Correlation-Sensitive Adaptive Sequence Detection

Aleksandar Kavčić   and   José M. F. Moura

Data Storage Systems Center

Department of Electrical and Computer Engineering

Carnegie Mellon University

5000 Forbes Avenue

Pittsburgh, PA 15213

Abstract

In high density magnetic recording, noise samples corresponding to adjacent signal samples are heavily correlated as a result of front-end equalizers, media noise, and signal nonlinearities combined with nonlinear filters to cancel them. This correlation significantly deteriorates the performance of detectors at high densities. In this paper, we propose a novel sequence detector that is correlation-sensitive and adaptive to the nonstationary signal sample statistics. We derive the correlation-sensitive maximum likelihood detector. It can be used with any Viterbi-like receiver (e.g., PRML, FDTS, MDFE) that relies on a tree/trellis structure. Our detector adjusts the metric computation to the noise correlation statistics. Because these statistics are nonstationary, we develop an adaptive algorithm that tracks the data correlation matrices. Simulation results are presented that show the applicability of the new correlation-sensitive adaptive sequence detector.

1 Introduction

In recent years, there has been a major shift in the design of signal detectors in magnetic recording. The traditional peak detectors (PD) [1] have been replaced by Viterbi-like detectors [2] in the form of partial response maximum likelihood (PRML) schemes [3, 4] or hybrids between tree/trellis detectors and decision feedback equalizers (DFE), like FDTS/DF [5], MDFE [6, 7] and RAM-RSE [8]. These methods were derived under the assumption of additive white Gaussian noise (AWGN) in the system. The resulting trellis/tree branch metrics are then computed as Euclidian distances.

It has long been observed that the noise in magnetic recording systems is neither white nor stationary [9, 10, 11, 12]. The nonstationarity of the media noise results from its signal dependent nature [13, 14, 15]. Combating media noise and its signal dependence has thus far been confined to modifying the Euclidian branch metric to account for these effects. Zeng and Moon [16], and Lee and Cioffi [17] have derived a branch metric computation method for combating the signal-dependent character of media noise. These authors ignore the correlation between noise samples. The effectiveness of this method has been demonstrated on real data in [18].

These methods do not take into consideration the correlation between noise samples in the readback signal. These correlations arise due to noise coloring by front-end equalizers, media noise, media nonlinearities, and magnetoresistive (MR) head nonlinearities. They cause significant performance degradation at high recording densities. Early work addressing noise correlation has been reported by Barbosa in [19]. We propose in this paper an *adaptive correlation-sensitive* maximum likelihood sequence detector. We derive the maximum likelihood sequence detector (MLSD) [20] without making the usual simplifying assumption that the noise samples are independent random variables. The trellis/tree branch metric computation becomes correlation-sensitive, being both signal-dependent and sensitive to correlations between noise samples. We refer to our method as the correlation-sensitive maximum likelihood sequence detector (CS-MLSD), or simply correlation-sensitive sequence detector (CS-SD).

Because the noise statistics are non-stationary, we adaptively compute the noise sensitive branch metrics by estimating the noise covariance matrices from the read-back data. These covariance matrices are different for each branch of the tree/trellis due to the signal dependent structure of the media noise. Since the channel characteristics in magnetic recording vary from track-to-track, these matrices are tracked (updated) on-the-fly, recursively using past samples and previously made detector decisions.

The block diagram of the newly proposed detector strategy is shown in Figure 1. Notice that the detector has a feedback structure. The first block is a Viterbi-like detector. The outputs of this block are the decisions and the delayed signal samples. They are used by the two feedback blocks. The first feedback block, labeled 'noise statistics tracker,' uses the past samples and detector decisions to update the noise statistics, i.e., to update the noise covariance matrices. The second block, labeled 'metric computation update,' uses the updated statistics to calculate the branch metrics needed in the Viterbi-like algorithm. Our new algorithm does not require replacing current detectors. It simply adds two new blocks in the feedback loop to adaptively estimate the branch metrics used in the Viterbi-like decoder.

The paper is organized as follows. Section II derives the general maximum likelihood correlation-sensitive sequence detection method and the corresponding branch metrics. In Section III, we propose an adaptive filter for tracking the statistics needed for the metric computation. Simulation performance results using data generated with a realistic stochastic zig-zag transition model are presented in Section IV. Section V concludes the paper.

2   Correlation-sensitive MLSD

In this section, we derive the correlation-sensitive maximum likelihood sequence detector (CS-MLSD). Assume that $N > 1$ channel bits (symbols), $a_1, a_2, \ldots, a_N$, are written on a magnetic medium. The symbols a_i, $i = 1, \ldots, N$, are drawn from an alphabet of four symbols, $a_i \in \{+, \oplus, -, \ominus\}$. The symbols '+' and '−' denote a positive and a negative transition, respectively.

The symbol '$\oplus$' denotes a written zero (no transition) whose nearest preceding non-zero symbol is a '$+$', while '$\ominus$' denotes a written zero whose nearest preceding transition is a negative one, i.e., '$-$'. We introduce this notation because a simple treatment of transitions as '1's and no transitions as '0's is blind to signal asymmetries (MR head asymmetries and base line drifts), which is inappropriate for our problem. In Figure 2 a sample waveform is illustrated. We exaggerate in Figure 2 the signal asymmetries and base line shifts to make our point clearer. The figure also shows the written symbols $a_1, \ldots, a_{18}$, as well as the samples $r_1, \ldots, r_{18}$ of the read-back waveform, sampled at the rate of one sample per symbol interval.

When the written sequence of symbols a_i, $i = 1, \ldots, N$, is read, the readback waveform is passed through a pulse-shaping equalizer and sampled one sample per symbol, resulting in the sequence of samples r_i, $i = 1, \ldots, N$. Due to the noise in the system, the samples r_i are realizations of random variables. The maximum likelihood detector determines the sequence of symbols a_i that has been written, by maximizing the likelihood function, i.e.,

$$\{\hat{a}_1, \ldots, \hat{a}_N\} = \arg \left[\max_{\text{all } a_i} f(r_1, \ldots, r_N | a_1, \ldots, a_N) \right]. \tag{1}$$

In (1), the likelihood function $f(r_1, \ldots, r_N | a_1, \ldots, a_N)$ is the joint probability density function (pdf) of the signal samples $r_1, \ldots, r_N$, conditioned on the written symbols $a_1, \ldots, a_N$. The maximization in (1) is done over all possible combinations of symbols in the sequence $\{a_1, \ldots, a_N\}$.

Due to the signal dependent nature of media noise in magnetic recording, the functional form of the joint conditional pdf $f(r_1, \ldots, r_N | a_1, \ldots, a_N)$ in (1) is different for different symbol sequences $a_1, \ldots, a_N$. Rather than making this distinction with more complex but cluttered notation, we keep the notation to a minimum by using simply the same symbol f to denote these different functions. We will frequently point out this distinction in our discussions throughout the text.

2.1   General Structure of CS-MLSD

By Bayes rule, the joint conditional pdf (likelihood function) is factored into a product of conditional pdfs

$$f(r_1, \ldots, r_N | a_i, \ldots, a_N) = \prod_{i=1}^{N} f(r_i | r_{i+1}, \ldots, r_N, a_1, \ldots, a_N). \tag{2}$$

To proceed and obtain more concrete results, we now exploit the nature of the noise and of the intersymbol interference in magnetic recording.

Finite correlation length. We start by assuming that the conditional pdfs in Equation (2) are independent of future samples after some length $L \geq 0$. We call L the correlation length of the noise. This independence leads to

$$f(r_i | r_{i+1}, \ldots, r_N, a_1, \ldots, a_N) = f(r_i | r_{i+1}, \ldots, r_{i+L}, a_1, \ldots, a_N). \tag{3}$$

Finite intersymbol interference. We further assume that the conditional pdf is independent of symbols that are *not* in the K-neighborhood of $r_i, \ldots, r_{i+L}$. The value of $K \geq 1$ is determined by the length of the intersymbol interference (ISI). For example, for PR4 we have $K = 2$, while for EPR4 $K = 3$. Define $K_l \geq 0$ as the length of the leading (anticausal) ISI and $K_t \geq 0$ as the length of the trailing (causal) ISI, such that $K = K_l + K_t + 1$. With this notation, we can write the conditional pdf in (3) as

$$f(r_i | r_{i+1}, \ldots, r_{i+L}, a_1, \ldots, a_N) = f(r_i | r_{i+1}, \ldots, r_{i+L}, a_{i-K_l}, \ldots, a_{i+L+K_t}). \tag{4}$$

Substituting (4) into (2) and applying the Bayes rule, we obtain the factored form of the likelihood function (conditional pdf)

$$f(r_1, \ldots, r_N | a_1, \ldots, a_N) = \prod_{i=1}^{N} f(r_i | r_{i+1}, \ldots, r_N, a_1, \ldots, a_N)$$

$$= \prod_{i=1}^{N} \frac{f(r_i, r_{i+1}, \ldots, r_{i+L} | a_{i-K_l}, \ldots, a_{i+L+K_t})}{f(r_{i+1}, \ldots, r_{i+L} | a_{i-K_l}, \ldots, a_{i+L+K_t})}. \tag{5}$$

The factored form of equation (5) is suitable for applying Viterbi-like dynamic programming detection techniques. Notice that equation (5) assumes anticausal factorization, i.e., it is derived by taking into account the effect of the samples $r_{i+1}, \ldots, r_{i+L}$ on r_i. If we take into account only the causal effects, we can derive the causal equivalent of (5) as $f(r_1, \ldots, r_N | a_1, \ldots, a_N) = \prod_{i=1}^{N} \frac{f(r_i, r_{i+1}, \ldots, r_{i+L} | a_{i-K_l}, \ldots, a_{i+L+K_t})}{f(r_i, \ldots, r_{i+L-1} | a_{i-K_l}, \ldots, a_{i+L+K_t})}$. In [21], we consider the causal formulation of the CS-MLSD. We could combine the causal and anticausal factorization and find the geometric mean of the two to form a causal-anticausal factorization. Since this only complicates derivations and does not provide further insight, we will work only with the anticausal Equation (5).

Maximizing the likelihood function in (5) is equivalent to minimizing its negative logarithm. Thus, the maximum likelihood detector is now

$$
\begin{aligned}
\{\hat{a}_1, \ldots, \hat{a}_N\} &= \arg \left[\min_{\text{all } a_i} \log \prod_{i=1}^{N} \frac{f(r_{i+1}, \ldots, r_{i+L} | a_{i-K_l}, \ldots, a_{i+L+K_t})}{f(r_i, r_{i+1}, \ldots, r_{i+L} | a_{i-K_l}, \ldots, a_{i+L+K_t})} \right] \\
&= \arg \left[\min_{\text{all } a_i} \sum_{i=1}^{N} \log \frac{f(r_{i+1}, \ldots, r_{i+L} | a_{i-K_l}, \ldots, a_{i+L+K_t})}{f(r_i, r_{i+1}, \ldots, r_{i+L} | a_{i-K_l}, \ldots, a_{i+L+K_t})} \right] \\
&= \arg \left[\min_{\text{all } a_i} \sum_{i=1}^{N} \mathcal{M}_i(r_i, r_{i+1}, \ldots, r_{i+L}, a_{i-K_l}, \ldots, a_{i+L+K_t}) \right]. \tag{6}
\end{aligned}
$$

The symbol $\mathcal{M}_i$ represents the branch metric of the trellis/tree in the Viterbi-like algorithm. The metric is a function of the observed samples $r_i, r_{i+1}, \ldots, r_{i+L}$. It is also dependent on the postulated sequence of written symbols $a_{i-K_l}, \ldots, a_{i+L+K_t}$, which ensures the signal-dependence of the detector. As a consequence, the branch metrics for every branch in the tree/trellis is based on its corresponding signal/noise statistics.

2.2   Specific Branch Metrics for the CS-MLSD

We next consider specific expressions for the branch metrics that result under different assumptions on the noise statistics.

Euclidian branch metric. In the simplest case, the noise samples are realizations of independent identically distributed Gaussian random variables with zero mean and variance σ^2. This is a white Gaussian noise assumption. This implies that the correlation distance is $L = 0$ and that the noise pdfs have the same form for all noise samples. The total ISI length is assumed to be $K = K_l + K_t + 1$, where K_l and K_t are the leading and trailing ISI lengths, respectively. The conditional signal pdfs are factored as

$$\frac{f(r_{i+1}, \ldots, r_{i+L} | a_{i-K_l}, \ldots, a_{i+L+K_t})}{f(r_i, r_{i+1}, \ldots, r_{i+L} | a_{i-K_l}, \ldots, a_{i+L+K_t})} = \sqrt{2\pi\sigma^2} \exp\left[\frac{(r_i - m_i)^2}{2\sigma^2}\right]. \tag{7}$$

Here the mean signal m_i is dependent on the written sequence of symbols. For example, for a PR4 channel, $m_i \in \{-1, 0, 1\}$, for details see, e.g., [3, 4]. The branch/tree metric is then the conventional Euclidian distance metric

$$\mathcal{M}_i = N_i^2 = (r_i - m_i)^2. \tag{8}$$

Variance dependent branch metric. We assume again that the noise samples are samples of independent Gaussian variables, but that their variance depends on the written sequence of symbols. The noise correlation length is still $L = 0$, but the variance of the noise samples is no longer constant for all samples. The variance is σ_i^2, where the index i denotes the dependence on the written symbol sequence. As for the Euclidian metric, we assume the total ISI length to be $K = K_l + K_t + 1$. The conditional signal pdf is factored to give

$$\frac{f(r_{i+1}, \ldots, r_{i+L} | a_{i-K_l}, \ldots, a_{i+L+K_t})}{f(r_i, r_{i+1}, \ldots, r_{i+L} | a_{i-K_l}, \ldots, a_{i+L+K_t})} = \sqrt{2\pi\sigma_i^2} \exp\left[\frac{(r_i - m_i)^2}{2\sigma_i^2}\right]. \tag{9}$$

The corresponding branch metric is

$$\mathcal{M}_i = \log\sigma_i^2 + \frac{N_i^2}{\sigma_i^2} = \log\sigma_i^2 + \frac{(r_i - m_i)^2}{\sigma_i^2}, \tag{10}$$

which is the metric presented in [16, 17].

Correlation-sensitive branch metric. In the most general case, the correlation length is $L > 0$. The leading and trailing ISI lengths are K_l and K_t, respectively. The noise is now considered to be both correlated and signal-dependent. We assume joint Gaussian noise pdfs. This assumption is well justified in magnetic recording since the experimental evidence shows that the dominant media noise modes have Gaussian-like histograms [22]. The conditional pdfs do not factor out in this general case, so we have the general form for the pdf

$$\frac{f(r_{i+1}, \ldots, r_{i+L} | a_{i-K_l}, \ldots, a_{i+L+K_t})}{f(r_i, r_{i+1}, \ldots, r_{i+L} | a_{i-K_l}, \ldots, a_{i+L+K_t})} = \sqrt{\frac{(2\pi)^{L+1} \det \mathbf{C}_i}{(2\pi)^L \det \mathbf{c}_i}} \frac{\exp\left[\underline{N}_i^{\mathrm{T}} \mathbf{C}_i^{-1} \underline{N}_i\right]}{\exp\left[\underline{n}_i^{\mathrm{T}} \mathbf{c}_i^{-1} \underline{n}_i\right]}. \tag{11}$$

The $(L+1) \times (L+1)$ matrix $\mathbf{C}_i$ is the covariance matrix of the data samples $r_i, r_{i+1}, \ldots, r_{i+L}$, when a sequence of symbols $a_{i-K_l}, \ldots, a_{i+L+K_t}$ is written. The matrix $\mathbf{c}_i$ in the denominator of (11) is the $L \times L$ lower principal submatrix of $\mathbf{C}_i = \begin{bmatrix} \cdot & \cdot \\ \cdot & \mathbf{c}_i \end{bmatrix}$. The $(L+1)$-dimensional vector $\underline{N}_i$ is the vector of differences between the observed samples and their expected values when the sequence of symbols $a_{i-K_l}, \ldots, a_{i+L+K_t}$ is written, i.e.,

$$\underline{N}_i = \begin{bmatrix} (r_i - m_i) & (r_{i+1} - m_{i+1}) & \cdots & (r_{i+L} - m_{i+L}) \end{bmatrix}^{\mathrm{T}}. \tag{12}$$

The vector $\underline{n}_i$ collects the last L elements of $\underline{N}_i$, $\underline{n}_i = \begin{bmatrix} (r_{i+1} - m_{i+1}) & \cdots & (r_{i+L} - m_{i+L}) \end{bmatrix}^{\mathrm{T}}$. With this notation, the general correlation-sensitive metric is

$$\mathcal{M}_i = \log \frac{\det \mathbf{C}_i}{\det \mathbf{c}_i} + \underline{N}_i^{\mathrm{T}} \mathbf{C}_i^{-1} \underline{N}_i - \underline{n}_i^{\mathrm{T}} \mathbf{c}_i^{-1} \underline{n}_i. \tag{13}$$

The computation of the metric written as in (13) requires roughly $4L(L+1)$ multiplications. In our recent work [21], we show that the metric in (13) can be implemented using an L-tap finite impulse response (FIR) filter, bringing the computational complexity down to only L multiplications. The well-behaved impulse response of FIR filters and the computational complexity only linearly proportional to L make this metric attractive for practical implementations even for fairly large L.

$\square$

In the derivations of the branch metrics (8), (10) and (13), we made no assumptions on

the exact Viterbi-type architecture, that is, the metrics can be applied to any Viterbi-type algorithm such as PRML [3, 4], FDTS/DF [5], RAM-RSE [8] or, MDFE [6, 7]. In this paper, in Section IV, we show results on how the chosen metric affects the performance of several class-4 partial response (PR4 and EPR4) detectors. Before considering that, we address the question of the on-the-fly estimation of the covariance matrices C_i that are needed in the branch metric computation in (13).

3   Adaptive statistics tracking

Computing the branch metrics in (10) or (13) requires knowledge of the signal statistics. These statistics are the mean signal values m_i in (12) as well as the covariance matrices C_i in (13). In magnetic recording systems, these statistics will generally vary from track to track. For example, the statistics that apply to a track at a certain radius will differ from those from another track at a different radius due to different linear track velocities at those radii. Also, the signal and noise statistics will be different if a head is flying slightly off-track or if it is flying directly over the track. The head skew angle is another factor that contributes to different statistics from track to track. These factors suggest that the system that implements the metric in (13) needs to be flexible to these changes. Storing the statistics for each track separately is out of the question because of the memory span required to accomplish this. A reasonable alternative is to use adaptive filtering techniques to track the needed statistics.

Tracking the mean signal values m_i is generally done so that these values fall on prespecified targets. An adaptive front-end equalizer is employed to force the signal sample values to their targets. This is certainly the case with partial response targets used in algorithms like PR4, EPR4, or EEPR4 where the target is prespecified to one of the class-4 partial responses. For example, in a PR4 system, the signal samples, if there is no noise in the system, fall on one of the three target values 1, 0, or -1. Typically this is done with an LMS-class (least mean-squares) algorithm that ensures that the mean of the signal samples is close to these target values. In decision feedback equalization (DFE) based detectors or hybrids between fixed delay

tree search and DFE, such as FDTS/DF [5] or MDFE [6, 7], the target response need not be prespecified. Instead, the target values are chosen on-the-fly by simultaneously updating the coefficients of the front-end and feed-back equalizers with an LMS-type algorithm, for details see [23].

When there are severe nonlinearities in the system (also referred to as nonlinear distortion or nonlinear ISI) a linear equalizer will generally not be able to place the signal samples right on target. Instead, the means of the signal samples will fall at a different value. For example, in a PR4 system, the response to a sequence of written symbols $\ldots, \ominus, +, \oplus, \ldots$ might result in mean sample target values $\ldots, 0, 1, 0.9, \ldots$, while a sequence of written symbols $\ldots, +, -, \ominus, \ldots$ might result in a sequence of mean sample values $\ldots, 0.95, -1.05, 0, \ldots$. Clearly, in this example, what should be a target value of 1 becomes either 1, 0.9, or 0.95 depending on the written sequence. Since we are talking about mean values and not about noisy samples, this deviation is due to nonlinearities in the system. There are two fixes for this problem. The first is to employ a nonlinear filter (neural network or Volterra series filter) that is capable of overcoming these nonlinear distortions, see [24, 25]. Although recently very popular, such a method introduces further correlation between noise samples due to the nonlinear character of the filter. The second fix is to track the nonlinearities in a feedback loop and use the tracked value in the metric computation. For example, let the response to a written symbol sequence $\ldots, \ominus, +, \oplus, \ldots$ be consistently $\ldots, 0, 1, 0.9, \ldots$. Then, rather than using the value 1 in the metric computation for the third target, we can track this behavior and use the value $m_i = 0.9$.

In the remainder of this paper, for simplicity, we assume that the front-end equalizer is placing the signal samples right on the desired target values and that there is no need for further mean corrections. We shift our focus to tracking the noise covariance matrices needed in the computation of the branch metrics (13).

Assume that the sequence of samples $r_i, r_{i+1}, \ldots, r_{i+L}$ is observed. Based on these as well as all other neighboring samples, after an appropriate delay of the Viterbi trellis, a decision is made that the most likely estimate for the sequence of symbols $a_{i-K_l}, \ldots, a_{i+L+K_t}$ is $\hat{a}_{i-K_l}, \ldots, \hat{a}_{i+L+K_t}$. Here L is the noise correlation length and $K = K_l + K_t + 1$ is the ISI

length. Let the current estimate for the $(L+1) \times (L+1)$ covariance matrix corresponding
to the sequence of symbols $\hat{a}_{i-K_l}, \ldots, \hat{a}_{i+L+K_t}$ be $\hat{\mathbf{C}}(\hat{a}_{i-K_l}, \ldots, \hat{a}_{i+L+K_t})$. We abbreviate this
symbol with the shorter notation, $\hat{\mathbf{C}}(\hat{a})$. If the estimate is unbiased, the expected value of the
estimate is

$$\mathrm{E}\hat{\mathbf{C}}(\hat{a}) = \mathrm{E}\left[\underline{N}_i\,\underline{N}_i^{\mathrm{T}}\right], \tag{14}$$

where $\underline{N}_i$ is the vector of differences between the observed samples and their expected values,
as defined in (12).

Note that once we have observed the samples $r_i, r_{i+1}, \ldots, r_{i+L}$, and once we have decided
that most likely they resulted from a series of written symbols $\hat{a}_{i-K_l}, \ldots, \hat{a}_{i+L+K_t}$, we know the
sequence of target (mean) values $m_i, m_{i+1}, \ldots, m_{i+L}$ that correspond to these samples. We use
them to compute the vector $\underline{N}_i$, with which we form the empirical rank-one covariance matrix
$\underline{N}_i\,\underline{N}_i^{\mathrm{T}}$.

In the absence of prior information, this rank-one matrix is our estimate for the covariance
matrix for the detected symbols. In a recursive adaptive scheme as we are proposing, we
use this rank-one data covariance estimate to update our current estimate of the covariance
matrix $\hat{\mathbf{C}}(\hat{a})$. A simple way to achieve this is provided by the recursive least-squares (RLS)
algorithm [26]. The RLS computes the next covariance matrix estimate $\hat{\mathbf{C}}'(\hat{a})$ as

$$\hat{\mathbf{C}}'(\hat{a}) = \beta(t)\hat{\mathbf{C}}(\hat{a}) + [1 - \beta(t)]\,\underline{N}_i\,\underline{N}_i^{\mathrm{T}}. \tag{15}$$

Here, $\beta(t)$, $0 < \beta(t) < 1$, is a forgetting factor. The dependence on t signifies that β is a
function of time. Equation (15) can be viewed as a weighted averaging algorithm, where the
data sample covariance $\underline{N}_i\,\underline{N}_i^{\mathrm{T}}$ is weighted by the factor $[1 - \beta(t)]$, while the previous estimate
is weighted by $\beta(t)$. The choice of $\beta(t)$ should reflect the nonstationarity degree of the noise.
For example, if the nonstationarity is small, $\beta(t)$ should be close to 1, while it should drop as
the nonstationarity level increases. The forgetting factor is typically taken time-dependent to
account for the start-up conditions of the RLS algorithm in (15). As more data is processed, a
steady-state is expected to be achieved and $\beta(t)$ is made to approach a constant value. If we

lack a good prior estimate $\hat{\mathbf{C}}(\hat{a})$, $\beta(t)$ should be close to zero, to weigh more the current data estimate. With time, $\beta(t)$ will increase and settle around a value close to 1. In our simulations in Section V, we chose a constant forgetting factor β, independent of time. In particular, we use $\beta = 0.95$.

Although our simulations have shown (15) to converge even when we pick the zero matrix as the initial condition, in practice, we want to have a good starting condition to obtain fast convergence. By dividing the disk in a number of sectors and picking the middle track as the representative track of the sector, the covariance matrix $\hat{\mathbf{C}}(\hat{a})$ for this track provides a good starting condition for all tracks in that sector. Convergence can be expected then to take a few hundred samples (bits).

The one-dimensional equivalent of equation (15) is

$$\hat{\sigma}^2_{\text{new}} = \beta\hat{\sigma}^2_{\text{old}} + [1 - \beta]\,N_i^2. \tag{16}$$

This equation can be used in conjunction with the metric in (10).

It is important to point out that, due to the signal-dependent character of the media noise, there will be a different covariance matrix to track for each branch in the tree/trellis of the Viterbi-like detector. Practical considerations of memory requirements, however, limit the dimensions of the matrices to be tracked. Fortunately, as we show in the next section, simple 2×2 matrices are enough to show substantial improvement in error rate performance.

Example. The following example illustrates how the algorithm in (15) works. Assume a PR4 target response with a simple trellis structure as shown in Figure 3. Notice that for PR4, the symbols can be equated to the trellis states, as is illustrated in Figure 3. The number next to each branch in Figure 3 represents the target value (mean sample value) for the corresponding path between states. The target values in PR4 can be one of three values -1, 0, or 1.

In this example we assume a noise correlation length of $L = 1$. We also assume that the leading and trailing ISI lengths are $K_l = 0$ and $K_t = 1$, respectively, to give the total ISI length

$K = K_l + K_t + 1 = 2$ for the PR4 response. Since $L = 1$, we will need to track signal covariance matrices of size $(L + 1) \times (L + 1) = 2 \times 2$. The number of these matrices equals the number of different combinations of two consecutive branches in the trellis. A simple count in Figure 3 reveals that this number is 16, since there are 4 nodes in the trellis and 2 branches entering and leaving each node.

Assume that, using the branch metric in (13), the Viterbi-like detector decides that the most likely written symbols a_i, a_{i+1}, a_{i+2}, equal $\{\hat{a}_i, \hat{a}_{i+1}, \hat{a}_{i+2}\} = \{\ominus, +, -\}$. This is illustrated in Figure 4, where the corresponding path through the trellis is highlighted. The noisy signal samples corresponding to the trellis branches are $r_i = 0.9$ and $r_{i+1} = -0.2$, which deviate slightly from their ideal partial response target values of 1 and 0, respectively.

Suppose that, prior to making the decision $\{\hat{a}_i, \hat{a}_{i+1}, \hat{a}_{i+2}\} = \{\ominus, +, -\}$, the estimate for the covariance matrix associated with this sequence of three symbols is

$$\hat{\mathbf{C}}(\ominus, +, -) = \begin{bmatrix} 0.5 & -0.2 \\ -0.2 & 0.8 \end{bmatrix}. \tag{17}$$

Let the forgetting factor be $\beta = 0.95$. To update the covariance matrix we first form the vector

$$\underline{N} = \begin{bmatrix} (r_i - 1) & (r_{i+1} - 0) \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} -0.1 & -0.2 \end{bmatrix}^{\mathrm{T}}. \tag{18}$$

We use the rank-one sample covariance matrix $\underline{N}\,\underline{N}^{\mathrm{T}}$ to find the covariance matrix update

$$\begin{aligned} \hat{\mathbf{C}}'(\ominus, +, -) &= \beta \hat{C}(\ominus, +, -) + (1 - \beta)\underline{N}\,\underline{N}^{\mathrm{T}} \\ &= \begin{bmatrix} 0.4755 & -0.189 \\ -0.189 & 0.7620 \end{bmatrix}. \end{aligned} \tag{19}$$

The matrix $\hat{\mathbf{C}}'(\ominus, +, -)$ becomes our estimate for the covariance matrix corresponding to this particular symbol sequence (trellis path) and we use it to compute the metrics (13) in the subsequent steps of the Viterbi-like algorithm.

4 Simulation results

In this section, we present simulation results using two partial response detection algorithms, namely PR4 and EPR4. To create realistic waveforms, corrupted by media noise, we used an efficient stochastic zig-zag model, the TZ-ZT model [15, 22]. These waveforms are then passed through the detectors. A Lindholm inductive head [27] is used for both writing and reading. Table 1 presents the recording parameters of the model. These recording parameters are chosen so that with a moderately low symbol density per $PW50$, we have a low number of transition widths a per symbol transition separation. Namely, at 3 symbols/$PW50$ we have a transition separation of only $2.9a$. We modeled the transition profile by an error function, where the transition width a denotes the distance from the transition center to the point where the magnetization equals $M_r/2$.

We write the symbols utilizing the (0,4) run length limited code. No error correction is applied, so the obtained error rates are not bit error rates, but (raw) *symbol* error rates.

We tested both the PR4 and EPR4 detectors using the following three different metric computation methods: the Euclidian metric (8), the variance dependent metric (10), also referred to as the C1 metric, and the 2×2 correlation sensitive metric (13), named the C2 metric for short. For a PR4 target response, the total ISI length is $K = K_l + K_t + 1 = 2$, where the leading and trailing ISI lengths are $K_l = 0$ and $K_t = 1$, respectively. The noise correlation length for the Euclidian and the C1 metrics is $L = 0$, while for the C2 metric the noise correlation length is $L = 1$. We refer to these three PR4 detectors as PR4(Euc), PR4(C1), and PR4(C2).

Similarly to the PR4 detectors, we tested three EPR4 detectors, EPR4(Euc), EPR4(C1) and EPR4(C2). The only difference between the PR4 detectors and the EPR4 detectors are the target response and the ISI length, which for the EPR4 target response equals $K = K_l + K_t + 1 = 3$, with $K_l = 1$ and $K_t = 1$.

The signal obtained by the TZ-ZT model is already corrupted with media noise. To this

signal we add white Gaussian noise to simulate the head and electronics noise in a real system. We quote the power of the additive white Gaussian noise as the signal to additive white Gaussian noise ratio, S(AWG)NR, which we obtain as

$$S(AWG)NR = 10 \log \frac{A_{iso}^2}{\sigma_n^2},$$ (20)

where A_{iso} is the mean (media noise free) amplitude of an isolated pulse and σ_n^2 is the variance of the additive white Gaussian noise. The noise distorted signal is first passed through a low-pass filter to clean out the noise outside the Nyquist band. The signal is then sampled at a rate of one sample per symbol and subsequently passed through a partial response shaping filter, either PR4 or EPR4. The partial response shaping filter is implemented as an adaptive FIR filter whose tap weights are adjusted using the LMS algorithm. Note that both filters add correlation to the noise. For the C1 and C2 metrics in (10) and (13), the RLS algorithms (15) and (16) are used to estimate the noise variances and covariance matrices for the branch metric computations. In both cases, the forgetting factor is set to $\beta = 0.95$.

All six detection algorithms were tested at three different recording densities.

Symbol separation of 4.4a. This recording density corresponds to a symbol density of 2 symbols/$PW50$, see Table 1.

Figure 5 shows the symbol error rate performance of the PR4 detectors for different additive noise SNRs. Keep in mind that the media noise is embedded in the system, which is why the x-axis on the graph is labeled as S(AWG)NR instead of simply SNR. Notice that, at this density, the PR4(Euc) and PR4(C1) detectors perform just about the same while the PR4(C2) detector outperforms them both by about 3dB. The reason for this is that the PR4 shaping filter averages noise samples from different symbols, which masks the signal dependent nature of the media noise. This is why there is not much to gain by using PR4(C1) over PR4(Euc). The PR4(C2) detector performs better because it partially removes the effects of noise correlation introduced by the PR4 shaping filter.

Figure 6 shows how the EPR4 detectors perform at this same density (symbol separation

4.4a). Here we see that the EPR4(C2) has the best performance, while EPR4(Euc) has the worst. The difference in performance at the error rate of 10^{-5} is only about 0.5dB between EPR4(Euc) and EPR4(C2). This is because the media noise power at this density is low and the signal is well matched to the target so the EPR4 shaping filter does not introduce unnecessary noise correlation.

Symbol separation of 3.5a. This recording density corresponds to a symbol density of 2.5 symbols/$PW50$.

Figure 7 shows the performance of the PR4 detectors at this density. The figure is very similar to Figure 5, except that the error rates have increased. This is again due to a mismatch between the original signal and the PR4 target response, which is why the PR4 shaping filter introduces correlation in the noise. PR4(C2) clearly outperforms the two other algorithms, showing the value of exploiting the correlation across signal samples.

Figure 8 shows the error rates obtained when using the EPR4 detectors. Due to a higher density, the media noise is higher than in the previous example with symbol separations of 4.4a. This is why the graph in Figure 8 has moved to the right by 2dB in comparison to the graph in Figure 6. While the required S(AWG)NR increased, the margin between the EPR4(Euc) and EPR4(C2) also increased from about 0.5dB to about 1dB, suggesting that the correlation-sensitive metric is more resilient to density increase. This is illustrated in Figure 9 where the S(AWG)NR required for an error rate of 10^{-5} is plotted versus the linear density for the three EPR4 detectors. From this figure we see that, for example, with an S(AWG)NR of 15dB, the EPR(Euc) detector operates at a linear density of about 2.2 symbols/$PW50$, while the EPR4(C2) detector operates at 2.4 symbols/$PW50$, thus achieving a gain of bout 10% in linear density.

Symbol separation of 2.9a. This recording density corresponds to a symbol density of 3 symbols/$PW50$. Due to a very low number of symbols per a, this is the density where the detectors significantly lose performance due to the percolation of magnetic domains also referred to as nonlinear amplitude loss or partial signal erasure.

Figures 10 and 11 show the performance of the PR4 and EPR4 families of detectors at

this density. The detectors with the C2 metric outperform the other two metrics. The error rates are quite high in all cases. This is because at the symbol separations of $2.9a$, nonlinear effects, such as partial erasure due to percolation of domains, start to dominate. These effects can only be undone with a nonlinear pulse shaping filter, which we have not employed here.

In conclusion, the experimental evidence shows that the correlation sensitive sequence detector outperforms the correlation insensitive detectors. We have also demonstrated that the performance margin between the correlation sensitive and the correlation insensitive detectors grows with the recording density. In other words, the performance of the correlation insensitive detector deteriorates faster than the performance of the correlation sensitive detector. Quantitatively, this margin depends on the amount of correlation in the noise passed through the system. Qualitatively, the higher the correlation between the noise samples, the greater will be the margin between the CS-SD and its correlation insensitive counter part.

5   Conclusion

We have developed the correlation sensitive adaptive sequence detector in this paper. The detector is a correlation sensitive maximum likelihood sequence detector. Because it is a sequence detector, it can be used with the basic architecture of any Viterbi-like sequence detector, see Figure 1. The basic difference between the correlation-sensitive sequence detector (CS-SD) and the conventional Viterbi-like sequence detector is that the branch metrics of the CS-SD are computed taking into account the correlation between noise samples. These branch metrics are computed using noise covariance matrices which are adaptively estimated from past signal samples using an RLS-type algorithm. The CS-SD retains the basic architecture of a Viterbi-like algorithm. The only visible difference between the CS-SD and the conventional Viterbi-like detector is that the CS-SD has two blocks in a feedback loop. These two blocks perform two tasks. The first block updates the noise correlation statistics, while the second one uses these statistics (the covariance matrices) to compute the correlation-sensitive tree/trellis

branch metrics.

The characteristics of the CS-SD make it attractive for use in high density recording systems, where correlations between noise samples can be expected. These correlations can be a result of media noise, front-end equalizers, and media and MR head nonlinearities. We presented simulation results showing how the CS-SD works when combined with two partial response detectors, PR4 and EPR4, in a magnetic recording system. We showed that the CS-SD outperforms its correlation-insensitive counterparts in all cases. The margin of this performance difference is dependent on the correlation between the noise samples. The higher their correlation, the larger the gain provided by the CS-SD. Because the media noise samples become more correlated at high densities, the achieved gain in performance by the CS-SD over the correlation-insensitive detectors increases with growing density.

References

[1] S. Nakagawa, K. Yokoyama, and H. Katayama, "A study of detection methods of NRZ recording," *IEEE Trans. Magn.*, vol. 16, pp. 1041–110, Jan. 1980.

[2] K. Han and R. Spencer, "Comparison of different detection techniques for digital magnetic recording channels," *IEEE Trans. Magn.*, vol. 31, pp. 1128–1133, March 1995.

[3] R. W. Wood and D. A. Petersen, "Viterbi detection of class IV partial response on a magnetic recording channel," *IEEE Trans. Comm.*, vol. COM-34, pp. 454–461, May 1986.

[4] H. K. Thapar and A. M. Patel, "A class of partial response systems for increasing storage density in magnetic recording," *IEEE Trans. Magn.*, vol. MAG-23, pp. 3666–3668, Sept. 1987.

[5] J. J. Moon and L. R. Carley, "Performance comparison of detection methods in magnetic recording," *IEEE Trans. Magn.*, vol. 26, pp. 3155–3170, Nov. 1990.

[6] J. G. Kenney and R. W. Wood, "Multi-level decision feedback equalization: An efficient realization of FDTS/DF," *IEEE Trans. Magn.*, vol. 31, pp. 1115–1120, March 1995.

[7] J. Hong, Y. Lee, H. Mutoh, Q. Sun, H. Ueno, J. Wang, and R. W. Wood, "An experimental MDFE detector," *IEEE Trans. Magn.*, pp. 2776–2778, Sept. 1997.

[8] C. Modlin, *Modeling, Detection, and Adaptive Signal Processing for Magnetic Disk Recording*. PhD thesis, Stanford University, Stanford, CA, Dec. 1996.

[9] R. A. Baugh, E. S. Murdock, and B. R. Natarajan, "Measurement of noise in magnetic media," *IEEE Trans. Magn.*, vol. MAG-19, pp. 1722–1724, Sept. 1983.

[10] R. N. Belk, K. P. George, and S. G. Mowry, "Noise in high performance thin-film longitudinal magnetic recording media," *IEEE Trans. Magn.*, vol. MAG-21, pp. 1350–1355, Sept. 1985.

[11] Y.-S. Tang, "Explicit formula for thin film disk noise based on zig-zag domain transitions," *IEEE Trans. Magn.*, vol. 24, pp. 3102–3104, Nov. 1988.

[12] J.-G. Zhu and N. H. Bertram, "Recording and transition noise simulations in thin film media," *IEEE Trans. Magn.*, vol. 24, pp. 2706–2708, Nov. 1988.

[13] J.-G. Zhu and H. Wang, "Noise characteristics of interacting transitions in longitudinal thin film media," *IEEE Trans. Magn.*, vol. 31, pp. 1065–1070, March 1995.

[14] J.-G. Zhu and X.-G. Ye, "Impact of medium noise on various partial response channels," *IEEE Trans. Magn.*, vol. 31, pp. 3087–3089, Nov. 1995.

[15] A. Kavčić and J. M. F. Moura, "Expedient media noise modeling: Isolated and interacting transitions," *IEEE Trans. Magn.*, vol. 32, pp. 3875–3877, Sept. 1996.

[16] W. Zeng and J. Moon, "Modified Viterbi algorithm for jitter-dominated $1 - D^2$ channel," *IEEE Trans. Magn.*, vol. MAG-28, pp. 2895–2897, Sept. 1992.

[17] I. Lee and J. M. Cioffi, "Performance analysis of the modified maximum likelihood sequence detector in the presence of data-dependent noise," in *Proc. 26th Asilomar Conference*, pp. 961–964, Oct. 1992.

[18] N. M. Zayed and L. R. Carley, "Comparison of equalization and detection for very high-density magnetic recording," in *IEEE INTERMAG Conference*, (New Orleans), Apr. 1997.

[19] L. C. Barbosa, "Towards the detection of signals in presence of signal-dependent noise," in *Proc. SPIE: Coding and Signal Processing for Information Storage*, vol. 2605, (Philadelphia), pp. 48–56, Oct. 1995.

[20] G. D. Forney Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Transactions on Information Theory*, vol. 18, pp. 363–378, March 1972.

[21] A. Kavčić and J. M. F. Moura, "The Viterbi algorithm and Markov noise memory," *IEEE Trans. Inform. Theory*, submitted Nov. 1997. 28 pages, accepted for publication in Jan. 2000 issue.

[22] A. Kavčić and J. M. F. Moura, "Experimental validation of the triangle zig-zag transition model," *IEEE Trans. Magn.*, vol. 33, pp. 2704–2706, Sept. 1997.

[23] K. C. Bracken and L. R. Carley, "Adaptive continuous-time equalization followed by FDT-S/DF sequence detection," *IEEE Trans. Magn.*, vol. 31, pp. 1089–1094, March 1995.

[24] E. Biglieri, E. Chiaberto, G. P. Maccone, and E. Viterbo, "Compensation of nonlinearities in high-density magnetic recording channels," *IEEE Trans. Magn.*, vol. 30, pp. 5079–5086, Nov. 1994.

[25] S. K. Nair and J. Moon, "Improved equalization for digital recording using nonlinear filtering and error confinement," *IEEE Trans. Magn.*, vol. 30, pp. 4221–4223, Nov. 1994.

[26] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs: Prentice-Hall, Inc., 1987.

[27] D. A. Lindholm, "Magnetic fields of finite track width heads," *IEEE Trans. Magn.*, vol. 13, pp. 1460–1462, Sept. 1977.

Parameter	Symbol	Value
media remanence	M_r	450kA/m
media coercivity	H_c	160kA/m
media thickness	δ	0.02μm
media cross-track correlation width	s	200Å
head-media separation	d	15nm
head field gradient factor	Q	0.8
head gap length	g	0.135μm
track width	TW	2μm
transition width parameter	a	0.019μm
percolation length	$L = 1.4a$	0.0266μm
50% pulse width	$PW50$	0.167μm

Table 1: Recording parameters used in simulations.
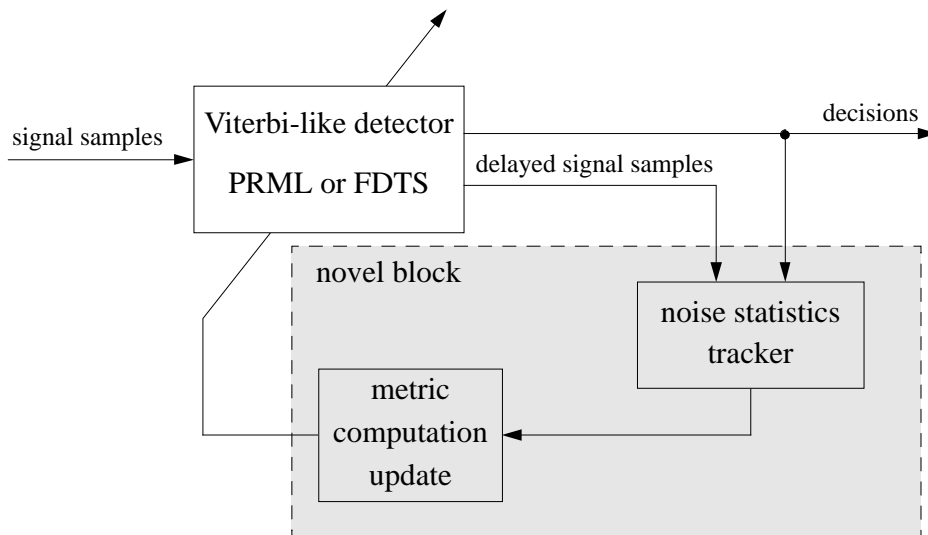
Figure 1: Block diagram of the Viterbi-like correlation-sensitive sequence detector (CS-SD).

BITS: 0 0 1 0 0 0 0 1 0 0 0 0 1 0 1 0 0 0

SAMPLES:

r_1 r_2 r_3 r_4 r_{17} r_{18}

SYMBOLS: $\ominus$ $\ominus$ + $\oplus$ $\oplus$ $\oplus$ $\oplus$ − $\ominus$ $\ominus$ $\ominus$ $\ominus$ + $\oplus$ − $\ominus$ $\ominus$ $\ominus$

a_1 a_2 a_3 a_4 a_{17} a_{18}
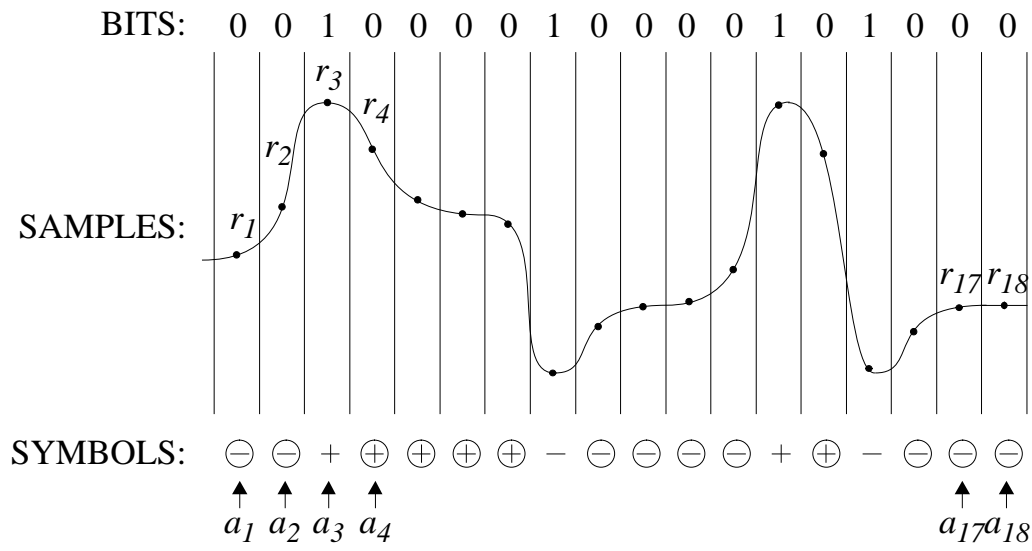
Figure 2: Sample signal waveform, its samples and written symbols.

a_i $\qquad a_{i+1}$

$E(r_i)$

+ 1 1 +

0

$\oplus$ 0 $\oplus$

0

$\ominus$ 0 $\ominus$

0 -1

− -1 −
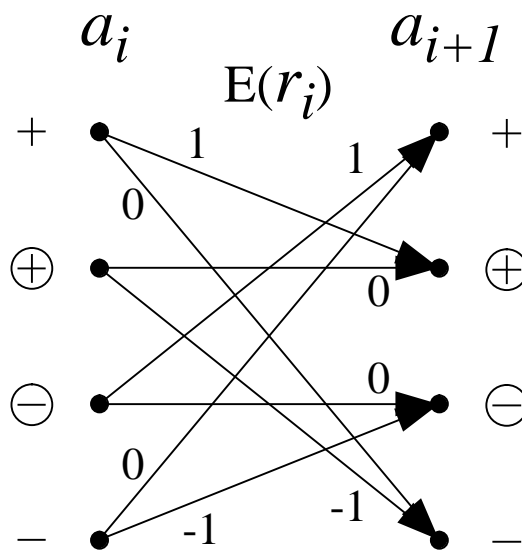
Figure 3: One cell of a PR4 trellis.

$$a_i \qquad a_{i+1} \qquad a_{i+2}$$
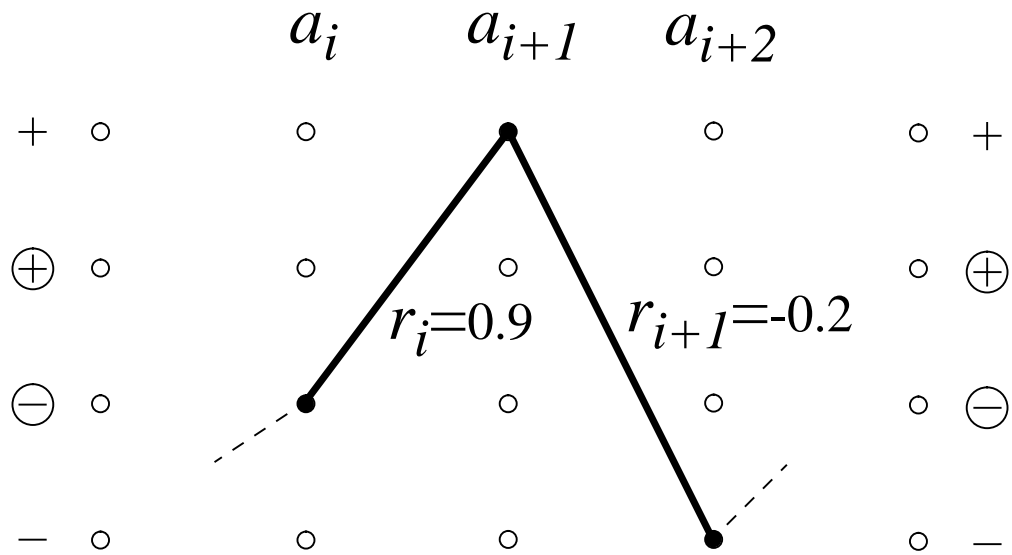
Figure 4: Detected path in a PR4 trellis.
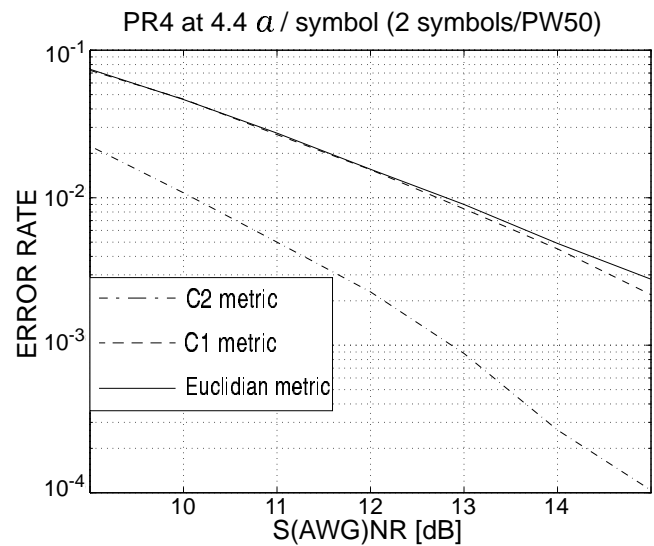
Figure 5: PR4 detection results at 4.4a/symbol.
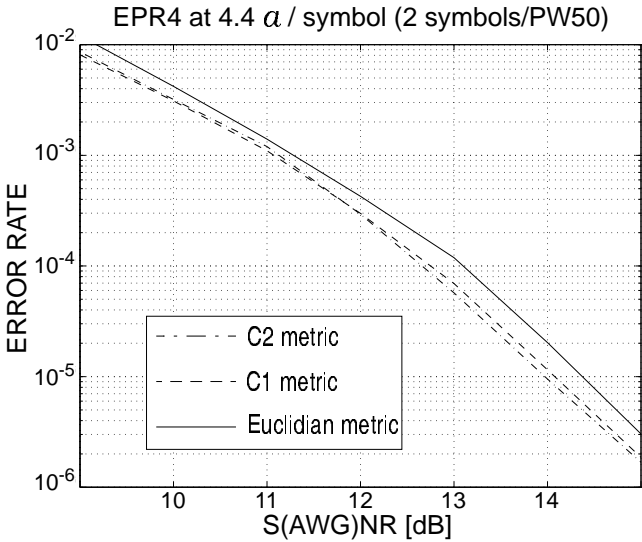
Figure 6: EPR4 detection results at 4.4a/symbol.
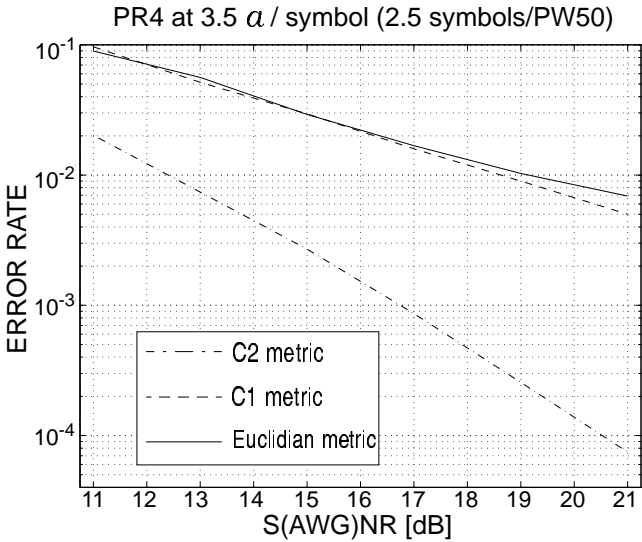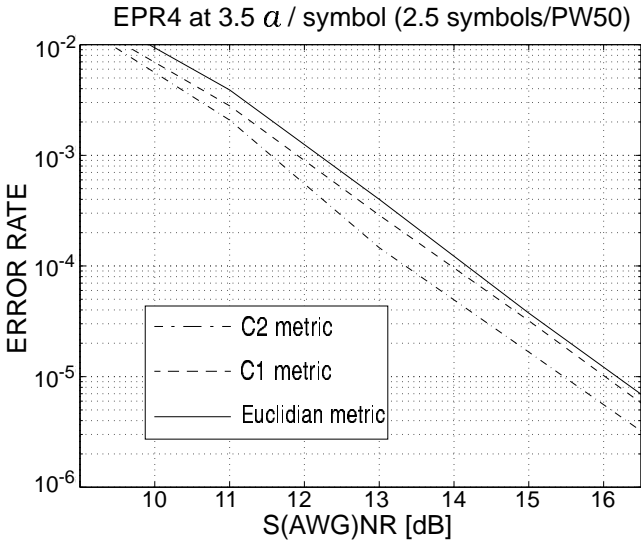

Figure 7: PR4 detection results at 3.5a/symbol.

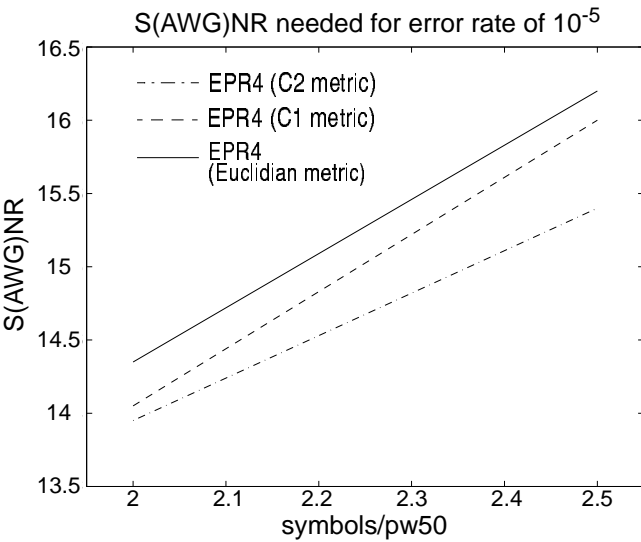Figure 8: EPR4 detection results at $3.5a/\text{symbol}$.

Figure 9: S(AWG)NR margins needed for error rate of 10^{-5} with EPR4 detectors.
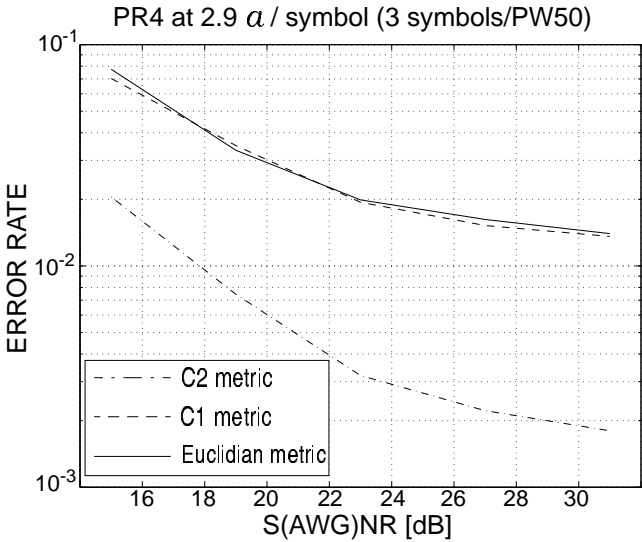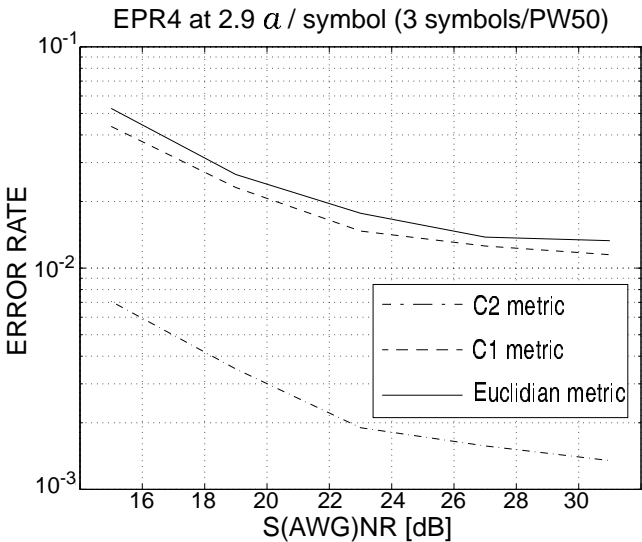
Figure 10: PR4 detection results at 2.9a/symbol.

Figure 11: EPR4 detection results at 2.9a/symbol.