

Coding Techniques for the Noisy Magnetic Recording Channel: A State-of-the-Art Report

KEES A. SCHOUHAMER IMMINK, SENIOR MEMBER, IEEE

Abstract—Coding techniques for improving the reliability of information storage on noisy magnetic recording channels are considered. It is assumed that the Lorentzian channel model applies and that the retrieved signal is perturbed with additive white Gaussian noise. The immunity against additive noise of state-of-the-art codes such as dc-free, runlength-limited, and trellis codes will be assessed.

I: INTRODUCTION

IF there is any truism in our recording world, it is this: Tomorrow it will be possible to store larger amounts of data on a smaller storage medium than it is today. Our recorders are only prototypes of what is to come. The evolution of technology in magnetic recording is the result of contributions from various fields of engineering such as mechanics, head and medium research, coding techniques, etc. In this paper, we will (attempt to) answer the question of what potential contribution may be expected from the use of sophisticated coding and detection techniques. An obvious approach would be to compute Shannon's channel capacity for the magnetic recorder. For two reasons this approach will not be followed in this paper. In the first place, computing the capacity of a magnetic recorder channel is a notoriously difficult task and at present an unsolved problem. Second, if we were able to find or estimate the channel capacity, it is not known how anything like the performance promised by Shannon can be achieved in practice. Therefore we prefer here to compute the performance of some classes of recording codes. Admittedly, this method does not offer the "ultimate" answer, but it does provide some conclusions which are sufficiently accurate for recording.

Recording codes are employed in magnetic and optical recording for a variety of reasons. By way of example, we would mention the use of recording codes to overcome problems involving servo systems used in track following, specific media properties such as over-write noise, the complexity of the detector, etc. It is worth emphasizing here that the communication channel aspects of a recorder are only a relatively small part of the overall trade-off of the various parameters of storage systems. Various coding techniques used at present to improve the reliability of information storage on magnetic recording channels will be assessed on the basis of the theoretical principles of communication. In particular, we consider the use of coded information in conjunction with maximum-likelihood sequence estimation (MLSE) detection, which has only recently attracted interest in magnetic recording literature [1]–[4]. The coding gain of the various code formats will be used as a figure of merit. This

Paper approved by the Editor for Coding Theory and Applications of the IEEE Communications Society. Manuscript received July 23, 1987; revised October 12, 1987. This paper was presented at the IEEE International Symposium on Information Theory, Kobe, Japan, June 19–24, 1988.

The author is with Philips Research Laboratories, 5600 JA Eindhoven, The Netherlands.

IEEE Log Number 8926988.

parameter gives the amount of saving in the required signal-to-noise ratio of the coded scheme as compared to the uncoded scheme assuming that the (recording) channel is perturbed with additive white Gaussian noise. Technical "details" such as complexity of the detector, clock regeneration, resistance to disturbances other than additive noise, etc., are ignored here. The reader is warned at this point not to exaggerate the importance of the coding gain parameter, as it has been shown that performance can deteriorate rapidly, for example, in the case of erroneous receiver knowledge about the actual channel characteristics [5].

The ordinary digital magnetic disk or tape recording operate according to the principle of saturation recording, i.e., two stable states of magnetization represent the binary data to be stored. The coding gain of the following classes of state-of-the-art codes, which are all based on two-level sequences, is determined as follows:

- dc-free block codes,
- runlength-limited codes,
- trellis codes.

dc-free (or dc-balanced) codes have been employed in communication systems since the early days of the introduction of PCM systems [6]. dc-free codes, as their name already suggests, have the property that the average power of the coded stream vanishes at zero frequency. It has often been conjectured (quantitative results are not available) that when the recording channel shows a null at zero frequency it is advantageous for the coded stream to exhibit a null at zero frequency as well. In other words, according to this belief, the power-density function of the coded stream should "match" the transfer function of the communication channel concerned. Other considerations, e.g., the ac coupling to a rotary head system, and simplicity of the detection circuitry also lead to the adoption of dc-free codes in practical embodiments. Simple recording codes such as the Manchester code were previously used in magnetic recording practice [7] and this is at present being employed in the digital audio track of the 8 mm video recorder [8]. Other current examples of dc-free channel codes can be found in the S-DAT and R-DAT digital audio recorders for domestic use [9] and experimental digital video recorders [10], [11].

Recording codes based on runlength-limited sequences have found almost universal application in optical and magnetic disk recording practice. Examples are the IBM 3380 disk drive where the rate $1/2$, $(2, 7)$ code is applied [12] and the compact disc which is based on the EFM code [13], [14]. Runlength-limited sequences are characterized by two parameters, $(d + 1)$ and $(k + 1)$, which represent the minimum and maximum runlength of the sequence, respectively. In sharp contrast with the considerable amount of literature available on the actual design of encoding the decoding embodiments of runlength-limited sequences, is the lack of knowledge on the actual performance of these sequences on real-life channels.

An entirely new breed of recording codes has recently been published in the literature. Examples of so-called trellis codes, which are designed to improve the coding gain on the partial-response channel with response $(1 - D^n)$ where $n = 1, 2, \dots$ and D is the unit-delay operator corresponding to one modulation (or channel symbol) interval, can be found in [15], [16]. As will be discussed in a moment, the response of an idealized magnetic recorder at asymptotically low information densities coincides with the response $(1 - D)$ of partial-response channels. We will take a closer look at the noise immunity of this new branch of recording codes at medium and high information density.

The outline of the paper is as follows. The basic scene of MLSE detection will be set in Section II. Thereafter, Section III deals with the description of the channel model of the idealized digital magnetic recorder. The Euclidean distance between dc-free codewords is then examined in Section IV. Finally, Sections V and VI deal with the performance of RLL and trellis codes, respectively.

II. PRELIMINARIES

Prior to the development of a quantitative description of the channel model some basics of maximum-likelihood sequence estimation (MLSE) will be considered. It is assumed that the binary user information with a bit rate of $1/T$ is translated into a coded channel sequence having a channel bit rate of $1/T_c$, $T_c \leq T$. The quotient

$$R = T_c/T \quad (1)$$

is, as usual, called the rate of the code. The recorded sequence \mathbf{a} consists of binary digits $a_i \in \{-1, 1\}$ that are generated each T_c second. The recorded channel sequence (or codeword) $\mathbf{a} = (a_1, \dots, a_N)$, taken to be of arbitrary length N , is a member of a predefined set S_c of codewords. Assuming a linear read-out mechanism, the retrieved signal $r(t)$ is of the form

$$r(t) = \sqrt{E_c} \sum_{i=1}^N a_i g(t - iT_c) + n_w(t) \quad (2)$$

where E_c is the received energy-per-channel-symbol, $g(t)$ is the channel waveform and $n_w(t)$ is additive white Gaussian noise with two-sided spectral density $N_0/2$. In other words, it is assumed that the only noise source originates from the head and read-electronics, and consequently medium noise is ignored. The channel waveform $g(t)$ is normalized in such a way that

$$\int_{-\infty}^{\infty} g^2(t) dt = 1. \quad (3)$$

When the retrieved signal is matched-filtered with a filter having a response $\sqrt{E_c}g(-t)$ and subsequently sampled at $t = kT_c$, the equivalent channel vector $\mathbf{r} = (r_1, \dots, r_N)$ can be written as

$$\mathbf{r} = E_c \mathbf{M} \mathbf{a} + \mathbf{n} \quad (4)$$

where \mathbf{M} is the $N \times N$ sampled autocorrelation matrix of $g(t)$ with entries

$$M_{i,j} = g_{i-j} = \int_{-\infty}^{\infty} g(t + iT_c)g(t + jT_c) dt \quad (5)$$

and \mathbf{n} is the noise vector $\mathbf{n} = (n_1, \dots, n_N)$ where n_i are zero mean, Gaussian distributed samples with autocorrelation function $E(n_j n_{j+k}) = E_c N_0 g_k/2$. The parameters g_i , $i \neq 0$, are termed intersymbol interference (ISI) coefficients. Note that $g_i = g - i$, and as a result of the normalization (3), $g_0 = 1$.

A bound for the probability that an error event starts at a particular instant was established by Forney [17]. Prior to presenting this expression of the error probability, the difference sequence will be defined.

Let \mathbf{a} and $\hat{\mathbf{a}}$ be distinct members of the codeword set S_c . The difference sequence $\mathbf{e} = (e_1, \dots, e_N)$ is defined by

$$e_i = (a_i - \hat{a}_i)/2.$$

Obviously $e_i \in \{-1, 0, 1\}$. The set of all difference sequences between any two distinct codewords of the set S_c is denoted by S_e .

For moderate and high signal-to-noise ratios the error probability $\Pr(\mathbf{e})$ is well approximated by the following lower bound:

$$\Pr(\mathbf{e}) > N_{\min} Q \left\{ d_{\min} \sqrt{\frac{2E_c}{N_0}} \right\} \quad (6)$$

where $Q(x)$ is the tail integral of the Gaussian distribution, defined by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-y^2/2} dy,$$

the minimum squared Euclidean distance d_{\min}^2 is defined by

$$d_{\min}^2 = \min_{\mathbf{e} \in S_e} \mathbf{e}^T \mathbf{M} \mathbf{e} \quad (7)$$

and N_{\min} is the (average) number of nearest-neighbor sequences. The lower bound (6) is approached asymptotically at high signal-to-noise ratios. In order to evaluate the code's immunity against additive noise, it is worth while to define what is called the asymptotic coding gain G

$$G = \frac{d_{\min}^2 E_c(\text{coded})}{d_{\min}^2 E_c(\text{uncoded})}. \quad (8)$$

The coding gain G expresses the saving in the required signal-to-noise ratio of the coded scheme as compared to the uncoded scheme, asymptotically for the same low error probability. The constant N_{\min} , though important, is in fact ignored. In principle, it is possible to compute the coding gain at more realistic values of the error probability and signal-to-noise ratio by taking into account the distance profile of the code. If, however, we keep in mind that the computer-assisted search used to find the asymptotic coding gain or minimum Euclidean distance of the various codes to be presented shortly takes hours of computing time on an IBM309 mainframe computer, then we may conclude that more refined criteria are out of the question. The asymptotic coding gain, in short coding gain, of various code formats will be used as a figure of merit of the resistance of the code to additive Gaussian noise.

Efficient and systematic procedures are available to design codes that achieve coding gain on the ISI-free channel. The complex nature of the Euclidean distance measure has hindered the design of codes when channels with intersymbol interference are involved. The computation of the minimum Euclidean distance is quite tedious, as the distance between *any* two codewords of the set has to be determined. As a result the computational burden can be enormous, even for relatively simple codes.

In the following sections, extensive use will be made of (7) and (8). Section III provides the channel model of the idealized digital magnetic recorder, so that \mathbf{M} can be computed. Thereafter, a description will be given of various code formats to be investigated.

III. CHANNEL MODEL

The previous section has considered, in highly brief terms, the steps necessary for assessing the quality of a recording

code when maximum-likelihood detection is employed. The subsequent analysis of the magnetic recorder channel is based on the Lorentzian channel model [18]. According to this model the step response of the read-out process is of the form

$$h(t) = \frac{v}{pw_{50}} \frac{1}{1 + (2vt/pw_{50})^2} \quad (9)$$

where v is the medium-to-head speed and pw_{50} determines the dispersivity of the recording channel. Proportionality constants not relevant for the study to follow have not been written down. The Lorentzian model is a convenient model for the step response but does not fit all experimental results. Its main attribute is that it has a single parameter. More realistic models, e.g., due to Wallace [18], have been used in our computations. Qualitatively, this model leads to similar results as obtained with the simple Lorentzian model.

The magnetic domains are assumed to be recorded as perfect full- T_c pulses $p_{T_c}(t)$ with constant amplitude $\sqrt{E_b}/T$, or

$$p_{T_c}(t) = \begin{cases} \sqrt{\frac{E_b}{T}}, & |t| < \frac{T_c}{2}, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

E_b denotes the energy-per-user bit. The channel waveform $g(t)$ of the combined recording in conjunction with the read-out process is given by

$$\sqrt{E_c}g(t) = \sqrt{\frac{E_b}{T}} \{h(t + t_c/2) - h(t - T_c/2)\}. \quad (11)$$

The normalized information density S is defined as

$$S = \frac{pw_{50}}{vT}. \quad (12)$$

This definition is given on the understanding that the user bit rate $1/T$ is constant in both the coded and uncoded situation. In the R-DAT digital audio system [9] using metal powder tape, we find, for example, a normalized density value of $S \approx 2$. On floppy and hard disks the normalized density S obtained is in general much lower.

Keeping in mind that

$$\int_{-\infty}^{\infty} h(t - \tau/2)h(t + \tau/2) dt = \frac{\pi}{4} \frac{v}{pw_{50}} \frac{1}{1 + (v\tau/pw_{50})^2}$$

and using definitions (1) and (12), the energy-per-channel-symbol E_c and the autocorrelation coefficients g_i are found by means of a straightforward analysis

$$g_k = \int_{-\infty}^{\infty} g(t)g(t + kT_c) dt = \frac{S^2(-3k^2R^2 + R^2 + S^2)(R^2 + S^2)}{k^6R^6 - 2k^4R^6 + 3k^4R^4S^2 + k^2R^6 + 3k^2R^2S^4 + R^4S^2 + 2R^2S^4 + S^6} \quad (13)$$

and

$$E_c/E_b = \frac{1}{T} \int_{-\infty}^{\infty} \{h(t + T_c/2) - h(t - T_c/2)\}^2 dt = \frac{\pi}{2} \frac{1}{T^2} \frac{R^2}{S(R^2 + S^2)}.$$

When the relative information density is low, i.e., $S \ll 1$, then, using (13) and (14), the ISI can be approximated as follows:

$$E_c/E_b = \frac{\pi}{2} \frac{1}{S},$$

and

$$\begin{cases} g_0 = 1 \\ g_1 = g_{-1} = -1/2 \\ g_i = 0, |i| \geq 2, \end{cases}$$

which coincides with a partial-response channel with response $(1 - D)$ corrupted with white additive noise [17].

We are now in the position, using (8) and (14), to write down a general expression for the coding gain

$$G = \frac{d_{\min}^2 E_c(\text{coded})}{E_c(\text{uncoded})} = d_{\min}^2(\text{coded}) R^2 \frac{1 + S^2}{R^2 + S^2} \quad (15)$$

where we have tacitly made the substitution

$$d_{\min}^2(\text{uncoded}) = 1.$$

This choice can be motivated as follows. Using a computer search based on (7), we have found that at low information densities the error performance is dominated by single errors of the form $\dots, 0, 1, 0, \dots$ and its inverse. The associated distance of this error sequence is unity. At a higher information density the error performance is dominated by error sequences of the form $\dots, 0, 1, -1, 1, 0, \dots$ and its inverse. Apparently, at higher density, and consequently more intersymbol interference, an error sequence consisting of symbols of alternating sign is received with less energy, or in other words, is more prone to error, than error sequences with single errors. The distance of the $\dots, 0, 1, -1, 1, 0, \dots$ error sequence is found using (13) and (7)

$$3g_0 - 4g_1 + 2g_2 = \frac{S^4 - 11S^2 + 180}{S^4 + 13S^2 + 36}.$$

So that

$$d_{\min}^2(\text{uncoded})(S) = \min \left\{ 1, \frac{S^4 - 11S^2 + 180}{S^4 + 13S^2 + 36} \right\}.$$

Solving $d_{\min}^2(\text{uncoded})(S_1) = 1$, we find that at an information density $S \geq S_1 = \sqrt{6} \approx 2.45$ $d_{\min}^2(\text{uncoded})(S) \leq 1$. As the information density $S_1 = 2.45$ is relatively large for practical situations, the equality $d_{\min}^2(\text{uncoded}) = 1$ is used for convenience in the entire information density range considered.

Two observations regarding (15) are of immediate interest.

When the information density is high, i.e., $S \gg 1$, then

$$G \approx R^2 d_{\min}^2(\text{coded}), \quad S \gg 1, \quad (16)$$

which reveals that at relatively high information densities, the penalty on a reduction of the code rate R is proportional to R^2 . Apparently, it is more difficult to profit by coding at such densities than on an ISI-free channel where the penalty is proportional to R [19].

When the relative information density is low, i.e., $S \ll 1$, then

$$G \approx d_{\min}^2 \text{ (coded)}, S \ll 1. \quad (17)$$

IV. CODING GAIN OF dc-FREE CODEWORDS

In this section, attention is devoted to the error performance of dc-free codewords in conjunction with MLSE detection. A codeword $\mathbf{x} = (x_1, \dots, x_N)$, $x_i \in \{-1, 1\}$, N even, is dc-free when $\sum_{i=1}^N x_i = 0$. Let S_N designate the set of all dc-free codewords of length N . Obviously, the cardinality M_N of the set S_N is given by the binomial coefficient

$$M_N = \binom{N}{N/2}.$$

The rate R of a dc-free code is given by

$$R = \frac{1}{N} \log_2 M_N.$$

The preceding relationships are best illustrated by the following simple example.

Example: Let $N = 2$. The number of codewords is $M_2 = 2$ and the code rate $R = 1/2$. The two codewords are $(+1, -1)$ and its inverse $(-1, +1)$. This code is sometimes called the "bi-phase" or Manchester code. The minimum squared Euclidean distance between the sequences $(+1, -1)$ and $(-1, +1)$ is found with (7) and (13)

$$d_{\min}^2 \text{ (coded)} = \mathbf{e}^T \begin{bmatrix} g_0 & g_1 \\ g_1 & g_0 \end{bmatrix} \mathbf{e} = 2(g_0 - g_1) = \frac{3}{S^2 + 1}.$$

The actual coding gain of the bi-phase code, designated by $G_{bi-\phi}$, is found with (15)

$$G_{bi-\phi} = \frac{d_{\min}^2 E_c \text{ (coded)}}{E_c \text{ (uncoded)}} = \frac{3}{4S^2 + 1}.$$

It can be noticed that at low information densities, say $S \ll 1$, the bi-phase code achieves a coding gain of 4.8 dB. Solving $G_{bi-\phi}(S_1) = 1$, we find that at an information density greater than $S_1 = 1/2\sqrt{2}$, the gain turns into an actual loss. ■

Since the computation of the coding gain for larger codeword lengths proved to be prohibitively complex for human beings, we resorted to an exhaustive search using a computer.

Fig. 1 shows the coding gain versus relative information density S of dc-free codes with codeword length $N = 2, \dots, 8$ as a parameter. It can be observed that little benefit is to be expected from the use of dc-free codes at medium and high relative information densities.

V. CODES BASED ON RUNLENGTH-LIMITED SEQUENCES

Runlength-limited sequences (RLL) are the state-of-the-art cornerstone of most data storage systems, whether their nature is magnetic or optical. RLL sequences possess the property that their minimum and maximum runlengths, i.e., the number of consecutive like symbols, are constrained between $(d + 1)$ and $(k + 1)$ where $d \geq 0$ and $d < k$ are predefined parameters. Many interesting theoretical properties of RLL sequences can be found in the pioneering work of Tang and Bahl [20]. A considerable amount of literature is available on practical constructions of RLL encoders and decoders. Survey papers given by Siegel [7], Kobayashi [21], and Immlink [22] describe properties and design techniques for RLL sequences.

A dk -limited sequence simultaneously satisfies the following two conditions.

a) d constraint—any two logical ones are separated by a run of at least d consecutive logical zeros.

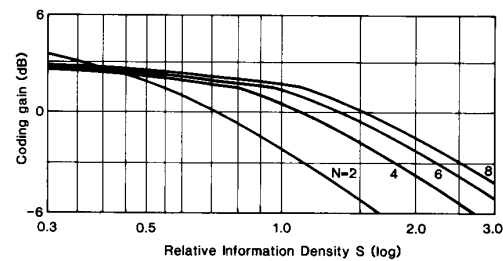


Fig. 1. Coding gain versus information density S of dc-free codes with codeword length N as a parameter.

b) k constraint—the length of any run of consecutive logical zeros is at most k .

Obviously, $k > d$. A sequence which only satisfies the d constraint is called a d -limited sequence. A d - or dk -limited is not the sequence to be recorded. The actually recorded runlength-limited sequence has at least $(d + 1)$ and at most $(k + 1)$ consecutive like symbols and is obtained by integrating modulo two a dk -limited sequence. The maximum runlength constraint is imposed to guarantee a clock pulse within some specified time. The minimum runlength constraint is imposed to control intersymbol interference and consequently has a bearing on the distortion of the received signal when the channel is bandwidth-limited [23]. Tang *et al.* [20] calculated the maximum rate, possible with the given d and k constraints, sometimes termed noiseless capacity, which is denoted by $C(d, k)$. The maximum rate $C(d, k)$ of dk sequences can be found with [20]

$$C(d, k) = \log_2 \lambda$$

where λ is given by the largest real root of

$$z^{k+2} - z^{k+1} - z^{k-d+1} + 1 = 0.$$

For d -constrained sequences λ is the largest real root of

$$z^{d+1} - z^d - 1 = 0.$$

Table I shows the relationship between the noiseless capacity $C(d) = C(d, \infty)$ of d -constrained sequences and the minimum runlength parameter d .

In practice, a rational number $m/n \leq C(d, k)$ is chosen for the rate of the code. To help keep the encoding and decoding hardware small, the integers m and n are often selected to be small. Franaszek [24] found that even with little hardware, practical codes could easily achieve rates of 90–95 percent of the noiseless capacity. Therefore, the noiseless capacity is used in the subsequent analysis as the rate of the runlength-limited code, i.e., $R = C(d)$, which frees the subsequent analysis from a specific embodiment of an RLL encoder and decoder. Since RLL sequences can be modeled as an output of a finite-state machine, the Viterbi detection of runlength-limited sequences can be established by a simple modification of the MLSE detector utilized for the detection of uncoded data [17], [5].

Fig. 2 shows the coding gain versus the information density of RLL sequences with the minimum runlength constraint $d = 1, 2, 3$ as a parameter. The coding gain has been determined numerically with an exhaustive computer search. It has been found that in the information-density range as depicted in Fig. 2, difference sequences of unity length dominate the error performance. Thus, in this region, using (15), the substitution $d_{\min}^2 \text{ (coded)}$ yields the coding gain of RLL sequences

$$G = \frac{d_{\min}^2 E_c \text{ (coded)}}{E_c \text{ (uncoded)}} = R^2 \frac{1 + S^2}{R^2 + S^2}$$

where $R = C(d)$. It can be concluded from Fig. 2 that in the

TABLE I
CAPACITY $C(d)$ OF RLL SEQUENCES

v	1	2	4	6
d_{free}^H	3	5	7	10
$H^1(D)$	11	101	10011	1011011
$H^2(D)$	01	111	11101	1111001
2^{v+1}	4	8	32	128
$G(\text{dB})$	3.01	4.77	6.02	6.98

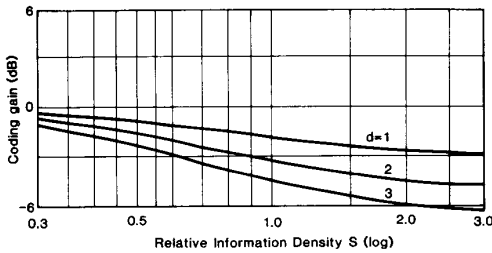


Fig. 2. Coding gain versus information density S of RLL sequences with minimum runlength d as a parameter.

entire information-density span shown, RLL sequences exhibit a clear coding loss. In other words, the distance properties between pairs of RLL sequences are rather poor. This conclusion coincides with the results of Pelchat and Geist [23], who concluded that these codes are bandwidth expansion codes in disguise.

VI. TRELLIS CODES

Examples of trellis codes for the partial-response channel with response $(1 - D^n)$ where $n = 1, 2, \dots$ and D is the unit-delay operator corresponding to one modulation (or channel symbol) interval, have recently been published [15], [16]. As previously discussed, the idealized magnetic recorder channel can be approximated at low densities by the partial-response channel with response $(1 - D)$. In this section, a closer look is taken at the performance of trellis codes designed for the $(1 - D)$ channel at medium and high information density where the channel response does not coincide with the partial-response $(1 - D)$ assumption. At medium and high information density the transfer function of the idealized recorder followed by properly designed linear filtering coincides with the $(1 - D^2)$ class IV partial-response channel [4]. If, however, the more sophisticated matched filtering is applied, this similarity does not take place. So that it may be expected—actual computations confirm this statement—that trellis codes designed for the ideal $(1 - D^2)$ do not improve the reliability of information storage on the Lorentzian modeled channel at medium and high information density.

Wolf and Ungerboeck studied the application of trellis coding techniques for improving the reliability of digital transmission over noisy partial-response channels. The authors concluded that the use of well-known binary convolutional codes in combination with a precoder actually the same device as utilized to transform a dk -limited sequence into an RLL sequence provides the answer to the coding problem. Precoding is a well-known technique usually used in partial-response systems to prevent error propagation [21]. The use of this fairly simple technique leads, for the configuration described, to coding gains which are smaller than the gains observed when convolutional codes are employed with compa-

table decoder complexity for ordinary ISI-free binary channels.

For the ideal $(1 - D)$ channel in conjunction with precoding, Wolf and Ungerboeck [15] showed the existence of a simple relation between the minimum squared Euclidean distance between distinct output sequences and the (free) Hamming distance of the (outer) code. With (17), we find in the information density region where the partial-response $(1 - D)$ assumption is valid that the coding gain equals the minimum squared Euclidean distance. Thus,

$$d_{\min}^2(\text{coded}) = G \geq \text{entier} \left(\frac{d_{\min}^H + 1}{2} \right), \quad S \ll 1 \quad (18)$$

where $\text{entier}(x)$ is the largest integer not greater than x and d_{\min}^H denotes the minimum (free) Hamming distance of the (convolutional) code employed.

Hamming-distance-increasing codes can now be specified, which in conjunction with the precoding technique, will lead to good Euclidean distance between output sequences. Though block codes have found widespread application in storage systems where large Reed-Solomon codes, in particular, hold undivided sway we will concentrate here on convolutional codes as this offers the opportunity to compare our results with those published in [15]. Table II shows the coding gain at asymptotically low information density, i.e., at an information density where the recording channel coincides with the partial-response $(1 - D)$ channel, of some selected convolutional codes with rate $R = 1/2$ (Note that as a result of the difference in definition of the coding gain, there is an $10 \log_{10} R = 3$ dB difference with ([15, Table III]). The constraint length of the convolutional encoder is denoted by v . The parity-check polynomials $H^1(D)$ and $H^2(D)$ are taken from [19]. The polynomials are specified in binary notation, e.g., $D^4 + D + 1$ is denoted by 10011.

Generally speaking, the $(1 - D)$ channel preceded by the precoder increases the number of states in the trellis diagram for the combined encoder and channel to 2^{v+1} . The number of decoder states 2^{v+1} gives an indication of the decoder complexity. Maximum-likelihood detection of the precoded sequence can be established in the receiver by a simple modification of the branch metrics of the Viterbi decoder. The coding gain of the various codes at medium and high information density is computed using an exhaustive search. It has been found that the performance depends on the order of switching at the convolutional encoder (i.e., a de-facto swapping of the parity-check polynomials), which demonstrates that the well-known optimum codes for the linear AWGN channel are no longer necessarily optimum for channels with severe intersymbol interference. As a consequence, a suitable choice of the configuration is quite important to achieve maximum performance. Unfortunately, this choice is quite difficult because no analytical tool has been found to determine the optimum code configuration.

Fig. 3 shows the coding gain of $R = 1/2$, trellis codes designed for the $(1 - D)$ channel as a function of the information density. It can be seen that the $(1 - D)$ trellis codes perform quite well at low information densities and their performance indeed coincides with Table II. It further becomes clear that this class of codes is not really successful when the bandwidth limitations play a dominant role. A comparison with Fig. 1 reveals that at an information density $S > 1$ they perform hardly better than the classical dc-free code with a codeword length $n = 8$.

VII. CONCLUDING REMARKS

In order to compare the immunity against additive noise of the various codes discussed so far, some typical examples of them are collected in the same figure. Fig. 4 shows the relationship between the coding gain and information density

TABLE II
CODING GAIN AT LOW DENSITY OF $R = 1/2$, $(1 - D)$ TRELLIS CODES

d	$C(d)$
0	1.00
1	0.69
2	0.55
3	0.47

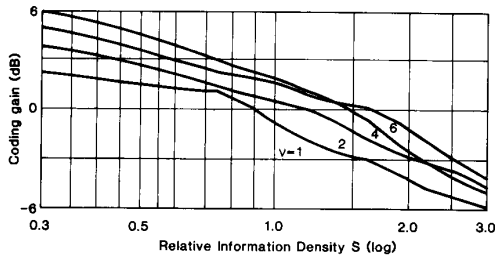


Fig. 3. Coding gain versus information density S of selected $(1 - D)$ trellis codes.

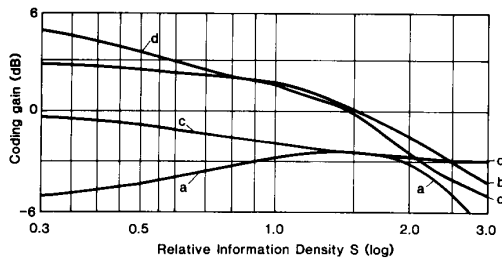


Fig. 4. Coding gain versus information density S of selected codes (a) partial-response class IV, (b) dc-free code $N = 8$ (c) RLL code $d = 1$, and (d) trellis code $R = 1/2$, $d_{\text{free}}^H = 7$.

of the following techniques: a) partial-response class IV detection of uncoded information, and MLSE detection of b) dc-free code $N = 8$, c) RLL code $d = 1$, and d) trellis code $R = 1/2$, $d_{\text{free}}^H = 7$. The relationship between the coding gain and information density of partial-response class IV detection of uncoded information has been added to offer a more complete picture, particularly as it provides the possibility to compare the performance of this relatively simple technique with that of the sophisticated and quite complex MLSE detection of coded sequences. The technique of the partial-response equalizer is well established, see, for example, [4], [25], so that we will not dwell on its details. Fig. 4 reveals that, for example, at an information density of $S = 2$, which is a value of practical interest, that the $R = 1/2$, $(1 - D)$ trellis codes by no means improve upon the simple dc-free block codes. It is further intriguing to observe that at this information density, all the selected state-of-the-art techniques perform equally well, within approximately ± 1 dB. This actually fuels the notion that at present there is no such thing as a "best" code or detection technique. One should, however, be careful in the conclusions. The sole criterion used in this paper to evaluate the performance of different coding techniques is the asymptotic coding gain, i.e., resistance to additive white noise. It is perhaps not wrong to conclude that the virtues of channel codes should be sought in overcoming problems in areas other than that of resistance to additive Gaussian noise.

By way of example, we mention the technical problems of a more deterministic nature involving the pilot tone generation for servo systems, specific media properties such as overwrite noise, nonlinear effects, cross-talk from adjacent tracks, the complexity of the detector, etc. Other problems have to do with particulate noise from the magnetic material and this noise is not white and may even be nonstationary, or signal time-base jitter due to tape-head contact dynamics. All these problems mentioned may have a more dramatic influence on the recorder performance than additive noise, or in summary, the results presented here may have limited usefulness in high density, high-speed recording systems.

VIII. CONCLUSIONS

We have considered the maximum-likelihood detection of various off-the-shelf recording codes employed on an idealized digital magnetic recording channel. When resistance to additive noise is of concern, only little benefit may be expected from the use of dc-free codes at medium and high information densities. It has been found that codes based on runlength-limited sequences exhibit a coding loss in the entire information density span examined. Rate = $1/2$, trellis codes designed for the $(1 - D)$ partial-response channel perform fairly well at the lower side of the information density range, but their performance rapidly deteriorates when the information density is pushed to the higher values more commonly employed in current systems. Our findings can easily be summarized: all the codes investigated show a substantial coding loss at the information densities which are relevant for practical implementation.

REFERENCES

- [1] H. Burkhardt, "An event-driven maximum-likelihood peak position detector for run-length-limited codes in magnetic recording," *IEEE Trans. Magn.*, vol. MAG-17, pp. 3337-3339, Nov. 1981.
- [2] C. H. Coleman, D. A. Lindholm, D. A. Petersen, and R. Wood, "High data rate magnetic recording in a single channel," *J. Inst. Electron. Radio Eng.*, vol. 55, pp. 229-236, June 1985.
- [3] R. W. Wood, "Viterbi reception of Miller-squared code on a tape channel," in *Proc. 4th Int. Conf. Video Data Record., IERE Conf. Proc. 54*, Southampton, Apr. 1982, pp. 333-343.
- [4] R. W. Wood and D. A. Petersen, "Viterbi detection of class IV partial response on a magnetic recording channel," *IEEE Trans. Commun.*, vol. COM-34, pp. 454-461, May 1986.
- [5] K. A. S. Immink, "Coding methods for high-density optical recording," *Phil. J. Res.*, vol. 41, pp. 410-430, 1986.
- [6] K. W. Cattermole, *Principles of Pulse Code Modulation*. London, England: Iliffe Books Ltd, 1969.
- [7] P. H. Siegel, "Recording codes for digital magnetic storage," *IEEE Trans. Magn.*, vol. MAG-21, pp. 1344-1349, Sept. 1985.
- [8] K. Nakano, H. Moriwaki, T. Takahashi, K. Akagiri, and M. Morio, "A new 8-bit PCM audio recording technique using an extension of the video track," *IEEE Trans. Consum. Elect.*, vol. CE-28, pp. 241-248, Aug. 1982.
- [9] H. Nakajima and M. Kosaka, "The DAT conference: Its activities and results," *IEEE Trans. Consum. Elect.*, vol. CE-32, pp. 404-415, Aug. 1986.
- [10] M. Morizono, H. Yoshida and Y. Hashimoto, "Digital video recording—Some experiments and future considerations," *SMPTE J.*, vol. 89, pp. 658-662, 1980.
- [11] M. A. Parker and F. A. Bellis, "Signal processing for an experimental 216 Mbit/s digital video tape recorder," in *Proc. 4th Int. Conf. Video Data Record., IERE Conf. Proc. 54*, Southampton, 1982, pp. 207-215.
- [12] J. S. Eggenberger and P. Hodges, "Sequential encoding and decoding of variable length, fixed rate data codes," U.S. Pat. 4 115 768, 1978.
- [13] G. Bouwhuis, J. Braat, A. Huijser, J. Pasman, G. van Rosmalen, and K. Schouhamer Immink, *Principles of Optical Disc Systems*. Boston, MA: Adam Hilger Ltd., 1985.
- [14] J. P. J. Heemskerck and K. A. S. Immink, "Compact disc: System aspects and modulation," *Phil. Tech. Rev.*, vol. 40, no. 6, pp. 157-164, 1982.

- [15] J. K. Wolf and G. Ungerboeck, "Trellis coding for partial-response channels," *IEEE Trans. Commun.*, vol. COM-34, pp. 765-773, Aug. 1986.
- [16] A. R. Calderbank, C. Heegard, and T. A. Lee, "Binary convolutional codes with application to magnetic recording," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 797-815, Nov. 1986.
- [17] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363-378, May 1972.
- [18] C. D. Mee and E. D. Daniel, *Magnetic Recording, Volume I, Technology*. New York: McGraw-Hill, 1987.
- [19] G. C. Clark and J. B. Cain, *Error-Correction Coding for Digital Communications*. New York: Plenum, 1981.
- [20] D. T. Tang and L. R. Bahl, "Block codes for a class of constrained noiseless channels," *Inform. Contr.*, vol. 17, pp. 436-461, 1970.
- [21] H. Kobayashi, "A survey of coding schemes for transmission or recording of digital data," *IEEE Trans. Commun.*, vol. COM-19, pp. 1087-1099, Dec. 1971.
- [22] K. A. S. Immink, "Properties and constructions of binary channel codes," Ph.D. dissertation, Eindhoven Univ. Technol., May 1985.
- [23] M. G. Pelchat and J. M. Geist, "Surprising properties of two-level bandwidth compaction codes," *IEEE Trans. Commun.*, vol. COM-23, pp. 878-883, Sept. 1975; Correction COM-24, pp. 479, Apr. 1976.
- [24] P. A. Franzaszek, "Sequence-state encoding for digital transmission," *Bell Syst. Tech. J.*, vol. 47, pp. 143-157, Jan. 1968.
- [25] H. Osawa, S. Tazaki, and S. Ando, "Performance analysis of partial-

response systems for non-return-to-zero recording," *IEEE Trans. Magn.*, vol. MAG-22, pp. 253-258, July 1986.



Kees A. Schouhamer Immink (SM'86) was born in Rotterdam, The Netherlands, in 1946. He received the B.S. degree from the Rotterdam Polytechnic in 1967 and the M.S. and Ph.D. degrees from the Eindhoven University of Technology in 1974 and 1985, respectively, all in electrical engineering.

He joined the Philips Research Laboratories, Eindhoven, in 1968, where his work involved the signal processing side of optical recording systems.

He was, among others, responsible for the design and development of channel coding techniques for the Compact Disc, Compact Disc Video and experimental erasable optical audio discs. Since 1986, he is a Senior Scientist of the magnetic recording group at the Philips Research.

Dr. Immink has written numerous papers in the field of coding techniques for optical and magnetic recording, and is coauthor of a book entitled *Principles of Optical Disc Systems*. He was awarded a Fellowship by the Audio Engineering Society in 1985 for his work in the area of optical laser disc and for detailed study on channel codes for the compact disc. He holds more than thirty patents, mainly in optical recording. He also is a Fellow of the IEE.