Generalized Weighted Sum Subtract Joint Detection for a Class of Multihead Multitrack Channels

Bing Fan[1], Paul H. Siegel[2], *Fellow, IEEE*, and Hemant K. Thapar[3], *Fellow, IEEE*

[1]Western Digital Corporation, San Diego, CA 92121 USA
[2]Center for Memory and Recording Research, University of California, San Diego, La Jolla, CA 92093 USA
[3]OmniTier, Inc., Milpitas, CA 95035 USA

Weighted sum subtract joint detector (WSSJD) was proposed by Fan *et al.* as an alternative way to implement maximum-likelihood (ML) detection for multihead multitrack (MHMT) systems. Different from the conventional ML detector, the trellis used in WSSJD is independent of the intertrack interference (ITI); therefore, WSSJD could efficiently adapt to the time-varying ITI environment. The channel transformation developed for WSSJD also naturally leads to a reduced-complexity implementation for the ML detector. In this paper, we generalize WSSJD to MHMT systems whose interference matrices satisfy *the WSSJD property*. The implementation details are given for a symmetric 3H2T system, as an example. Other interference matrices where WSSJD is applicable are also presented.

***Index Terms*— Bit-patterned magnetic recording (BPMR), intersymbol interference (ISI), intertrack interference (ITI), maximum-likelihood sequence estimation (MLSE), multihead multitrack (MHMT), two-dimensional magnetic recording (TDMR).**

I. Introduction

MAGNETIC recording systems with multihead multitrack (MHMT) joint signal processing were recognized in the early 1990s as a technique to potentially improve the areal density (AD) of hard disk drives (HDDs) [5]. Since then, they have been extensively studied [6]–[9]. In 2008, Wood *et al.* [10] proposed two-dimensional magnetic recording (TDMR), which brought the concept of MHMT into reality. TDMR aims to achieve an AD of about 10 Tbit/in^2, yet staying with the conventional magnetic recording disks. In TDMR, the data tracks are heavily squeezed to increase track density, and signals from multiple tracks are processed simultaneously to combat the intertrack interference (ITI). An array reader consisting of multiple sensors is used to improve the system throughput [11]–[13].

Recent progress in TDMR with shingled writing boosts the research on practical 2-D signal processing techniques [10], [14], [15]. Many sophisticated 2-D signal processing methods have been previously explored in the literature. One common approach is to treat the readback signals as a 2-D image and adopt a turbo-structured detector to iteratively process data row-by-row and column-by-column [16]–[19]. These techniques achieve promising performance, but their high computational complexity and relatively large processing time could prevent practical application at this point.

In contrast, an MHMT detector has relatively lower complexity and smaller delay that could allow its adoption in the first generation TDMR system. Barbosa [5] implemented a maximum-likelihood (ML) detector for an ideal 2H2T symmetric interfering magnetic recording system. Theoretical bounds on the achievable performance of an ML detector for several MHMT channels were derived in [20]. Recently, MHMT detectors have been simulated on more complex channels, which better approximate the real HDDs. In [21], the ML detector along with a joint-track equalizer was simulated on a channel model for bit-patterned media storage (BPMR). Wang and Kumar [22] derive a symbol-based maximum *a posteriori* detector for a BPMR channel consisting of three heads and two tracks. A pattern-dependent Viterbi detector is given in [23] for the array-reader system. Shi and Barry [24] propose a multitrack detector with 2-D pattern-dependent noise prediction, which significantly outperforms the conventional 2-D Viterbi detector when the channel noise is pattern-dependent. More realistic HDD channel models and system optimization have been studied in [25] and [26]. In [27], the implementation cost for realizing the optimal MHMT detection is studied. Many efforts have been devoted to finding low-cost suboptimal detection methods. In [13], the readback signals from multiple heads are equalized to a 1-D target to reduce the detection complexity. An ITI cancelation scheme [28]–[30] is another effective technique to combat ITI with low implementation cost.

In this paper, we assume a rather simplified MHMT model, which captures both intersymbol interference (ISI) and ITI. We aim to address two fundamental problems with the ML detector. First, it is generally inefficient to implement ML detection in a conventional manner on MHMT models with varying ITI. Since the construction of ML trellis depends on ITI, a change in ITI estimates requires to recalculate the output labels, which could lead to intolerable delay. Second, as the number of tracks and channel memory increase, the computational complexity of ML detection exponentially grows so that implementing ML detector could be infeasible. Searching for suboptimal solutions that could significantly reduce the complexity without degrading performance too

Manuscript received June 1, 2018; revised August 28, 2018; accepted October 31, 2018. Date of publication December 25, 2018; date of current version January 18, 2019. Corresponding author: B. Fan (e-mail: bing.j.fan@gmail.com).
Digital Object Identifier 10.1109/TMAG.2018.2879629

0018-9464 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

much is pressing. In [2], we proposed a variation of the ML detector, called the weighted sum subtract joint detector (WSSJD), for a symmetric nHnT channel. We showed that the ITI interfering channels in nHnT can be transformed to n parallel subchannels, whose joint trellis is independent of ITI. We then considered the case when ITI is unknown and proposed an architecture with gain loops to efficiently track the ITI estimates. A reduced-complexity implementation was presented in [3] and [4]. We showed, both by simulations and theoretical analysis, that trellis with much fewer states could achieve a near optimal performance.

The system considered in [1]–[4] is limited to nHnT channel with symmetric ITI. In this paper, we extend our work to more general channels. To begin with, the symmetric 3H2T system, which is also studied in [11] and [12], is assumed. We generalize the channel decomposition, complexity reduction, and ITI estimation techniques in [1]–[4] to the 3H2T system. We then show that, if the interference matrix of the MHMT system satisfies a specific property, WSSJD can be applied. Several classes of matrices that have this property are identified and discussed.

The rest of this paper is organized as follows. In Section II, we summarize the notation. In Section III, we introduce the channel model of the 3H2T system and analyze the achievable performance of the ML detector. In Section IV, we generalize the WSSJD to the 3H2T system. We discuss three techniques: channel decomposition, reduced-complexity implementation, and ITI estimation. In Section V, we introduce the WSSJD property for interference matrices, which is a necessary condition for an MHMT system amenable to apply WSSJD. Several examples of such systems are given. We also discuss the application of WSSJD to other potential 3H2T models based on the work in [11]. We conclude this paper in Section VI.

II. Notations

A length-L data sequence is represented by a polynomial $x(D) = \sum_{k=0}^{L-1} x_k D^k$. The inner product of two data sequences of length L is calculated by

$$\langle x(D), y(D)\rangle = \sum_{k=0}^{L-1} x_k y_k \tag{1}$$

and the norm of a data sequence is

$$\|x(D)\| = \sqrt{\langle x(D), x(D)\rangle} = \sqrt{\sum_{k=0}^{L-1} x_k^2}. \tag{2}$$

We use a capitalized bold symbol to represent a column vector of data sequences, e.g., $\boldsymbol{X}(D) = [x^{(1)}(D), \ldots, x^{(n)}(D)]^\top$. The coefficients of $\boldsymbol{X}(D)$ at time k form a column vector $\boldsymbol{x}_k = [x_k^{(1)}, \ldots, x_k^{(n)}]^\top$. The multiplication of $\boldsymbol{X}(D)$ with a data sequence $h(D)$ is equivalent to element-wise convolutions, i.e., $\boldsymbol{X}(D)h(D) = [x^{(1)}(D)h(D), \ldots, x^{(n)}(D)h(D)]^\top$.

Matrices are denoted by boldface symbols, with subscripts indicating the matrix size. For instance, $\mathbf{A}_{m,n}$ is an $m \times n$ matrix. We use $\mathbf{A}_n$ for short when $m = n$. A matrix element is referred by a lowercase symbol, e.g., a_{ij} on row i and column j in $\mathbf{A}_{m,n}$. The $n \times n$ identity matrix is represented by $\mathbf{I}_n$. The transpose of $\mathbf{A}$ is denoted by $\mathbf{A}^\top$. The ith power of $\mathbf{A}$ is written as $\mathbf{A}^i$, and $\mathbf{A}^0 = \mathbf{I}$.

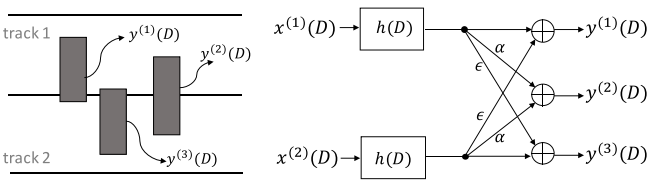


Fig. 1. Three-head/two-track model with ISI and ITI.

III. Three-Head/Two-Track Channel Model

We consider a 3H2T system with ISI along the down-track direction and ITI in the cross-track direction, as shown in Fig. 1. The channel inputs, $\boldsymbol{X}(D) = [x^{(1)}(D), x^{(2)}(D)]^\top$, are data sequences stored on track 1 and track 2. We assume that both $x_k^{(1)}$ and $x_k^{(2)}$ are independent identically distributed (i.i.d) and randomly chosen from $\{+1, -1\}$ with equal probability. We also assume that there is no phase offset between $x^{(1)}(D)$ and $x^{(2)}(D)$.

Three heads are used to readback data simultaneously. Head 1 and head 3 are symmetrically placed around the boundary between two tracks, with an offset toward tracks 1 and 2, respectively. With this placement, head 1 senses a majority signal from track 1 (the target track) and a small interference from track 2 (the side track), while the case is reversed for head 3. Head 2 is placed symmetrically over the boundary and senses equal level signals from two tracks. We use this head ordering because the resulting interference matrix possesses a desirable symmetry property. Let $\boldsymbol{R}(D) = [r^{(1)}(D), r^{(2)}(D), r^{(3)}(D)]^\top$ be the vector of discretized readback signals from heads 1, 2, and 3, respectively. The channel model can be compactly written as

$$\boldsymbol{R}(D) = \mathbf{A}_{3,2}\boldsymbol{X}(D)h(D) + \boldsymbol{\Omega}(D) \tag{3}$$

where the electronic noise at the three heads, $\boldsymbol{\Omega}(D) = [\omega^{(1)}(D), \omega^{(2)}(D), \omega^{(3)}(D)]^\top$, is i.i.d and follows the Gaussian distribution, $\omega_k^{(i)} \sim \mathcal{N}(0, \sigma^2)$, $i = 1, 2, 3$. In this model, we assume that the data tracks are ideal partial response channels, with the same channel polynomial $h(D) = \sum_0^\nu h_k D^k$. The interference matrix $\mathbf{A}_{3,2}$ characterizes the ITI sensed at the three heads. Based on the head alignment, we assume that $\mathbf{A}_{3,2}$ has the form

$$\mathbf{A}_{3,2} = \begin{bmatrix} 1 & \epsilon \\ \alpha & \alpha \\ \epsilon & 1 \end{bmatrix} \tag{4}$$

where ϵ and α are variables whose values may change for various reasons. For example, the values of ϵ and α are expected to be very different at inner tracks and outer tracks, since the track pitch changes. Moreover, in some cases, accurate values of these variables may be unknown at the receiver, and a parameter estimation process is needed. Due to these uncertainties, typical of real drives, we use variables ϵ and α to represent the ITI parameters. In this paper, ϵ is restricted to the range $[0, 0.5)$, and $\alpha > 0$. A matrix whose elements do not contain variables is called a *deterministic matrix*.

At the receiver, the ML estimates of the input sequences $\boldsymbol{X}(D)$ are obtained by passing the outputs $\boldsymbol{R}(D)$ through a

2-D joint Viterbi detector. Let $\boldsymbol{X}(D)$ and $\hat{\boldsymbol{X}}(D)$ denote the correct inputs and the estimated inputs, respectively. Their difference $\boldsymbol{E}(D) = \boldsymbol{X}(D) - \hat{\boldsymbol{X}}(D)$, if not zero, is called an error event. It is well known that the error event probability of a Viterbi detector can be approximated by $K \cdot Q(d_{\min}/2\sigma)$, where K is a constant related to the frequency of the dominant error events, σ^2 is the noise variance, and the Q-function is the tail probability of the standard Gaussian distribution. The parameter $d_{\min}$ is the minimum distance taken over all possible error events. For the 3H2T channel given by (3) and (4), we find

$$d^2_{\min,\mathbf{A}_{3,2}} = \begin{cases} (1+\alpha^2+\epsilon^2)d_0^2, & \text{if } \alpha^2 > 1-4\epsilon+\epsilon^2 \\ 2(1-\epsilon)^2 d_0^2, & \text{otherwise} \end{cases} \tag{5}$$

where d_0^2 is the squared minimum distance of the ISI channel $h(D)$. The derivation of (5) is presented in the Appendix. The case of $\alpha = 0$ is equivalent to the 2H2T system with interference matrix [2]

$$\mathbf{A}_2 = \begin{bmatrix} 1 & \epsilon \\ \epsilon & 1 \end{bmatrix}. \tag{6}$$

The incorporation of ITI parameters presents a barrier to the realization of an ML detector. To implement the 2-D joint Viterbi detector, a trellis that simultaneously tracks $x^{(1)}(D)$ and $x^{(2)}(D)$ is constructed and stored in memory. Notice that the output labels of trellis branches, given by $\mathcal{L}_{\text{out}} = \boldsymbol{Y}_k$, where

$$\boldsymbol{Y}(D) = \mathbf{A}_{3,2}\boldsymbol{X}(D)h(D) \tag{7}$$

require the knowledge of α and ϵ. However, these two parameters are generally unknown to the receiver and subject to change. Storing trellises for different parameters would consume a large memory, while recomputing output labels during the detection process could incur an additional delay. Using static estimates could avoid these issues, but results in performance loss due to sub-optimality.

IV. WSSJD for 3H2T System

We generalize WSSJD to the 3H2T system, as an alternative way to implement the ML detector. We show that by means of channel decomposition, the 3H2T system can be transformed to two parallel subchannels, whose joint trellis is free from ITI parameters. Then, a gain loop structure could be deployed to adaptively track the ITI parameters. Complexity reduction techniques will also be discussed. For convenience, the conventional ML detector directly designed for system (3) is referred to as "the ML detector"; however, we emphasize that WSSJD also gives the ML solutions.

A. Channel Decomposition

The interference matrix $\mathbf{A}_{3,2}$ can be factored as

$$\mathbf{A}_{3,2} = \mathbf{U}\Lambda\mathbf{V}^\top \tag{8}$$

where

$$\mathbf{V} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{9}$$

is a 2×2 deterministic matrix. The 3×2 matrix $\mathbf{U}$ has the form

$$\mathbf{U} = \left[\lambda_1^{-1}\boldsymbol{u}_1, \lambda_2^{-1}\boldsymbol{u}_2, \boldsymbol{u}_3\right] \tag{10}$$

where

$$\boldsymbol{u}_1 = \left[\frac{1+\epsilon}{2}, \alpha, \frac{1+\epsilon}{2}\right]^\top, \quad \boldsymbol{u}_2 = \left[\frac{1-\epsilon}{2}, 0, \frac{\epsilon-1}{2}\right]^\top \tag{11}$$

and $\lambda_1 = \|\boldsymbol{u}_1\|$ and $\lambda_2 = \|\boldsymbol{u}_2\|$ are the normalization factors. The third column vector $\boldsymbol{u}_3$ has the unit length and is orthogonal to both $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$. Therefore, $\mathbf{U}$ is a unitary matrix, i.e., $\mathbf{U}^\top\mathbf{U} = \mathbf{U}\mathbf{U}^\top = \mathbf{I}_3$. Matrix Λ is a 3×2 diagonal matrix whose diagonal elements are λ_1 and λ_2

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \\ 0 & 0 \end{bmatrix}. \tag{12}$$

Notice that the factorization (8) is equivalent to the singular value decomposition (SVD) of $\mathbf{A}_{3,2}$, up to a scaling factor. The pseudoinverse of Λ is denoted as

$$\Lambda^\dagger = \begin{bmatrix} \lambda_1^{-1} & 0 & 0 \\ 0 & \lambda_2^{-1} & 0 \end{bmatrix}. \tag{13}$$

To decompose the channel, we substitute (8) for the matrix $\mathbf{A}_{3,2}$

$$\boldsymbol{R}(D) = \mathbf{U}\Lambda\mathbf{V}^\top\boldsymbol{X}(D)h(D) + \boldsymbol{\Omega}(D). \tag{14}$$

Since $\mathbf{U}$ is unitary and $\Lambda^\dagger\Lambda = \mathbf{I}_2$, we can multiply both the sides by $\Lambda^\dagger\mathbf{U}^\top$ to obtain

$$\Lambda^\dagger\mathbf{U}^\top\boldsymbol{R}(D) = \mathbf{V}^\top\boldsymbol{X}(D)h(D) + \Lambda^\dagger\mathbf{U}^\top\boldsymbol{\Omega}(D). \tag{15}$$

Equation (15) is *the transformed 3H2T channel*. Define

$$\begin{aligned} \boldsymbol{Z}(D) &= \mathbf{V}^\top\boldsymbol{X}(D) \\ &= [x^{(1)}(D) + x^{(2)}(D), x^{(1)}(D) - x^{(2)}(D)]^\top \\ &= [z^{(1)}(D), z^{(2)}(D)]^\top \end{aligned} \tag{16}$$

to be the new channel inputs. Notice that $z_k^{(i)} \in \{-2, 0, +2\}$ is ternary. The new channel outputs are

$$\begin{aligned} \bar{\boldsymbol{R}}(D) &= \Lambda^\dagger\mathbf{U}^\top\boldsymbol{R}(D) \\ &= \left[\lambda_1^{-2}\boldsymbol{u}_1^\top\boldsymbol{R}(D), \lambda_2^{-2}\boldsymbol{u}_2^\top\boldsymbol{R}(D)\right]^\top \\ &= [\bar{r}^{(1)}(D), \bar{r}^{(2)}(D)]^\top. \end{aligned} \tag{17}$$

The transformed noise components

$$\bar{\boldsymbol{\Omega}}(D) = \Lambda^\dagger\mathbf{U}^\top\boldsymbol{\Omega}(D) \tag{18}$$

satisfy a zero-mean multivariate Gaussian distribution, with covariance matrix

$$\begin{aligned} E\left[\bar{\boldsymbol{\omega}}_k\bar{\boldsymbol{\omega}}_k^\top\right] &= E\left[\Lambda^\dagger\mathbf{U}^\top\boldsymbol{\omega}_k\boldsymbol{\omega}_k^\top\mathbf{U}(\Lambda^\dagger)^\top\right] \\ &= \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \end{aligned} \tag{19}$$

where $\sigma_1^2 = \sigma^2\lambda_1^{-2}$ and $\sigma_2^2 = \sigma^2\lambda_2^{-2}$.

B. WSSJD

After channel decomposition, the resulting system (15) consists of two parallel subchannels

$$\bar{r}^{(1)}(D) = z^{(1)}(D)h(D) + \bar{\omega}^{(1)}(D) \tag{20}$$

$$\bar{r}^{(2)}(D) = z^{(2)}(D)h(D) + \bar{\omega}^{(2)}(D). \tag{21}$$

We call (20) the *sum channel* and (21) the *subtract channel*, since $z^{(1)}(D)$ and $z^{(2)}(D)$ are basically the sum and the difference of the original channel inputs $x^{(1)}(D)$ and $x^{(2)}(D)$.

WSSJD searches for the joint ML solution to (20) and (21)

$$\begin{aligned} \boldsymbol{Z}^*(D) &= \arg\max_{\mathbf{Z}(D)} \log \Pr(\bar{\boldsymbol{R}}(D)|\boldsymbol{Z}(D)) && (22)\\ &= \arg\max_{\mathbf{Z}(D)} \log \Pr(\bar{r}^{(1)}(D)|z^{(1)}(D)) \\ &\quad + \log \Pr(\bar{r}^{(2)}(D)|z^{(2)}(D)) && (23)\\ &= \arg\min_{\mathbf{Z}(D)} \lambda_1^2 \|\bar{r}^{(1)}(D) - z^{(1)}(D)h(D)\|^2 \\ &\quad + \lambda_2^2 \|\bar{r}^{(2)}(D) - z^{(2)}(D)h(D)\|^2. && (24) \end{aligned}$$

Here, (23) comes from the fact that the noises on the two subchannels are independent. The scale factors in (24), λ_1^2 and λ_2^2, are the *channel weights* due to the effect of unequal noise powers in two subchannels.

The WSSJD trellis constructed to decode $\boldsymbol{Z}(D)$ contains 4^ν states, each associated with four outgoing edges. It has the same computational complexity as the ML detector. In contrast to the ML detector whose trellis labels vary with ITI, the WSSJD trellis is deterministic. Namely, the trellis does not require knowledge of α and ϵ, remaining the same even though α and ϵ may change. This property would be preferable in hardware realization. Another difference from the ML detector is that, to retain optimality, the branch metrics must be scaled by λ_1^2 and λ_2^2 whose values depend on α and ϵ. This additional step is easy to realize in practice.

Assume that $\bar{\mathbf{E}}(D) = [\bar{e}^{(1)}(D), \bar{e}^{(2)}(D)]$ are the error events on the sum and subtract channels, respectively. The distance function is calculated by

$$d^2_{\text{WSSJD}}(\bar{\mathbf{E}}(D)) = \lambda_1^2 \|\bar{e}^{(1)}(D)h(D)\|^2 + \lambda_2^2 \|\bar{e}^{(2)}(D)h(D)\|^2. \tag{25}$$

The squared minimum distance

$$d^2_{\min} = \min_{\bar{\mathbf{E}}(D)} d^2_{\text{WSSJD}}(\bar{\mathbf{E}}(D)) \tag{26}$$

has the same closed form as (5).

C. Complexity Reduction

Reduced-state sequence estimation (RSSE), proposed in [31], allows one to make a tradeoff between performance and complexity in trellis-based detection and decoding. Instead of relying on the complete ML trellis, RSSE uses *subset trellis*, in which the *subset states* are disjoint groups of the original ML states. Suppose J_k, $k = 1, \ldots, \nu$, are set partitionings of the input symbols. A subset state is represented by

$$\boldsymbol{s}_i = [a_{i-1}(1), a_{i-2}(2), \ldots, a_{i-\nu}(\nu)] \tag{27}$$

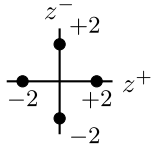


Fig. 2. Input constellation of the WSSJD system contains four symbols.

TABLE I
ESPDs BETWEEN INPUT SYMBOLS OF DECOMPOSED 3H2T SYSTEM

$(\boldsymbol{z}, \hat{\boldsymbol{z}})$	$d^2_{\text{ESPD}}(\boldsymbol{z}, \hat{\boldsymbol{z}})$
$([+2,0],[-2,0])$	$\Delta_1^2 = 8(1+\epsilon)^2 + 16\alpha^2$
$([0,+2],[0,-2])$	$\Delta_2^2 = 8(1-\epsilon)^2$
$([+2,0],[0,+2])$	$\Delta_3^2 = 4(1+\epsilon^2) + 4\alpha^2$
$([+2,0],[0,-2])$	
$([-2,0],[0,+2])$	
$([-2,0],[0,-2])$	

where $a_{i-k}(k)$ is a subset index in J_k. The *configuration* of the resulting subset trellis is a vector $[|J_1|, |J_2|, \ldots, |J_\nu|]$, and the number of subset states is $\prod_{k=1}^{\nu} |J_k|$.

To obtain a valid subset trellis, the sequence of set partitionings $\{J_k\}_{k=1}^{\nu}$ is restricted to the condition that J_k is a further partition of the subsets in J_{k+1} for all k. A *set partitioning tree* can be used to assist the set partitioning selection process. In Fig. 4, we present two examples of the set partitioning tree. The set partitionings $\{J_k\}_{k=1}^{\nu}$ are chosen from four levels of the tree, with the restriction that J_{k+1} should not be a lower level than J_k.

A general design rule for a good set partitioning tree is to maximize the minimum intra-subset distance at each level. For the transformed 3H2T system, since the two subchannels have different noise powers, we use *effective symbol pair distance* (ESPD), instead of Euclidean distance, to evaluate distances between input symbols. The ESPD between two symbols $\boldsymbol{z}$ and $\hat{\boldsymbol{z}}$ is defined by

$$d^2_{\text{ESPD}}(\boldsymbol{z}, \hat{\boldsymbol{z}}) = \lambda_1^2 \cdot (z^{(1)} - \hat{z}^{(1)})^2 + \lambda_2^2 \cdot (z^{(2)} - \hat{z}^{(2)})^2. \tag{28}$$

The input constellation is shown in Fig. 2. Table I summarizes ESPDs for all pairs of input symbols. Compared with the ESPDs in the 2H2T system (see [4, Table I]), the additional head in 3H2T only affects Δ_1^2 and Δ_3^2. As shown in Fig. 3(a), for $\alpha = 0.5$, the curves of Δ_2^2 and Δ_3^2 cross near $\epsilon = 0.2$. For $\alpha = 1$ in Fig. 3(b), $\Delta_3^2 > \Delta_2^2$ for all $\epsilon > 0$. We also plot the squared minimum distance of the corresponding 2H2T system (with $h(D) = 1 + D - D^2 - D^3$) for comparison.

We want to construct a set partitioning tree where the minimum intra-subset ESPD increases from the top level to the bottom. Since the ordering of Δ_1^2, Δ_2^2, and Δ_3^2 is not fixed, we propose two possible set partitioning trees that can be used in different ITI environments. The constructions are shown in Fig. 4. Tree-A [see Fig. 4(a)] was originally proposed in [4] for the 2H2T channel. Under the assumption that $\Delta_2^2 \geqslant \Delta_3^2$, the minimum intra-subset ESPDs from the top to the bottom are Δ_3^2, Δ_2^2, Δ_1^2, and ∞ in a sequential order. Tree-B [see Fig. 4(b)] only differs from tree-A at level L_2. It is designed for the case when $\Delta_2^2 < \Delta_3^2$, so that the minimum ESPDs

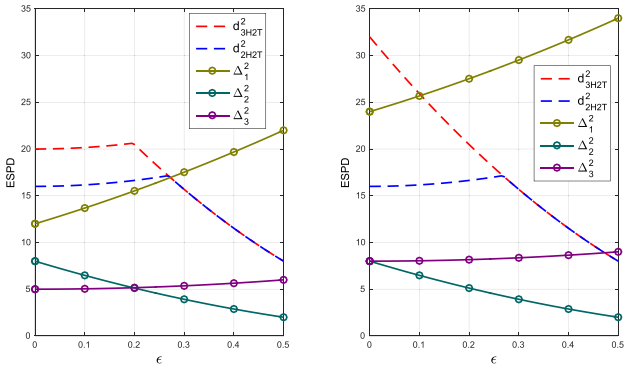


Fig. 3. ESPDs for 3H2T system with (a) $\alpha = 0.5$ and (b) $\alpha = 1$.

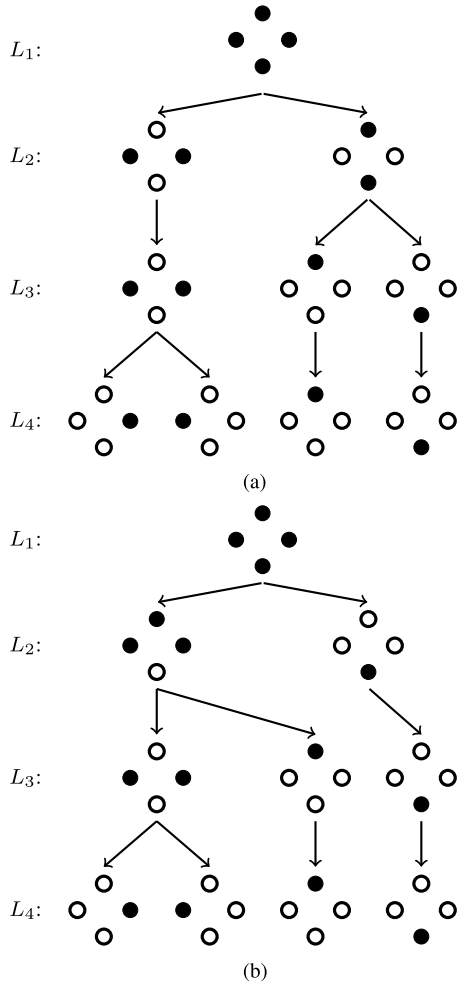


Fig. 4. Two set partitioning trees designed for the 3H2T system. Each tree has four levels, $\{L_i\}_{i=1}^{4}$. Four input symbols form a diamond-shaped constellation. (a) Tree-A. (b) Tree-B.

become Δ_2^2 followed by Δ_3^2, Δ_1^2, and ∞, from the top to the bottom. The performance of RSSE based on tree-A has been simulated and analyzed for the 2H2T system [4]. In this paper, we aim to show the performance improvement of using tree-B rather than tree-A in the 3H2T system where α is large.

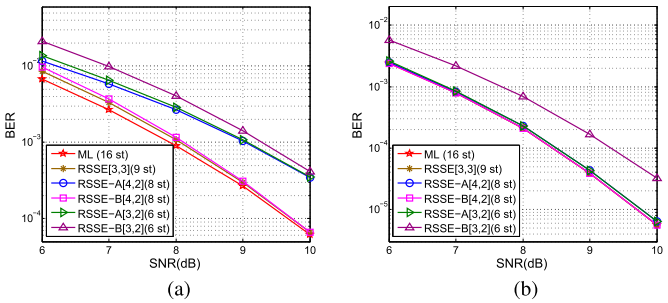


Fig. 5. Performance comparison of RSSE trellises with different configurations. The 3H2T systems are set to $\alpha = 1$ and $\epsilon = 0.3$, but with different channel targets. (a) $h(D) = 1 + 2D + D^2$. (b) $h(D) = 4 + 7D + D^2$.

TABLE II
ERROR EVENT ANALYSIS FOR DIFFERENT SUBSET TRELLIS CONFIGURATIONS IN FIG. 5(a)

	$\min d_{\mathrm{r}}^2(\bar{e}^{(1)}, \bar{e}^{(2)})$	$\bar{e}^{*(1)}, \bar{e}^{*(2)}$
ML	15.68	$[0,0,0,0], [4,-4,0,0]$
RSSE-A$[4,2]$	11.76	$[0,0,0], [4,-4,0]$
RSSE-B$[4,2]$	15.68	$[0,0,0,0], [4,-4,0,0]$
RSSE-A$[3,2]$	11.76	$[0,0,0], [4,-4,0]$
RSSE-B$[3,2]$	12.28	$[-2,4], [-2,0]$

We construct subset trellises based on tree-A and tree-B and simulate RSSE for the 3H2T system where $\alpha = 1$ and $\epsilon = 0.3$. In Fig. 3(b), we see that $\Delta_3^2 > \Delta_2^2$ for all $\epsilon > 0$, so we expect the bit error rate (BER) performance of the trellises based on tree-B to be lower than that of the trellises constructed from tree-A. The simulation results on channels with two different length-3 channel polynomials are plotted in Fig. 5. The PR2 channel with $h(D) = 1 + 2D + D^2$ is a linear phase, while the channel with target $h(D) = 4+7D+D^2$ is a minimum phase channel which is also considered in [11]. The set partitioning tree used to construct a subset trellis is indicated by "A" or "B" in the legends. For example, the term RSSE-A[4, 2] represents the subset trellis constructed from tree-A with configuration [4, 2]. Since tree-A and tree-B have the same 3-subset partitioning (level L_3), they result in the same subset trellis with [3, 3]. We represent it by RSSE[3, 3] for short. The signal-to-noise ratio (SNR) is defined by

$$\mathrm{SNR(dB)} = 10 \log \frac{\|h(D)\|^2}{2\sigma^2}. \tag{29}$$

It is shown in Fig. 5(a) that RSSE[3, 3] with nine states and RSSE-B[4, 2] with eight states perform very close to the ML detector, which uses 16 states, especially when SNR is high. In contrast to RSSE-B[4, 2], RSSE-A[4, 2], also with eight states, suffers from large performance loss. RSSE-A[3, 2] has a similar performance as RSSE-A[4, 2]. An interesting observation is that RSSE-B[3, 2] is even worse than RSSE-A[3, 2], which is different from our expectation.

We explain the BER performances through error event analysis. Due to the use of subset states, certain decoding paths will merge earlier in the subset trellis than in the ML trellis. These "early merged" paths can result in *early merged error events* [4] that possibly decrease the minimum distance of the system. In Table II, we present the minimum

distances for different trellises, with the same system setup as in Fig. 5(a). For each trellis configuration, we also give an example of the error events, which achieve the minimum distance. Given $\bar{e}^{(1)}$ and $\bar{e}^{(2)}$, which are error events in the vector form, the minimum squared distance of using subset trellis is calculated by [4]

$$d_{\mathrm{r}}^2(\bar{e}^{(1)}, \bar{e}^{(2)}) = \lambda_1^2 \sum_{i=0}^{l-1} \left(\sum_{k=0}^{\nu} h_k \bar{e}_{i-k}^{(1)} \right)^2 + \lambda_2^2 \sum_{i=0}^{l-1} \left(\sum_{k=0}^{\nu} h_k \bar{e}_{i-k}^{(2)} \right)^2 \quad (30)$$

where l is the length of the error pattern.

It is shown in Table II that RSSE-B[4, 2] does not produce additional early merged error events whose distance is smaller than the minimum distance of the ML detector. In contrast, the minimum distances of RSSE-A[4, 2] and RSSE-A[3, 2] are smaller than that of the ML detector. Therefore, their BER performances are close and subject to be worse than the ML detector. Different from RSSE-B[4, 2], the performance of RSSE-B[3, 2] is dominated by early merged error events. Although it has a slightly larger $\min d_{\mathrm{r}}^2$ than RSSE-A[3, 2], the dominant early merged error events for RSSE-B[3, 2] appear more frequently so that its BER performance is worse than that of RSSE-A[3, 2].

Compared with the PR2 channel, the RSSE algorithm generally performs better on channel $h(D) = 4 + 7D + D^2$, as indicated in Fig. 5(b). It is shown that RSSE[3, 3], RSSE-A[4, 2], RSSE-B[4, 2], and RSSE-A[3, 2] essentially achieve the same performance as the ML detector. RSSE-B[3, 2], however, has roughly 0.8 dB loss at BER $= 10^{-4}$. Our error event analysis shows that RSSE-B[3, 2] is dominated by the same error events as in the PR2 channel, and has a much smaller minimum distance than that of the ML detector.

We also present the performance of RSSE when $h(D) = 1 + D - D^2 - D^3$ in Fig. 6. We consider the cases when ITI is low ($\epsilon = 0.1$) and high ($\epsilon = 0.3$). In Fig. 6(a), where $\epsilon = 0.1$, RSSE[4, 3, 3] with 36 states could nearly achieve the same performance as the ML detector. The performance loss of RSSE[3, 3, 3] with 27 states is within 0.2 dB at BER $= 10^{-5}$. We observe that the performances of subset trellises constructed from tree-A and tree-B are very close. It is reasonable since when ITI is low, the difference between Δ_2^2 and Δ_3^2 is small. In Fig. 6(b), where $\epsilon = 0.3$, the BER curves of the ML detector, RSSE[4, 3, 3], RSSE[3, 3, 3], and RSSE-B[4, 3, 2] are essentially overlapped. RSSE-B[3, 3, 2] has approximately 0.2 dB performance loss at BER $= 10^{-5}$. RSSE-A[4, 3, 2] and RSSE-A[3, 3, 2] have the same performance. We can see that for the case of high ITI, the trellises constructed from tree-B generally achieve better performance than those constructed from tree-A.

In [4], we have presented how to use error event diagram to search for the dominant error events for a given RSSE trellis on the 2H2T channel. This technique can be easily extended to the 3H2T system. Once we know the dominant error events that mostly degrade the RSSE performance, it would be possible to design proper constrained codes to avoid such error events [32].

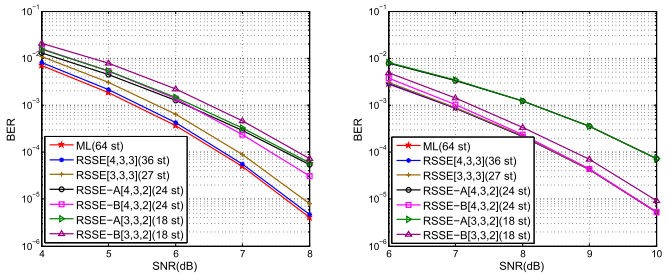


Fig. 6. Performance comparison of RSSE trellises on 3H2T systems with $\alpha = 1$ and $h(D) = 1 + D - D^2 - D^3$. (a) $\epsilon = 0.1$. (b) $\epsilon = 0.3$.

D. ITI Estimation

In general, the receiver may not have the information about α or ϵ. We show that gain loops can be incorporated in WSSJD to adaptively estimate the ITI parameters. Once we know the α and ϵ estimates, they are fed into WSSJD to evaluate the weighted branch metrics. First, notice that, in the transformed system, the subtract channel (21) has the same form as the one we obtained in the 2H2T system [2]

$$\frac{1}{1-\epsilon}(r^{(1)}(D) - r^{(3)}(D)) = z^{(2)}(D)h(D) + \bar{\omega}^{(2)}(D). \quad (31)$$

The parameter $1/(1-\epsilon)$ is a normalization factor. Let $g_k^{(2)}$ be the estimate of $1/(1-\epsilon)$ at time k. In [2], we have presented a gain loop for tracking $g_k^{(2)}$, so the details are omitted here. We are more interested in the sum channel, which is different from [2].

For the sum channel, we again model it as a gain factor estimation problem

$$\frac{1}{(1+\epsilon)^2 + 2\alpha^2}[(1+\epsilon)[r^{(1)}(D) + r^{(3)}(D)] + 2\alpha r^{(2)}(D)] = z^{(1)}(D)h(D) + \bar{\omega}^{(1)}(D). \quad (32)$$

Let $g_k^{(1)}$ be the estimate of $(1/((1+\epsilon)^2 + 2\alpha^2))$ at time k, and let $f_k^{(1)}$ and $f_k^{(2)}$ be the estimators of $(1+\epsilon)$ and 2α, respectively. Based on the least mean squares adaptive algorithm, we estimate $g_k^{(1)}$ by the following steps:

$$\bar{r}_k^{(1)} = g_{k-1}^{(1)} \left[f_{k-1}^{(1)} \left(r_k^{(1)} + r_k^{(3)} \right) + f_{k-1}^{(2)} r_k^{(2)} \right] \quad (33)$$

$$e_{k-\delta}^{(1)} = \bar{y}_{k-\delta}^{(1)} - \bar{r}_{k-\delta}^{(1)} \quad (34)$$

$$g_k^{(1)} = g_{k-1}^{(1)} + \beta\, \bar{y}_{k-\delta}^{(1)}\, e_{k-\delta}^{(1)}. \quad (35)$$

Here, parameter δ represents a small time delay, and $\bar{y}_{k-\delta}^{(1)}$ is the instantaneous estimate of the noiseless output of the sum channel. The step-size parameter β is used to control the convergence speed. The estimates of $f_k^{(1)}$ and $f_k^{(2)}$ are updated once we obtain the value of $g_k^{(1)}$. In some applications, the value of α is fixed and known. For this case, $f_k^{(2)}$ is set to 2α, and $f_k^{(1)}$ is updated by

$$f_k^{(1)} = \sqrt{\frac{1}{g_k^{(1)}} - 2\alpha^2}. \quad (36)$$

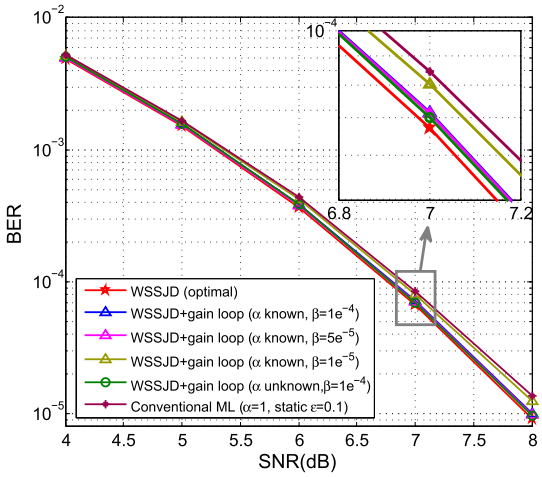


Fig. 7. Performance comparison of WSSJD with gain loops on the 3H2T channel. $h(D) = 4 + 7D + D^2$.

If α is unknown, we can use $g_k^{(2)}$ from the subtract channel and set

$$f_k^{(1)} = 2 - \frac{1}{g_k^{(2)}}, \quad f_k^{(2)} = \sqrt{2\left(\frac{1}{g_k^{(1)}} - \left(f_k^{(1)}\right)^2\right)}. \tag{37}$$

The updated $f_k^{(1)}$, $f_k^{(2)}$, and $g_k^{(1)}$ are used in the next iteration to calculate (33)–(35).

Recall that the path metric should be properly weighted in WSSJD. We replace λ_1^2 and λ_2^2 by their estimators $(g_k^{(1)})^{-1}$ and $(g_k^{(2)})^{-2}$, and the branch metric of the trellis edge from state $\boldsymbol{s}_{k-1}$ to $\boldsymbol{s}_k$ is calculated by

$$m(\boldsymbol{s}_{k-1}, \boldsymbol{s}_k) = \left(g_{k-1}^{(1)}\right)^{-1}\left[\bar{r}_k^{(1)} - \sum_{i=0}^{\nu} h_i z_{k-i}^{(1)}\right]^2 + \left(g_{k-1}^{(2)}\right)^{-2}\left[\bar{r}_k^{(2)} - \sum_{i=0}^{\nu} h_i z_{k-i}^{(2)}\right]^2. \tag{38}$$

Remark 1: We estimate the gain factor instead of directly estimating the unknown coefficients of $r^{(1)}(D)$, $r^{(2)}(D)$, and $r^{(3)}(D)$ because the latter optimization problem has many global minima.

In Fig. 7, we show some illustrative simulation results. The 3H2T system has channel target $h(D) = 4 + 7D + D^2$. Parameter α has a fixed value, $\alpha = 1$, and ϵ is assumed to be time varying

$$\epsilon(k) = \epsilon_0 + 0.1\sin(4\pi(k/L)). \tag{39}$$

In our simulation, ϵ_0 is set to 0.1, and the sector length $L = 4096$. Hence, ϵ sinusoidally varies within the range $[0, 0.2]$. The time delay δ is set to be 5. The optimal performance is obtained by implementing WSSJD with known α and $\epsilon(k)$. We then simulated the WSSJD algorithm incorporating a gain loop for ITI estimation. The initial values of $g_0^{(1)}$, $g_0^{(2)}$ and $f_0^{(1)}$, $f_0^{(2)}$ were obtained by training the algorithm with one sample sector. For the case when α is known at the receiver, we simulated the gain loop structure with different values of step size β. We observe that both $\beta = 5 \times 10^{-5}$ and $\beta = 1 \times 10^{-4}$ yield the WSSJD performance that is close to optimal. We also simulated WSSJD with gain loops when both α and ϵ are unknown. It is surprising to see that a lack of knowledge of α does not appreciably degrade the performance. For comparison purposes, we also show simulation results for a conventional ML detector with static $\alpha = 1$ and $\epsilon = 0.1$; the reason for using a static value of ϵ is the ITI dependence of the ML trellis. We can see that WSSJD with gain loop estimates outperforms the static conventional ML detector.

V. Generalized WSSJD

In this section, we generalize the WSSJD algorithm to an extended MHMT family of channels. We show that, if the interference matrix satisfies a specific property, the transformed channel can be detected on a deterministic trellis whose branch labels do not depend on ITI.

A. WSSJD Property

Consider a general mHnT system given by

$$\boldsymbol{R}(D) = \mathbf{A}_{m,n}\boldsymbol{X}(D)h(D) + \boldsymbol{\Omega}(D) \tag{40}$$

where $\boldsymbol{X}(D)$ consists of input sequences stored on n tracks, $\boldsymbol{R}(D)$ corresponds to the readback sequences from m heads, and $\boldsymbol{\Omega}(D)$ is a length-m vector of Gaussian noise sequences. We assume that the entries of $\mathbf{A}_{m,n}$ may depend on the ITI parameters of the system.

Suppose the SVD of $\mathbf{A}_{m,n}$ has the form of

$$\mathbf{A}_{m,n} = [\mathbf{U}_{(1)} \quad \mathbf{U}_{(2)}] \cdot \begin{bmatrix} \Lambda_p & \mathcal{O} \\ \mathcal{O} & \mathcal{O} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{V}_{(1)}^\top \\ \mathbf{V}_{(2)}^\top \end{bmatrix} \tag{41}$$

where $\mathbf{U}_{(1)}$ is $m \times p$, $\mathbf{U}_{(2)}$ is $m \times (m-p)$, $\mathbf{V}_{(1)}$ is $n \times p$, and $\mathbf{V}_{(2)}$ is $n \times (n-p)$. For full rank $\mathbf{A}_{m,n}$, $p = \min(m, n)$; otherwise, $p < \min(m, n)$. The submatrix $\Lambda_p = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ is a diagonal matrix of singular values of $\mathbf{A}_{m,n}$.

Let $\Lambda_p^{-1} = \mathrm{diag}(\lambda_1^{-1}, \ldots, \lambda_p^{-1})$ be the inverse of Λ_p. Substituting $\mathbf{A}_{m,n}$ by (41) and reorganizing (40), we can transform the original mHnT channel into

$$\Lambda_p^{-1}\mathbf{U}_{(1)}^\top\mathbf{R}(D) = \mathbf{V}_{(1)}^\top\mathbf{X}(D)h(D) + \Lambda_p^{-1}\mathbf{U}_{(1)}^\top\Omega(D). \tag{42}$$

Define $\bar{\mathbf{R}}(D) = \Lambda_p^{-1}\mathbf{U}_{(1)}^\top\mathbf{R}(D)$ as the new outputs, $\mathbf{Z}(D) = \mathbf{V}_{(1)}^\top\mathbf{X}(D)$ as the new inputs, and $\bar{\Omega}(D) = \Lambda_p^{-1}\mathbf{U}_{(1)}^\top\Omega(D)$ as the new noise components. The transformed channel model is described by

$$\bar{\mathbf{R}}(D) = \mathbf{Z}(D) + \bar{\Omega}(D). \tag{43}$$

It can be verified that the noise components in the new system are independent

$$E\left[\bar{\boldsymbol{\omega}}_k\bar{\boldsymbol{\omega}}_k^\top\right] = E\left[\Lambda_p^{-1}\mathbf{U}_{(1)}^\top\boldsymbol{\omega}_k\boldsymbol{\omega}_k^\top\mathbf{U}_{(1)}\Lambda_p^{-1}\right] = \sigma^2\Lambda_p^{-2}. \tag{44}$$

The transformed channel consists of p parallel subchannels, each of which is mathematically modeled as

$$\lambda_i^{-1}\left(\sum_{j=1}^{m} u_{ij}r^{(j)}(D)\right) = z^{(i)}(D)h(D) + \bar{\omega}^{(i)}(D) \tag{45}$$

for $i = 1, \ldots, p$. The ML solution of (43) is the vector of input sequences that minimize the *weighted* path metric

$$\boldsymbol{Z}^*(D) = \arg\min_{\boldsymbol{Z}(D)} \sum_{i=1}^{p} \lambda_i^2 \|\bar{r}^{(i)}(D) - z^{(i)}(D)h(D)\|^2. \quad (46)$$

A WSSJD trellis is constructed to track $\boldsymbol{Z}(D)$. From (42), we see that $\boldsymbol{Z}(D)$ is obtained by applying a linear transformation $V_{(1)}^{\top}$ to the original input $\boldsymbol{X}(D)$. Therefore, if $\mathbf{V}_{(1)}$ is a deterministic matrix, then the resulting WSSJD trellis will be independent of the ITI parameters.

Definition 1: An MHMT interference matrix $\mathbf{A}_{m,n}$ satisfies the WSSJD property if its SVD has a deterministic matrix $\mathbf{V}_{(1)}$.

The practical benefit of a system whose interference matrix satisfies the WSSJD property is that its ML detector can be implemented with a WSSJD trellis whose branch labels are independent of the ITI parameters. When these parameters vary, for example, from track-to-track and disk-to-disk, we can incorporate gain loops to adaptively train the detector. For instance, assume that the entries in $\mathbf{A}_{m,n}$ are the functions of an unknown parameter ϵ. Mathematically, the ith subchannel is given by

$$\lambda_i^{-1}(\epsilon) \cdot \left(\sum_j u_{ij}(\epsilon) r^j(D) \right) = z^i(D)h(D) + \bar{\omega}_i. \quad (47)$$

The estimators of $\lambda_i(\epsilon)^{-1}$ and $u_{ij}(\epsilon)$, denoted $g^{(i)}$ and $\hat{u}_{ij}$, respectively, can be computed by the following procedure.

1) Treat $g^{(i)}$ as the gain factor, and estimate it by a gain loop.
2) Solve for $\hat{\epsilon}$ from $g^{(i)}$ by assuming $g^{(i)} = \lambda_i^{-1}(\hat{\epsilon})$.
3) Update the estimates $\hat{u}_{ij} = u_{ij}(\hat{\epsilon})$ using the new estimate $\hat{\epsilon}$.
4) Weight the branch metric of the ith subchannel by $(g^{(i)})^{-2}$.
5) Repeat step 1)–4) using new estimates g^i and $\hat{u}_{ij}$.

B. WSSJD Channels

Propositions 1 and 2 give sufficient conditions for an interference matrix to satisfy the WSSJD property.

Proposition 1: Let $\mathbf{T}_n$ be an arbitrary $n \times n$ symmetric deterministic matrix. If an $n \times n$ matrix $\mathbf{A}_n$ can be written as

$$\mathbf{A}_n = \sum_{i=0}^{N} \gamma_i \mathbf{T}_n^i \quad (48)$$

where γ_i, $i = 0, \ldots, N$, are variables, and N is an arbitrary nonnegative integer, then $\mathbf{A}_n$ satisfies the WSSJD property.

Proof: The symmetric matrix $\mathbf{T}_n$ can be factorized as $\mathbf{T}_n = \mathbf{V}_n \Sigma_n \mathbf{V}_n^{\top}$, where $\mathbf{V}_n$ is a unitary matrix, and Σ_n is diagonal. Since $\mathbf{V}_n^{\top}\mathbf{V}_n = \mathbf{I}_n$, we have $\mathbf{T}_n^i = \mathbf{V}_n \Sigma_n^i \mathbf{V}_n^{\top}$. Therefore, (48) becomes

$$\mathbf{A}_n = \sum_{i=0}^{N} \gamma_i \mathbf{V}_n \Sigma_n^i \mathbf{V}_n^{\top} = \mathbf{V}_n \left(\sum_{i=0}^{N} \gamma_i \Sigma_n^i \right) \mathbf{V}_n^{\top}. \quad (49)$$

Since $\mathbf{V}_n$ is deterministic, $\mathbf{A}_n$ satisfies the WSSJD property. ■

Proposition 2: Let $\mathbf{T}_n$ be an arbitrary $n \times n$ symmetric deterministic matrix. If an $m \times n$ matrix $\mathbf{A}_{m,n}$ satisfies

$$\mathbf{A}_{m,n}^{\top}\mathbf{A}_{m,n} = \sum_{i=0}^{N} \gamma_i \mathbf{T}_n^i \quad (50)$$

where γ_i, $i = 0, \ldots, N$, are variables, and N is an arbitrary nonnegative integer, then $\mathbf{A}_{m,n}$ satisfies the WSSJD property.

Proof: Suppose the SVD of $\mathbf{A}_{m,n}$ is $\mathbf{A}_{m,n} = \mathbf{U}_m \Sigma \mathbf{V}_n^{\top}$. Then,

$$\mathbf{A}_{m,n}^{\top}\mathbf{A}_{m,n} = \mathbf{V}_n(\Sigma^{\top}\Sigma)\mathbf{V}_n^{\top}. \quad (51)$$

As shown in the proof of Proposition 1, $\mathbf{V}_n$ is the deterministic unitary matrix in the eigendecomposition of $\mathbf{T}_n$. Hence, $\mathbf{A}_{m,n}$ satisfies the WSSJD property. ■

We construct several matrices that satisfy the WSSJD property, based on Propositions 1 and 2. In the following examples, we assume $\mathbf{T}_n$ to be the $n \times n$ symmetric matrix of the form:

$$\mathbf{T}_n = \begin{bmatrix} 0 & 1 & & & \mathcal{O} \\ 1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 0 & 1 \\ \mathcal{O} & & & 1 & 0 \end{bmatrix}. \quad (52)$$

On real HDDs, the dominant ITI often comes from the adjacent tracks, so (52) is a reasonable assumption for $\mathbf{T}_n$ to capture this phenomenon. The eigendecomposition of $\mathbf{T}_n$ has been well studied [33]. Examples 1–5 are constructed based on $\mathbf{T}_n$. The ITI parameters are represented by Greek letters.

Example 1: An interference matrix of the form

$$\mathbf{A}_n = \begin{bmatrix} \alpha & \epsilon & & & \mathcal{O} \\ \epsilon & \alpha & \epsilon & & \\ & \ddots & \ddots & \ddots & \\ & & \epsilon & \alpha & \epsilon \\ \mathcal{O} & & & \epsilon & \alpha \end{bmatrix} \quad (53)$$

satisfies the WSSJD property. It is easy to see that $\mathbf{A}_n = \alpha \mathbf{I}_n + \epsilon \mathbf{T}_n$.

Example 2: Consider the case when $n = 3$ and $N = 2$. Since

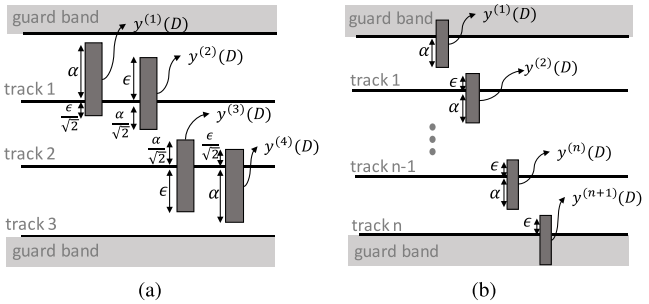
$$\mathbf{T}_3^2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad (54)$$

by Proposition 1, we know that $\mathbf{A}_3$ of the form

$$\mathbf{A}_3 = \alpha \mathbf{I}_3 + \beta \mathbf{T}_3 + \gamma \mathbf{T}_3^2 = \begin{bmatrix} \alpha+\gamma & \beta & \gamma \\ \beta & \alpha+2\gamma & \beta \\ \gamma & \beta & \alpha+\gamma \end{bmatrix} \quad (55)$$

satisfies the WSSJD property. We do not need to consider the case when $N \geqslant 3$ since

$$\mathbf{T}_3^i = \begin{cases} 2^{\lfloor i/2 \rfloor} \mathbf{T}_3, & i \text{ is odd} \\ 2^{i/2-1} \mathbf{T}_3^2, & i \text{ is even.} \end{cases} \quad (56)$$



(a) (b)

Fig. 8. Head alignments corresponding to the interference matrices in (a) Example 4 and (b) Example 5.

Example 3: It is obvious to see that $\mathbf{A}_{3,2}$ given in (4) satisfies

$$\mathbf{A}_{3,2}^{\top}\mathbf{A}_{3,2} = (1+\alpha^2+\epsilon^2)\mathbf{I}_2 + (\alpha^2+2\epsilon)\mathbf{T}_2. \tag{57}$$

Example 4: The 4×3 matrix $\mathbf{A}_{4,3}$ given by

$$\mathbf{A}_{4,3} = \begin{bmatrix} \alpha & \epsilon/\sqrt{2} & 0 \\ \epsilon & \alpha/\sqrt{2} & 0 \\ 0 & \alpha/\sqrt{2} & \epsilon \\ 0 & \epsilon/\sqrt{2} & \alpha \end{bmatrix} \tag{58}$$

satisfies $\mathbf{A}_{4,3}^{\top}\mathbf{A}_{4,3} = (\alpha^2+\epsilon^2)\mathbf{I}_3 + (\sqrt{2}\alpha\epsilon)\mathbf{T}_3$.

Example 5: Suppose $\mathbf{A}_{n+1,n}$ is of the form

$$\mathbf{A}_{n+1,n} = \begin{bmatrix} \alpha & & & \mathcal{O} \\ \epsilon & \alpha & & \\ & \ddots & \ddots & \\ & & \epsilon & \alpha \\ \mathcal{O} & & & \epsilon \end{bmatrix}. \tag{59}$$

Since $\mathbf{A}_{n+1,n}^{\top}\mathbf{A}_{n+1,n} = (\alpha^2+\epsilon^2)\mathbf{I}_n + (\alpha\epsilon)\mathbf{T}_n$, $\mathbf{A}_{n+1,n}$ satisfies the WSSJD property.

Fig. 8(a) and (b) shows possible physical head configurations for Examples 4 and 5, respectively. The amount of ITI sensed by each head, expressed in terms of the ITI parameters, is indicated in Fig. 8(a) and (b).

The ITI parameters are labeled in Fig. 8(a) and (b) to indicate the amount of ITI sensed by each head. Guard bands are added to prevent crosstalk between different track groups.

C. Discussion on Feasible Interference Matrices

Varnica *et al.* [11] investigate the optimal placement of three heads that achieves the lowest BER. They show that, when ITI is high, the optimal placement is to put two heads symmetrically over two tracks, and place the third head at a similar off-track as either one of the other two. As mentioned in [11], this asymmetric placement could be derived from the distance analysis of the simplified model. We assume the same interference matrix as in [11]

$$\mathbf{B}_{3,2} = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \tag{60}$$

where $\alpha \in [0.5, 1]$, and $\beta \in [0.5, 1]$. It indicates that a large α value leads to small ITI in the first two heads. Assume that the ISI channel is $h(D)$. The squared distance of a given error event $\boldsymbol{E}(D)$ is

$$\begin{aligned} &d^2(\boldsymbol{E}(D)) \\ &= \|\mathbf{B}_{3,2}\boldsymbol{E}(D)h(D)\|^2 \\ &= (\alpha^2+(1-\alpha)^2+\beta^2)\|e^{(1)}(D)h(D)\|^2 \\ &\quad + (\alpha^2+(1-\alpha)^2+(1-\beta)^2)\|e^{(2)}(D)h(D)\|^2 \\ &\quad + [4\alpha(1-\alpha)+2\beta(1-\beta)]\langle e^{(1)}(D)h(D), e^{(2)}(D)h(D)\rangle. \end{aligned} \tag{61}$$

The minimum squared distance considering all single-track error events is

$$d_s^2 = \left[2\left(\alpha-\frac{1}{2}\right)^2 + (\beta-1)^2 + \frac{1}{2}\right] d_0^2 \tag{62}$$

which can be achieved by setting $e^{(1)}(D) = 0$ and $e^{(2)}(D)$ to be the non-zero error event that leads to d_0^2 on $h(D)$.

For double-track error events, the minimum squared distance becomes

$$d_d^2 = \left[8\left(\alpha-\frac{1}{2}\right)^2 + 4\left(\beta-\frac{1}{2}\right)^2\right] d_0^2. \tag{63}$$

To achieve (63), choose $e^{(1)}(D)$ to be the dominant error event on channel $h(D)$, and set $e^{(2)}(D) = -e^{(1)}(D)$.

The overall minimum squared distance

$$d_{\min,\mathbf{B}_{3,2}}^2 = \min d^2(\boldsymbol{E}(D)) = \min\left\{d_s^2, d_d^2\right\}. \tag{64}$$

Consider the effect of varying α and β in the range of [0.5, 1]. Increasing α will increase both d_s^2 and d_d^2. Increasing β will decrease d_s^2, but increase d_d^2. The difference between d_d^2 and d_s^2 is

$$d_d^2 - d_s^2 = 6\left(\alpha-\frac{1}{2}\right)^2 + 3\left(\beta-\frac{1}{3}\right)^2 - \frac{5}{6}. \tag{65}$$

When α is close to 1, i.e., the ITI is small, it is easy to see $d_d^2 > d_s^2$, so the single-track error events dominate. To make $d_{\min}^2$ larger, a small value of β is preferred since it leads to larger d_s^2. In contrast, when α decreases to 0.5, i.e., the ITI is large, the double-track error events become more severe. Therefore, a large β value could help to improve the overall performance. In Fig. 9, we plot the minimum distance found for different values of β. Our analysis agrees with the simulation results in [11]. It is shown in [11, Fig. 13] that the optimal placement of the third head is to set $\beta = 0.5$ for small ITI and $\beta = 1$ for large ITI. In addition, the 3H2T system generally performs better when ITI is small.

We now examine the head placements that are amenable to the application of WSSJD. When $\beta = 0.5$, the matrix $\mathbf{B}_{3,2}$ is equivalent to $\mathbf{A}_{3,2}$ in (4) up to a scaling and row permutation. When $\beta = 1$, the matrix

$$\mathbf{B}_{3,2} = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \\ 1 & 0 \end{bmatrix} \tag{66}$$

does not satisfy the WSSJD property. However, it can be partitioned into a 2×2 submatrix that has the desired property, and a 1×2 deterministic matrix. To apply WSSJD, only the

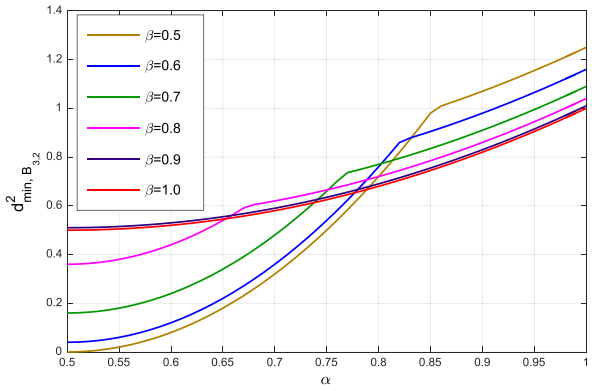


Fig. 9. Minimum squared distance of the 3H2T system with $\mathbf{B}_{3,2}$. Assume $d_0 = 1$.

2×2 submatrix needs to be transformed. Varnica *et al.* [11] also considered an interference matrix of the form

$$\mathbf{B}_{3,2} = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{bmatrix}. \tag{67}$$

To apply WSSJD to this channel, we can average the readback signals from heads 1 and 3. Then, the resulting channel becomes a symmetric 2H2T system that has been studied in [2].

VI. Conclusion

The WSSJD algorithm is an alternative way to implement the ML detection for the nHnT channel. Its advantage in practical implementations is that the branch labels in a WSSJD trellis are independent of the ITI parameters. In this paper, we extend WSSJD to the generalized MHMT channels whose interference matrices satisfy a structural property we call the WSSJD property. It is shown that gain loop structures can be incorporated into WSSJD to adaptively estimate the unknown ITI parameters. We mainly focus on the performance of WSSJD in the context of 3H2T systems, since they are potential candidates for implementation in the new generation of TDMR drives. Simulation results show that WSSJD with gain-loop-based ITI estimation can achieve a near optimal performance in a time-varying ITI environment. We also analyze the head alignment for the 3H2T system assumed in [11] from the minimum distance perspective. Finally, we show that WSSJD can be applied with a simple modification to many feasible interference matrices derived in [11] .

Appendix
Minimum Distance of 3H2T System

We consider the channel model given by (3) and (4). Let $\boldsymbol{E}(D)$ be an error event. The squared minimum distance is calculated by

$$\begin{aligned} & d^2(\boldsymbol{E}(D)) \\ &= \|\mathbf{A}_{3,2}\boldsymbol{E}(D)h(D)\|^2 \\ &= \|e^{(1)}(D)h(D) + \epsilon \cdot e^{(2)}(D)h(D)\|^2 \\ &\quad + \alpha^2\|e^{(1)}(D)h(D) + e^{(2)}(D)h(D)\|^2 \\ &\quad + \|\epsilon \cdot e^{(1)}(D)h(D) + e^{(2)}(D)h(D)\|^2 \\ &= (1+\alpha^2+\epsilon^2)(\|e^{(1)}(D)h(D)\|^2 + \|e^{(2)}(D)h(D)\|^2) \\ &\quad + (4\epsilon + 2\alpha^2)\langle e^{(1)}(D)h(D), e^{(2)}(D)h(D)\rangle \end{aligned} \tag{68}$$

where $\|\cdot\|$ calculates the norm of the vector and $\langle\cdot,\cdot\rangle$ is the inner product of two sequences. We find $\min d^2(\boldsymbol{E}(D))$ by considering the single-track error events and double-track error events separately.

1) *Single-Track Error Events:* Since the system is symmetric, it is sufficient to consider the case when $e^{(2)}(D) = 0$. Under this assumption, (68) reduces to

$$\begin{aligned} d^2(e^{(1)}(D)) &= (1+\alpha^2+\epsilon^2)\|e^{(1)}(D)h(D)\|^2 \\ &\geqslant (1+\alpha^2+\epsilon^2)d_0^2 \end{aligned} \tag{69}$$

where d_0 is the minimum distance measured on the single-track ISI channel with $h(D)$.

2) *Double-Track Error Events:* We have

$$\begin{aligned} & \langle e^{(1)}(D)h(D), e^{(2)}(D)h(D)\rangle \\ &\geqslant -|\langle e^{(1)}(D)h(D), e^{(2)}(D)h(D)\rangle| \\ &\geqslant -\|e^{(1)}(D)h(D)\| \cdot \|e^{(2)}(D)h(D)\| \\ &\geqslant -\frac{1}{2}(\|e^{(1)}(D)h(D)\|^2 + \|e^{(2)}(D)h(D)\|^2). \end{aligned} \tag{70}$$

The second inequality is based on the Cauchy–Schwarz inequality. Using (70), we can get a lower bound of (68)

$$\begin{aligned} & d^2(e^{(1)}(D), e^{(2)}(D)) \\ &\geqslant (\epsilon^2 - 2\epsilon + 1)(\|e^{(1)}(D)h(D)\|^2 + \|e^{(2)}(D)h(D)\|^2) \\ &\geqslant 2(1-\epsilon)^2 d_0^2. \end{aligned} \tag{71}$$

To obtain $d^2_{\min} = 2(1-\epsilon)^2\, d_0^2$, choose $e^{(1)}(D)$ to be the error event that leads to d_0^2 on channel $h(D)$, and set $e^{(2)}(D) = -e^{(1)}(D)$. Therefore, $d^2_{\min} = 2(1-\epsilon)^2\, d_0^2$ is an achievable lower bound.

Comparing (69) and (71) and taking the minimum, we conclude that

$$d^2_{\min,3H2T} = \begin{cases} (1+\alpha^2+\epsilon^2)d_0^2, & \text{if } \alpha^2 > 1 - 4\epsilon + \epsilon^2 \\ 2(1-\epsilon)^2 d_0^2, & \text{otherwise.} \end{cases} \tag{72}$$

Acknowledgment

This work was supported by the National Science Foundation under Grant CCF-1405119 and Grant CCF-1619053.

References

[1] 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.

[2] B. Fan, H. K. Thapar, and P. H. Siegel, "Multihead multitrack detection for next generation magnetic recording. Part I: Weighted sum subtract joint detection with ITI estimation," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1635–1648, Apr. 2017.

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

[4] B. Fan, H. K. Thapar, and P. H. Siegel, "Multihead multitrack detection for next generation magnetic recording. Part II: Complexity reduction—Algorithms and performance analysis," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1649–1661, Apr. 2017.

[5] 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.

[6] M. P. Vea and J. M. F. Moura, "Magnetic recording channel model with intertrack interference," *IEEE Trans. Magn.*, vol. 27, no. 6, pp. 4834–4836, Nov. 1991.
[7] P. A. Voois and J. M. Cioffi, "Multichannel signal processing for multiple-head digital magnetic recording," *IEEE Trans. Magn.*, vol. 30, no. 6, pp. 5100–5114, Nov. 1994.
[8] P. A. Voois and J. M. Cioffi, "Upper bounds on achievable storage density: A two-dimensional approach," *IEEE Trans. Magn.*, vol. 33, no. 1, pp. 844–854, Jan. 1997.
[9] 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.
[10] 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.
[11] N. Varnica, R. Radhakrishnan, S. K. Chilappagari, M. Khatami, and M. Öberg, "Comparison of two-reader and three-reader 2-D magnetic recording systems," *IEEE Trans. Magn.*, vol. 52, no. 2, pp. 1–8, Feb. 2016.
[12] G. Mathew, E. Hwang, J. Park, G. Garfunkel, and D. Hu, "Capacity advantage of array-reader-based magnetic recording (ARMR) for next generation hard disk drives," *IEEE Trans. Magn.*, vol. 50, no. 3, pp. 155–161, Mar. 2014.
[13] M. R. Elidrissi, K. S. Chan, and Z. Yuan, "A study of SMR/TDMR with a double/triple reader head array and conventional read channels," *IEEE Trans. Magn.*, vol. 50, no. 3, pp. 24–30, Mar. 2014.
[14] R. Wood, R. Galbraith, and J. Coker, "2-D magnetic recording: Progress and evolution," *IEEE Trans. Magn.*, vol. 51, no. 4, pp. 1–7, Apr. 2015.
[15] Y. Shiroishi *et al.*, "Future options for HDD storage," *IEEE Trans. Magn.*, vol. 45, no. 10, pp. 3816–3822, Oct. 2009.
[16] M. Marrow and J. K. Wolf, "Iterative detection of 2-dimensional ISI channels," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Apr. 2003, pp. 131–134.
[17] 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.
[18] 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.
[19] 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.
[20] E. Soljanin and C. N. Georghiades, "On coding in multi-track, multihead, disk recording systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, vol. 4, Houston, TX, USA, Nov. 1993, pp. 18–22.
[21] 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.
[22] 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.
[23] H. Xia, L. Lu, S. Jeong, L. Pan, and J. Xiao, "Signal processing and detection for array reader magnetic recording system," *IEEE Trans. Magn.*, vol. 51, no. 11, pp. 1–4, Nov. 2015.
[24] S. Shi and J. R. Barry, "Multitrack detection with 2D pattern-dependent noise prediction," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.
[25] J. R. Barry *et al.*, "Optimization of bit geometry and multi-reader geometry for two-dimensional magnetic recording," *IEEE Trans. Magn.*, vol. 52, no. 2, pp. 1–7, Feb. 2016.
[26] K. S. Chan *et al.*, "User areal density optimization for conventional and 2-D detectors/decoders," *IEEE Trans. Magn.*, vol. 54, no. 2, pp. 1–12, Feb. 2018.
[27] 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.
[28] M. Fujii and N. Shinohara, "Multi-track iterative ITI canceller for shingled write recording," in *Proc. Int. Symp. Commun. Inform. Tech. (ISCIT)*, Tokyo, Japan, Oct. 2010, pp. 1062–1067.
[29] 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)*, Budapest, Hungary, Jun. 2013, pp. 4342–4346.
[30] 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.
[31] M. V. Eyuboglu and S. U. H. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. COM-36, no. 1, pp. 13–20, Jan. 1988.
[32] B. H. Marcus, R. M. Roth, and P. H. Siegel. (Oct. 2001). *An Introduction to Coding for Constrained Systems*. [Online]. Available: http://cmrr-star.ucsd.edu/psiegel/book_draft
[33] 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.