UDC 681.327.64

Optimal codes for digital magnetic recording

J. C. MALLINSON, M.A., M.Inst.P.*

and

J. W. MILLER, B.S.*

Based on a paper presented at the Conference on Video and Data Recording held in Birmingham from 20th to 22nd July 1976

SUMMARY

Channel coding is universally employed in digital magnetic recording to match certain properties of the coded data to the characteristics of the recorder channel. In this paper, attention is focused primarily on the digital sum variation and d.c. content of various codes. The older, more familiar examples, such as NRZ, Manchester, enhanced-NRZ and Miller are compared with the more recently developed group, zero modulation and M² codes.

* *Ampex Corporation, 401 Broadway, Redwood City, California 94063.*

1 Introduction

Three distinct classes of binary digital coding are discussed frequently in connection with magnetic recording. *Source coding* is used to reduce the quantity of data to be recorded, as in differential p.c.m. or delta modulation, or to render recorder defects less significant, as in Fourier, Hadamard or other transformations. Conversely, in *error detection and correction coding* the quantity of data is increased by the inclusion of redundant information which permits a purely logical determination and treatment of error conditions; examples range from simple parity checking schemes to Hamming and other highly complex cyclic codes. In *channel coding*, the principal subject of this paper, the data to be recorded are modified to obtain the highest density permitted by the limiting characteristics of the magnetic recording channel.[1]

Whilst it is clear that many differing criteria could be used in the design of such channel codes, two points of view are most prevalent. Using well-founded tenets of communication theory, it is argued that in order to maximize, on the average, the storage or transfer of information in the recorder channel, the power spectral density of the channel code, for random input data, should match the transfer function of the recorder.[2] Unfortunately, however, a code based upon statistical considerations alone remains vulnerable to specific worst case or so-called 'pathological' data sequences for which the error rates may be greatly in excess of the average. This fact, which is of particular concern in the recording of digital data for computers, has led increasingly to an alternative approach which may be termed worst case design of channel codes.

Upon recalling that magnetic recorders are absolutely incapable of reproducing very low frequency or d.c. waveforms, it is perhaps not surprising that most of the important advances in worst case channel code design have involved modification to the low frequency or d.c. spectral content. Indeed the evolution of channel codes for magnetic recording may be regarded rather logically as a gradual process of reduction of the d.c. content without concomitant increases in bandwidth.

After reviewing the magnetic recording channel limitations and certain definitions concerning codes, in this paper we follow the evolution of the following codes: NRZ, Manchester, ENRZ, GC, Miller, ZM and M^2. Attention is directed primarily at the digital sum variation and maximum d.c. content.

2 Magnetic Recording Channel Limitations

A magnetic recorder is a band-pass, highly non-linear communication channel which suffers both amplitude instability and timing errors. Each of these factors constrains the selection of channel codes.

As a band-pass channel the recorder will reproduce neither very low frequency, long wavelength nor very high frequency, short wavelength waveforms. Transmission of d.c. is precluded for several reasons; no magnetic field emanates from d.c. magnetized media and, therefore, no read-head flux is induced; the read-head output voltage is due to time domain differentiation of the head flux and the possible use of either coupling (rotary) transformers or capacitors. The absence of read-head d.c. flux is absolute in all current designs and will only be changed by the unlikely advent of designs in which the recording medium actually threads the head coil.[3] As a practical matter, the long wavelength limit in current recorders is directly related to the overall physical dimensions of the read head. Short wavelength response may be limited by the read-head gap null and the existence of extreme spacing losses. Practical limits in current precision recorders are long and short wavelengths of 1–2 cm and 0·5–1·0 μm respectively. The band-pass characteristic favours codes with small or zero d.c. content and a high density ratio.

At long and medium wavelengths a recording channel may be rendered linear by the use of a.c. bias. However, at short wavelengths this technique is no longer effective and both a.c. biased and unbiased recording yield identical non-linear responses.[4] On the other hand, when binary symmetric levels are recorded on erased media with proper pre-equalization the output becomes a linear transform of the input and linear post-equalization is then effective in correcting the channel distortions. The channel non-linearity and the output amplitude instability, due principally to variations in head–medium spacing, operate against the use of multi-level or partial response coding.

Timing errors arise for two main reasons: improper corrections, or equalization of the channel distortions and variations in the head–medium relative velocity. This necessitates run length controlled channel codes which are self-clocking.

3 Code Parameter Definitions

It is useful to define several parameters of the channel codes; in general our definitions follow those of Patel.[5]

Suppose that, on average, x data bits are encoded into y binary digits; the ratio x/y is called the rate of the code. All codes considered below have rates between one half and unity; lower rate codes require higher frequency clock and detection circuits.

Following conventional practice we assume that the ones and zeroes in the coded sequence are recorded on the medium by the presence and absence of magnetic polarity transitions respectively. The shortest run length of zeroes between consecutive ones is d digits; this determines the minimum distance $(d+1)$ between transitions and hence the highest transition density. The longest run of zeroes between consecutive ones is k digits; the greater k, the worse become the self-clocking properties of the code.

The digital sum variation (d.s.v.) is the running integral of the area beneath the coded sequence; in computing the d.s.v. the binary levels are assumed to be ± 1. If the d.s.v. of the code can grow indefinitely the code has d.c. content; if the d.s.v. is bounded the code is d.c. free.

A convenient measure of code efficiency is the density ratio, DR, which is given by

$$DR = \frac{\text{data density}}{\text{highest transition density}} = \frac{x}{y}(d+1)$$

For data bits arriving at time intervals of T, the minimum time interval between media transitions is $DR.T$. The density ratios of all the codes treated below fall between one half and unity. As a general rule, the high frequency response of the channel must extend somewhat above the Nyquist frequency corresponding to the maximum transition density. This frequency is $(2DR.T)^{-1}$; the larger the DR, the lower the bandwidth of the code.

4 The Codes

Seven distinct channel codes are in current use in magnetic recording. In Fig. 1 we show typical waveforms corresponding to all the codes. In Table 1 we give the values of all the code parameters defined in the previous Section.

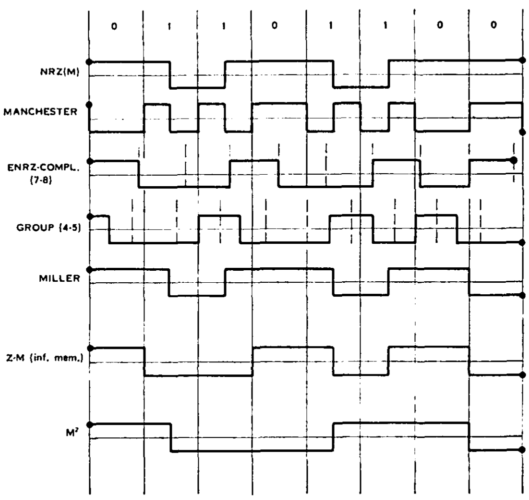


Fig. 1. Code waveforms.

4.1 Non-return to Zero

NRZ is the archetypal code in which data are generally supplied to the recorder. Two main classes exist; in NRZ (mark), ones in the data stream are recorded (i.e. marked) as magnetic transitions in the middle of the data bit interval and zeroes are ignored.[6] In NRZ (level), zeroes and ones are recorded as positive and negative

Table 1. Code parameters

	Rate	d	k	DR	Maximum d.s.v.	Maximum d.c. content
NRZ(M)	1	0	∞	1	$\pm\infty$	± 1
Manchester	1/2	0	1	1/2	$\pm T$	0
ENRZ-C	7/8	0	7	7/8	$\pm\infty$	$\pm 3/4$
Group	4/5	0	2	4/5	$\pm\infty$	$\pm 2/5$
Miller	1/2	1	3	1	$\pm\infty$	$\pm 1/3$
Z-M(INF.)	1/2	1	3	1	$\pm 3/2T$	0
M^2	1/2	1	5	1	$\pm 3/2T$	0

levels so that transitions only occur whenever the data bits change. NRZ is not a run-length-limited code. For long strings of zeroes the d.s.v. is unbounded and the maximum d.c. content is unity; in consequence the 'eye' pattern† can become completely closed rendering simple binary threshold detection impossible. NRZ(M) is used in 800 bits/in IBM compatible longitudinal recorders.

4.2 Manchester

Manchester coding and its variants are known by many names: bi-phase, double-frequency, phase-encoding, frequency-shift-keying and, quite inaccurately, frequency-modulation.[7] Here the coding rules are identical to NRZ(M) with the addition of extra transitions at the beginning of every data bit interval. The resultant half-rate code is run-length limited and was the first known code with bounded d.s.v. and zero d.c. content. Unfortunately these advantages are achieved at the cost of a *DR* of one half. The Nyquist frequency is accordingly twice that of NRZ and this excessive bandwidth requirement has limited Manchester coding to low density applications. Manchester coding has not led to significant evolutionary developments. Manchester coding is used in 1600 bits/in IBM compatible data recorders.

4.3 Enhanced-NRZ

Two variations of ENRZ have been described; both may be regarded as simple, but mainly ineffective, modifications of NRZ(L).[8] The basic idea is to add extra transitions to NRZ(L) in order to make it run-length limited (i.e. k bounded); this assures clock extraction for any input data. In both versions an extra interval is inserted after seven code bits. In ENRZ-parity, the eighth bit is odd parity on the previous seven code bits; in ENRZ-complement the eighth bit is the complement of the seventh code bit. A single bit memory is required to implement both codes. In both cases the d.s.v. remains unbounded and the maximum value of d.c. is 3/4. For worst case input data the eye pattern amplitude can close to 1/4; this corresponds to a loss in effective signal-to-noise ratio of 12 dB, which is incompatible with high density recording. Whilst it is clear that the d.c. content could be reduced by introducing the extra bit more frequently; this would reduce the *DR* and increase the bandwidth excessively. ENRZ has been used in, for example, certain high density recorders for instrumentation applications.

4.4 Group Coding

Group coding, which is also called run-length-coded-NRZ, is a generic term identifying the technique of slicing the incoming data into blocks or groups and transforming these data groups into longer code words which are then recorded as NRZ(M). The transformation may be performed according to algebraic rules or by using a look-up table or dictionary. These alternatives should strictly be called group (after group theory) and block coding respectively but this distinction is not always honoured. The advantage of transforming the data words into larger code words is that undesirable code words may be rejected and only favourable ones retained. For example, in the only known application of group coding, in 6250 bits/in IBM compatible longitudinal recorders, four-bit data words become five-bit code words; a four-data-bit memory is thus required.[9] The 16 (out of 32) chosen code words are shown in Table 2, wherein it will be seen that the possibilities with long strings of zeroes (large k) have been rejected; this makes clock extraction more certain and keeps the d.s.v. of each code word within the limits $\pm 2T$. When concatenating five-bit code words, the maximum d.s.v. value can grow indefinitely and the maximum d.c. content is 2/5. It is obvious that, by selecting a code word length sufficiently greater than the data word, only code words with a d.s.v. value of zero need be used; this possibility would be accompanied, of course, by significant increases in the bandwidth necessary.

† Eye pattern: a time synchronized oscilloscope display of the coded waveform recovered from the recorder which is frequently used in assessing the subsequent detectability of the data.

4.5 Miller

Miller code is known by two other names: delay-modulation and (even more confusingly) modified-frequency-modulation (MFM).[10] The coding rules are: data ones are coded, as in NRZ(M), with transitions in the middle of the bit cell, isolated data zeroes are ignored (or delayed) and transitions are inserted at the beginning of the bit cell between pairs of data zeroes. The great virtue of Miller code is that, since isolated zeroes are ignored, the *DR* remains unity as in NRZ; the bandwidth requirements of Miller code are, consequently, little greater than in NRZ. The penalties include a rate of one half which necessitates a double frequency clock, the inability to recover clocking until a 101 data pattern occurs, a loss of effective signal-to-noise ratio of 3 dB since transitions have to be identified both at the beginning and the middle of bit cells and the requirement

Table 2. Four-five group code

Data words	Code words	d.s.v.†
0000	11001	$-2T$
0001	11011	$-1T$
0010	10010	$+1T$
0011	10011	$+2T$
0100	11101	$+1T$
0101	10101	0
0110	10110	$+2T$
0111	10111	$+1T$
1000	11010	0
1001	01001	$+1T$
1010	01010	$-1T$
1011	01011	0
1100	11110	$-1T$
1101	01101	$-2T$
1110	01110	0
1111	01111	$-1T$

† Assuming starting conditions of zero d.s.v. and a negative level.

for a single bit memory for encoding. That these difficulties are not considered too severe may be judged from the widespread application of Miller code; it is used in all recent IBM disk recorders and until recently in all Ampex high density digital longitudinal recorders. An examination of Miller code leads to the conclusion that the maximum d.c. content is 1/3; the eye pattern is thus constricted by 3·5 dB on this account. When comparing ENRZ with Miller we may, therefore, expect a difference in effective signal-to-noise ratio of $12-(3+3{\cdot}5) = 5{\cdot}5$ dB. This value is in close accord with the experimental determination of the difference in 'recording margin' reported in a companion paper.[11] Several minor variations of the Miller theme are employed (they include the even more confusingly styled modified-modified-frequency-modulation (MMFM) or Wood codes) but the minor changes involved do not render them d.c. free.

4.6 Zero-Modulation

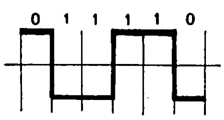
ZM codes were first described in 1975 by Patel who states, 'Zero-modulation was designed for, and is used in, magnetic recording'.[5] The crucial idea in ZM is to modify Miller code so that it becomes d.c. free and yet retains unambiguous code patterns which allow non-error-propagating decoding. Any data stream may be considered a series of sequences of two types:

(a) 011110; n ones bounded by zeroes, $n \geqslant 0$

and (b) 111111; m ones.

Under Miller coding rules, sequences of type (a) with n even and non-zero have a non-zero d.s.v. which, upon concatenation with interleaving type (b), m even sequences, can grow indefinitely. All other sequences, type (a) with n odd or zero or type (b) have zero d.s.v. In the sequences with non-zero d.s.v., ZM encodes the zeroes in the Miller manner; the ones, however, are encoded as though they were zeroes but with alternate transitions deleted. The example in Fig. 2 makes the d.s.v. of the sequence zero. The omission of alternate transitions between pairs of 'ones coded as zeroes' makes this pattern distinct for decoding. Note that since

Fig. 2. Zero-modulation coding.

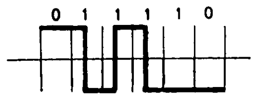


the d.s.v. returns to zero at the very end of the sequence, the final data zero must not be counted again for the next sequence; an example appears in Fig. 1. All other data sequences are, of course, coded by the normal Miller rules. Unfortunately this coding change is rather difficult to implement because complete sequences have to be altered. In principle, therefore, infinite look-forward and look-backward memories are required to identify the sequence boundaries and to provide the ZM modified code sequence if, indeed, n turns out even. Patel recognized this difficulty and suggested that the necessary memory could be limited by blocking the data into groups of f bits followed by a zeroes parity bit. This guarantees that all modified sequences fall within an $(f+1)$ bit block. Naturally, however, the addition of extra code bits reduces the density ratio, DR, and increases the bandwidth requirements of the code.

4.7 M²

M² is another recent modification, invented by one of the authors (J. W. Miller), of the Miller code. This proprietary Ampex code is used in the current high density longitudinal recorders discussed in a companion paper.[11] M² is also used in the recently announced Electronic-Still-Store, the world's first commercial digital video recorder. In M², as in ZM, the basic Miller code is modified so that it becomes d.c. free. Again, this is achieved by modifying the sequences which have non-zero d.s.v. To limit the memory required, however, the changes are introduced only at the end of the sequences. In the sequences with non-zero d.s.v., M² encodes all the ones, excepting the last, in the normal Miller manner; the final one is simply ignored. For example, in Fig. 3 the d.s.v. returns to zero after the final data one and consequently, in contrast to the situation prevailing in ZM, the final data zero is to be counted again for the next

Fig. 3. M² coding.



sequence. All other sequences are coded by standard Miller rules. M² coding produces characteristic transition free runs of $2\frac{1}{2}$ and 3 data intervals which do not occur in Miller code; this ensures unique decoding. The most significant advantages of M² over ZM are that only 3 bits of memory are needed and that a density ratio, DR, of unity is maintained. The bandwidth requirements of this d.c. free code are, in consequence, very little greater than in NRZ. Negligible base-line wander or eye-pattern closure occurs with M². The effective worst-pattern signal-to-noise ratio expected is 3·5 dB better than in Miller; in high density digital recording this difference becomes extremely significant.

5 Conclusions

Driven by the continuing requirement for higher recording densities, a steady improvement in channel codes for digital magnetic recording has taken place. The most important developments have been guided by a philosophy of 'worst-case' design and have been reviewed in this paper. Particular emphasis has been placed in reducing or eliminating the d.c. content of codes without incurring appreciable increases in the bandwidths necessary. Whereas group coding has the potential to be d.c. free, its present realizations are not. On the other hand, the most recently announced codes, ZM and M², appear to have achieved the goals completely. Both are d.c. free and need little more bandwidth than NRZ. Even though M² is scarcely more difficult to implement than Miller, it is expected that the current pace of intense channel code development will continue. More refined and complicated codes will most probably be discovered; for example, by encoding longer sequences it may be possible to increase the code rate (x/y) or to match the code spectrum more closely to the magnetic recorder channel. The ultimate viability of such developments will depend largely upon the cost and complexity of the associated encoder–decoder circuitry.

6 References

1. Kiwimagi, R. G. *et al.*, 'Channel coding for digital recording', *IEEE Trans. on Magnetics*, **MAG-10**, No. 3, pp. 515–8, September 1974.
2. Knoll, A. L., 'Spectrum analysis of digital magnetic recording waveforms', *IEEE Trans. on Electronic Computers*, **EC-16**, No. 6, pp. 732–43, December 1967.
3. Mallinson, J. C., 'On recording head field theory', *IEEE Trans.*, **MAG-10**, No. 3, pp. 773–5, September 1974.
4. Mallinson, J. C. and Bertram, H. N., 'Write processes in high density recording', *IEEE Trans.*, **MAG-9**, No. 3, pp. 329–31, September 1973.
5. Patel, A. M., 'Zero-modulation encoding in magnetic recording', *IBM J. Res. Dev.*, **19**, No. 4, pp. 366–78, July 1975.
6. A.N.S.I. Standard X 3.22–1973, 'Recorded Magnetic Tape for Information Interchange' (800 char/in, NRZI).
7. A.N.S.I. Standard X 3.39–1973, 'Recorded Magnetic Tape for Information Interchange' (1600 char/in, PE).
8. Wells, J. B., 'High density digital magnetic tape recording using enhanced-NRZ coding', Conf. on Video and Data Recording, July 1973, pp. 113–8 (IERE Conference Proceedings No. 26.)
9. A.N.S.I. Standard X 3.54 (Proposed), 'Recorded Magnetic Tape for Information Interchange' (6250 char/in, group coded recording).
10. Cullum, C. D., 'Encoding and signal processing', in 'Advances in Magnetic Recording', Annals New York Acad. Sciences, Vol. 189, pp. 52–62, 1972.
11. Spitzer, C. F., Jensen, T. A. and Utschig, J. M., 'High bit-rate, high density magnetic tape recording', Conf. on Video and Data Recording, July 1976, pp. 147–60. (IERE Conference Proceedings No. 35.)

Manuscript first received by the Institution on 29th March 1976 and in final form on 19th November 1976. (Paper No. 1762/Com. 145.)

The Authors

John Mallinson read natural philosophy at University College, Oxford, graduating in 1953. He joined Amp, Inc., Harrisburg, Pennsylvania, in 1956 to work on the theory and design of all-magnetic logic elements and six years later moved to Ampex Corporation, Redwood City, California, to investigate fundamental considerations in magnetic tape recording. As manager of the Basic Technology Section of the Research Department, he directed the activities of a group working on magnetic recording theory, micro-magnetics, communication theory, and the exploration of advanced concepts in various areas of recording. He holds several patents as a result of his work. Mr. Mallinson is currently managing a High Bit Rate Recording Group in the Data Products Division of Ampex, which is investigating several techniques of recording gigabit per second data at extremely high spatial densities.

Jerry Miller received the B.S. degree in electrical engineering from Heald College, San Francisco in 1959, where he remained until the end of 1960, lecturing in a variety of engineering subjects. He joined Ampex in 1961 and worked for the next seven years on the development of instrumentation tape recorder systems of both longitudinal and rotary-head designs. He then took part in the development of the TBM memory system, first as system engineer and subsequently as engineering manager. Since 1974 he has been with Company's Research Department studying high density magnetic recording channel noise and signal characteristics. Mr. Miller is the holder of several issued and pending design patents and has published three previous papers dealing with recording methods and systems.