Multihead Multitrack Detection for Next Generation Magnetic Recording, Part I: Weighted Sum Subtract Joint Detection With ITI Estimation

Bing Fan, *Student Member, IEEE*, Hemant K. Thapar, *Fellow, IEEE*, and Paul H. Siegel, *Fellow, IEEE*

***Abstract*— Multitrack detection with array-head reading is a promising technique proposed for next generation magnetic storage systems. The multihead multitrack (MHMT) system is characterized by intersymbol interference in the downtrack direction and intertrack interference (ITI) in the crosstrack direction. Constructing the trellis of a MHMT maximum likelihood (ML) detector requires knowledge of the ITI, which is generally unknown at the receiver. Furthermore, in a time-varying ITI environment, updating ML trellis labels using adaptively-generated ITI estimates could incur significant delay. In this paper, we propose one approach to solve these issues. The proposed detector uses a different trellis structure whose output labels are independent of the ITI level, with ITI-dependence appearing only in a scale factor used to suitably weight the computed path metrics in order to retain ML optimality. The detector formulation facilitates the design of a gain loop structure that can track the time-varying ITI and provide ITI estimates to adaptively adjust the weights in the path metric evaluation. Simulation results show that the proposed detector architecture with ITI estimation offers a substantial performance advantage over ML detection using a static ITI estimate.**

***Index Terms*— Shingled magnetic recording (SMR), intertrack interference (ITI), intersymbol interference (ISI), adaptive estimation, maximum-likelihood sequence estimation (MLSE).**

I. Introduction

With the development of information networks and data centers, the demand for ultra-high capacity storage devices is continually increasing. For hard disk drive (HDD) storage, several promising technologies have been proposed to increase the areal density. Among these options, bit patterned magnetic recording (BPMR), heat assisted magnetic recording (HAMR), and microwave assisted magnetic recording (MAMR) are technologies that require significant modification of the recording media and/or the read-write transducer to push beyond the superparamagnetic limit [1]. In contrast, two dimensional magnetic recording (TDMR) combined with shingled writing, has been proposed as an approach to advancing areal density using conventional media and recording head technology, relying instead upon more sophisticated signal processing techniques to decode data under severe noise and interference conditions [2].

One particular challenge faced by future generations of magnetic recording systems will be how to deal with the cross-talk between densely-packed data tracks. It is conceivable that in the near term, the width and separation of recording tracks will be significantly reduced, with a relatively smaller reduction in read head size [2]–[4]. Therefore, the read head will sense signals from neighboring tracks when reading from a target track, causing substantial intertrack interference (ITI) [5]. The performance of a single-head single-track (SHST) detector in such a scenario with ITI was studied in [6]. The results show that the additional signal distortion caused by ITI can severely degrade the performance of disk drives using conventional SHST detection methods.

The next-generation magnetic recording channel, which is characterized by both ITI in the crosstrack direction and intersymbol interference (ISI) in the downtrack direction, can be well approximated by the two-dimensional ISI (2D-ISI) model, in which the impulse response of the read head is represented by a matrix [7]. Since maximum likelihood (ML) detection on a general 2D-ISI channel is quite complex – in fact, it has been shown to be NP-hard even for some elementary 2D-ISI channels [8] – considerable effort has been spent on designing practical suboptimal detectors. For example, the authors in [9] propose a one-dimensional (1D) multi-strip base Bahl, Cocke, Jelinek and Raviv (BCJR) detector for a 2×2 impulse response matrix. Their algorithm, which passes soft information from column to column, could achieve performance that is approximately $\frac{2}{3}$dB from ML. The detector proposed in [10] adopts a turbo structure, where the soft information is iteratively shared between a row-by-row BCJR detector and another column-by-column BCJR detector. Simulation results show that the performance of the proposed detector is near optimal. In [11], the performance of a similar turbo detector was then examined on a more realistic magnetic recording channel derived from a Voronoi grain model. Several other turbo-structured detectors have also been explored in [12]–[14]. Although the row-column turbo detector achieves promising performance on the 2D-ISI channel, it generally suffers from high computational

Manuscript received May 27, 2016; revised October 31, 2016 and January 4, 2017; accepted January 5, 2017. Date of publication January 16, 2017; date of current version April 14, 2017. This work was supported in part by the National Science Foundation under Grant CCF-1405119, and the Center for Memory and Recording Research (formerly, Center for Magnetic Recording Research) at UC San Diego. The associate editor coordinating the review of this paper and approving it for publication was L. Dolecek.

B. Fan and P. H. Siegel are with the Center for Memory and Recording Research, Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093, USA (e-mail: bifan@ucsd.edu; psiegel@ucsd.edu).

H. K. Thapar is with OmniTier, Inc., Santa Clara, CA 95054, USA (e-mail: hemantkthapar@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2017.2652471

complexity and long delay. In [15] the authors construct a factor graph for the 2D-ISI channel and design a general belief propagation (GBP) detector to combat both the 2D-ISI and the "overwriting" effect in TDMR. Although GBP has higher complexity, its parallelized information processing could lead to shorter detection delay.

Another approach to approximating the magnetic recording channel is the multihead multitrack (MHMT) model, where the ISI is generally more severe than ITI. In an MHMT system, multiple read heads scan a group of data tracks simultaneously. Guard gaps are generally added between neighboring groups to prevent cross-talk [16], [17]. Detector design for MHMT has been studied for decades. Iterative ITI cancellation, which removes ITI from each single-track readback signal before detection, was explored in [18], [19]. The simulation results in [19] show that, under a specified ITI level, the ITI cancellation scheme could offer 3dB gain compared to the conventional SHST detector. Another ITI cancellation algorithm that uses soft decisions was recently studied in [20]. Unlike ITI cancellation, which tends to reduce the problem to a 1D detection task, an MHMT detector processes multiple readback waveforms and decodes multiple data tracks simultaneously. An optimal MHMT detector was proposed in [21], and then theoretically analyzed in [22]. An iterative detection/decoding scheme for a two-track channel model with two heads was simulated in [23]. In [24], the authors studied the performance and implementation cost of the MHMT ML detector for a higher-order MHMT system. Similar structures were also analyzed for BPMR in [3], [17]. As shown in these works, the MHMT detector can better combat ITI for a large group of channels compared to the SHST scheme. However, processing data from multiple heads exponentially increases the computational complexity of the MHMT detector.

In this two-part paper, we adopt the MHMT model and aim to address two challenging problems that can potentially make ML detection for the MHMT channel impractical – adaptive ITI estimation and implementation complexity. Our model is similar to those used in [19], [21], [23], [25]. We assume that there are n data tracks representing an independent band, with n read heads, all with the same physical properties, evenly spaced over the band. The channel seen by each head from the track to which it is associated is assumed to be a partial-response ISI channel with additive Gaussian noise. The model for ITI assumes that the interfering signal sensed by a read head from a neighboring track is a scaled version of the signal sensed by the head directly over the neighboring track. Of particular interest is the case where ITI derives only from directly adjacent tracks. We recognize that this ITI model is a relatively simple approximation to the actual magnetic recording channel, but it nonetheless serves as useful motivation for the investigation of potentially useful detection algorithms and architectures.

In Part I, we consider the problem of estimating the side track response. In [6], the authors propose a least mean square (LMS) adaptive algorithm to estimate the off-track interference for the SHST system. For the MHMT detector, the knowledge of side track response is crucial because the construction of MHMT trellis requires complete channel responses from side tracks. In addition, updating ITI estimate could possibly delay the detection process since the output labels of the MHMT trellis need to be recalculated. To solve this issue, we propose a novel detection method, weighted sum-subtract joint detection (WSSJD), whose trellis labels do not change with ITI. Then we reformulate the ITI estimation as a channel gain control problem, and design gain loops to adaptively track the value of ITI. The design and the performance of the proposed detector for the simplest case, which is the 2H2T system, have been introduced in [26].

This paper extends [26] in the following ways. We generalize the proposed 2H2T WSSJD algorithm to the general n-head, n-track (nHnT) case. By taking the eigenvalue decomposition of the channel interference matrix, and applying coordinate transformations in both the input and the output spaces, the cross-interfering channels are transformed into n separate sub-channels. The parameter that indicates the ITI level appears in the gain factor of each resulting sub-channel, and thus can be estimated by gain loops. The trellis of the transformed system remains the same under varying ITI levels, so adaptive ITI level estimates can be used. Simulation results on 2H2T and 3H3T systems show that under time-varying ITI conditions, WSSJD with adaptively estimated ITI outperforms the traditional ML detector for which practical considerations dictate the use of a static ITI estimate. We also analyze the performance advantage of the nHnT detection system over ITI-free single-track detection by examining the distance properties of each system.

Another important issue, which will be discussed in Part II of this work, is the increased computational complexity of the MHMT detector. We find that the transformation developed in WSSJD offers a natural set-partition principle for the input constellation that can be used in a reduced-state sequence estimation (RSSE) architecture. This concept was partially presented in [27] and [28]. In Part I, we present the performance of reduced WSSJD with gain loops for comparison. The implementation details and the theoretical performance analysis of the reduced algorithm will be presented in Part II.

The paper is organized as follows. In Section II we introduces the nHnT model and our assumptions in deriving the model. The ML detector and its performance approximation in terms of the minimum distance parameter are also reviewed. In Section III we introduce the WSSJD algorithm for the 2H2T system, and analyze its performance by examining its minimum distance. A suboptimal implementation is also discussed. In Section IV we study the ITI sensitivity of the detectors. We also analyze the effect of ITI mismatch on different types of error events. A gain loop structure is then proposed to adaptively estimate the ITI level for use by WSSJD. The proposed WSSJD algorithm is generalized in Section V. We show how the WSSJD transformation decomposes the nHnT system into parallel sub-channels. Pseudocode of the generalized WSSJD algorithm is also provided. Bit-error-rate simulation results for the 2H2T and 3H3T channels are presented in Section VI, where we consider both static and adaptive ITI environments. The paper concludes in Section VII.

II. n-Head n-Track Channel Model

In next-generation magnetic recording using shingled magnetic recording, the data tracks will be organized into bands, each consisting of a number of narrow, closely space tracks, with a small gap between bands to prevent interference [16].

Consider a band of n tracks. Let $x^i(D)$ denote the bipolar data sequence recorded on the i-th track, $x^i(D) = \Sigma_{k=0}^{N} x_k^i D^k$, with $x_k^i \in \{-1, +1\}$. We assume that x_k^i is i.i.d. and equiprobable, and $x^i(D), i \in \{1, \ldots, n\}$ are independent sequences. We also assume that there is no phase offset during the writing, i.e., the data sequences are perfectly aligned.

During the readback process, n heads are evenly placed over n tracks, with one head designated per track. The signal from each head is passed through a matched filter, a sampler, and then equalized to the target dipulse response represented by a polynomial $h(D) = h_0 + h_1 D + \cdots + h_\nu D^\nu$ that reflects the intersymbol interference (ISI). The interference from neighboring tracks is additive, and the interfering by a read head from a neighboring track is assumed to be a scaled version of the read signal that would be sensed by the head directly over the neighboring track in the absence of any intertrack interference. Let $r^i(D)$ denote the sampled readback samples corresponding to the signal from the head corresponding to the i-th track, $i = 1, \ldots, n$. The resulting n-head n-track (nHnT) system is described by

$$\mathbf{R}(D) = \mathbf{A}_n \mathbf{X}(D) h(D) + \mathbf{\Omega}(D), \tag{1}$$

where

$$\mathbf{X}(D) = [x^1(D), \cdots, x^n(D)]^\top \tag{2}$$

is the input vector, and

$$\mathbf{R}(D) = [r^1(D), \cdots, r^n(D)]^\top \tag{3}$$

is the output vector. The vector of electronic noise components is denoted by

$$\mathbf{\Omega}(D) = [\omega^1(D), \cdots, \omega^n(D)]^\top. \tag{4}$$

We assume that the noise samples are independent and Gaussian distributed, with zero mean and variance σ^2. The term $\mathbf{X}(D)h(D) = [x^1(D)h(D), \cdots, x^n(D)h(D)]^\top$ denotes the vector of noiseless ISI channel outputs, and $\mathbf{A}_n$ is an $n \times n$ interference matrix. In our model, we assume that only adjacent tracks interfere. This assumption is reasonable since in most cases the ITI from the adjacent tracks is the dominant one. Additionally, by assuming the physical uniformity and symmetry of read heads, $\mathbf{A}_n$ can be modeled as a symmetric tridiagonal Toeplitz matrix

$$\mathbf{A}_n = \begin{bmatrix} 1 & \epsilon & & & \\ \epsilon & 1 & \ddots & 0 & \\ & \ddots & \ddots & \ddots & \\ & 0 & \ddots & 1 & \epsilon \\ & & & \epsilon & 1 \end{bmatrix}, \tag{5}$$

where $\epsilon \in [0, 0.5]$ represents the ITI level determined by the distance between the head and the adjacent track.

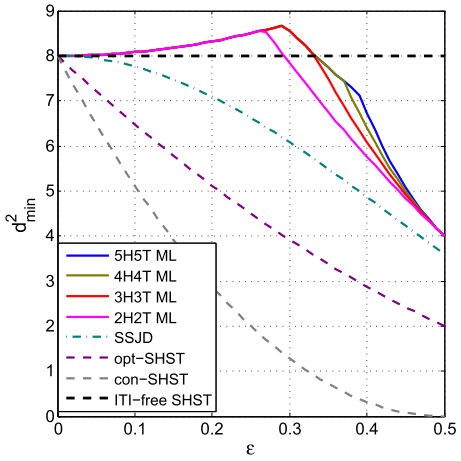


Fig. 1. Minimum squared-distance comparison of different detectors on channel $h(D) = 1 + D$ with $d_0^2 = 8$.

Given the received sequences $\mathbf{R}(D)$, the ML detector chooses $\mathbf{X}^*(D)$ that satisfies

$$\begin{aligned} \mathbf{X}^*(D) &= \arg\max_{\mathbf{X}(D)} \Pr(\mathbf{R}(D)|\mathbf{X}(D)) \\ &= \arg\min_{\mathbf{X}(D)} \|\mathbf{R}(D) - \mathbf{A}_n \mathbf{X}(D) h(D)\|^2, \end{aligned} \tag{6}$$

where $\|\cdot\|^2$ denotes the squared Euclidean norm. For instance, $\|\mathbf{X}(D)\|^2$ is calculated by

$$\|\mathbf{X}(D)\|^2 = \sum_i \|x^i(D)\|^2 = \sum_{i,j} (x_j^i)^2. \tag{7}$$

The trellis constructed in the ML detector contains $2^{n\nu}$ states, each of which is associated with 2^n edges.

Let $x^i(D)$ and $\hat{x}^i(D)$ be the correct and estimated input sequences corresponding to track i, respectively. An error event occurs if for some track i, $e^i(D) = x^i(D) - \hat{x}^i(D)$ is not zero. For an error event

$$\boldsymbol{e}(D) = [e^1(D), \cdots, e^n(D)]^\top, \tag{8}$$

the distance parameter is calculated by

$$d^2(\boldsymbol{e}(D)) = \|\mathbf{A}_n \boldsymbol{e}(D) h(D)\|^2 = \sum_{i=1}^{n} \|y^i(D)\|^2, \tag{9}$$

where

$$y^1(D) = [e^1(D) + \epsilon e^2(D)] h(D), \tag{10}$$

$$y^n(D) = [e^n(D) + \epsilon e^{n-1}(D)] h(D), \tag{11}$$

$$y^i(D) = [e^i(D) + \epsilon e^{i-1}(D) + \epsilon e^{i+1}(D)] h(D), \quad i \in [2, n-1]. \tag{12}$$

The minimum distance parameter of the channel is obtained by minimizing $d^2(\boldsymbol{e}(D))$ over all possible $\boldsymbol{e}(D)$. In Fig. 1, we plot the minimum distances of the nHnT ML detector for $n = 2, 3, 4, 5$, assuming $h(D) = 1 + D$. It can be observed that in a large region of ϵ, the nHnT ML detectors have a greater minimum distance parameter than the ITI-free SHST ML detector. The distance parameters associated with the other detectors plotted in Fig. 1 will be discussed in Section III-C.

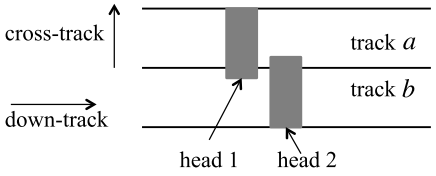


Fig. 2. Schematic of a two-head two-track recording system.

It is well known that the error event probability of a trellis-based detector can be approximated as $\mathrm{P_e} \propto Q(\frac{d_{\min}}{2\sigma})$, where $Q(\cdot)$ is the tail probability of the standard Gaussian distribution, $d_{\min}$ is the minimum distance, and σ is the standard deviation of the additive Gaussian noise. The performance of the detector can be accurately predicted by analyzing its minimum distance.

We notice that the calculation of trellis output labels, $\mathbf{Y}(D) = \mathbf{A}_n\mathbf{X}(D)h(D)$, requires knowledge of the ITI level ϵ. This can introduce disadvantages in the hardware realization of the conventional ML detector. With time-varying ITI levels, the detector in effect has to recalculate and update the branch output labels whenever ϵ changes. If ϵ changes frequently, this process could incur a substantial increase in hardware complexity and delay. Therefore, using a conventional approach to implementation, MHMT ML detection is essentially limited to a static value of ϵ. Indeed, in the prior literature, studies of MHMT ML detection have always assumed a static ITI environment [21]–[23]. On a hard disk drive, however, ϵ generally varies spatially due to mechanical effects such as head skew and flying height variation. Thus, adaptive estimation of ϵ will be necessary, potentially introducing the complexity and delay issues just mentioned if conventional ML detection is used.

In the following sections, we propose a novel detection architecture that is amenable to incorporating adaptive ITI estimation while retaining the optimality of ML detection. The proposed detector uses a different trellis diagram from the conventional ML detector. For convenience, we refer to the latter as the "ML trellis" even though both detectors produce the same ML output sequences.

III. WSSJD on 2H2T System

The weighted sum-subtract joint detection (WSSJD) algorithm differs from the conventional ML detector in two respects. First, it has a "sum-subtract" preprocessor before the Viterbi detector. Second, it uses weighted branch metrics in detection. In this section, we assume ϵ to be known. This condition will be relaxed in Section IV where we show that ϵ acts as a gain factor that can be estimated by means of a first-order gain loop.

We begin the presentation of WSSJD by first analyzing the 2H2T case. A schematic of the 2H2T system is shown in Fig. 2. Let $x^a(D)$ and $x^b(D)$ be the input sequences corresponding to track a and track b, respectively. The outputs of the corresponding read heads are

$$\begin{aligned} r^a(D) &= y^a(D) + \omega^a(D), \\ r^b(D) &= y^b(D) + \omega^b(D), \end{aligned} \tag{13}$$

where

$$\begin{aligned} y^a(D) &= x^a(D)h(D) + \epsilon x^b(D)h(D), \\ y^b(D) &= \epsilon x^a(D)h(D) + x^b(D)h(D), \end{aligned} \tag{14}$$

are the noiseless outputs. As given in [22], the minimum distance parameter of the 2H2T ML detector is

$$d^2_{\min,\,\mathrm{ML}} = \begin{cases} (1+\epsilon^2)d_0^2 & \text{if } 0 \leqslant \epsilon \leqslant 2-\sqrt{3} \\ 2(1-\epsilon)^2 d_0^2 & \text{if } 2-\sqrt{3} \leqslant \epsilon \leqslant 1/2 \end{cases} \tag{15}$$

where d_0 is the minimum distance of the single-track detector on channel $h(D)$ when there is no ITI. The single track error events are the dominant error patterns at low ITI, while the double track error events are the dominant ones at high ITI. The operating point that gives the highest minimum distance, or the best performance of the ML detector, is at $\epsilon = 2-\sqrt{3}$.

A. Sum-Subtract Preprocessing

Instead of directly processing $r^a(D)$ and $r^b(D)$, the WSSJD method first calculates the weighted sum and difference, $r^+(D)$ and $r^-(D)$, given by

$$\begin{aligned} r^+(D) &= \frac{1}{1+\epsilon}\,[\,r^a(D) + r^b(D)\,], \\ r^-(D) &= \frac{1}{1-\epsilon}\,[\,r^a(D) - r^b(D)\,], \end{aligned} \tag{16}$$

respectively. Defining new input signals by

$$\begin{aligned} z^+(D) &= x^a(D) + x^b(D), \\ z^-(D) &= x^a(D) - x^b(D), \end{aligned} \tag{17}$$

and the corresponding noiseless output signals by

$$\begin{aligned} y^+(D) &= z^+(D)h(D), \\ y^-(D) &= z^-(D)h(D), \end{aligned} \tag{18}$$

we can rewrite (16) as

$$\begin{aligned} r^+(D) &= y^+(D) + \omega^+(D), \\ r^-(D) &= y^-(D) + \omega^-(D). \end{aligned} \tag{19}$$

The new noise components,

$$\begin{aligned} \omega^+(D) &= \frac{1}{1+\epsilon}\,(\omega^a(D) + \omega^b(D)), \\ \omega^-(D) &= \frac{1}{1-\epsilon}\,(\omega^a(D) - \omega^b(D)) \end{aligned} \tag{20}$$

satisfy $\omega_k^+ \sim \mathcal{N}(0, \frac{2\sigma^2}{(1+\epsilon)^2})$, $\omega_k^- \sim \mathcal{N}(0, \frac{2\sigma^2}{(1-\epsilon)^2})$. Furthermore,

$$E[\,\omega_k^+\,\omega_k^-\,] = \frac{1}{1-\epsilon^2}(E[\,\omega_k^{a\,2}\,] - E[\,\omega_k^{b\,2}\,]) = 0, \tag{21}$$

which implies that $\omega^+(D)$ and $\omega^-(D)$ are uncorrelated and, therefore, independent.

We can think of $r^+(D)$ and $r^-(D)$ as the noisy outputs obtained by passing each of $z^+(D)$ and $z^-(D)$ through a channel $h(D)$, but with different noise powers. These two channels are called the "sum channel" and the "subtract channel", respectively. Notice that the new inputs z_k^+ and z_k^- have a three-level alphabet, $\mathcal{B} = \{-2, 0, 2\}$. There is a

TABLE I
MAPPING BETWEEN (x_k^a, x_k^b) AND (z_k^+, z_k^-)

x_k^a	x_k^b	z_k^+	z_k^-
1	1	2	0
1	-1	0	2
-1	1	0	-2
-1	-1	-2	0

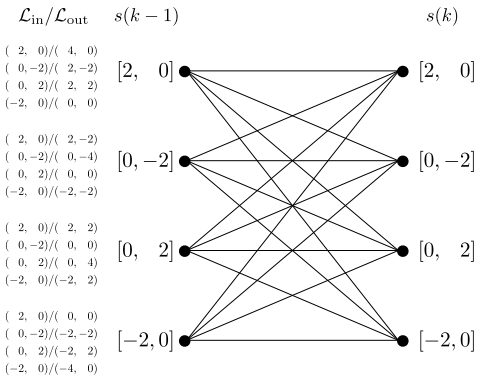


Fig. 3. WSSJD trellis for channel $h(D) = 1 + D$.

one-to-one mapping between (z_k^+, z_k^-) and (x_k^a, x_k^b), as shown in Table I.

Since $r^+(D)$ and $r^-(D)$ are obtained from separate channels, one can independently detect $z^+(D)$ and $z^-(D)$, and then map (z_k^+, z_k^-) to (x_k^a, x_k^b) according to Table I. This corresponds to solving for

$$\begin{aligned}\hat{z}^+(D) &= \arg\max_{z^+(D)} \log \Pr(r^+(D) \mid z^+(D)) \\ &= \arg\min_{z^+(D)} \|r^+(D) - z^+(D)\|^2, \\ \hat{z}^-(D) &= \arg\max_{z^-(D)} \log \Pr(r^-(D) \mid z^-(D)) \\ &= \arg\min_{z^-(D)} \|r^-(D) - z^-(D)\|^2. \end{aligned} \tag{22}$$

This computation has complexity $O(2 \cdot 3^\nu)$. However, it is suboptimal. From Table I we see that $z^+(D)$ and $z^-(D)$ are not independent, e.g., $z_k^+ = 2$ forces z_k^- to be 0. Independent detection ignores this correlation and produces some undecodable $(\hat{z}_k^+, \hat{z}_k^-)$ pairs. Optimal detection must jointly consider both the sum channel and the subtract channel, determining

$$\begin{aligned}&\hat{z}^+(D), \hat{z}^-(D) \\ &= \arg\max_{z^+(D), z^-(D)} \log \Pr(r^+(D), r^-(D) \mid z^+(D), z^-(D)). \end{aligned} \tag{23}$$

The Viterbi algorithm (VA) is applied to solve (23). The WSSJD trellis has the same number of states as the ML trellis. Each branch connects an initial state $s(k) = [z_{k-\nu}^+ \cdots z_{k-1}^+, z_{k-\nu}^- \cdots z_{k-1}^-]$ to a terminal state $s(k+1) = [z_{k-\nu+1}^+ \cdots z_k^+, z_{k-\nu+1}^- \cdots z_k^-]$ with input label $\mathcal{L}_{\text{in}} = (z_k^+, z_k^-)$ and output label $\mathcal{L}_{\text{out}} = (y_k^+, y_k^-)$. Fig. 3 shows a WSSJD trellis for the channel $h(D) = 1 + D$. The text to the left of each state lists the branch labels in the form of input/output. Note that, unlike the ML trellis, the WSSJD trellis labels are independent of ϵ.

B. Weighted Branch Metric

Since the sum and the subtract channel have different noise powers, WSSJD computes a weighted sum of their individual distance metrics, $\|r^+(D) - y^+(D)\|^2$ and $\|r^-(D) - y^-(D)\|^2$. The optimal choice of the weights is found by evaluating (23),

$$\begin{aligned}&\hat{z}^+(D), \hat{z}^-(D) \\ &= \arg\max_{z^+(D), z^-(D)} \log \Pr(r^+(D), r^-(D) | z^+(D), z^-(D)) \\ &= \arg\max_{z^+(D), z^-(D)} \log \Pr(r^+(D) | z^+(D)) + \log \Pr(r^-(D) | z^-(D)) \\ &= \arg\min_{z^+(D), z^-(D)} \frac{\|r^+(D) - y^+(D)\|^2}{2\sigma^2/(1+\epsilon)^2} + \frac{\|r^-(D) - y^-(D)\|^2}{2\sigma^2/(1-\epsilon)^2} \\ &= \arg\min_{z^+(D), z^-(D)} (1+\epsilon)^2 \|r^+(D) - y^+(D)\|^2 \\ &\quad + (1-\epsilon)^2 \|r^-(D) - y^-(D)\|^2. \end{aligned} \tag{24}$$

Let $M_k(s)$ denote the survivor path metric for state s at time k. The path metric corresponding to the extension along a branch from state s to s' is

$$\begin{aligned} M_{k+1}(s') = M_k(s) &+ (1+\epsilon)^2 (r_k^+ - y_k^+)^2 \\ &+ (1-\epsilon)^2 (r_k^- - y_k^-)^2. \end{aligned} \tag{25}$$

The term $m_k(s, s') = (1+\epsilon)^2 (r_k^+ - y_k^+)^2 + (1-\epsilon)^2 (r_k^- - y_k^-)^2$ is the weighted branch metric.

Since the sum-subtract transformation is bijective, we have

$$\begin{aligned} &\Pr(r^+(D), r^-(D) | z^+(D), z^-(D)) \\ &\quad = \Pr(r^a(D), r^b(D) | x^a(D), x^b(D)). \end{aligned} \tag{26}$$

Therefore, WSSJD gives the ML solution.

Assume $(z^+(D), z^-(D))$ are the correct input sequences. WSSJD outputs wrong estimates $(\hat{z}^+(D), \hat{z}^-(D))$ if

$$\begin{aligned} &\Pr(r^+(D), r^-(D) | z^+(D), z^-(D)) \\ &\quad < \Pr(r^+(D), r^-(D) | \hat{z}^+(D), \hat{z}^-(D)). \end{aligned} \tag{27}$$

Let $e^+(D) = z^+(D) - \hat{z}^+(D)$ and $e^-(D) = z^-(D) - \hat{z}^-(D)$ be the error event. Notice that the alphabet of e_k^+ and e_k^- is $\{\pm 4, \pm 2, 0\}$, and e_k^+ and e_k^- are not independent, e.g., $e_k^+ = 4$ implies $e^- = 0$. The probability of having $(e^+(D), e^-(D))$ is approximated by $Q(\frac{d_{\text{WSSJD}}(e^+(D), e^-(D))}{2\sigma})$, where

$$\begin{aligned} &d^2_{\text{WSSJD}}(e^+(D), e^-(D)) \\ &= \frac{(1+\epsilon)^2 \|e^+(D)h(D)\|^2 + (1-\epsilon)^2 \|e^-(D)h(D)\|^2}{2} \end{aligned} \tag{28}$$

is the effective distance parameter defined for WSSJD. Although the labels on the WSSJD trellis are independent of ϵ, the distance metric used to estimate the WSSJD performance has to incorporate the effect of the signal-to-noise ratio (SNR) differences in the sum and subtract channels. Evaluating (28) for all possible error events shows that WSSJD has the same minimum distance parameter as the ML detector.

C. Performance Loss From Neglecting Branch Metric Weighting Factors

We refer to the detector that ignores the weighting factors, i.e., that uses

$$m_k(s, s') = (r_k^+ - y_k^+)^2 + (r_k^- - y_k^-)^2 \tag{29}$$

as the branch metric, as the sum-subtract joint detector. The corresponding sum-subtract joint detection (SSJD) method is suboptimal. The performance loss of SSJD is reflected in its minimum distance parameter. Let $(z^+(D), z^-(D))$ and $(\hat{z}^+(D), \hat{z}^-(D))$ be the correct and estimated sequences. The error event probability is

$$\begin{aligned}&\Pr(\|r^+(D) - z^+(D)h(D)\|^2 + \|r^-(D) - z^-(D)h(D)\|^2 \\ &\quad > \|r^+(D) - \hat{z}^+(D)h(D)\|^2 + \|r^-(D) - \hat{z}^-(D)h(D)\|^2) \\ &\quad = Q\left(\frac{d_{\text{SSJD}}(e^+(D), e^-(D))}{2\sigma}\right),\end{aligned} \tag{30}$$

where

$$d_{\text{SSJD}}(e^+(D), e^-(D)) = \frac{\|e^+(D)h(D)\|^2 + \|e^-(D)h(D)\|^2}{\sqrt{\frac{2\|e^+(D)h(D)\|^2}{(1+\epsilon)^2} + \frac{2\|e^-(D)h(D)\|^2}{(1-\epsilon)^2}}}. \tag{31}$$

Since $e^+(D)$ and $e^-(D)$ are not independent, we express them as $e^+(D) = e^a(D) + e^b(D)$ and $e^-(D) = e^a(D) - e^b(D)$ to find $d^2_{\text{min, SSJD}}$. To simplify the notation, let $A(D) = e^a(D)h(D)$ and $B(D) = e^b(D)h(D)$. We have

$$\begin{aligned}&d^2_{\text{SSJD}}(e^a(D), e^b(D)) \\ &\quad = \frac{(1+\epsilon)^2(1-\epsilon)^2(\|A(D)\|^2 + \|B(D)\|^2)^2}{(1+\epsilon^2)(\|A(D)\|^2 + \|B(D)\|^2) - 4\epsilon\langle A(D), B(D)\rangle}.\end{aligned} \tag{32}$$

Consider the case of a single-track error event, e.g., assume $e^b(D) = 0$. Then

$$\begin{aligned}d^2_{\text{SSJD}}(e^a(D), 0) &= \frac{(1+\epsilon)^2(1-\epsilon)^2}{1+\epsilon^2}\|A(D)\|^2 \\ &\geqslant \frac{(1+\epsilon)^2(1-\epsilon)^2}{1+\epsilon^2}d_0^2\end{aligned} \tag{33}$$

with equality achieved when $e^a(D)$ gives the minimum distance d_0^2 on channel $h(D)$.

For the case of a double-track error event we have

$$\begin{aligned}d^2_{\text{SSJD}}(e^a(D), e^b(D)) &\geqslant (1-\epsilon)^2(\|A(D)\|^2 + \|B(D)\|^2) \\ &\geqslant 2(1-\epsilon)^2 d_0^2,\end{aligned} \tag{34}$$

where we have used the fact that

$$\begin{aligned}-\langle A(D), B(D)\rangle &\leqslant \|A(D)\|\,\|B(D)\| \\ &\leqslant \frac{1}{2}(\|A(D)\|^2 + \|B(D)\|^2).\end{aligned} \tag{35}$$

Equality is achieved in (34) when $e^a(D) = -e^b(D)$ and both $e^a(D)$ and $e^b(D)$ lead to the minimum distance d_0 on channel $h(D)$. Comparison between (33) and (34) shows that, in contrast to WSSJD, the minimum distance of SSJD is always dominated by single-track error events. Therefore

$$d^2_{\text{min, SSJD}} = \frac{(1+\epsilon)^2(1-\epsilon)^2}{1+\epsilon^2}d_0^2. \tag{36}$$

In Fig. 1 we plot the squared minimum distance parameters for several detectors as a function of ϵ. Recall that WSSJD on the 2H2T model has the same d^2_{min} as the 2H2T ML detector. Two SHST detectors [25] are included for comparison purposes. The optimal SHST detector estimates the data on a track using the model for channel ISI and the interference induced by the side track. Its minimum distance is dominated by double-track error events, leading to

$$d^2_{\text{min, opt-SHST}} = (1-\epsilon)^2 d_0^2. \tag{37}$$

The conventional SHST detector considers only the channel ISI and treats the side track interference as additional electronic noise, and is therefore suboptimal. For the channel with $h(D) = 1 + D$, the minimum distance has a closed form expression given by

$$d^2_{\text{min, con-SHST}} = (1-2\epsilon)^2 d_0^2. \tag{38}$$

Finally, the ITI-free SHST corresponds to the SHST channel model with no ITI. The performance of the ITI-free SHST detector can be viewed as the performance of an ideal ITI cancellation scheme in which the side track response is perfectly removed.

We summarize the proposed algorithms as follows. The branch labels of the trellis constructed for WSSJD or SSJD are independent of the ITI level ϵ. This independence is the key property underlying the proposed architecture for combining WSSJD or SSJD with adaptive estimation of ϵ. In WSSJD, the ITI level ϵ is used to weight the branch metrics, thereby ensuring that it achieves ML performance. The SSJD algorithm does not weight the branch metrics, thus producing suboptimal decisions.

IV. Adaptive ITI Level Estimation

A. ITI Sensitivity

To evaluate the sensitivity of the various detectors to a small change in the ITI level, we introduce a small offset into our performance simulations. Suppose the nominal ITI level is ϵ_0, while the true ITI level is adjusted by an offset $\Delta\epsilon$. The new noiseless channel outputs become

$$\begin{aligned}y^a(D) &= x^a(D)h(D) + (\epsilon_0 + \Delta\epsilon)\,x^b(D)h(D), \\ y^b(D) &= x^b(D)h(D) + (\epsilon_0 + \Delta\epsilon)\,x^a(D)h(D).\end{aligned} \tag{39}$$

If the detectors use the nominal level ϵ_0, rather than the true level $\epsilon = \epsilon_0 + \Delta\epsilon$, mismatch will lead to a degradation in performance.

Fig. 4 shows the simulated bit error rate (BER) as a function of the mismatch $\Delta\epsilon$ for WSSJD, SSJD, and the conventional ML detector on the channel with $h(D) = 1 + D$ at SNR = 10dB, with $\epsilon_0 = 0.1$ and $\epsilon_0 = 0.3$, respectively. We see that the minimum BER occurs at $\Delta\epsilon \approx 0$ for $\epsilon_0 = 0.1$ and at $\Delta\epsilon = -0.02$ for $\epsilon_0 = 0.3$. However, the BER performance does not vary significantly in the interval $|\Delta\epsilon| \leqslant 0.02$ in either case. It is evident that the BER curves are not symmetric about $\Delta\epsilon = 0$. When $\epsilon_0 = 0.1$, the slope of the BER curve in the region $\Delta\epsilon < 0$ is slightly higher than that in the region $\Delta\epsilon > 0$. On the other hand, when $\epsilon_0 = 0.3$, the asymmetry

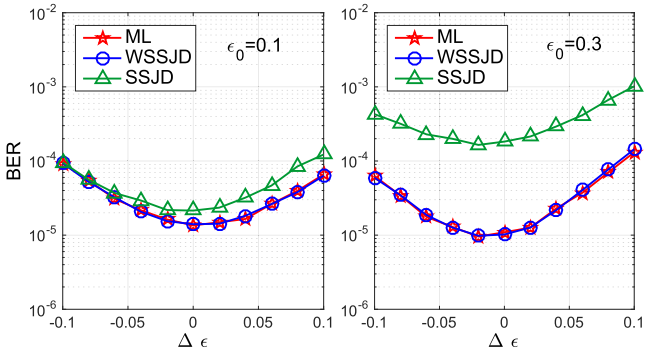


Fig. 4. ITI sensitivity of different detectors on channel $h(D) = 1 + D$ at $\epsilon_0 = 0.1$ (left) and $\epsilon_0 = 0.3$ (right). SNR=10dB.

is reversed, and the difference between the slopes is more significant.

Fig. 1 suggests that the observed behaviors are due to minimum distance properties of the mismatched detectors. To see this, consider the ML detector as an example. The probability of having an error event $(e^a(D), e^b(D))$ when the system has an offset $\Delta\epsilon$ is

$$\mathrm{P_e} = Q\left(\frac{1}{2\sigma} d(e^a, e^b, x^a, x^b)\right) = Q\left(\frac{1}{2\sigma}(d_{\text{ideal}} + d_{\text{mism}})\right), \tag{40}$$

where

$$d_{\text{ideal}} = \sqrt{\|\mathcal{A}(D)\|^2 + \|\mathcal{B}(D)\|^2}, \tag{41}$$

$$d_{\text{mism}} = 2\Delta\epsilon \frac{\langle \mathcal{A}(D), x^b(D)h(D)\rangle + \langle \mathcal{B}(D), x^a(D)h(D)\rangle}{\sqrt{\|\mathcal{A}(D)\|^2 + \|\mathcal{B}(D)\|^2}}, \tag{42}$$

$$\mathcal{A}(D) = e^a(D)h(D) + \epsilon \cdot e^b(D)h(D),$$
$$\mathcal{B}(D) = e^b(D)h(D) + \epsilon \cdot e^a(D)h(D). \tag{43}$$

Compared to the ideal case, d_{mism} is the additional effect caused by the mismatch. Notice that with the existence of mismatch, the distance parameter is now dependent on the input sequence $(x^a(D), x^b(D))$. In addition, having mismatch does not always decrease the distance. Some sequence combinations could lead to larger distance than the ideal case. The error event probability is dominated by the sequence combination (e^a, e^b, x^a, x^b) that leads to the smallest value of $d_{\text{ideal}} + d_{\text{mism}}$. Finding such a combination is not an easy task because $(e^a(D), e^b(D))$ and $(x^a(D), x^b(D))$ are not independent. For example, $e^a_k = 2$ forces x^a_k to be 1. Due to this correlation, it is hard to obtain an explicit expression for the minimum distance of a general channel polynomial. But for the channel with $h(D) = 1 + D$, we show in the Appendix that the minimum distance of the single-track error events is

$$d_{\mathrm{s}}^2 = \begin{cases} \dfrac{8(1+\epsilon_0^2 - 2\Delta\epsilon)^2}{1+\epsilon_0^2} & \text{if } \Delta\epsilon > 0 \\ \dfrac{8[1+\epsilon_0^2 + (2+2\epsilon_0)\Delta\epsilon]^2}{1+\epsilon_0^2} & \text{if } \Delta\epsilon < 0. \end{cases} \tag{44}$$

With the additional assistance of computer search, we also show that the minimum distance produced by double-track

TABLE II

SEQUENCES ACHIEVING d_{MIN} IN (46) UNDER POSITIVE/NEGATIVE OFFSETS FOR (a) SINGLE-TRACK ERROR EVENTS, AND (b) DOUBLE-TRACK ERROR EVENTS

(a) single track error events

$\Delta\epsilon < 0$	$\boldsymbol{e}^a = \cdots, \quad 0, \quad 0, \quad 2, \quad 0, \quad 0, \cdots$ $\boldsymbol{x}^a = \cdots, x^a_{k-2}, -1, +1, -1, x^a_{k+2}, \cdots$ $\boldsymbol{x}^b = \cdots, x^b_{k-2}, -1, -1, -1, x^b_{k+2}, \cdots$
$\Delta\epsilon > 0$	$\boldsymbol{e}^a = \cdots, \quad 0, \quad 0, \quad 2, \quad 0, \quad 0, \cdots$ $\boldsymbol{x}^a = \cdots, x^a_{k-2}, +1, +1, +1, x^a_{k+2}, \cdots$ $\boldsymbol{x}^b = \cdots, x^b_{k-2}, +1, +1, +1, x^b_{k+2}, \cdots$

(b) double track error events

$\Delta\epsilon < 0$	$\boldsymbol{e}^a = \cdots, \quad 0, \quad 0, \quad 2, \quad 0, \quad 0, \cdots$ $\boldsymbol{e}^b = \cdots, \quad 0, \quad 0, -2, \quad 0, \quad 0, \cdots$ $\boldsymbol{x}^a = \cdots, x^a_{k-2}, -1, +1, -1, x^a_{k+2}, \cdots$ $\boldsymbol{x}^b = \cdots, x^b_{k-2}, +1, -1, +1, x^b_{k+2}, \cdots$
$\Delta\epsilon > 0$	$\boldsymbol{e}^a = \cdots, \quad 0, \quad 0, \quad 2, \quad 0, \quad 0, \cdots$ $\boldsymbol{e}^b = \cdots, \quad 0, \quad 0, -2, \quad 0, \quad 0, \cdots$ $\boldsymbol{x}^a = \cdots, x^a_{k-2}, +1, +1, +1, x^a_{k+2}, \cdots$ $\boldsymbol{x}^b = \cdots, x^b_{k-2}, -1, -1, -1, x^b_{k+2}, \cdots$

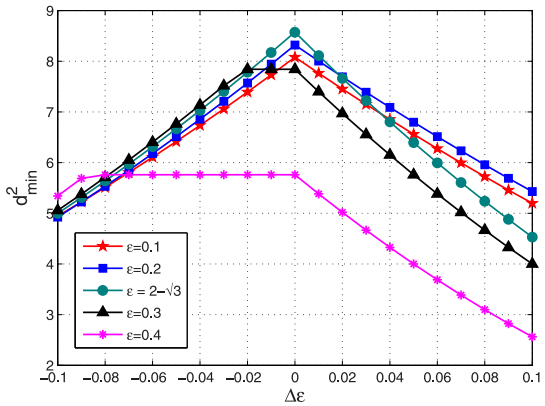


Fig. 5. Minimum distance parameter of the ML detector at different levels of mismatch for channel $1 + D$.

error events is

$$d_{\mathrm{d}}^2 = \begin{cases} 16[(1-\epsilon_0) - 2\Delta\epsilon]^2 & \text{if } \Delta\epsilon > 0 \\ 16(1-\epsilon_0)^2 & \text{if } \Delta\epsilon < 0. \end{cases} \tag{45}$$

The distance values, d_{s}^2 and d_{d}^2, are achieved by the single track error events and double track error events that minimize d_{ideal}, respectively. Table II gives examples of sequence combinations that achieve d_{s}^2 and d_{d}^2. The overall minimum distance of the system is

$$d_{\text{min}}^2 = \min\{d_{\mathrm{s}}^2, d_{\mathrm{d}}^2\}. \tag{46}$$

In summary, the asymmetry of the BER curve about $\Delta\epsilon = 0$ is due to the correlation between $(e^a(D), e^b(D))$ and $(x^a(D), x^b(D))$. The reason that minimum BER points for $\epsilon_0 = 0.1$ and $\epsilon_0 = 0.3$ occur at different values of the mismatch $\Delta\epsilon$ is because at $\epsilon_0 = 0.1$ the system performance is largely dominated by the single track error events, while at $\epsilon_0 = 0.3$ the double track error events are dominant. Fig. 5 shows the overall minimum distance d^2_{min} as a function of the mismatch $|\Delta\epsilon| \leqslant 1$ for several values of ϵ between 0.1 and 0.4. Comparing these results to the BER curves in

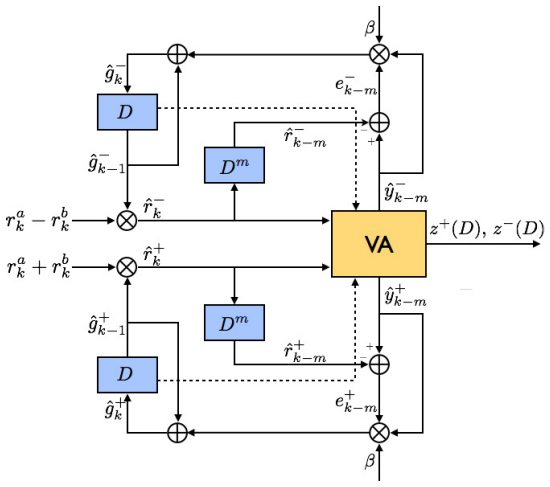
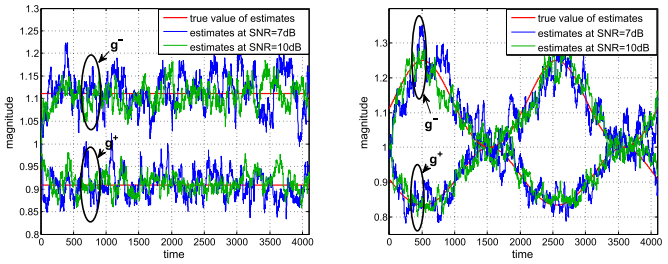


Fig. 6. WSSJD with gain loops to adaptively estimate the ITI level.



(a) $\epsilon = 0.1$ (b) $\epsilon(k) = 0.1 + 0.1\sin(4\pi k/N)$

Fig. 7. Adaptive estimation of g^+ and g^- over one sector of $N = 4096$ bits for the channel $h(D) = 1 + D$ at different SNRs. In (a), $\epsilon = 0.1$. In (b), ϵ is a sinusoidal function of time.

Fig. 4, we find that for $\epsilon_0 = 0.1$, a positive mismatch produces a higher $d^2_{\min}$ than a negative offset of the same magnitude. For $\epsilon_0 = 0.3$, this situation is reversed, and in a small range of negative offsets, $-0.02 \leqslant \Delta\epsilon \leqslant 0$, the mismatch does not reduce the minimum distance of the system. In this case, the mismatch also reduces the probability of worst case error events, leading to a shift of the minimum BER to the negative side.

B. Gain Loop

Recall that in the sum-subtract preprocessing, ϵ appears in the gain factors that normalize $r^+(D)$ and $r^-(D)$. We rewrite (16) as

$$\begin{aligned} r^+(D) &= g^+ [\, r^a(D) + r^b(D) \,], \\ r^-(D) &= g^- [\, r^a(D) - r^b(D) \,], \end{aligned} \tag{47}$$

where g^+, g^- are the gain factors with true values $\frac{1}{1+\epsilon}$ and $\frac{1}{1-\epsilon}$, respectively. There are several well-known ways to estimate g^+ and g^-[29]. In our work, we adopt the LMS adaptive algorithm due to its simplicity and good convergence properties. For $\hat{g}^+$, the updating rule is given by

$$\hat{r}_k^+ = \hat{g}_{k-1}^+ (r_k^a + r_k^b), \tag{48}$$

$$e_k = \hat{y}_k^+ - \hat{r}_k^+, \tag{49}$$

$$\hat{g}_k^+ = \hat{g}_{k-1}^+ + \beta\, \hat{y}_k^+\, e_k. \tag{50}$$

The step-size parameter β controls the convergence speed. Larger β leads to faster convergence, but also results in larger error variance. Note that $\hat{y}_k^+$ represents the instantaneous hard decision made by the Viterbi detector. The use of hard decisions can potentially lead to growing estimation error at low SNR. To mitigate this, one can introduce a small delay $m \geqslant 1$ to get more accurate tentative decisions.[1] In this case, (49) and (50) become

$$e_{k-m} = \hat{y}_{k-m}^+ - \hat{r}_{k-m}^+, \tag{51}$$

$$\hat{g}_k^+ = \hat{g}_{k-1}^+ + \beta\, \hat{y}_{k-m}^+\, e_{k-m}. \tag{52}$$

[1]The use of soft decisions would lead to even further improvement in the gain estimation, at the cost of increased complexity.

Similarly, $\hat{g}_k^-$ can be estimated in the same manner. The estimates $\hat{g}_k^+$ and $\hat{g}_k^-$ will be fed back to the Viterbi detector to evaluate path metrics, i.e.,

$$M_{k+1}(s') = M_k(s) + \hat{g}_{k-1}^+{}^2 (r_k^+ - y_k^+)^2 + \hat{g}_{k-1}^-{}^2 (r_k^- - y_k^-)^2. \tag{53}$$

Fig. 6 shows a complete block diagram for WSSJD with adaptive gain estimation. The system contains two separate gain loops for $\hat{g}_k^+$ and $\hat{g}_k^-$. While a combined loop for estimating $\hat{g}_k^+$ and $\hat{g}_k^-$ can provide a better estimate for ϵ, using separate loops achieves similar performance in a more efficient way.

In our simulations, $\hat{g}_0^+$ and $\hat{g}_0^-$ are initially set to 1. At time k, $r_k^a + r_k^b$ and $r_k^a - r_k^b$ are normalized by the previously estimated $\hat{g}_{k-1}^+$ and $\hat{g}_{k-1}^-$, respectively. The resulting signals $\hat{r}_k^+$ and $\hat{r}_k^-$ are sent to the Viterbi detector. The path metric of each trellis state is calculated by using weights $\hat{g}_{k-1}^+$ and $\hat{g}_{k-1}^-$. The Viterbi detector picks the most likely path, and makes an instantaneous decision on $\hat{y}_{k-m}^+$ and $\hat{y}_{k-m}^-$. The error signal is calculated to update $\hat{g}_k^+$ and $\hat{g}_k^-$. Note that SSJD can also work with these gain loops, without feeding $\hat{g}_k^+$ and $\hat{g}_k^-$ to the path metric evaluation.

In Fig. 7, we track the values of g_k^+ and g_k^- estimated by gain loops in one sector of $N = 4096$ bits on the channel with $h(D) = 1 + D$ at high and low SNRs. The true values of the gain factors, $g^+ = \frac{1}{1+\epsilon}$ and $g^- = \frac{1}{1-\epsilon}$, are also plotted for comparison. The step-size β is set to 0.005, and the delay unit $m = 5$. In Fig. 7(a), ϵ has the fixed value 0.1, while in Fig. 7(a) ϵ slowly varies around the value 0.1. We see that gain factors are well tracked by the gain loops. The estimates at high SNR (10dB) show better convergence than those in a low SNR (7dB) environment.

V. WSSJD on General ITI Channel

A. Decomposition of Interference Matrix

To generalize WSSJD to the nHnT model, consider the eigen-decomposition of $\mathbf{A}_n$ in (5),

$$\mathbf{A}_n = \mathbf{V}_n \mathbf{\Lambda}_n \mathbf{V}_n^\top, \tag{54}$$

where $\mathbf{V}_n$ is an $n \times n$ matrix whose columns are the eigenvectors of $\mathbf{A}_n$, and $\mathbf{\Lambda}_n$ is a diagonal matrix whose diagonal elements are the corresponding eigenvalues. The eigenvalues

and eigenvectors of the symmetric tridiagonal Toeplitz matrix have a known closed form [29] [30]. If we define

$$\hat{\mathbf{T}}_n = \begin{bmatrix} 0 & 1 & & O \\ 1 & 0 & \ddots & \\ & \ddots & \ddots & 1 \\ O & & 1 & 0 \end{bmatrix}, \tag{55}$$

then

$$\mathbf{A}_n = \mathbf{I}_n + \epsilon \hat{\mathbf{T}}_n = \mathbf{V}_n(\mathbf{I}_n + \epsilon \hat{\mathbf{\Lambda}}_n)\mathbf{V}_n^\top, \tag{56}$$

where $\mathbf{I}_n$ is an $n \times n$ identity matrix, and $\hat{\mathbf{\Lambda}}_n$ is the diagonal matrix containing the eigenvalues of $\hat{\mathbf{T}}_n$. Therefore, the columns of $\mathbf{V}_n$ are also the eigenvectors of $\hat{\mathbf{T}}_n$, and $\mathbf{\Lambda}_n = \mathbf{I}_n + \epsilon \hat{\mathbf{\Lambda}}_n$. In fact, both $\hat{\mathbf{\Lambda}}_n$ and $\mathbf{V}_n$ have closed forms: the k^{th} eigenvalue of $\hat{\mathbf{T}}_n$ is

$$\hat{\lambda}_k = 2\cos\left(\frac{k\pi}{n+1}\right), \tag{57}$$

and the j^{th} element in the k^{th} eigenvector $\boldsymbol{v}_k$ is

$$v_{jk} = \sqrt{\frac{2}{n+1}} \sin\left(\frac{kj\pi}{n+1}\right). \tag{58}$$

Note that $\mathbf{V}_n$ is independent of ϵ.

Example 1:. For the case $n = 2$,

$$\mathbf{\Lambda}_2 = \begin{bmatrix} 1+\epsilon & 0 \\ 0 & 1-\epsilon \end{bmatrix}, \quad \mathbf{V}_2 = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}.$$

Example 2:. For the case $n = 3$,

$$\mathbf{\Lambda}_3 = \begin{bmatrix} 1+\sqrt{2}\epsilon & 0 & 0 \\ 0 & q1 & 0 \\ 0 & 0 & 1-\sqrt{2}\epsilon \end{bmatrix},$$

$$\mathbf{V}_3 = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{2}}{2} & \frac{1}{2} \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \\ \frac{1}{2} & -\frac{\sqrt{2}}{2} & \frac{1}{2} \end{bmatrix}.$$

B. Channel Decomposition and Generalized WSSJD

The nHnT model is given by (1). Using the decomposition of $\mathbf{A}_n$ in (54), we can express the channel output as

$$\mathbf{R}(D) = \mathbf{V}_n \mathbf{\Lambda}_n \mathbf{V}_n^\top \mathbf{X}(D)h(D) + \mathbf{\Omega}(D). \tag{59}$$

Reorganizing (59) gives

$$\mathbf{\Lambda}_n^{-1}\mathbf{V}_n^\top \mathbf{R}(D) = \mathbf{V}_n^\top \mathbf{X}(D)\, h(D) + \mathbf{\Lambda}_n^{-1}\mathbf{V}_n^\top \mathbf{\Omega}(D). \tag{60}$$

Let $\bar{\mathbf{X}}(D) = \mathbf{V}_n^\top \mathbf{X}(D)$, $\bar{\mathbf{R}}(D) = \mathbf{\Lambda}_n^{-1}\mathbf{V}_n^\top \mathbf{R}(D)$ and $\bar{\mathbf{\Omega}}(D) = \mathbf{\Lambda}_n^{-1}\mathbf{V}_n^\top \mathbf{\Omega}(D)$ be the vectors of transformed input sequences, received sequences and noises, respectively. The new channel model becomes

$$\bar{\mathbf{R}}(D) = \bar{\mathbf{X}}(D)h(D) + \bar{\mathbf{\Omega}}(D), \tag{61}$$

which is composed of n parallel channels. The j^{th} channel is obtained by considering the j^{th} row of both sides of (61), which gives

$$\bar{r}^j(D) = \bar{x}^j(D)h(D) + \bar{\omega}^j(D), \tag{62}$$

where

$$\bar{r}^j(D) = \frac{1}{1+\epsilon\hat{\lambda}_j} \sum_{i=1}^{n} v_{ij}\, r^i(D), \tag{63}$$

$$\bar{x}^j(D) = \sum_{i=1}^{n} v_{ij}\, x^i(D), \tag{64}$$

$$\bar{w}^j(D) = \frac{1}{1+\epsilon\hat{\lambda}_i} \sum_{i=1}^{n} v_{ij}\, \omega^i(D). \tag{65}$$

Several properties of the transformed channel model can be observed:

1) The noise components in $\bar{\mathbf{\Omega}}(D)$ are independent. Let $\boldsymbol{\omega}_k$ and $\bar{\boldsymbol{\omega}}_k$ be length-n vectors of the original and transformed noise samples at time k, i.e., the coefficients of D^k in the sequences $\boldsymbol{\omega}(D)$ and $\bar{\boldsymbol{\omega}}(D)$, respectively. Then

$$E[\bar{\boldsymbol{\omega}}_k \bar{\boldsymbol{\omega}}_k^\top] = E[\mathbf{\Lambda}_n^{-1}\mathbf{V}_n^\top \boldsymbol{\omega}_k \boldsymbol{\omega}_k^\top \mathbf{V}_n \mathbf{\Lambda}_n^{-1}] = (\sigma\, \mathbf{\Lambda}_n^{-1})^2, \tag{66}$$

which is a diagonal matrix. So the components of $\bar{\boldsymbol{\omega}}_k$ are uncorrelated and Gaussian, therefore independent. Furthermore, the noise power of the j^{th} channel is σ^2/λ_j^2.

2) After the transformation, the inputs of different component channels have different alphabets. For the j^{th} component channel, the alphabet Σ_j is

$$\Sigma_j = \left\{ \sum_{i=1}^{n} v_{ij} x_i | x_i \in \{+1, -1\} \right\}. \tag{67}$$

3) The j^{th} component channel corresponds to transmitting $\bar{x}^j(D)$ through the ISI channel $h(D)$ and adding electronic noise of power σ^2/λ_j^2. Since the inputs to different channels are correlated, a joint trellis is needed to search for the optimal decision. The new trellis states can be found by applying the one-to-one mapping to the conventional ML states. Therefore, the WSSJD trellis has 2^{nv} states.
4) Since $\mathbf{V}_n$ is determined once n is given, the WSSJD trellis is well-defined, and the branch labels are also independent of ϵ.

The optimal decision $\bar{\mathbf{X}}^*(D)$ satisfies

$$\begin{aligned} \bar{\mathbf{X}}^*(D) &= \arg\max_{\bar{\mathbf{X}}(D)} \log \Pr(\bar{\mathbf{R}}(D)|\bar{\mathbf{X}}(D)) \\ &= \arg\min_{\bar{\mathbf{X}}(D)} \sum_{j=1}^{n} \lambda_j^2 \, \|\bar{r}^j(D) - \bar{x}^j(D)h(D)\|^2. \end{aligned} \tag{68}$$

It is easy to see that WSSJD gives the optimal ML solution.

For a given error event $\bar{\boldsymbol{e}}(D) = [\bar{e}^1(D), \ldots, \bar{e}^n(D)]$, where $\bar{e}^j(D)$ is the error sequence on the j^{th} component channel, the distance parameter is given by

$$d^2(\bar{\boldsymbol{e}}(D)) = \sum_j \lambda_j^2 \, \|\bar{e}^j(D)h(D)\|^2. \tag{69}$$

Algorithm 1 WSSJD With Gain Loop on nHnT

function $\hat{\mathbf{X}}(D)$ =WSSJD($\mathbf{R}(D), \epsilon_0$)
Initialize:
$M(0) = 0$,
$M(p) = \infty$ for $p = 1, \cdots, 2^{nv} - 1$ ▷ path metric
Ψ is a $2^{nv} \times N$ zero matrix ▷ path history
$\mathbf{G} = (\mathbf{I}_n + \epsilon_0 \hat{\mathbf{\Lambda}}_n)^{-1}$ ▷ gain factors
Begin:
for $k = 1$ to N **do**
  $\bar{\boldsymbol{r}}_k = \mathbf{G}\mathbf{V}_n^\top \boldsymbol{r}_k$
  for $p = 0$ to $2^{nv} - 1$ **do**
    for each predecessor state q_i of p
    $M_i = M(q_i) + (\bar{\boldsymbol{r}}_k - \boldsymbol{y}^{(q_i,p)})^\top \mathbf{G}^{-2}(\bar{\boldsymbol{r}}_k - \boldsymbol{y}^{(q_i,p)})$
    update $M(p) = \min_i M_i$
    $\Psi(p,k) = q_i$ ▷ extend survivor path
  end for
  if $k > m$ **then**
    $p^* = \arg\min_p M(p)$
    for $j = 1$ to δ **do**
      $p^* = \Psi(p^*, k - j + 1)$;
    end for ▷ trace back path history
    $\boldsymbol{e} = \boldsymbol{y}^{(\Psi(p^*,k-\delta),p^*)} - \bar{\boldsymbol{r}}_{k-\delta}$
    $\mathbf{G} = \mathbf{G} + \beta \text{diag}(\boldsymbol{y}^{(\Psi(p^*,k-\delta),p^*)})\text{diag}(\boldsymbol{e})$
  end if
end for
end

C. Gain Loops

As shown in (62)-(65), for each channel the ITI level ϵ appears in a gain factor normalizing $\sum_{i=1}^{n} v_{ij}\, r^i(D)$ such that its expectation is $\bar{x}(D)h(D)$. Gain loops can be used to adaptively estimate these gain factors.

Let g_k^j denote the gain factor estimated for the j^{th} channel at time k. Then $E[g_k^j] = \frac{1}{1+\epsilon\hat{\lambda}_j}$. Again, we use the LMS adaptive algorithm to update g_k^j, according to the equations

$$\hat{r}_k^j = g_{k-1}^j \sum_i v_{ij} r_k^i, \tag{70}$$

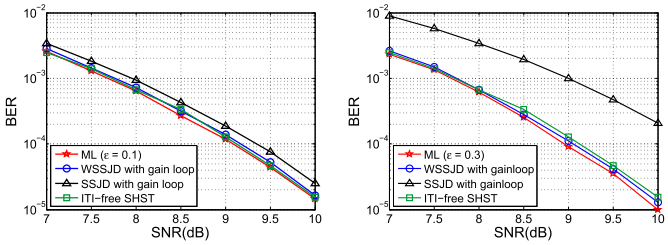
$$\hat{e}_{k-m}^j = \hat{y}_{k-m}^j - \hat{r}_{k-m}^j, \tag{71}$$

$$g_k^j = g_{k-1}^j + \beta \hat{y}_{k-m}^j \hat{e}_{k-m}^j, \tag{72}$$

where $\hat{y}_{k-m}^j$ is the instantaneous hard decision on the noiseless output of the j^{th} ISI channel at time $k-m$. To find $\hat{y}_{k-m}^j$, we identify the trellis state which currently has the smallest path metric, and trace back the path history for m time slots to obtain the corresponding channel output. The gain factors g_k^j are also used in weighting the path metric, to ensure the ML performance of the detector.

Algorithm 1 summarizes the procedures to implement WSSJD with gain loops on the nHnT channel. The algorithm makes use of the following notation:

1) N is the length of one frame; ϵ_0 is the initial value of ϵ.
2) $\mathbf{G}$ is a diagonal matrix with g_k^j as the j^{th} diagonal element.
3) $\boldsymbol{r}_k$ is a column vector of the received signals from $nHnT$ channel at time k; $\bar{\boldsymbol{r}}_k$ is the vector of outputs from the transformed channel.
4) $M(p)$ is the accumulated path metric at state p.
5) $\boldsymbol{y}^{(q,p)}$ is a column vector of the trellis output label from state q to p.
6) diag($\boldsymbol{v}$) is a diagonal matrix with the elements of the vector $\boldsymbol{v}$ along the diagonal.



(a) $\epsilon = 0.1$ (b) $\epsilon = 0.3$

Fig. 8. BER vs. SNR of different detectors with static ITI level (a) $\epsilon = 0.1$ and (b) $\epsilon = 0.3$.

VI. Simulation Results

In this section, we present BER performance simulation results for WSSJD and SSJD on the nHnT channel model for several values of n. We plot the BER as a function of channel SNR, where we define SNR as

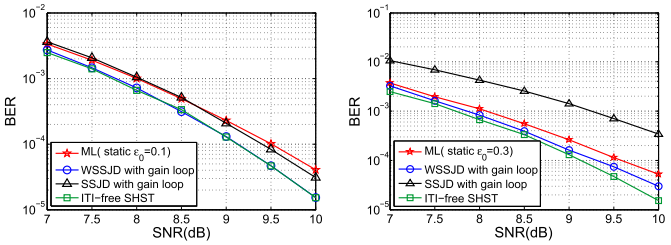
$$\text{SNR(dB)} = 10 \log \frac{\|h(D)\|^2}{2\sigma^2}. \tag{73}$$

Note that, in contrast to some studies such as [12], our definition of SNR does not take the energy $\epsilon^2\|h(D)\|^2$ in the cross-track signal into account. Since we use separate figures for the results corresponding to different static ITI values, this definition should not lead to any confusion. It is also better suited for presenting the results obtained when the ITI is characterized by a small variation around a fixed nominal ITI level.

A. 2H2T System

We simulate WSSJD and SSJD with gain loops on the 2H2T system with channel polynomial $h(D) = 1 + D$. We set $\beta = 0.008$ and $m = 5$. The initial values of gain factors g_0^+ and g_0^- are obtained by passing training samples through the system.

We first test the proposed detectors and gain loop structure when ϵ is fixed. In Fig. 8, we compare the BER performances of WSSJD, SSJD, and the conventional 2H2T ML detector, for $\epsilon = 0.1$ and $\epsilon = 0.3$. The frame size is $N = 4096$ bits. In our setup, the ML detector knows the value of ϵ, while WSSJD and SSJD adaptively estimate gain factors using the gain loop structure shown in Fig. 6. Therefore, the static ML detector provides the optimal BER performance. It is observed that the BER curve of WSSJD with a gain loop almost coincides with that of the static ML detector. This indicates that the LMS adaptive algorithm provides sufficiently accurate estimates of



(a) (b)

Fig. 9. BER vs. SNR of different detectors with ϵ slowly varying about the mean value (a) $\epsilon_0 = 0.1$ and (b) $\epsilon_0 = 0.3$.

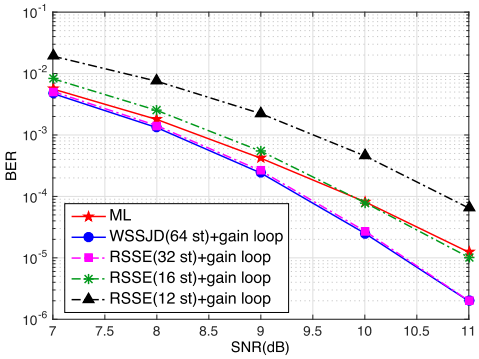


Fig. 10. Performance of reduced complexity implementations of WSSJD with gain loop on 2H2T EPR4 channel. ϵ is sinusoidally varying with mean value $\epsilon_0 = 0.1$.

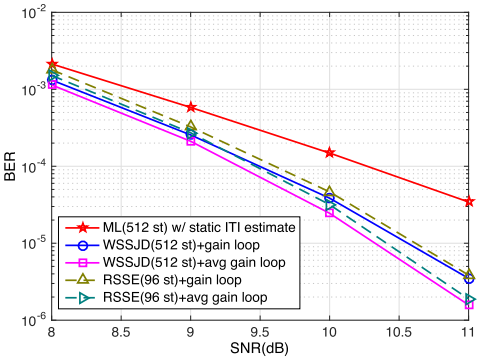


Fig. 11. BER performance of WSSJD on 3H3T EPR4 channel ($h(D) = 1 + D - D^2 - D^3$) under sinusoidally varying ITI with mean value $\epsilon_0 = 0.1$.

gain factors. The recursive least squares (RLS) algorithm can speed up the convergence of the gain loop at the expense of higher complexity, but the BER performance improvement would be negligible. As expected from the minimum distance plots in Fig. 1, the performance loss suffered by SSJD relative to WSSJD is more severe at $\epsilon = 0.3$ than at $\epsilon = 0.1$. Although not shown here, simulation results for the frame error rate vs. SNR correlate well with the BER results.

Next, we test the performance of the detectors by assuming a dynamic ITI model in which ϵ changes slowly with respect to the location k. Specifically, we set

$$\epsilon(k) = \epsilon_0 + 0.1 \sin(4\pi k/N), \tag{74}$$

where $N = 4096$ and ϵ_0 is the mean value. In this case, we compare the adaptive WSSJD algorithm, in which gain loops are used to track the value of ϵ, with the ML detector which uses the static value ϵ_0. The simulation results, shown in Figs. 9, show that the adaptive WSSJD outperforms the static ML detectiion by about 0.3-0.5dB at high SNR.

In both cases, the performance of the ITI-free SHST detector is plotted for comparison. It is interpreted as the performance of the ideal ITI cancellation scheme, where the ITI is completely removed from the readback signals of each track.

The trellis complexity of the MHMT ML detector can be prohibitively complex for practical purposes. We address this problem in [27] and [28] by using a reduced-state trellis with the RSSE detection algorithm. Moreover, the gain loop structure for ITI estimation in WSSJD can be directly applied in the RSSE setting. Fig. 10 shows BER simulation results for WSSJD with several RSSE implementations of various trellis-complexities on the extended class-IV partial response (EPR4) channel, with channel polynomial $h(D) = 1 + D - D^2 - D^3$. We see that, as expected, WSSJD with a 64-state trellis and adaptive ITI estimation outperforms the static ML detector with the same number of states. Results are also shown for RSSE using 12, 16, and 32 trellis states, along with gain estimation loops. We can see that RSSE can achieve the same performance as WSSJD using only one-half the number of trellis states (32 vs. 64).

B. 3H3T System

In Fig. 11, we plot the BER performance of several detectors on the 3H3T channel. We assume that each component channel is equalized to the EPR4 target. The WSSJD detector therefore requires 512 trellis states. As the matrix decomposition described in Example 2 indicates, the ITI estimation for the transformed 3H3T system requires only 2 gain loops, one for each of the first and the third component channels. We observe in the figure that WSSJD with ITI estimation outperforms the static ML algorithm by about 1dB when the BER is in the range $[10^{-5}, 10^{-4}]$. The performance is further improved by averaging the two gain factors to get a better estimate of ϵ. Finally, we see that an RSSE implementation of WSSJD with only 96 states achieves performance comparable to that of the full 512-state WSSJD implementation.

VII. Conclusion

We study MHMT detection as a potential candidate for next-generation magnetic recording. We assume a simplified symmetric nHnT model in which only adjacent tracks interfere. The ITI is assumed to be linear and additive, and its amplitude is controlled by a scalar. In this paper, we consider how to estimate the ITI and how to efficiently use the new estimates to improve the detector performance. By means of a channel transformation, we decompose the original nHnT system into n separate subchannels, each of which has a channel gain factor that depends on the ITI level. Based on this transformed

system, we propose a novel detection method which we call weighted sum-subtract joint detection (WSSJD). The new method achieves ML performance. However, in contrast to the conventional MHMT ML detector trellis, the branch labels of the WSSJD trellis do not depend on the ITI level. Instead, the value of ITI level is incorporated into a weighting factor used to compute the path metrics. A simple gain loop structure is described, permitting efficient adaptive estimation of the gain factors and their straightforward incorporation into the WSSJD algorithm. Simulation results show that when ITI is time-varying, WSSJD with adaptive ITI estimation outperforms the ML detection using static ITI estimates. Minimum distance analysis is presented, providing a theoretical basis for the comparison of different MHMT detectors. The WSSJD technique is also amenable to a reduced complexity implementation based upon RSSE, which is the focus of Part II of this work.

APPENDIX
MINIMUM DISTANCE ANALYSIS FOR ITI SENSITIVITY

In this section we give the derivation of (44) and (45).

A. Single Track Error Events

Assume $e^b(D) = 0$. The distance components reduce to

$$d_{\text{ideal}} = \sqrt{(1+\epsilon_0^2)\|e^a(D)h(D)\|^2},$$
$$d_{\text{mism}} = 2\Delta\epsilon \times \frac{\left\langle e^a(D)h(D), x^b(D)h(D)\right\rangle + \epsilon_0 \left\langle e^a(D)h(D), x^a(D)h(D)\right\rangle}{\sqrt{(1+\epsilon_0^2)\|e^a(D)h(D)\|^2}}.$$

We bound d_{mism} as follows:

$$\begin{aligned}
&\left\langle e^a(D)h(D), x^b(D)h(D)\right\rangle \\
&= \sum\nolimits_n \left(\sum\nolimits_m x^b_{n-m}h_m\right)\left(\sum\nolimits_m e^a_{n-m}h_m\right) \\
&\leqslant \left|\sum\nolimits_n \left(\sum\nolimits_m x^b_{n-m}h_m\right)\left(\sum\nolimits_m e^a_{n-m}h_m\right)\right| \\
&\leqslant \sum\nolimits_n \left|\sum\nolimits_m x^b_{n-m}h_m\right| \cdot \left|\sum\nolimits_m e^a_{n-m}h_m\right| \\
&\leqslant M_h \sum\nolimits_n \left|\sum\nolimits_m e^a_{n-m}h_m\right| \\
&= 2M_h \sum\nolimits_n \left|\sum\nolimits_m \frac{e^a_{n-m}}{2}h_m\right| \\
&\leqslant 2M_h \sum\nolimits_n \left(\sum\nolimits_m \frac{e^a_{n-m}}{2}h_m\right)^2 \\
&= \frac{M_h}{2}\|e^a(D)h(D)\|^2, \qquad (75)
\end{aligned}$$

where $M_h = \sum_m |h_m| = 2$ for the channel $h(D) = 1 + D$. Using a similar derivation, we can show

$$\left\langle e^a(D)h(D), x^b(D)h(D)\right\rangle \geqslant -\frac{M_h}{2}\|e^a(D)h(D)\|^2. \qquad (76)$$

To find the bounds for $\langle e^a(D)h(D), x^a(D)h(D)\rangle$, note that

$$\begin{aligned}
&\left\langle e^a(D)h(D), x^a(D)h(D)\right\rangle \\
&= \sum\nolimits_k (e^a_{k-1} + e^a_k)(x^a_{k-1} + x^a_k) \\
&= \sum_{k=k_1}^{k_2+1} (e^a_{k-1}x^a_{k-1} + e^a_{k-1}x^a_k + e^a_k x^a_{k-1} + e^a_k x^a_k) \qquad (77) \\
&\geqslant \sum_{k=k_1}^{k_2+1} \left(|e^a_{k-1}| - |e^a_{k-1}| - |e^a_k| + |e^a_k|\right) \qquad (78) \\
&= 0.
\end{aligned}$$

The inequality in (78) follows from the fact that x^a_k always has the same sign as e^a_k, so $e^a_k x^a_k = |e^a_k|$. Choosing x^a_k to have the opposite sign to e^a_{k-1} leads to the lower bound $e^a_{k-1}x^a_k \geqslant -|e^a_{k-1}|$.

The upper bound derived in (75) is also applicable to $\langle e^a(D)h(D), x^a(D)h(D)\rangle$. Therefore,

$$0 \leqslant \left\langle e^a(D)h(D), x^a(D)h(D)\right\rangle \leqslant \|e^a(D)h(D)\|^2. \qquad (79)$$

Combining (75) and (79), and using $\|e^a(D)h(D)\|^2 \geqslant 8$ for the channel $h(D) = 1 + D$, we can bound the single-track minimum distance in the two cases of $\Delta\epsilon > 0$ and $\Delta\epsilon < 0$ as follows:

$$\begin{aligned}
d_{\text{s}} &= d_{\text{ideal}} + d_{\text{mism}} \\
&\geqslant \left(\sqrt{1+\epsilon_0^2} - \frac{2\Delta\epsilon}{\sqrt{1+\epsilon_0^2}}\right)\|e^a(D)h(D)\| \\
&\geqslant \frac{2\sqrt{2}(1+\epsilon_0^2 - 2\Delta\epsilon)}{\sqrt{1+\epsilon_0^2}}, \quad \text{if } \Delta\epsilon > 0,
\end{aligned}$$

and

$$\begin{aligned}
d_{\text{s}} &= d_{\text{ideal}} + d_{\text{mism}} \\
&\geqslant \left(\sqrt{1+\epsilon_0^2} + \frac{2\Delta\epsilon(1+\epsilon_0)}{\sqrt{1+\epsilon_0^2}}\right)\|e^a(D)h(D)\| \\
&\geqslant \frac{2\sqrt{2}(1+\epsilon_0^2 + 2(1+\epsilon_0)\Delta\epsilon)}{\sqrt{1+\epsilon_0^2}}, \quad \text{if } \Delta\epsilon < 0.
\end{aligned}$$

Table II(a) gives examples of error events that achieve these bounds.

B. Double Track Error Events

In this case, both $e^a(D)$ and $e^b(D)$ are non-zero at some locations. To find an achievable bound on $d_{\text{ideal}} + d_{\text{mism}}$, we assume $\Delta\epsilon \ll 1$. Therefore the distance increment/decrement caused by the mismatch will not be as significant as the distance in the ideal case. The minimum value of d_{ideal} given by (41) is $4(1-\epsilon_0)$, achieved by the error sequences of the form

$$\begin{aligned}
\boldsymbol{e}^a &= [0, \cdots, 0, e^a_{k_1}, \cdots, e^a_{k_2}, \cdots, 0], \\
\boldsymbol{e}^b &= [0, \cdots, 0, e^b_{k_1}, \cdots, e^b_{k_2}, \cdots, 0],
\end{aligned}$$

with $e^a_{k+1} = -e^a_k$ for $k_1 \leqslant k \leqslant k_2 - 1$, and $e^b_k = -e^a_k$ for $k_1 \leqslant k \leqslant k_2$. The assumption on $\Delta\epsilon$ suggests that we focus on these error events. We use d^*_{mism} to denote the minimum distance parameter attained by this subset of double track error events.

We can express $\langle \mathcal{A}(D), x^b(D)h(D)\rangle$ by

$$\begin{aligned}
&\langle \mathcal{A}(D), x^b(D)h(D)\rangle \\
&= \sum_{k=k_1}^{k_2+1} [e^a_k + e^a_{k-1} + \epsilon_0(e^b_k + e^b_{k-1})](x^b_k + x^b_{k-1}) \\
&= (e^a_{k_1} + \epsilon_0 e^b_{k_1})(x^b_{k_1} + x^b_{k_1-1}) + (e^a_{k_2} + \epsilon_0 e^b_{k_2})(x^b_{k_2+1} + x^b_{k_2}) \\
&= -|e^a_{k_1}| + \epsilon_0 |e^b_{k_1}| + (e^a_{k_1} + \epsilon_0 e^b_{k_1}) x^b_{k_1-1} \\
&\quad - |e^a_{k_2}| + \epsilon_0 |e^b_{k_2}| + (e^a_{k_2} + \epsilon_0 e^b_{k_2}) x^b_{k_2+1}. \qquad (80)
\end{aligned}$$

Upper and lower bounds for (80) can be found by carefully choosing values for $x^b_{k_1-1}$ and $x^b_{k_2+1}$. If $x^b_{k_1-1}$ and $x^b_{k_2+1}$ have the same sign as $e^a_{k_1}$ and $e^a_{k_2}$, respectively, (80) achieves the maximum value 0. If $x^b_{k_1-1}$ and $x^b_{k_2+1}$ have the same sign as $e^b_{k_1}$ and $e^b_{k_2}$, respectively, (80) achieves the minimum value $8(\epsilon_0 - 1)$. Similarly, we have

$$8(\epsilon_0 - 1) \leqslant \langle \mathcal{B}(D), x^a(D)h(D)\rangle \leqslant 0. \qquad (81)$$

We conclude that in the case of $\Delta\epsilon > 0$ and $\Delta\epsilon < 0$,

$$\begin{aligned}
d^*_{\text{d}} \stackrel{\text{def}}{=} d_{\text{ideal}} + d^*_{\text{mism}} &\geqslant 4(1-\epsilon_0) + \frac{2\Delta\epsilon}{4(1-\epsilon_0)} \cdot 16(\epsilon_0 - 1) \\
&= 4(1 - \epsilon_0 - 2\Delta\epsilon), \quad \text{if } \Delta\epsilon > 0,
\end{aligned}$$

and

$$d^*_{\text{d}} \stackrel{\text{def}}{=} d_{\text{ideal}} + d^*_{\text{mism}} \geqslant 4(1-\epsilon_0), \quad \text{if } \Delta\epsilon < 0.$$

Notice that these bounds are derived for a subset of double track error events which achieve $\min d_{\text{ideal}}$, examples of which are given in Table II.

We compared the values of $d^2_{\text{min}} = \min\{d^2_{\text{s}}, d^2_{\text{d}}\}$ obtained by exhaustive computer search with $\min\{(d^*_{\text{s}})^2, (d^*_{\text{d}})^2\}$, and they agreed at all points plotted in Fig. 5. This supports the assumption that the simplification in our analysis of double track error events does not affect the d^2_{min} computation.

REFERENCES

[1] Y. Shiroishi *et al.*, "Future options for HDD storage," *IEEE Trans. Magn.*, vol. 45, no. 10, pp. 3816–3822, Oct. 2009.

[2] R. Wood, M. Williams, A. Kavcic, and J. Miles, "The feasibility of magnetic recording at 10 Terabits per square inch on conventional media," *IEEE Trans. Magn.*, vol. 45, no. 2, pp. 917–923, Feb. 2009.

[3] S. Karakulak, P. H. Siegel, J. K. Wolf, and H. N. Bertram, "Joint-track equalization and detection for bit patterned media recording," *IEEE Trans. Magn.*, vol. 46, no. 9, pp. 3639–3647, Sep. 2010.

[4] J. Park, B. Lengsfield, R. Galbraith, R. Wood, and S. Fu, "Optimization of magnetic read widths in two-dimensional magnetic recording based on micromagnetic simulations," *IEEE Trans. Magn.*, vol. 51, no. 11, pp. 1–4, Nov. 2015.

[5] P. W. Nutter, I. T. Ntokas, and B. K. Middleton, "An investigation of the effects of media characteristics on read channel performance for patterned media storage," *IEEE Trans. Magn.*, vol. 41, no. 11, pp. 4327–4334, Nov. 2005.

[6] B. G. Roh, S. U. Lee, J. Moon, and Y. Chen, "Single-head/single-track detection in interfering tracks," *IEEE Trans. Magn.*, vol. 38, no. 4, pp. 1830–1838, Jul. 2002.

[7] E. Hwang, R. Negi, and B. V. K. V. Kumar, "Signal processing for near 10 Tbit/in^2 density in two-dimensional magnetic recording (TDMR)," *IEEE Trans. Magn.*, vol. 46, no. 6, pp. 1813–1816, Jun. 2010.

[8] E. Ordentlich and R. M. Roth, "Two-dimensional maximum-likelihood sequence detection is NP hard," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 7661–7670, Dec. 2011.

[9] M. Marrow and J. K. Wolf, "Iterative detection of 2-dimensional ISI channels," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Mar. 2003, pp. 131–134.

[10] Y. Chen and S. G. Srinivasa, "Joint self-iterating equalization and detection for two-dimensional intersymbol-interference channels," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3219–3230, Aug. 2013.

[11] M. Mehrnoush, B. J. Belzer, K. Sivakumar, and R. Wood, "Turbo equalization for two dimensional magnetic recording using Voronoi model averaged statistics," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 9, pp. 2439–2449, Sep. 2016.

[12] Y. Wu, J. A. O'Sullivan, N. Singla, and R. S. Indeck, "Iterative detection and decoding for separable two-dimensional intersymbol interference," *IEEE Trans. Magn.*, vol. 39, no. 4, pp. 2115–2120, Jul. 2003.

[13] T. Cheng, B. J. Belzer, and K. Sivakumar, "Row-column soft-decision feedback algorithm for two-dimensional intersymbol interference," *IEEE Signal Process. Lett.*, vol. 14, no. 7, pp. 433–436, Jul. 2007.

[14] Y. Chen, P. Njeim, T. Cheng, B. J. Belzer, and K. Sivakumar, "Iterative soft decision feedback zig-zag equalizer for 2D intersymbol interference channels," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 2, pp. 167–180, Feb. 2010.

[15] S. M. Khatami and B. Vasić, "Generalized belief propagation detector for TDMR microcell model," *IEEE Trans. Magn.*, vol. 49, no. 7, pp. 3699–3702, Jul. 2013.

[16] A. Amer, J. Holliday, D. D. E. Long, E. L. Miller, J. Paris, and T. Schwarz, "Data management and layout for shingled magnetic recording," *IEEE Trans. Magn.*, vol. 47, no. 10, pp. 3691–3697, Oct. 2011.

[17] Y. Wang and B. V. K. V. Kumar, "Multi-track joint detection for shingled magnetic recording on bit patterned media with 2-D sectors," *IEEE Trans. Magn.*, vol. 52, no. 7, pp. 1–7, Jul. 2016.

[18] N. Kumar, J. Bellorado, M. Marrow, and K. K. Chan, "Inter-track interference cancelation in presence of frequency offset for shingled magnetic recording," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 4342–4346.

[19] M. Fujii and N. Shinohara, "Multi-track iterative ITI canceller for shingled write recording," in *Proc. Int. Symp. Commun. Inf. Technol. (ISCIT)*, Tokyo, Japan, Oct. 2010, pp. 1062–1067.

[20] E. B. Sadeghian and J. R. Barry, "Soft intertrack interference cancellation for two-dimensional magnetic recording," *IEEE Trans. Magn.*, vol. 51, no. 6, pp. 1–9, Jun. 2015.

[21] L. C. Barbosa, "Simultaneous detection of readback signals from interfering magnetic recording tracks using array heads," *IEEE Trans. Magn.*, vol. 26, no. 5, pp. 2163–2165, Sep. 1990.

[22] E. Soljanin and C. N. Georghiades, "Multihead detection for multitrack recording channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2988–2997, Nov. 1998.

[23] X. Ma and L. Ping, "Iterative detection/decoding for two-track partial response channels," *IEEE Commun. Lett.*, vol. 8, no. 7, pp. 464–466, Jul. 2004.

[24] N. Zheng, K. S. Venkataraman, A. Kavcic, and T. Zhang, "A study of multitrack joint 2-D signal detection performance and implementation cost for shingled magnetic recording," *IEEE Trans. Magn.*, vol. 50, no. 6, pp. 1–6, Jun. 2014.

[25] E. Soljanin and C. N. Georghiades, "On coding in multi-track, multi-head, disk recording systems," in *Proc. IEEE Houston Global Telecommun. Conf. Including Commun. Theory Mini-Conf. Tech. Program Conf. Rec. (GLOBECOM)*, vol. 4, Nov. 1993, pp. 18–22.

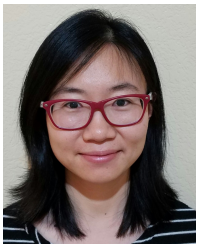
[26] B. Fan, H. K. Thapar, and P. H. Siegel, "Multihead multitrack detection in shingled magnetic recording with ITI estimation," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 425–430.

[27] B. Fan, H. K. Thapar, and P. H. Siegel, "Multihead multitrack detection with reduced-state sequence estimation in shingled magnetic recording," in *Proc. IEEE Magn. Conf. (INTERMAG)*, Beijing, China, May 2015, p. 1.

[28] B. Fan, H. K. Thapar, and P. H. Siegel, "Multihead multitrack detection with reduced-state sequence estimation," *IEEE Trans. Magn.*, vol. 51, no. 11, pp. 1–4, Nov. 2015.

[29] G. D. Smith, *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, 3rd ed. Oxford, U.K.: Clarendon, 1986.

[30] S. Noschese, L. Pasquini, and L. Reichel, "Tridiagonal Toeplitz matrices: Properties and novel applications," *Numer. Linear Algebra Appl.*, vol. 20, no. 2, pp. 302–326, Mar. 2013.



Bing Fan (S'14) received the B.S. degree in electrical and computer engineering from Shanghai Jiao Tong University, Shanghai, China, in 2011. She is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of California San Diego, La Jolla, CA, USA. Her research interests include coding and signal processing in storage systems. She is affiliated with the Center for Memory and Recording Research (formerly, Center for Magnetic Recording Research).

Hemant K. Thapar (M'78–SM'98–F'98) received the Ph.D. degree in electrical engineering from Purdue University. He was an Adjunct Lecturer of Electrical Engineering with Santa Clara University from 1984 to 2004 and a Visiting Research Scientist with the Center for Memory and Recording Research, University of California San Diego, from 2013 to 2014. During his early career, he held various technical and management positions with Bell Telephone Laboratories, Holmdel (1979–1984) and with IBM Corporation, San Jose (1984–1994), where he was involved in the areas of circuit-switched networking, data communications, and data storage. He is a founder and the CEO of OmniTier, Inc., a start-up company developing infrastructure application solutions for cloud data centers. He was previously co-founder and CEO of two successful startups, Link-A-Media Devices and DataPath Systems. He is co-recipient of three best paper awards for his work on high-speed data transmission and high-density data storage. He served as a Guest Editor for two special issues of the IEEE Transactions devoted to data storage technologies. He serves on the Board of Directors of Assia Corporation, a private company enabling next-generation broadband technology and products; and on the Advisory Boards of the School of Engineering, Santa Clara University and the Ambala College of Engineering and Applied Sciences, India.

Paul H. Siegel (M'82–SM'90–F'97) received the S.B. and Ph.D. degrees in mathematics from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1975 and 1979, respectively. He held a Chaim Weizmann Post-Doctoral Fellowship with Courant Institute, New York University, New York, NY, USA. He was with IBM Research Division, San Jose, CA, USA, from 1980 to 1995. He joined the faculty of the University of California San Diego, La Jolla, CA, USA, in 1995, where he is currently a Professor of Electrical and Computer Engineering with the Jacobs School of Engineering. He is affiliated with the Center for Memory and Recording Research, where he holds an Endowed Chair and served as a Director from 2000 to 2011. His research interests include information theory and communications, particularly coding and modulation techniques, with applications to digital data storage and transmission. He is a member of the National Academy of Engineering. He was a member of the Board of Governors of the IEEE Information Theory Society from 1991 to 1996 and from 2009 to 2014. He was the 2015 Padovani Lecturer of the IEEE Information Theory Society. He was a recipient of the 2007 Best Paper Award in Signal Processing and Coding for Data Storage from the Data Storage Technical Committee of the IEEE Communications Society. He was a co-recipient of the 1992 IEEE Information Theory Society Paper Award and the 1993 IEEE Communications Society Leonard G. Abraham Prize Paper Award. He served as a Co-Guest Editor of the 1991 Special Issue on Coding for Storage Devices of the IEEE TRANSACTIONS ON INFORMATION THEORY. He served as an Associate Editor of Coding Techniques of the IEEE TRANSACTIONS ON INFORMATION THEORY from 1992 to 1995, and as an Editor-in-Chief from 2001 to 2004. He was also a Co-Guest Editor of the 2001 two-part issue on The Turbo Principle: From Theory to Practice and the 2016 issue on Recent Advances in Capacity Approaching Codes of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.