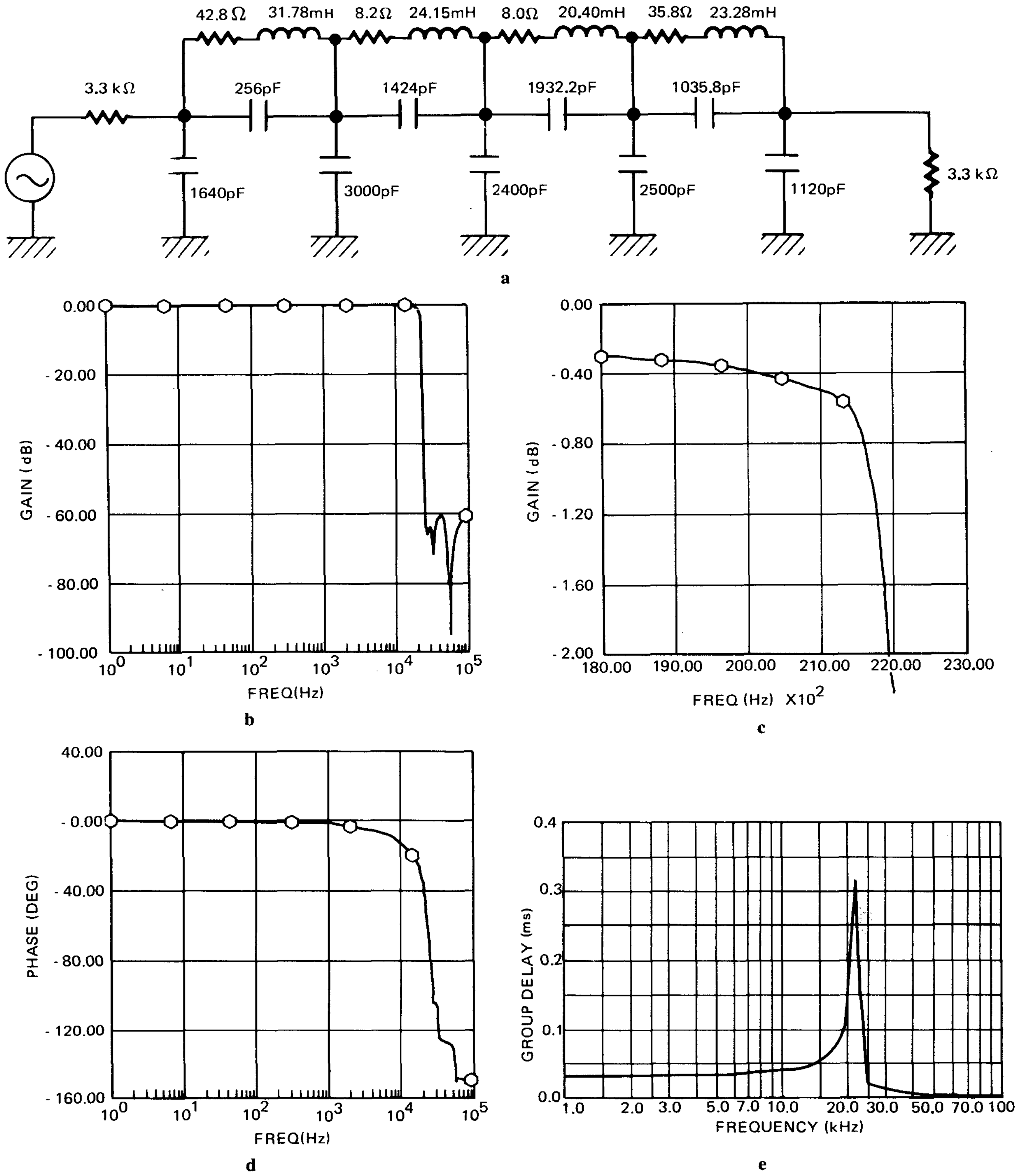

Fig. 2. Circuit and typical characteristics of 9-pole elliptic filter.

The master clock frequency f_M, which also provides the serial bit clock frequency, must be N times the least common multiple of f_s and f_{eq}, where N is the number of bits for one word (in Table I, N is assumed to be 32). Therefore,

$$f_M = N \cdot n \cdot f_s = N \cdot m \cdot f_{eq}. \tag{5}$$

H_{PCM} is the number of horizontal lines in one field, where PCM data are put in [see Eq. (2)].

Judging from the design of the antialiasing low-pass filter, the higher the sampling frequency, the more desirable. However, even if the sampling frequency is acceptable, there are some cases where the master clock frequency is too high as in no. 1, no. 2, and no. 3 shown in Table I.

In view of the above, the sampling frequency shown in no. 4 is very useful with 7.04895 MHz for the master clock. Fig. 4 shows the block diagram of frequency dividing from the master clock frequency (7.04895 MHz

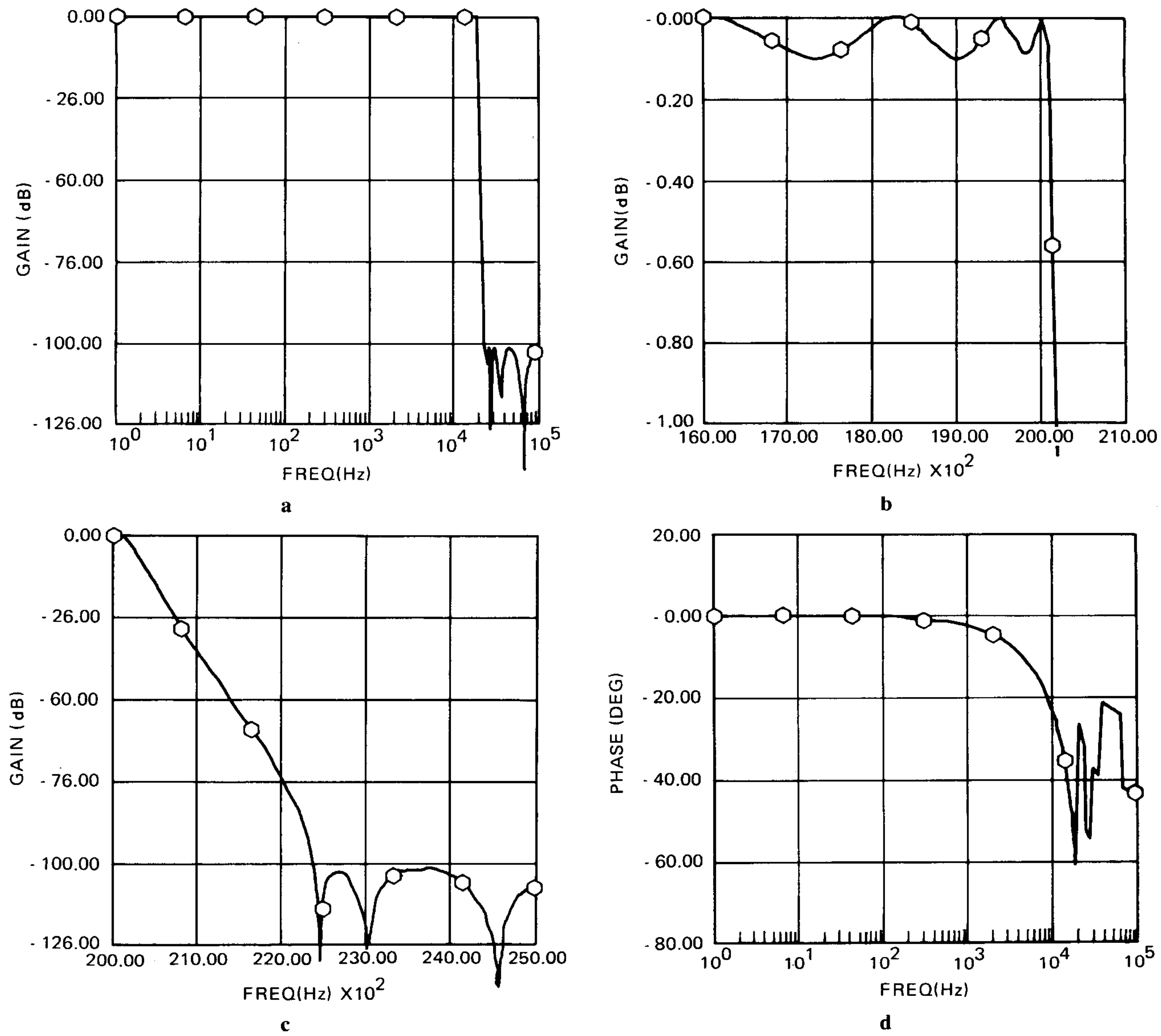

Fig. 3. Typical characteristics of 13-pole elliptic filter.

× 2).

In the standard video signal, a front porch, horizontal synchronizing signal, and back porch are necessary, and herein we will use an equivalent of 2 bits (about 1.13 μs) for the front porch and an equivalent of 8 bits (about 4.5 μs) each for the horizontal synchronizing signal and the back porch.

Table 1. Candidates for sampling frequency.

Number	f_s (kHz)	m	n	f_M (MHz)	H_{PCM}
1	44.59540	248	175	249.7342	248
2	44.41558	247	175	248.7273	247
3	44.23576	246	175	247.7203	246
4	44.05594	7	5	7.048950	245
5	43.87612	244	175	245.7063	244
6	43.69630	243	175	244.6993	243
7	43.51648	242	175	243.6923	242
8	43.33666	241	175	242.6853	241
9	43.15684	48	35	48.33561	240
10	42.97702	239	175	240.6713	239
11	42.79720	34	25	34.23776	238
12	42.61738	237	175	238.6573	237
13	42.43756	236	175	237.6503	236
14	42.25774	47	35	47.32867	235
15	42.07792	234	175	235.6364	234

As a result, we can put data of 94 bits in 1H, and the bit transmission rate of the pseudo-video signal will be 1.762238 megabits per second.

In all these calculations we employed

$$f_H = 15.734264 \text{ kHz}$$

where f_H is the frequency of horizontal synchronization of NTSC color.

The above discussions are based on the NTSC color frequency, and if we use the black and white frequency that is 0.1% higher, the sampling frequency becomes 44.1 kHz. Then 1764 words are allocated in one frame of PAL or SECAM video and 3,675 words for two frames of movie film.

1.4 Interleaving: 1 block = 92 Words

The problem of dropouts is important in the magnetic recording and reproduction of PCM. If a dropout occurs, the reproducing codes will become errors, and the most complete method to compensate for them is the use of error-correcting codes. However, the errors in codes due to dropouts will be error bursts. About 30% redundancy is considered necessary for correcting the errors completely, although we carry out a proper interleave. Here, if one word consists of 26 bits (we consider this value the minimum necessary), the possible redundancy is 5 to 6 bits per word, or about 18%.

Therefore, even if we adopt correction codes, we cannot expect effective results, and so we consider it proper to detect errors by error-detecting codes, and to interpolate with mean values.

Using interpolation with mean values, error words are dispersed at random, and errors of two words or more must not occur in succession. For this purpose we use word interleaving.

The length and method of interleaving blocks should be determined by surveying the actual condition of dropouts.

a) *Measurement of dropout* [5]. We measured the dropouts of home-use VTRs with a dropout-measuring machine. We observed many differences among tapes with regard to the number of dropouts, but little variation was found in the length of dropouts.

A portion of the measured data is shown in the following fig' res. Fig. 5 shows the distribution of the length of dropou when six rolls of tape sampled from six different production lots were measured for 1 hour. Fig. 6 shows the variation in dropouts among 1-hour tapes.

As is apparent in the above measurements, as the dropout length decreases, the occurrence rate increases exponentially, and a long dropout equivalent to 7H–9H will occur on rare occasions. However, it can be said that no dropout occurs exceeding 10H. So we can say that there will be no problem if we design interleaving to compensate for dropouts equivalent to 15H.

We determine the block length of interleaving based on the above. We have also carried out computer simulation of dropouts by a Gilbert model, and the result matched the measured values well [6], [7].

b) *Method of interleaving*. Fig. 7 shows three interleaving methods, and it also indicates error distribution where a burst error occurs due to dropouts. The figure shows in each system the initial order of words in the upper tier, the order of words modified by interleaving in the middle tier, and the reorder of words by deinterleaving in the lower tier.

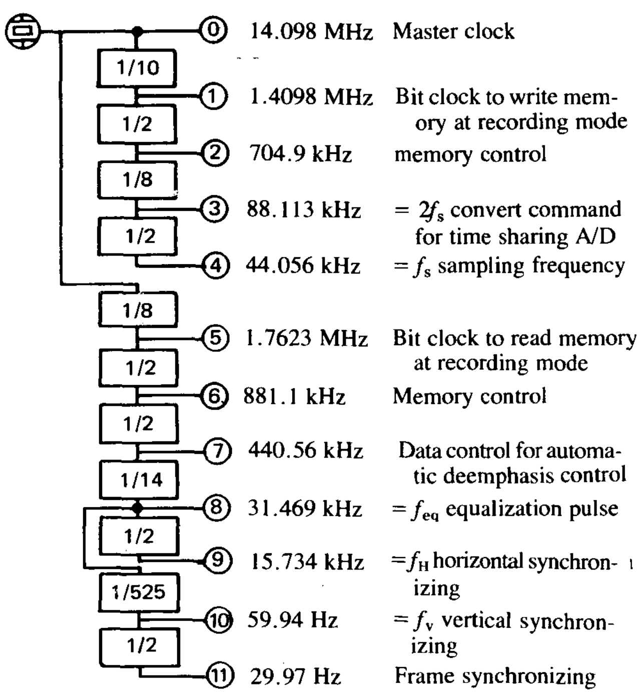

Fig. 4. Block diagram of frequency division.

Dropout Length	Total Number of Dropouts (1 h × 6 rolls)	Relative Frequency
~ ⅓ H	29750	0.64295
⅓ H ~ ⅔ H	13749	0.29714
⅔ H ~ 1 H	1913	0.04134
1 H ~ 1⅓ H	532	0.01150
1⅓ H ~ 1⅔ H	147	0.00318
1⅔ H ~ 2 H	59	0.00128
2 H ~ 2⅓ H	31	0.00067
2⅓ H ~ 2⅔ H	22	0.00048
2⅔ H ~ 3 H	15	0.00032
3 H ~ 3⅓ H	11	0.00024
3⅓ H ~ 3⅔ H	9	0.00019
3⅔ H ~ 4 H	3	0.00006
4 H ~ 4⅓ H	7	0.00015
4⅓ H ~ 4⅔ H	7	0.00015
4⅔ H ~ 5 H	3	0.00006
5 H ~ 5⅓ H	1	0.00002
5⅓ H ~ 5⅔ H	2	0.00004
5⅔ H ~ 6 H	4	0.00009
6 H ~ 6⅓ H	3	0.00006
6⅓ H ~ 6⅔ H		
6⅔ H ~ 7 H		
7 H ~ 7⅓ H		
7⅓ H ~ 7⅔ H	1	0.00002
7⅔ H ~ 8 H	1	0.00002
8 H ~ 8⅓ H		
8⅓ H ~ 8⅔ H		
8⅔ H ~ 9 H		
9 H ~ 9⅓ H	1	0.00002
9⅓ H ~ 9⅔ H		
9⅔ H ~ 10 H		

Fig. 5. Distribution of dropout.

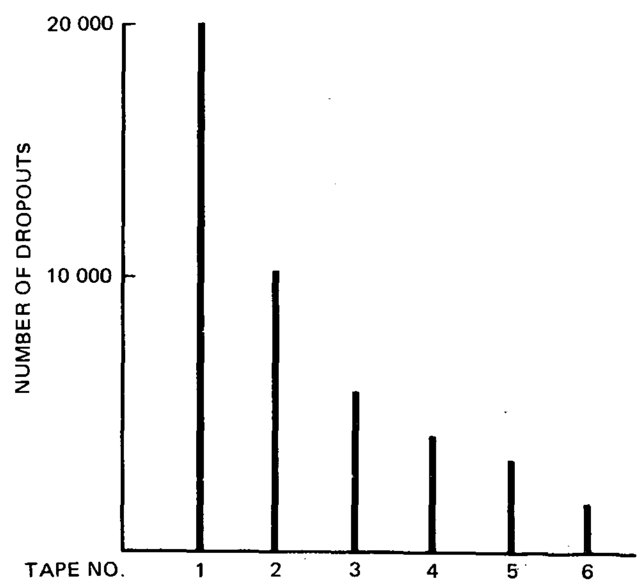

Fig. 6. Difference in dropouts due to different tapes.

According to Fig. 7, an error burst of n words or less can be replaced by random errors which do not continue for two words or more.

Errors occur once in two words in system 1, once in three words in system 2, and once in four words in system 3.

c) *Limit of detection of mean-value interpolation* [8]. A hearing test was carried out in order to find out the difference of the minimum detectable signal frequency (pure tone) for three types of interleaving methods, where dropouts are compensated by mean-value interpolation.

As shown in Fig. 8, we found that we could detect interpolation for pure tones of 1 kHz or over, but we could hardly observe any difference due to the three kinds of interleaving, so we regard it as proper to use system 1 which simplifies the interleaving circuit and memory.

This is because when compensating for error bursts of the same length, we can minimize the block of interleaving.

By the preceding we will divide the 735 words that come during one vertical field into eight blocks, and will carry out the interleaving with 92 words in each block.

At this time the length of dropout which we can compensate is 15H or 1410 bits.

Fig. 9 shows the schematic diagram of the interleaving method, where the numbers express the sequence of the data.

1.5 Redundancy Bit: 16 Bits (During 1H)

Having determined the sampling frequency, it is possible to put data of 94 bits into 1H.

In these data we will determine the number of bits that are to be allocated as redundancy bits for detecting errors. It is essential to put three words into 1H, and the redundancy bits that are allocated to one word are 5–6 bits.

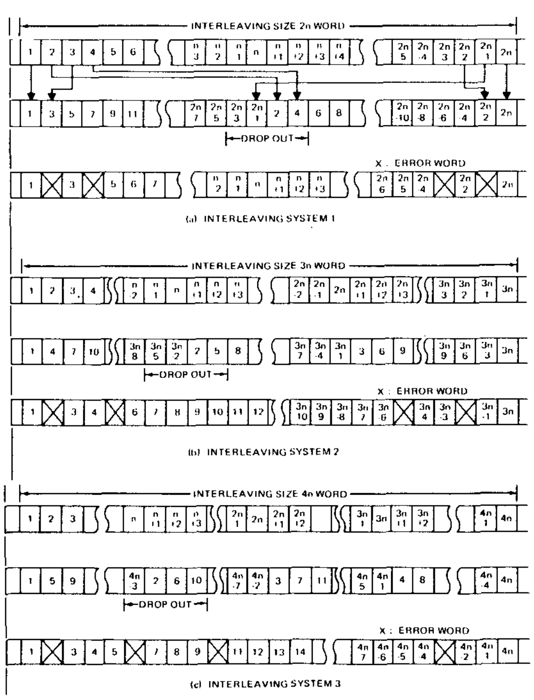

Fig. 7. Three types of interleaving.

The detection probability for a cyclic redundancy check (CRC) as the error-detecting code is about 98%, and is too low, judging from the number of dropouts which are actually occurring.

However, if we add redundancy bits by consolidating three words, we can have 16 redundant bits for every three 26-bit words (detection probability 99.998%), and when we consider a 28-bit word, we have 10 redundant bits (detection probability 99.9%).

As is apparent in Fig. 6, although the number of dropouts differs greatly among tapes, it would be sufficient to expect 20 000 dropouts per hour, and accordingly the detection probability needs to be 99.995% or more.

Therefore, we consider one word to consist of 26 bits, with 16 redundant bits in 1H.

Also, we use CRC as the error-detecting code, and in line with the recommendation by CCITT, the generator

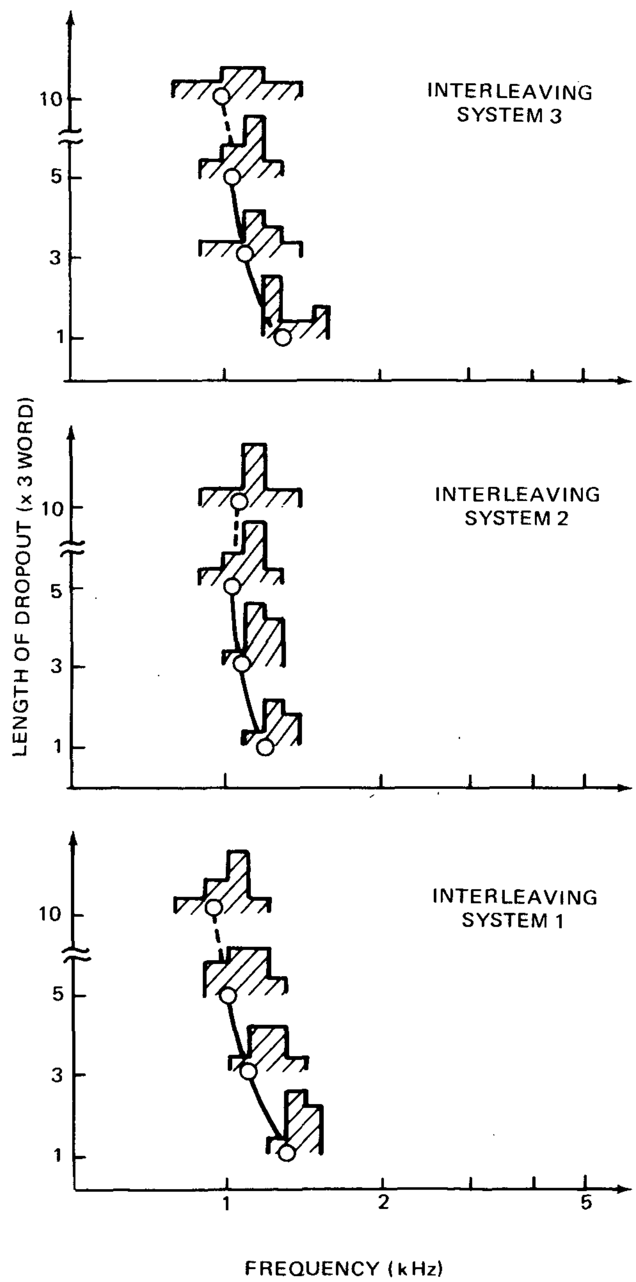

Fig. 8. Limit of detection between interpolation.

polynomial $G(x)$ is

$$G(x) = x^{16} + x^{12} + x^5 + 1. \tag{6}$$

1.6 Tape Format

Based on 44.05594 kHz for the sampling frequency, the number of horizontal lines containing PCM data in one field is 245H, leaving vacant the equivalent of about 5H.

However, at present, 1H for the system control comprises the signal which shows the start of the PCM data (in one field), and the other 1H is being used for the automatic detection of emphasis.

The remaining 3H can be used for automatic detection, etc., such as in the case where the polygonal line quantization system (companding system) is changed. Fig. 10 shows the signals to be recorded.

2. STANDARD A

Many methods are available for PCM recording and reproducing machines which are being researched and developed for professional use at present.

They are roughly broken down into those using rotary heads and those using fixed heads of multitracks. The ones using rotary heads mainly utilize existing VTRs.

The second standard being proposed in this article, as previously mentioned, employs rotary heads and is to be used with VTRs for professional use (mainly U-matic).

2.1 Number of Channels: 2

If we use a professional VTR, there is enough bandwidth for the video signal band to be recorded and reproduced, and four channels are possible. But since it is difficult to record and reproduce each channel independently, it is advisable to use surplus bandwidth for improvement of resolution and compensation for dropouts.

Therefore we will use two channels as in standard B.

2.2 Form of Recording Signal : NRZ

We will employ NRZ for the same reasons as standard B.

2.3 Sampling Frequency: 44.05594 kHz

Between the PCM recording and reproducing machines of standard A and standard B, sampling frequency must be common because data might be dubbed as digital signals. Therefore we will use 44.05594 kHz in accordance with standard B.

2.4 Compensation for Errors : Correction by Error-Correcting Codes

As a result of the determination of the sampling frequency, the number of words runs to three words times two channels[4] as in standard B.

However, when using professional VTRs, the number of bits that can be put in 1H will become twice as many as for consumer VTRs due to differences in frequency characteristics, jitter, signals-to-noise ratio, etc.

Therefore even though one word consists of 16 bits we can add 50% redundancy and employ an error-correcting code.

Because the system is for professional use, we must not lower the sound quality by dubbing or editing, and thus error correction is essential.

Many kinds of error-correction codes have been devised, but existing error-correction codes are not optimum for this application.

We must transmit high-quality information on a low-quality transmission line in the audio PCM recording and reproducing machine, and we have to use a combination of other error-detecting codes with interpolation circuits,

[4] Note that in standard A one word is considered to be for one channel for the convenience in forming a cross word code (see Fig. 11).

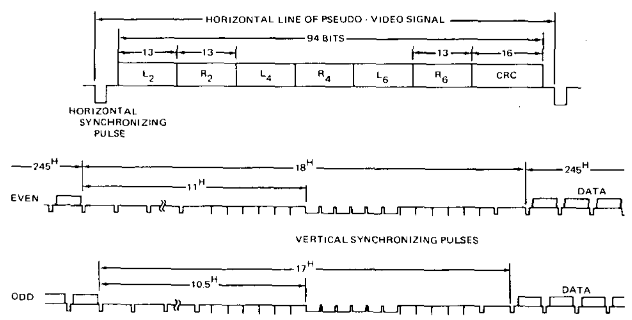

Fig. 10. Recording and reproducing signal.

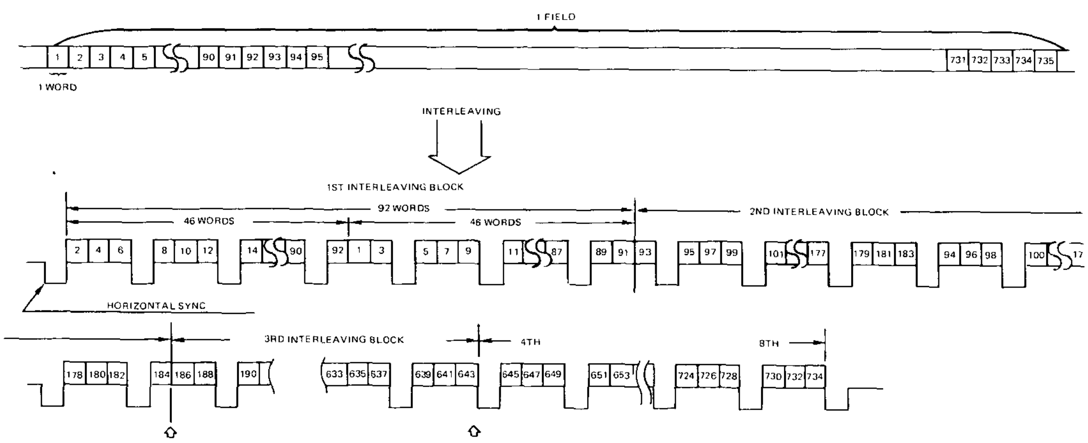

Fig. 9. Interleaving system for standard B. 1 field = 8 interleaving blocks = 7 × 92 words + 91 words = 735 words

etc., in case of errors beyond correction capacity.

To solve these problems, cross word codes were devised suitable for audio use PCM recording and reproducing machines [9]. These are described as follows.

(1) *Definition of cross word codes*. Representing a transmission vector as V, sindrome vector as S, check vector as H, error vector as E, and receiving vector as U, one example of this code is shown in the following formula:

$$HV = O \tag{7}$$
$$HU = HE = S \tag{8}$$
$$U = V + E \tag{9}$$

$$H = \begin{bmatrix} 110000100000 \\ 001100010000 \\ 000011001000 \\ 101010000100 \\ 010101000010 \\ 000000111001 \end{bmatrix} \tag{10}$$

2) *Error-correcting capacity*. If we use the check vector shown in Eq. (10), we can correct a burst error of four words, and simultaneously we can correct a random error of any two words out of a block of 12 words. In case of a three-word error, correction is possible up to 46.4%.

In view of the above, we use error correction by cross word codes, in which one word is regarded as 16 bits.

Fig. 11 shows the block configuration of the cross word code and its interleaving systems, where the letters L and R represent left and right channels, respectively, and the subscript shows the sequence of data. The letter C means check word for the cross word code. The block consists of three lines: the first and the third lines mainly consist of information words, while the second line is their parity. The interleaving is carried out by ordering every line of each interleaving block as shown in Fig. 11.

By virtue of interleaving, this code can completely correct burst errors less than 2240 bits (11.7 horizontal lines) in one interleaved block of 6720 bits (35 horizontal lines). In addition, burst errors less than 4480 bits (23.3 horizontal lines) can be compensated by means of one-word interpolation.

3. CONCLUSION

We have described the background for PCM recording and reproducing machines using VTRs, and proposed a system of converting PCM signals into video signals. Technical limitations are very tight, and flexibility is lacking. Nevertheless, this kind of system, utilizing unmodified VTRs, is very useful, and is strongly recommended for the following reasons.

1) The system is very economical. An adapter costs only $2000, and this will be greatly reduced within a few years. Any VTR, including old-fashioned black and white machines, can be used. One VTR is usable for both video and hi-fi audio. This is economically attractive for consumers and for broadcasting companies.

2) Various systems developed for VTR can be used for PCM systems. This includes editing machines, synchronized playback machines, remote-control systems, and SMPTE time-code devices. PCM recording or playback synchronized with video or another PCM system can be easily carried out.

3) Existing transmission systems and distribution networks for video signals can be used for PCM audio signals without any modification.

4) There is a possibility of PCM broadcasting over ordinary TV channels using existing facilities. No technical problems are expected.

Figs. 12 and 13 show examples of PCM systems described above. The PCM-1 (Fig. 12) uses standard B, while the PCM-1600 (Fig. 13) uses standard A. Both have already been marketed.

The authors will be gratified if this proposal is helpful in commercializing PCM recording and reproducing machines.

ACKNOWLEDGMENT

The authors would like to express sincere gratitude to Dr. Heitaro Nakajima, Director of Sony Audio Development Laboratory, and all those concerned with its Audio Division, including Chief Nagami, for their guidance and discussions. They also want to thank Mr. L. S. Silverman, Sony Corporation of America, for his help in improving the English.

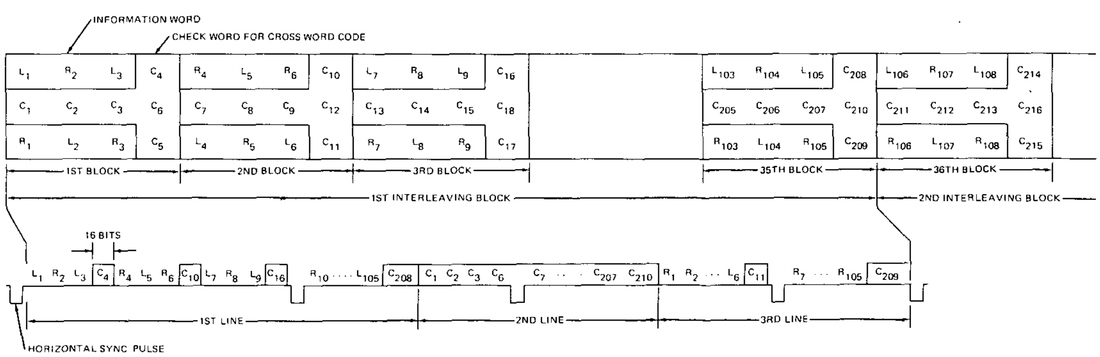

Fig. 11. Interleaving system for standard A.

1 video field = 7 interleaving blocks = 245 blocks = 735 information words × 2 channels
1 interleaving block = 35 blocks = 105 information words × 2 channels

Fig. 12. PCM-1 and a VTR (Betamax).

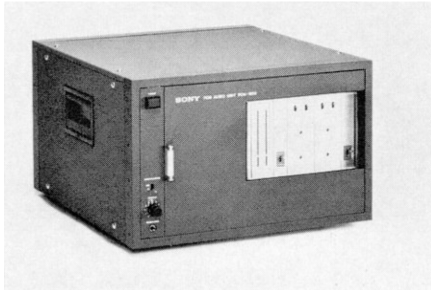

Fig. 13. PCM-1600 for use with a U-matic VTR.

REFERENCES

[1] A. Iga *et al.*, "A Consumer PCM Audio Unit Connectable to Home-Use Video Tape Recorder," presented at the Conference of the Acoustical Society of Japan, October 1977, paper 3-2-9.

[2] Y. Tsuchiya *et al.*, "Professional PCM Audio Unit Using Error Correcting Code," presented at the Conference of the Acoustical Society of Japan, October 1977, paper 3-2-10.

[3] T. Doi *et al.*, "On Several Standards of Conversion Form of PCM Signals into Standard Video Signals," presented at the Conference of the Acoustical Society of Japan, October 1977, paper 3-2-15.

[4] G. V. Jacoby, "A New Look-Ahead Code for Increased Data Density," presented at INTERMAG, June 1977.

[5] A. Iga *et al.*, "A Measurement Method of Drop-Outs of Audio PCM Recorder," presented at the Conference of the Acoustical Society of Japan, October 1977, paper 3-2-12.

[6] G. Fukuda *et al.*, "Drop-Out Statistical Model on PCM Magnetic Sound Recording," presented at the Conference of the Acoustical Society of Japan, April 1977, paper 4-6-5.

[7] G. Fukuda *et al.*, "Drop-Out Simulations on PCM Magnetic Sound Recording (2)," presented at the Conference of the Acoustical Society of Japan, October 1977, paper 3-2-11.

[8] T. Sonoda *et al.*, "A Listening Test of Dropout-Compensation of PCM Audio Recorder," presented at the Conference of the Acoustical Society of Japan, October 1977, paper 3-2-13.

[9] T. Doi, "A Class of Error Correcting Code for a Transmission System with High Error Rate," presented at the Conference of the Acoustical Society of Japan, October 1977, paper 3-2-14.

THE AUTHORS

Toshitada Doi received a B.E. degree in 1964 from the Tokyo Institute of Technology, Tokyo, Japan, and the Dr. Eng. degree in 1972 from the Tohoku University, Sendai, Japan.

Since 1964, he has been working for Sony Corporation. His major fields of research are antennas, circuit theories, computer-aided design of integrated circuits, large-scale tidal generation of electricity, psychoacoustics, and digital audio systems.

Dr. Doi is a member of the IEEE, the IECE, the Acoustical Society of Japan and the Oceanographical Society of Japan.

•

Yoshikazu Tsuchiya received a B.E. degree in 1966 from the Waseda University, Tokyo, Japan. He joined Sony Corporation in 1966 and has been working on designing tape recorders, integrated circuits, and on the development of digital audio systems.

Mr. Tsuchiya is a member of the IECE and the Acoustical Society of Japan.

•

Akira Iga received a B.E. degree in 1970 from Kansai University, Osaka, Japan. In 1973, he joined Sony Corporation. He has been working on research in the area of quadraphony, and digital audio systems.

Mr. Iga is a member of the IECE and the Acoustical Society of Japan.