Data Recovery for Multilayer Magnetic Recording

Kheong Sann Chan[1], Ahmed Aboutaleb[2], Krishnamoorthy Sivakumar[2], Benjamin Belzer[2], Roger Wood[2], *Life Fellow, IEEE*, and Susanto Rahardja[3], *Fellow, IEEE*

[1]Nanjing Institute of Technology, Nanjing 211167, China
[2]School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99164 USA
[3]School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710129, China

Multilayer magnetic recording (MLMR) is a next-generation technology that has the potential to enhance the longevity of magnetic recording technology by storing data on multiple layers that are stacked on top of one another. The challenges associated with the development of MLMR fall into two main categories: the challenges of writing to multiple layers and the challenges of reading back from multiple layers. As there is already a fair amount of existing work addressing the former, in this article we target the latter by building and testing detection schemes that are able to separate the signals from multiple layers that are mixed together during the readback. In this initial work, we consider a two-layer system. The resolution of the bottom layer is worse than that of the top due to the increased separation of the bottom layer from the reader. We assume a geometry of two single-width tracks on the top-layer straddling a double-width track on the bottom layer and with intersymbol interference (ISI) response on the bottom layer having half the amplitude and double the length of that on the top-layer. We examine the performance of two candidate detection algorithms, a Viterbi-based algorithm operating on a trellis, and a least-squared (LS) algorithm that attempts to invert the effect of the channel's interference. We evaluate the error rate performance of these detection schemes using a common channel model that sums the ISI and intertrack interference (ITI) components from each of the layers. The main result from this work is that 14.5% density gains over the conventional perpendicular magnetic recording (PMR) and 7.6% gains over two-dimensional magnetic recording (TDMR) could be possible for the novel MLMR system.

Index Terms— **Channel model, least-squared (LS) detector, microwave-assisted magnetic recording (MAMR), multilayer magnetic recording (MLMR), Viterbi.**

I. Introduction

THE magnetic recording industry writes data onto the surface of a recording medium in circular tracks with magnetization at 90° to the medium's surface with a technology known as perpendicular magnetic recording (PMR). In [1], it was predicted that PMR might be able to achieve about 1 Terabit per square inch (Tbpsi), a prediction that has been largely confirmed by the technology development since then. Subsequently in 2009, two-dimensional magnetic recording (TDMR) [2], [3] was proposed to carry the industry upward beyond 1 Tbpsi. Other candidate technologies also proposed in this same time-period include heat-assisted magnetic recording (HAMR) [4], [5], microwave-assisted magnetic recording (MAMR) [6]–[8], and bit-patterned media recording (BPMR) [9], [10].

Strongly competing with the magnetic recording technology in the hard disk drive (HDD), is flash memory technology [11], [12] in the solid state drive (SSD). Flash technology has made great strides recently by arranging the flash memory cells not only on the 2-D surface but also by stacking many cells vertically upward, taking advantage of the third dimension. This approach has allowed flash technology to come close to the HDD's traditional cost per byte advantage. A similar approach has not been viewed as being helpful for the HDD. Magnetic recording technology on which the HDD is based, currently only deposits a single recording layer onto the medium. The separation between the recording layer and the read/write head has a great influence on the density at which data can be stored on the media grains. When multiple layers are deposited, the lower layers are further from the read/write head and are, therefore, expected to hold less data than the top layer. Nevertheless, reasonable density gains may still be possible using appropriate coding and detection schemes. A preliminary investigation in [13] has suggested that gains of around 17% may be possible. In that work, one isolated track in the top layer over one isolated track in the bottom layer was assumed, with just one reader, and with different downtrack bit densities on the bottom layer. In this article, we introduce the impact of the cross-track dimension as well. We examine a geometry with two side-by-side tracks in the top-layer straddling a single, double-width track on the bottom layer, again with varying downtrack densities on the bottom layer, and we investigate two and three-reader configurations. We also assume a write process that obeys linear superposition and media noise that is stationary, as it enables us to focus on the main contribution of the work: the development of the readback channel. This means that the effects of imperfect writing such as nonlinear transition-shift, transition noise, and the coloration of the media noise are not included. However, this simplification in the model applies to both the novel and reference systems that are compared in this work. We also assume synchronization in the writing of the top and bottom layers. In practice, we do not expect this to be a serious problem as the same patterns can be written through both layers that would allow the synchronization of the writing to the top and bottom layers.

Manuscript received June 21, 2019; revised August 4, 2019; accepted August 18, 2019. Date of current version November 18, 2019. Corresponding author: K. S. Chan (e-mail: kheongsann@ieee.org).

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

Digital Object Identifier 10.1109/TMAG.2019.2937692

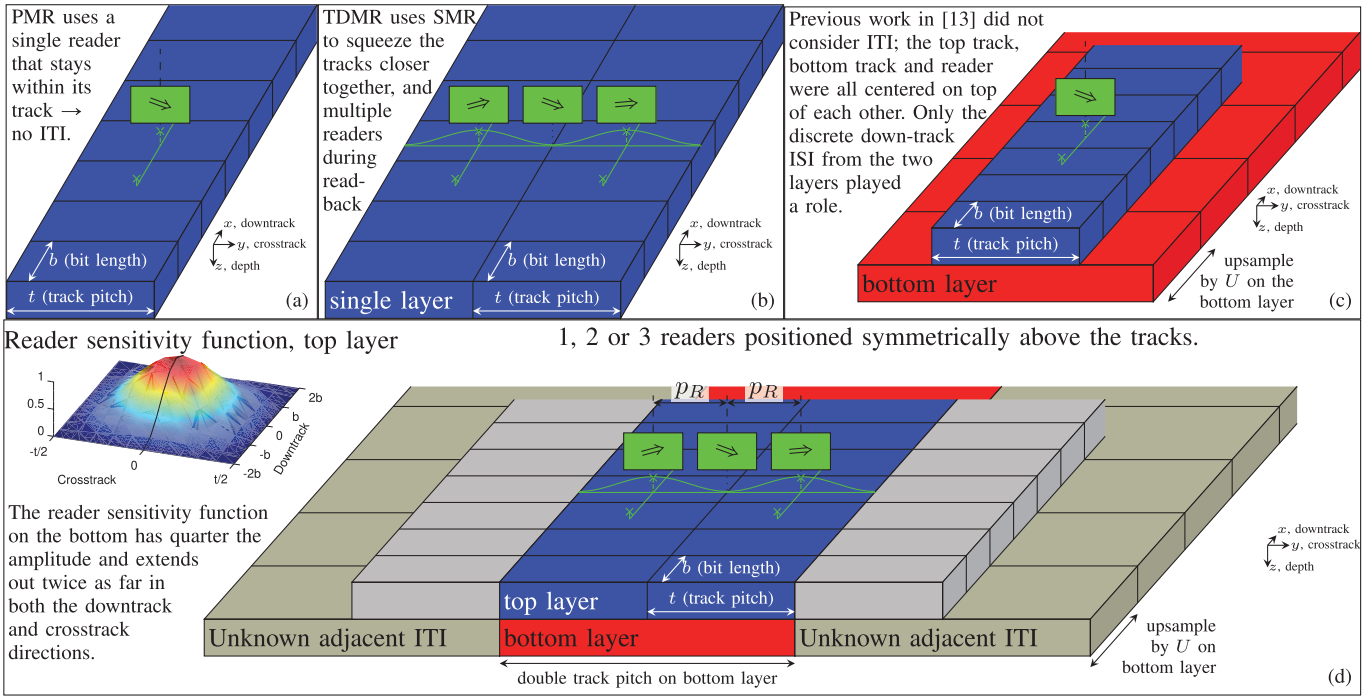

Fig. 1. Comparison of PMR, TDMR, previous work in [13], and the current work. In this work, we attempt to jointly detect the bits from the two tracks in the top layer and single track in the bottom layer.

The architectural evolution from 1-D to 2-D and now to the 3-D recording is depicted in Fig. 1. The original 1-D PMR technology is shown in Fig. 1(a), in which the reader width is maintained at approximately 60% of the track width to ensure that intertrack interference (ITI) is negligible, and therefore, 1-D detection schemes perform adequately. The proposal for TDMR [in Fig. 1(b)] was to squeeze down the track pitch using the shingled magnetic recording (SMR), a technique that overlaps successively written tracks in a shingle block. This enables the writing of tracks narrower than the magnetic write width (MWW). To achieve the equivalent of reading tracks narrower than the magnetic read width (MRW), it was proposed to use 2-D coding and detection schemes to mitigate the ITI caused by the partial reading of the adjacent track [14]. However, TDMR could not circumvent the decrease in the number of grains per bit cell as the track width shrank, leading to a decrease in SNR and reduced capacity.

The traditional method of increasing areal density (AD) is to produce media with smaller grains, and that has tighter distributions. Smaller grains lead to the ability to pack larger quantities of grains onto the same area of the medium, whereas tighter distributions lead to higher quality grains that are able to support higher SNRs at fewer grains per bit cell, with BPMR being the ultimate goal in the evolution toward increasing grain quality. However, at a given level of technology, typically the grain size (or quantity) and grain quality vary inversely with each other: smaller grains tend to produce larger distributions. By adding a second layer, we can potentially double the quantity of grains at the same level of grain quality.

Fig. 1(c) shows the geometry of the setup in [13], where the target was to build and test the performance of a detector that is able to separate the signals from the two layers. ITI was not considered with only a single isolated track on the top layer centered over a single isolated track on the bottom layer, and the reader centered over both tracks. Similar to the PMR system, only downtrack intersymbol interference (ISI) played a role in this scenario. We adopt a naming convention for the various architectures that identify the number of layers, tracks, and readers used in the architecture. In this convention, the architecture in Fig. 1(c) is referred to as 2L2T1R, for two layers, two tracks, and one reader.

The geometry used in this work has two layers, three tracks, and two or three readers, that we refer to as two-layer, three-track, three-reader (2L3T3R) architecture. It is shown in Fig. 1(d), with two tracks on the top layer straddling a single double-width track on the bottom as the home tracks containing the target data bits. There are also four tracks containing interfering bits adjacent to these three home tracks, on both sides and on both layers. We investigate situations with one, two, or three readers placed symmetrically over the tracks, that pick up the signals from the target tracks as well as the interfering ITI tracks.

In this article, we propose a new generalized channel model for multilayer magnetic recording (MLMR), that can turn on and off various features of the recording, such as individual readers, tracks, and the ITI from the adjacent tracks. We develop and test the performance of two detection schemes to operate over the MLMR channel: the least squares (LSs) and the Viterbi detectors. Using the model and detection schemes, we investigate the impact of varying the density on the lower layer, and the cross-track positions of the read heads on the overall achievable user bit density (UBD).

Algorithm 1 Iterative LSs for Bit Sequences Estimation in MLMR Channels. Note That for Each SNR Value Locally Optimal K, K_i, ϵ, and ϵ_i Are Determined Beforehand via Offline Training to Minimize Individual BERs for the Upper and Lower Layers

We are given an operating SNR and, hence, σ_n.
Pick K and K_i corresponding to the operating SNR.
Set to zero any σs such that $\sigma^2 < \max\{K\sigma_n^2, \epsilon\}$ and any σ_is such that $\sigma_i^2 < \max\{K_i\sigma_n^2, \epsilon_i\}$ $(i = L, U)$.
Compute $\mathbf{H}^\dagger = \mathbf{V}\Sigma^{-1}\mathbf{U}^T\mathbf{H}^T\mathbf{C}_\mathbf{Z}^{-1}$ and $\mathbf{H}_i^\dagger = \mathbf{V}_i\Sigma_i^{-1}\mathbf{U}_i^T\mathbf{H}_i^T\mathbf{C}_\mathbf{Z}^{-1}$.
(Steps 1 to 4 can be performed offline if the operating SNR is known.)
for $j = 1$ to Number of Iterations **do**
  if $j = 1$ **then**
    $\hat{\mathbf{A}}_{(j)}^T = [\hat{\mathbf{A}}_{\mathrm{L},(j)}^T, \hat{\mathbf{A}}_{\mathrm{U},(j)}^T, \hat{\mathbf{A}}_{\mathrm{ITI},(j)}^T] = \mathrm{sign}\{\mathbf{H}^\dagger\mathbf{R}\}$
  else
    $\mathbf{R}_2 = \mathbf{R} - \mathbf{H}_L^{\mathrm{eff}}\hat{\mathbf{A}}_{\mathrm{L},(j-1)}$
    $\hat{\mathbf{A}}_{\mathrm{U},(j)} = \mathrm{sign}\{\mathbf{H}_U^\dagger\mathbf{R}_2\}$
    $\mathbf{R}_1 = \mathbf{R} - \mathbf{H}_U^{\mathrm{eff}}\hat{\mathbf{A}}_{\mathrm{U},(j-1)}$
    $\hat{\mathbf{A}}_{\mathrm{L},(j)} = \mathrm{sign}\{\mathbf{H}_L^\dagger\mathbf{R}_1\}$
    if BER $\leq$ Desired Threshold **then** break
    end if
  end if
end for

II. Channel Model

In this work, we focus on the readback of MLMR, and we assume that the writing process has been appropriately handled. This could potentially have been accomplished through a multilayer MAMR writing scheme such as described in [15] and [16]. The block diagram that we implement to model the readback channel is shown in Fig. 2. There are three input bit sequences: $a_{2L}[k]$, $a_1[k]$, and $a_{2R}[k]$ that are written onto the top-left, bottom, and top right tracks, respectively, with $a_1[k]$ being written at $1/U$ the density of a top-layer track, where U is an integer parameter chosen by the user. Each track contributes its signal to each of three possible readers labeled L (left), C (center), and R (right) that are above the tracks at positions p_R, p_C, and p_L. In this work, we assume symmetrical reader positioning at $-p_R$, 0, and p_R, respectively. We also investigate the performance of the two-reader scenario, omitting the central reader, to see how much gain is obtained from introducing the third reader. One of the investigations performed in this work is the impact of varying the cross-track reader location p_R; however, for most of the simulations, the readers are in the default "home position", with each reader centered over one of the tracks at positions $p_L = -t/2$, $p_C = 0$, and $p_R = +t/2$, shown in Fig. 1 and the inset of Fig. 2.

A. Evaluation of the ISI Coefficients Weights: β_i

Each of the three data-streams contributes to each of the (up to) three readers through nine possible linear ISI filters in the paths labeled 1 through 9 in Fig. 2. The basic form of the discrete ISI components are the same as those used in [13], with $\mathbf{h}_1 = [0.0732, 0.25, 0.4268, 0.5, 0.4268, 0.25, 0.0732]$ for the bottom layer and $\mathbf{h}_2 = [0.5, 1, 0.5]$, for the top layer, respectively, that correspond to samples from a raised-cosine pulse. These are the channel coefficients after integration over a wide track, equalization, and sampling. There is an underlying assumption here that the overall system design is such that the response can be equalized to a relatively short target while, simultaneously, the total noise presented to the Viterbi detector is approximately white. This ensures that performance is close to the maximum likelihood (ML). As a simplification, we assume the noise from the head and from the media layers are all white and stationary. In reality, the head noise as seen at the detector would be slightly blue and the media noise would be slightly red in order that the total noise be white. Also, the media noise would be signal dependent. These simplifications are acknowledged shortcomings of the model.

In this work, we continue the usage of the nomenclature from [13] of having subscript 1 and 2 refer to the bottom and top layers, respectively. The basic downtrack responses $\mathbf{h}_1$ and $\mathbf{h}_2$ are weighted by coefficients β_i that depend on the cross-track position of the particular read head, equal to the area under the cross-track response that overlaps the profile of the rectangular track being sensed. In this work, we assume *discrete* downtrack coefficients in $\mathbf{h}_1$ and $\mathbf{h}_2$, but a *continuous* cross-track profile, with the underlying shape being a raised-cosine pulse in both cases as depicted in Fig. 1. The reason for this difference in the way we treat the downtrack and cross-track profiles is that the downtrack readback is sampled at discrete time instances prior to being processed by the digital back-end channel, whereas the reader can pick up the signal from anywhere in a continuum of cross-track locations $y \in \mathbb{R}$. To evaluate the β_i coefficients, we define the continuous cross-track profile functions

$$\begin{aligned} h_2(y) &= 0.5(1+\cos(2\pi y)), \quad -0.5 \leq y \leq 0.5 \\ h_1(y) &= 0.25(1+\cos(\pi y)), \quad -1.0 \leq y \leq 1.0 \end{aligned} \tag{1}$$

as two raised-cosine pulses that are zero outside their defined ranges, representing the continuous cross-track response functions on the top and bottom layers, normalized to the track pitch t of the top layer. Note that the resolution on the lower layer (h_1) is worse by a factor of two compared to the resolution on the of the upper layer (h_2). We define cross-track weighting functions $\beta_2(y)$ and $\beta_1(y)$ as the fraction of the area under the reader sensitivity function (RSF) overlapping the track when the track and reader centers are offset by the normalized value of y, for the top and bottom layers, respectively. They are

$$\beta_2(y) = \frac{\int_{-0.5}^{0.5} h_2(y'-y)dy'}{\int_{-\infty}^{\infty} h_2(y'-y)dy'} \text{ and } \beta_1(y) = \frac{\int_{-1}^{1} h_1(y'-y)dy'}{\int_{-\infty}^{\infty} h_1(y'-y)dy'} \tag{2}$$

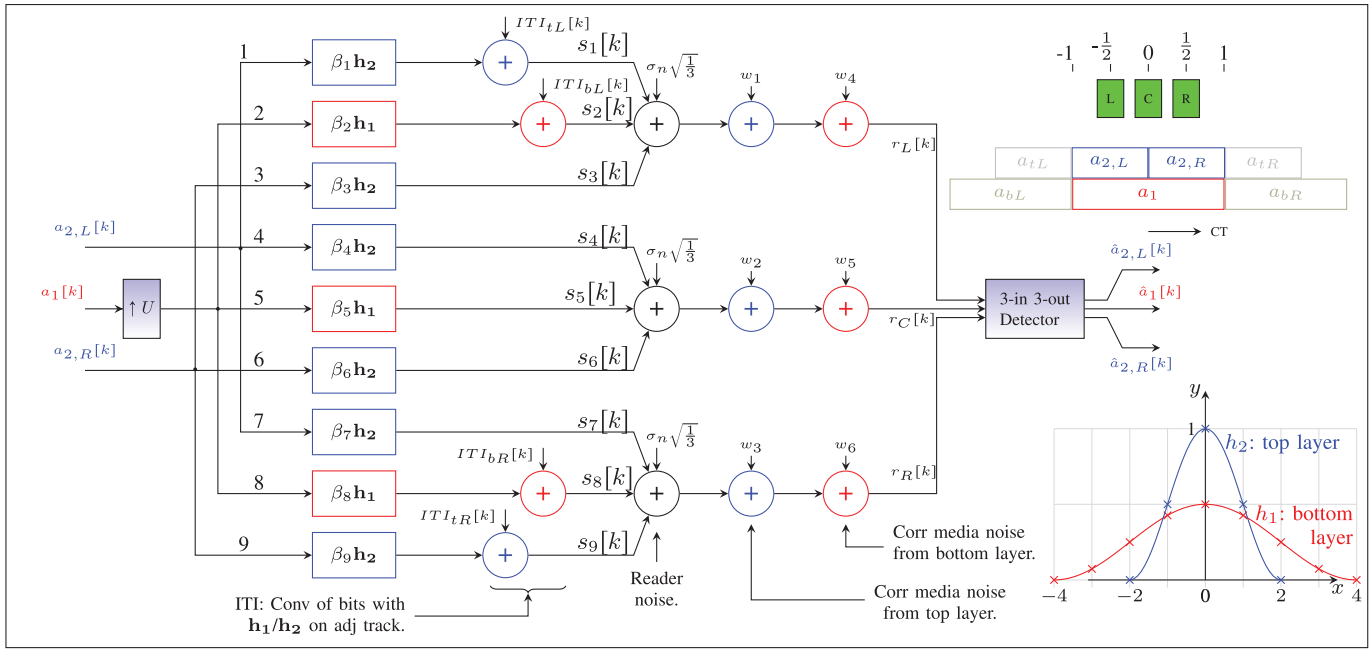

Fig. 2. Channel block diagram for the MLMR system, modeled as the summation of linear ISI components from each track to each reader, plus the noise from each media layer and noise from each reader. The full three-reader configuration is shown.

respectively. $\beta_2(y)$ and $\beta_1(y)$ can be evaluated to

$$\beta_2(y) = \begin{cases} -\frac{\sin(2\pi y)}{2\pi} + y + 1, & -1 \le y \le 0 \\ \frac{\sin(2\pi y)}{2\pi} - y + 1, & 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\beta_1(y) = \begin{cases} -\frac{\sin(\pi y)}{2\pi} + \frac{y}{2} + 1, & -2 \le y \le 0 \\ \frac{\sin(\pi y)}{2\pi} - \frac{y}{2} + 1, & 0 \le y \le 2 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

and are shown in Fig. 3. From $\beta_2(y)$ and $\beta_1(y)$, the β_i coefficients in Fig. 2 can be easily derived with appropriate choices of y in each case. The values of β_i for three readers in their home positions are

$$\begin{bmatrix} \beta_1 & \beta_2 & \beta_3 \\ \beta_4 & \beta_5 & \beta_6 \\ \beta_7 & \beta_8 & \beta_9 \end{bmatrix} = \begin{bmatrix} 1.0000 & 0.9092 & 0.0000 \\ 0.5000 & 1.0000 & 0.5000 \\ 0.0000 & 0.9092 & 1.0000 \end{bmatrix}. \quad (4)$$

B. Modeling of the ITI

Following the downtrack ISI filters in Fig. 2 are four adders that introduce the unwanted ITI from the outside unknown tracks that also depend on the readers' cross-track positions. The adjacent-track ITI filters are similarly versions of $\mathbf{h}_1$ and $\mathbf{h}_2$ scaled by a factor equal to the area of overlap of the cross-track profile, with the interfering track. These can be evaluated to

$$\gamma_2(y) = \begin{cases} \frac{\sin(2\pi y)}{2\pi} + y - \frac{1}{2}, & 0.5 \le y \le 1.5 \\ 0, & -0.5 \le y \le 0.5 \\ -\frac{\sin(2\pi y)}{2\pi} - y - \frac{1}{2}, & -1.5 \le y \le -0.5 \\ 1, & \text{otherwise} \end{cases}$$

$$\gamma_1(y) = \begin{cases} -\frac{\sin(\pi y)}{2\pi} + \frac{y}{2}, & 0 \le y \le 2 \\ \frac{\sin(\pi y)}{2\pi} - \frac{y}{2}, & -2 \le y \le 0 \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

and are also shown in Fig. 3. The four values of the γ coefficients in the readers' home positions are

$$\begin{bmatrix} \gamma_{tL} & \gamma_{tR} \\ \gamma_{bL} & \gamma_{bR} \end{bmatrix} = \begin{bmatrix} 0.0000 & 0.0000 \\ 0.09085 & 0.09085 \end{bmatrix} \quad (6)$$

respectively. To obtain the ITI sequences ITI_{tL}, ITI_{bL}, ITI_{tR}, and ITI_{bR} in Fig. 2, adjacent-track bit sequences $a_{tL}[k]$, $a_{bL}[k]$, $a_{tR}[k]$, and $a_{bR}[k]$ are generated and convolved with the weighted ISI filters: $\gamma_{tL}\mathbf{h}_2$, $\gamma_{bL}\mathbf{h}_1$, $\gamma_{bR}\mathbf{h}_1$, and $\gamma_{tR}\mathbf{h}_2$, respectively.

C. Reader Noise and Media Noise

To incorporate the media and reader noise, we adopt the same approach as in [13]. In the reference PMR system with a single reader over a single track on a single layer, approximately two-thirds of the noise power comes from the medium while approximately one-third comes from the reader. The reader noise and top-layer media noise rms remain as in the conventional case at $\sigma_n(1/3)^{1/2}$ and $\sigma_n(2/3)^{1/2}$, respectively, where the value of σ_n is related to the SNR of our simulations by

$$\sigma_n^2 = \frac{\sum_k h_2^2[k]}{10^{\text{SNR}/10}} = \frac{1.5}{10^{\text{SNR}/10}}. \quad (7)$$

Throughout this article, this reference value of SNR is used and it always refers to the single-head, single-layer configuration. The media noise power of the bottom layer is a quarter that of the top due to a 2× loss in resolution in both the downtrack and cross-track directions. It is also further scaled

down by a factor of U to account for the lower bit density on the bottom layer, giving a bottom-layer media noise rms of $\sigma_n(2/3)^{1/2}/(4U)^{1/2} = \sigma_n/(6U)^{1/2}$.

The media noise sequences w_1, w_2, and w_3 on the top layer (and w_4, w_5, and w_6 on the bottom layer) are individually modeled as white in this work, but the noise within each layer as seen by any two readers at different positions, is correlated between the readers. This correlation is caused by the overlap of the RSFs over the same set of grains on each layer. The noise correlation between two readers, located at normalized positions p_1 and p_2 (with $p_2 > p_1$) both with cross-track RSF $h(y)$ that ranges from $-t/2$ to $t/2$, due to the overlap of their sensitivity functions on the same set of grains is given by

$$A(p_1, p_2) = \frac{\int_{p_2-\frac{t}{2}}^{p_1+\frac{t}{2}} h(y-p_1)h(y-p_2)dy}{\int_{-\frac{t}{2}}^{+\frac{t}{2}} h^2(y)dy}. \tag{8}$$

The correlation is dependent only on the separation of the two readers $y = p_2 - p_1$ ($y > 0$) and when substituting in (1) and (8) can be evaluated to

$$\begin{aligned} A_1(y) &= \frac{\sin(\pi y)}{2\pi} - \frac{(y-2)[\cos(\pi y)+2]}{6}, \quad 0 \le y \le 2 \\ A_2(y) &= \frac{\sin(2\pi y)}{2\pi} - \frac{(y-1)[\cos(2\pi y)+2]}{3}, \quad 0 \le y \le 1 \end{aligned} \tag{9}$$

for the bottom and top layers, respectively. These top and bottom correlations are shown in Fig. 4. In the readers' home positions, the noise correlation matrices evaluate to

$$\begin{aligned} \mathbf{A}_1 &= \frac{\sigma_n^2}{6U}\begin{bmatrix} 1.0000 & 0.6592 & 0.1667 \\ 0.6592 & 1.0000 & 0.6592 \\ 0.1667 & 0.6592 & 1.0000 \end{bmatrix} \\ \mathbf{A}_2 &= \frac{2\sigma_n^2}{3}\begin{bmatrix} 1.0000 & 0.1667 & 0.0000 \\ 0.1667 & 1.0000 & 0.1667 \\ 0.0000 & 0.1667 & 1.0000 \end{bmatrix} \end{aligned} \tag{10}$$

on the bottom and top layers, respectively. Within the model of the recording channel, we wish to create three correlated bit streams w_1, w_2, and w_3 on the top layer and three correlated bit streams w_4, w_5, and w_6 on the bottom layer with cross correlation coefficients as given in (10). This is accomplished by multiplying the unity variance white noise sequences n_1, n_2, and n_3 and n_4, n_5, and n_6 by matrices $\mathbf{C}_2$ and $\mathbf{C}_1$, respectively. The matrices that create the desired correlated noise are

$$\mathbf{C}_1 = \sqrt{\mathbf{A}_1} \text{ and } \mathbf{C}_2 = \sqrt{\mathbf{A}_2} \tag{11}$$

respectively. The symmetric square root of matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ can be obtained using a standard technique of solving for the eigenvalues and eigenvectors of $\mathbf{A}_1$ and $\mathbf{A}_2$, which allows us to write $\mathbf{A}_i = \mathbf{V}_i \Lambda_i \mathbf{V}_i^T$, where $\mathbf{V}_i$ is an orthonormal matrix holding the eigenvectors of $\mathbf{A}_i$ in each column and Λ_i is a diagonal matrix holding its eigenvalues on the diagonal, from which we can get $(\mathbf{A}_i)^{1/2} = \mathbf{V}_i(\Lambda_i)^{1/2}\mathbf{V}_i^T$. The correlated

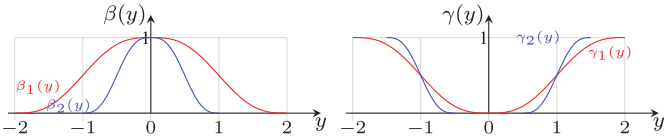

Fig. 3. $\beta(y)$ and $\gamma(y)$ functions, are the percentage overlap of the response function with the home tracks and ITI tracks, respectively.

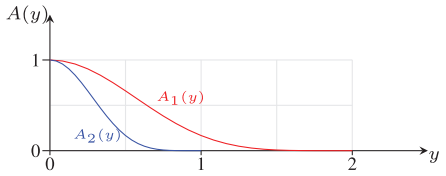

Fig. 4. Noise correlation functions for the top and bottom layers, for two readers separated by a cross-track distance y.

noise sequences in Fig. 2 can be obtained from the matrices $\mathbf{C}_2$ and $\mathbf{C}_1$ as

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \mathbf{C}_2 \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} \text{ and } \begin{bmatrix} w_4 \\ w_5 \\ w_6 \end{bmatrix} = \mathbf{C}_1 \begin{bmatrix} n_4 \\ n_5 \\ n_6 \end{bmatrix} \tag{12}$$

respectively. In the readers' home positions at SNR=10 dB, $\sigma_n^2 = 0.15$, and $U = 1$, the matrices become

$$\begin{aligned} \mathbf{C}_1 &= \begin{bmatrix} 0.1472 & 0.0577 & 0.0028 \\ 0.0577 & 0.1354 & 0.0577 \\ 0.0028 & 0.0577 & 0.1472 \end{bmatrix} \\ \mathbf{C}_2 &= \begin{bmatrix} 0.3151 & 0.0265 & -0.0011 \\ 0.0265 & 0.3140 & 0.0265 \\ -0.0011 & 0.0265 & 0.3151 \end{bmatrix} \end{aligned} \tag{13}$$

respectively. As well as introducing the correct correlation, these matrices adjust the variances of the noise to the appropriate levels. To achieve ML performance, this cross-track noise correlation that is introduced in the channel must be removed before the signals are presented to the joint Viterbi detector.

A key aspect of the MLMR model is the separability of the sensitivity function in the downtrack and cross-track dimensions, that avoids the need for a 2-D convolution and allows the use of simpler 1-D downtrack ISI filters adjusted by a cross-track weighting coefficient. The model also incorporates unwanted ITI from the two outside neighboring tracks on the upper and lower layers, as well as reader and media noise in proportions typical of that in a real recording system. The media noise injected into the readers in our model is correlated when the RSFs intersect, to account for them responding to the same grains in the overlapping regions. The model also accounts for a lower linear density in the downtrack direction by an integer amount U, and a constant factor of half in the cross-track dimension, on the bottom layer.

III. Channel Detectors

Based on the channel model previously described and shown in Fig. 2, readback signals $r_L[k]$, $r_C[k]$, and $r_R[k]$ can be generated for each reader, which are subsequently passed to the detector module. Our model also allows us to enable/disable individual readers and individual tracks. Turning on/off individual readers corresponds to choosing which of the $r_L[k]$, $r_C[k]$, and $r_R[k]$ signals are passed to the detector.

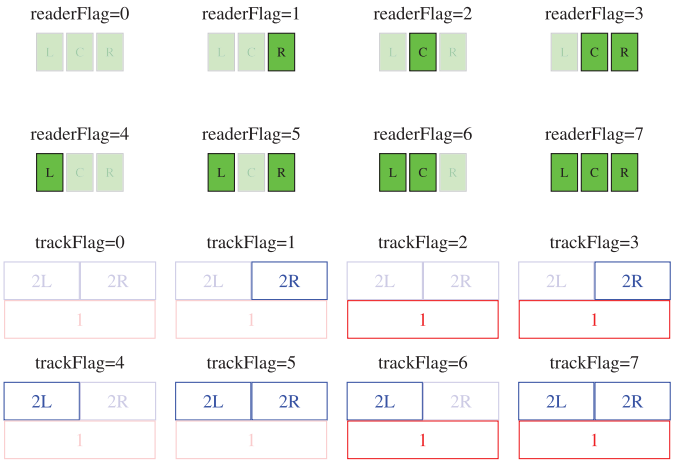

Fig. 5. Combinations of readers and tracks turned on/off for the different values of readerFlag/trackFlag.

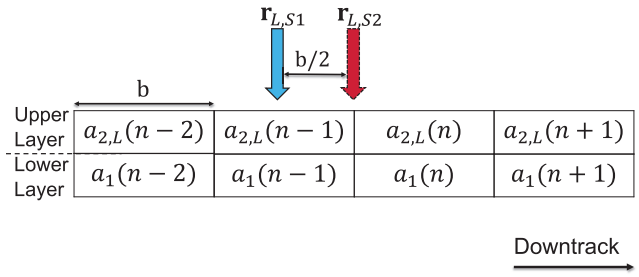

Fig. 6. For $U = 1$ and looking at the drown-track cross section showing the left upper layer track and the lower layer track, shown is the position of the read head for obtaining the second sample $\mathbf{r}_{L,S2}$ with respect to the read head position for obtaining the first sample $\mathbf{r}_{L,S1}$. $\mathbf{r}_{L,S1}$ and $\mathbf{r}_{L,S2}$ are measured $b/2$ bit length apart.

Meanwhile turning on/off individual tracks corresponds to enabling/disabling the $a_{2,L}[k]$, $a_1[k]$, and $a_{2,R}[k]$ signal paths in Fig. 2. In the simulations we will show results for varying both the numbers of readers and tracks. We use two 3-bit flags called trackFlag and readerFlag that determine the status of which tracks/readers are turned on/off, respectively. These reader/track flags are described in Fig. 5. The main situations that we examine are trackFlag=5, 7 and readerFlag=5, 7 corresponding to the double and triple reader or track configurations. A convenient shorthand that we also use is (#readerFlag,#trackFlag) to refer to a particular architecture, for example (#7,#7) refers to the full system with three-reader on three tracks. The signals $r_L[k]$, $r_C[k]$ and $r_R[k]$ from the channel model are fed into two detectors: the LS detector and Viterbi detector that we now describe.

A. Least Squares Detector

It is desirable to compare the performance of the Viterbi detector with a low complexity detector; in this section, we propose a LS detector for the MLMR channel. From the system model in Fig. 2, we can derive a LS detector for the bit sequences of interest. Convolution can be represented via matrix multiplication of the respective bit sequences by an appropriately constructed Toeplitz matrix [17, eq. (2.27)]. Let $\mathbf{H}_1$ and $\mathbf{H}_2$ be such Toeplitz matrices constructed from $\mathbf{h}_1$ and $\mathbf{h}_2$, respectively, and $\mathbf{G}$ denote an oversampling matrix of rate U. Let the vectors of readings from the left, center, and right readers be denoted by $\mathbf{r}_L$, $\mathbf{r}_C$, and $\mathbf{r}_R$, and similarly define the respective vectors for the bit sequences and noise variables in Fig. 2. Then, letting N be the block size, the $3N \times 1$ readings vector $\mathbf{R}$ can be written as shown in bottom of the next page (14), in terms of the $(4N + 3N/U) \times 1$ matrix $\mathbf{A}$, comprised of the three wanted bit streams and four interfering bit streams, a known $3N \times (4N+3N/U)$ deterministic matrix, $\mathbf{H}$, describing the channel response, and a $3N \times 1$ noise vector $\mathbf{Z}$. The $N \times 1$ upsampled lower layer bit sequences can be written in terms of the lower layer $N/U \times 1$ data bit sequence, $\mathbf{a}'_1$, and the ITI generating bit sequences $\mathbf{a}'_{1,bL}$ and $\mathbf{a}'_{1,bR}$ as

$$\begin{aligned} \mathbf{a}_1 &= \mathbf{G}\mathbf{a}'_1 \\ \mathbf{a}_{1,bL} &= \mathbf{G}\mathbf{a}'_{1,bL} \\ \mathbf{a}_{1,bR} &= \mathbf{G}\mathbf{a}'_{1,bR} \end{aligned} \tag{16}$$

respectively, and the block components of $\mathbf{Z}$ are defined as

$$\begin{aligned} \mathbf{z}_L &\triangleq \mathbf{n}_1 + \mathbf{w}_1 + \mathbf{w}_4 \\ \mathbf{z}_C &\triangleq \mathbf{n}_2 + \mathbf{w}_2 + \mathbf{w}_5 \\ \mathbf{z}_R &\triangleq \mathbf{n}_3 + \mathbf{w}_3 + \mathbf{w}_6. \end{aligned} \tag{17}$$

Also, the $3N \times 3N$ covariance of $\mathbf{Z}$, denoted by $\mathbf{C}_\mathbf{Z}$ is given by (15), as shown at the bottom of the next page, where $\otimes$ denotes the Kronecker product. The binary LS estimator of $\mathbf{A}$, denoted $\hat{\mathbf{A}}$, is a $7N \times 1$ vector given by

$$\hat{\mathbf{A}} = \text{sign}\{\mathbf{H}^\dagger \mathbf{R}\} \tag{18}$$

$$= \text{sign}\left\{\left(\mathbf{H}^T \mathbf{C}_\mathbf{Z}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^T \mathbf{C}_\mathbf{Z}^{-1} \mathbf{R}\right\} \tag{19}$$

where $\mathbf{H}^\dagger$ is the $7N \times 3N$ pseudoinverse matrix

$$\mathbf{H}^\dagger \triangleq \left(\mathbf{H}^T \mathbf{C}_\mathbf{Z}^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^T \mathbf{C}_\mathbf{Z}^{-1}. \tag{20}$$

It can be shown that $(\mathbf{H}^T \mathbf{C}_\mathbf{Z}^{-1} \mathbf{H})^{-1}$ can be computed using [17, eq. (4.7)]

$$\left(\mathbf{H}^T \mathbf{C}_\mathbf{Z}^{-1} \mathbf{H}\right)^{-1} = \mathbf{V} \left[\begin{array}{c|c} \Sigma^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \mathbf{U}^T \tag{21}$$

where $\mathbf{H}^T \mathbf{C}_\mathbf{Z}^{-1} \mathbf{H} = \mathbf{U} \Sigma \mathbf{V}^T$ is a singular-value decomposition (SVD), hence, $\mathbf{U}$ and $\mathbf{V}$ are $7N \times 7N$ unitary matrices, and Σ is a diagonal matrix of singular values. Numerical experiments have shown the LS detector is sensitive to noise. To improve the robustness of the LS detector to noise, we discard all singular values $\sigma \in \Sigma$ such that $\sigma^2 < K\sigma_n^2$ or $\sigma^2 < \epsilon$ when computing (21), where the finite positive parameter K and the small positive constant ϵ can be tuned via an offline training stage to improve the bit error rate (BER). Note that the sparsity of the matrices $\mathbf{H}$ and $\mathbf{C}_\mathbf{Z}^{-1}$ can further be exploited to make the SVD computation more efficient and minimize their storage requirements. It is noted that this formulation implicitly accounts for the correlation between readers of the media noise. In contrast, the Viterbi detector requires an explicit step to first decorrelate the noise.

Further, the bottom-layer density-reduction factor U affects the relative accuracy of the estimates of the upper and lower layers' bits. When U is large (small), the lower (upper) layer bit estimates are more accurate than the upper (lower) layer

bit estimates. Thus, we introduce iterative steps to the LS detector whereby we subtract the effect of the more reliable layer bit estimates from the readings and re-estimate the other layer's bits. An offline training stage can identify the cutoff lower layer oversampling factor $Q = 1, 2, \ldots$, such that if $U \geq Q$, the lower layer's bit estimates are more reliable than the upper layer's bit estimates. Note that $\mathbf{A}$ can be partitioned into the lower and upper layers' bit sequences, denoted by $\mathbf{A}_{\mathrm{L}}$ and $\mathbf{A}_{\mathrm{U}}$, respectively, as $\mathbf{A}^T = [\mathbf{A}_{\mathrm{L}}^T, \mathbf{A}_{\mathrm{U}}^T, \mathbf{A}_{\mathrm{ITI}}^T]$, where $\mathbf{A}_{\mathrm{L}} = \mathbf{a}'_1$ is an $N/U \times 1$ column vector, $\mathbf{A}_{\mathrm{U}}^T \triangleq [\mathbf{a}_{2,L}^T, \mathbf{a}_{2,R}^T]$ is a $2N \times 1$ column vector, and $\mathbf{A}_{\mathrm{ITI}}^T \triangleq [\mathbf{a}_{2,tL}^T, \mathbf{a}_{2,tR}^T, \mathbf{a}_{1,bL}^T, \mathbf{a}_{1,bT}^T]$ is a $(2N+2N/U) \times 1$ column vector. To remove the effect of the more reliable estimates from the readings, the effective $3N \times N/U$ and $3N \times 2N$ matrices $\mathbf{H}_L^{\mathrm{eff}}$ and $\mathbf{H}_U^{\mathrm{eff}}$ are constructed from the block components of $\mathbf{H}$ that premultiply the lower and upper layers' bit estimates, respectively. Then, we similarly construct $3N \times 3N/U$ and $3N \times 4N$ matrices $\mathbf{H}_L$ and $\mathbf{H}_U$ from $\mathbf{H}$ for the lower and upper layers, respectively, except that we also include the ITI generating components of $\mathbf{H}$. Let $\mathbf{H}_L^{\dagger}$ and $\mathbf{H}_U^{\dagger}$ denote the corresponding $3N \times 3N/U$ and $3N \times 4N$ pseudoinverse matrices as in (20) that premultiply the appropriately processed readings to estimate the lower and upper layer bits, respectively. Then, the iterative LS algorithm is given in Algorithm 1. Notice that the impact of ITI is not subtracted from the readings because the estimates of the ITI generating bits is generally very poor for SNRs of interest. But the components of $\mathbf{H}$ generating ITI are taken into account in the re-estimation so that the LS detector reduces the impact of ITI on the detection of the bit sequences of interest.

The system in (14) uses $\sim 3N$ readings, but is attempting to estimate $2N + N/U$ unknowns ($2N$ upper layer bits and N/U lower layer bits), without considering the length $2N$ and $2N/U$ ITI generating bit sequences at the upper and lower layers, respectively. In general, $\mathbf{H}$ is not full rank. Hence, the LS detector would benefit from additional readings. One method for generating more readings without increasing the complexity of the MLMR architecture is by sampling twice per bit in the downtrack direction as shown in Fig. 6. The first sample sequence $\mathbf{r}_{L,S1}$ is obtained when the reader is positioned over the center of a top-layer bit, and the second sample sequence $\mathbf{r}_{L,S2}$ is obtained when the reader is positioned in between two bits. The sampling of the sensitivity function in Fig. 1 to obtain the discrete ISI components is illustrated in Fig. 7. The sampled ISI coefficients are therefore, $\mathbf{h}'_1 = [0.0190, 0.1543, 0.3457, 0.48100.4810, 0.3457, 0.1543, 0.0190]$ and $\mathbf{h}'_2 = [0.1464, 0.8536, 0.8536, 0.1464]$, for the lower and upper layers, respectively, at the read head position for measuring the second sample. When the second sample is obtained, the noise bandwidth doubles. Hence, from (7), the noise variance when the second sample is measured is given by

$$\sigma_n^2 = \frac{3}{10^{\mathrm{SNR}/10}}. \tag{22}$$

The addition of the second sample provides the LS detector with $6N$ readings while maintaining the number of unknown bits of interest to $2N+N/U$. This results in an overdetermined $\mathbf{H}$, which our numerical investigations have shown improves the BER performance of the LS detector considerably even though the noise bandwidth has doubled.

Since the pseudoinverses $\mathbf{H}^{\dagger}$ and $\mathbf{H}_i^{\dagger}$ ($i \in \{L, U\}$) can be computed offline, the real-time implementation complexity of the LS detector is $3N$ scalar multiplications and $3N-1$ scalar additions per data bit per iteration with only a single sample

$$\underbrace{\begin{bmatrix}\mathbf{r}_L\\ \hdashline \mathbf{r}_C\\ \hdashline \mathbf{r}_R\end{bmatrix}}_{\triangleq \mathbf{R}} = \underbrace{\left[\begin{array}{ccccccc} \beta_1\mathbf{H}_2 & \beta_2\mathbf{H}_1\mathbf{G} & \beta_3\mathbf{H}_2 & C_{tL}\mathbf{H}_2 & C_{tR}\mathbf{H}_2 & C_{bL}\mathbf{H}_1\mathbf{G} & C_{bR}\mathbf{H}_1\mathbf{G}\\ \hdashline \beta_4\mathbf{H}_2 & \beta_5\mathbf{H}_1\mathbf{G} & \beta_6\mathbf{H}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \hdashline \beta_7\mathbf{H}_2 & \beta_8\mathbf{H}_1\mathbf{G} & \beta_9\mathbf{H}_2 & C_{tL}\mathbf{H}_2 & C_{tR}\mathbf{H}_2 & C_{bL}\mathbf{H}_1\mathbf{G} & C_{bR}\mathbf{H}_1\mathbf{G}\end{array}\right]}_{\triangleq \mathbf{H}} \underbrace{\begin{bmatrix}\mathbf{a}_{2,L}\\ \hdashline \mathbf{a}'_1\\ \hdashline \mathbf{a}_{2,R}\\ \hdashline \mathbf{a}_{2,tL}\\ \hdashline \mathbf{a}_{2,tR}\\ \hdashline \mathbf{a}'_{1,bL}\\ \hdashline \mathbf{a}'_{1,bR}\end{bmatrix}}_{\triangleq \mathbf{A}} + \underbrace{\begin{bmatrix}\mathbf{z}_L\\ \hdashline \mathbf{z}_C\\ \hdashline \mathbf{z}_R\end{bmatrix}}_{\triangleq \mathbf{Z}} \tag{14}$$

$$\mathbf{C}_{\mathbf{Z}} = \sigma_n^2 \begin{bmatrix} 1+\frac{1}{6U} & \frac{2C_t(p_C-p_L)}{3}+\frac{C_b(p_C-p_L)}{6U} & \frac{C_t(p_R-p_L)}{3}+\frac{C_b(p_R-p_L)}{6U}\\ \frac{2C_t(p_C-p_L)}{3}+\frac{C_b(p_C-p_L)}{6U} & 1+\frac{1}{6U} & \frac{2C_t(p_R-p_C)}{3}+\frac{C_b(p_R-p_C)}{6U}\\ \frac{2C_t(p_R-p_L)}{3}+\frac{C_b(p_R-p_L)}{6U} & \frac{2C_t(p_R-p_C)}{3}+\frac{C_b(p_R-p_C)}{6U} & 1+\frac{1}{6U}\end{bmatrix} \otimes \mathbf{I}_{N\times N} \tag{15}$$

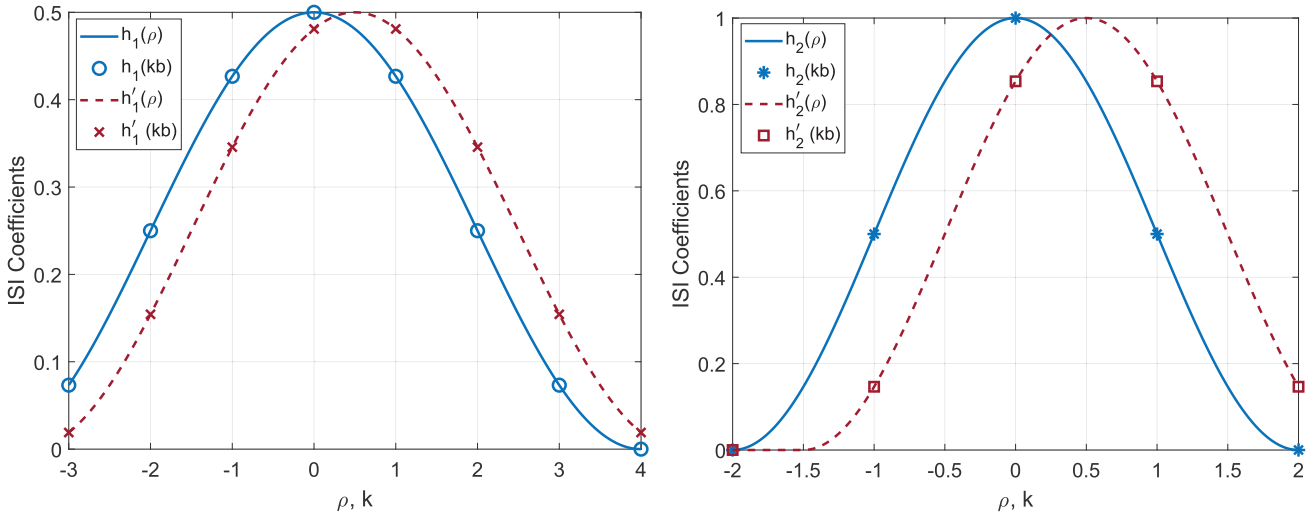

Fig. 7. Sampling of the sensitivity functions for the lower and upper layers when obtaining the first and second readings. (a) Lower layer. (b) Upper layer.

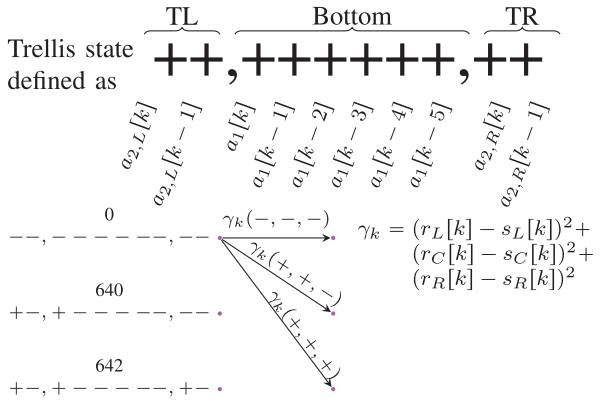

Fig. 8. State definition and transitions for the three-track/three-reader Viterbi detector, with three sample transitions depicted.

per top-layer bit, and $6N$ scalar multiplications and $6N-1$ scalar additions per data bit per iteration with two samples per top-layer bit. Note that the complexity of the LS detector remains approximately constant as the length of the ISI filters increases.

B. Viterbi Detector

The Viterbi detector for the 2L3T3R architecture is based on the conventional Viterbi detector, but with a redefinition of the states and the metrics associated with the transitions between states. Each state contains 3 bit sequences of length, N_h-1 concatenated together one after the other, where N_h is the length of the ISI coefficients for the particular subchannel. Since we are using three-tap ISI channels for the top-layer bit streams and a seven-tap ISI channel for the bottom-layer bit-stream, there are 2 bits for each of the two top layers and 6 bits for the bottom giving 10 bits and 1024 states in total. This is independent of the number of readers. The state definition together with some example trellis transitions is depicted in Fig. 8.

Fig. 8 shows the definition of a single state for our 2L3T3R architecture and part of a trellis with a transition from the all-zero state to three of the possible eight candidate next states. As there are potentially three new input bits per clock cycle, there are potentially eight new branches at each time instant k. However, if there is a repetition-constraint on the bottom layer that enforces U consecutive bits of $a_1[k]$ to be all identical, then there will be cases when there are only two new bits in some clock-cycles, and therefore, only four branches emanating from each state at those time instances. Therefore, the trellis definition changes with U. In the scenarios where some of the tracks are omitted, the corresponding subportion of the trellis state definition is also removed, again leading to fewer states in the trellis.

Besides a modification to the trellis state definitions, that account for the bits being written to the different tracks, there is also a modification to be made to the branch metric computation as depicted in Fig. 8. This modification accounts for the fact that there are now multiple readers contributing signals to the detector from $r_L[k]$, $r_C[k]$, and $r_R[k]$. While the normal Viterbi detector uses the Euclidean distance (squared) between the received waveform and the candidate noise-free waveforms as the branch metric, the 2L3T3R Viterbi uses the sum of the Euclidean distances squared from each readback waveform to the corresponding noise-free signal associated with the branch. Given the known channel parameters, the noise-free signals are

$$s_L[k] = \sum_{i=0}^{2} \beta_1 h_2[i] a_{2,L}[k-i] + \sum_{i=0}^{6} \beta_2 h_1[i] a_1[k-i] + \sum_{i=0}^{2} \beta_3 h_2[i] a_{2,R}[k-i]$$

$$s_C[k] = \sum_{i=0}^{2} \beta_4 h_2[i] a_{2,L}[k-i] + \sum_{i=0}^{6} \beta_5 h_1[i] a_1[k-i]$$

$$+\sum_{i=0}^{2}\beta_6 h_2[i]a_{2,R}[k-i]$$

$$s_R[k]=\sum_{i=0}^{2}\beta_7 h_2[i]a_{2,L}[k-i]+\sum_{i=0}^{6}\beta_8 h_1[i]a_1[k-i]+\sum_{i=0}^{2}\beta_9 h_2[i]a_{2,R}[k-i] \quad (23)$$

respectively. In the full trellis with three bit streams and no repetition code on the lower layer ($U = 1$) there are 1024 states with eight branches per state. When $U \neq 1$, there will be four branches per state, except at the time instances that are divisible by U, when there will again be eight branches per state.

C. Noise Decorrelation Between Readers

The noise correlation between the readers, caused by the overlap in the RSFs, described in (13) can cause some performance loss in the Viterbi detector. Prior to detection, we can multiply the signals $r_L[k]$, $r_C[k]$, and $r_R[k]$ by a cross-track noise-decorrelating (ND) matrix to remove the correlation and improve the performance. The ND matrix is the inverse of the square root of the total noise correlation matrix at the detector input which consists of the noise from three sources: the top layer, the bottom layer, and the reader. The total noise correlation at the detector input and corresponding ND matrix will be

$$\mathbf{C}=\mathbf{A}_1+\mathbf{A}_2+\frac{\sigma_n^2}{3}\mathbf{I} \text{ and } \mathbf{D}=\mathbf{C}^{-\frac{1}{2}} \quad (24)$$

respectively. The noise correlation and decorrelating matrices in the readers home positions (after normalization) are

$$\mathbf{C}=\begin{bmatrix}1.0000 & 0.1894 & 0.0238\\ 0.1894 & 1.0000 & 0.1894\\ 0.0238 & 0.1894 & 1.0000\end{bmatrix}$$
$$\mathbf{D}=\begin{bmatrix}1.014 & -0.0974 & 0.0017\\ -0.0974 & 1.028 & -0.0974\\ 0.0017 & -0.0974 & 1.014\end{bmatrix} \quad (25)$$

respectively. The matrix, $\mathbf{D}$, de-correlates the noise, while leaving the noise magnitudes unchanged. For two-reader configurations, the middle row and middle column, are omitted. It is noted that while the correlating matrices transform noise with unity power into noise that is suitable for combining with the signal at a certain SNR, they depend on σ_n, while the decorrelating matrix in (25) operates on the combined signal+noise waveforms and therefore, does not.

In addition to whitening the noise, the above ND matrix also changes the values of the cross-track dependent β coefficients which need to be modified in the detector. The beta coefficients need to be updated as follows:

$$\begin{aligned}\beta_1' &= D_{1,1}\beta_1+D_{1,2}\beta_4+D_{1,3}\beta_7\\ \beta_2' &= D_{1,1}\beta_2+D_{1,2}\beta_5+D_{1,3}\beta_8\\ \beta_3' &= D_{1,1}\beta_3+D_{1,2}\beta_6+D_{1,3}\beta_9\\ \beta_4' &= D_{2,1}\beta_1+D_{2,2}\beta_4+D_{2,3}\beta_7\\ \beta_5' &= D_{2,1}\beta_2+D_{2,2}\beta_5+D_{2,3}\beta_8\\ \beta_6' &= D_{2,1}\beta_3+D_{2,2}\beta_6+D_{2,3}\beta_9\\ \beta_7' &= D_{3,1}\beta_1+D_{3,2}\beta_4+D_{3,3}\beta_7\\ \beta_8' &= D_{3,1}\beta_2+D_{3,2}\beta_5+D_{3,3}\beta_8\\ \beta_9' &= D_{3,1}\beta_3+D_{3,2}\beta_6+D_{3,3}\beta_9. \end{aligned} \quad (26)$$

The values for β' after noise decorrelation for the readers in their home position are

$$\begin{bmatrix}\beta_1' & \beta_2' & \beta_3'\\ \beta_4' & \beta_5' & \beta_6'\\ \beta_7' & \beta_8' & \beta_9'\end{bmatrix}=\begin{bmatrix}0.9652 & 0.8260 & -0.0470\\ 0.4165 & 0.8508 & 0.4165\\ -0.0470 & 0.8260 & 0.9652\end{bmatrix} \quad (27)$$

respectively.

IV. Simulation Results and Discussion

In this section, we show the simulation results for the LS and Viterbi detectors described in Section III for the various scenarios.

A. LS Detector Results

We demonstrate the BER performance of the LS detector. For the nominal positions of three read heads, $p_L = -t/2$, $p_C = 0$, and $p_R = t/2$, the bit sequences vector becomes $\mathbf{A} = \text{col}[\mathbf{a}_{2,L}, \mathbf{a}_1', \mathbf{a}_{2,LR}, \mathbf{a}_{1,bL}', \mathbf{a}_{1,bR}']^T$. Also, from (3) and (4), $\mathbf{H}$ simplifies to

$$\mathbf{H}=\left[\begin{array}{c|c|c|c|c}\mathbf{H}_2 & 0.9092\mathbf{H}_1\mathbf{G} & \mathbf{0} & 0.0908\mathbf{H}_1\mathbf{G} & \mathbf{0}\\ \hline 0.5\mathbf{H}_2 & \mathbf{H}_1\mathbf{G} & 0.5\mathbf{H}_2 & \mathbf{0} & \mathbf{0}\\ \hline \mathbf{0} & 0.9092\mathbf{H}_1\mathbf{G} & \mathbf{H}_2 & \mathbf{0} & 0.0908\mathbf{H}_1\mathbf{G}\end{array}\right] \quad (28)$$

and, from (10), $\mathbf{C_Z}$ becomes

$$\mathbf{C_Z}=\sigma_n^2\begin{bmatrix}1+\frac{1}{6U} & \frac{1}{9}+\frac{1+\pi}{12\pi U} & \frac{1}{36U}\\ \frac{1}{9}+\frac{1+\pi}{12\pi U} & 1+\frac{1}{6U} & \frac{1}{9}+\frac{1+\pi}{12\pi U}\\ \frac{1}{36U} & \frac{1}{9}+\frac{1+\pi}{12\pi U} & 1+\frac{1}{6U}\end{bmatrix}\otimes\mathbf{I}_{N\times N}. \quad (29)$$

Algorithm 1 is then implemented to obtain the BER versus SNR performance of the iterative LS detector. We examine the performance of the LS detector for different values of U, number of readers, and with and without the second sample sequences. Three iterations of the LS algorithm are implemented, which give most of the possible improvements in the BER with the LS detector. Since the error rates on the two upper layer tracks are similar, the upper layer BERs shown represent their average. To examine the impact of possible degradation on the BERs due to the superposition of the signals from the two layers, we also simulate the BERs for a one-layer reference with two upper layer tracks and no lower layer track. The LS simulations use $N = 1200$ bits, 1000 testing blocks, and 100 training blocks to tune the parameters mentioned in Algorithm 1.

Fig. 9(a) shows the BER performance of the LS detector when only two readers, the left and right, are available, with

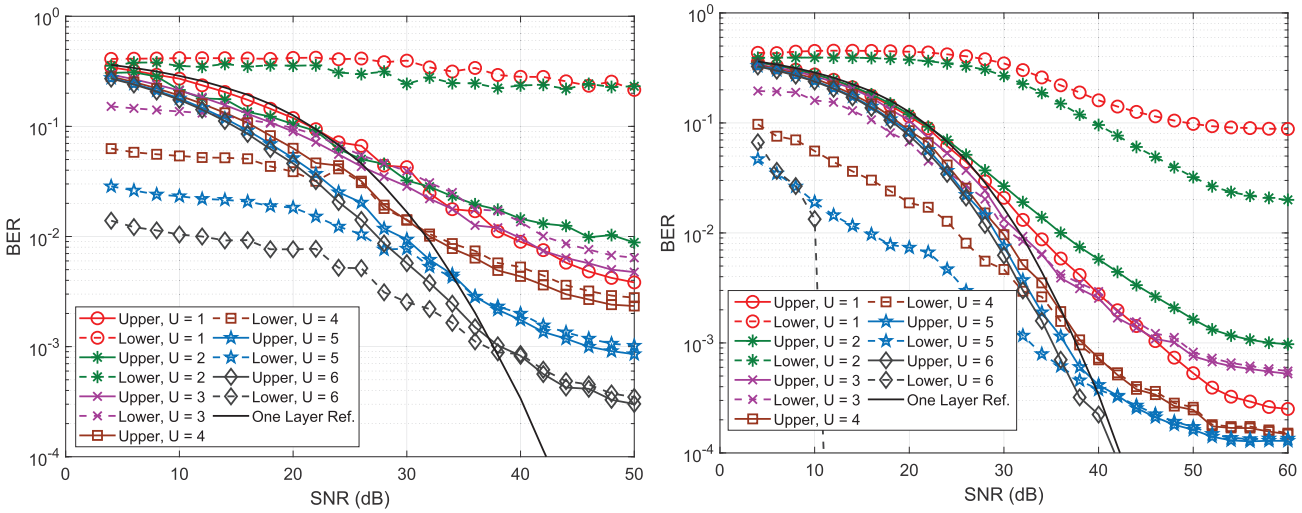

Fig. 9. BER performance of the LS detector is shown with one reading per top-layer bit. (a) Two readers. (b) Three readers.

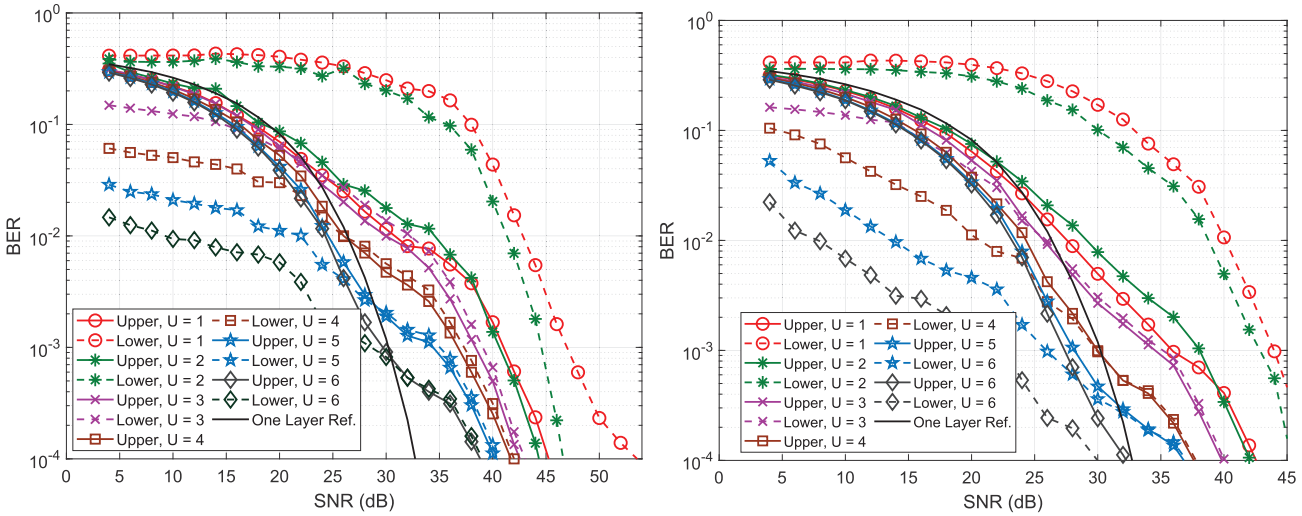

Fig. 10. BER performance of the LS detector with two readings per top-layer bit. (a) Two readers. (b) Three readers.

one reading per upper layer bit, and different values of U. Hence, the readings vector consists only of the components $\mathbf{r}_L$ and $\mathbf{r}_R$. Such a system provides $2N$ readings and requires estimating $2N + N/U$ unknown bits. Hence, the system is under-determined. From Fig. 9(a), the LS detector is not able to achieve BERs less than 10^{-2} for the lower layer track with $U = 1$ and $U = 2$, and the upper layer requires 10 dB more of SNR than the one-layer reference to achieve a BER of 10^{-2}. For $U \geq 3$, BERs less than 10^{-2} are achieved on the lower and upper layers with respective error floors. A BER of 10^{-2} on the upper and lower layers are achieved by 42, 32, 30, and 28 dB for $U = 3, \ldots, 6$, respectively, whereas the one-layer reference achieves a BER of 10^{-2} by 32 dB.

In Fig. 9(b), all three readers in the system model in Figs. 1 and 2 are available. For $U = 1, 2$, the BERs on the lower layer reach error floors of about 10^{-1} and 2×10^{-2}, respectively. For $U = 3$ and $U = 4, 5, 6$, BERs of 10^{-2} on the upper layer are achieved by 32, 30 dB, respectively, which are within 2 dB of the one-layer reference, indicating that a further increase in U beyond three results in negligible improvement in the BERs of the upper layer for a BER of 10^{-2}. But increasing U decreases the eventual error floor on the upper layer bit estimates. On the lower layer, BERs of 10^{-2} are attained by 32, 26, 16, and 12 dB for $U = 3, \ldots, 6$, respectively. At $U = 6$, the upper and lower layers' BERs can reach 10^{-4}.

Fig. 10(a) shows the BER versus SNR performance of the LS detector when the sampling rate is doubled and two readings per upper layer bit are obtained. Though such a scheme doubles the noise bandwidth as discussed, it provides more readings than unknowns for the LS detector, making the LS solution better-behaved. More precisely, with two readers the LS detector has $4N$ readings from which it estimates $2N + N/U$ unknown bits. Unlike with the single sample scenario, a BER of 10^{-4} is achievable on the upper and lower

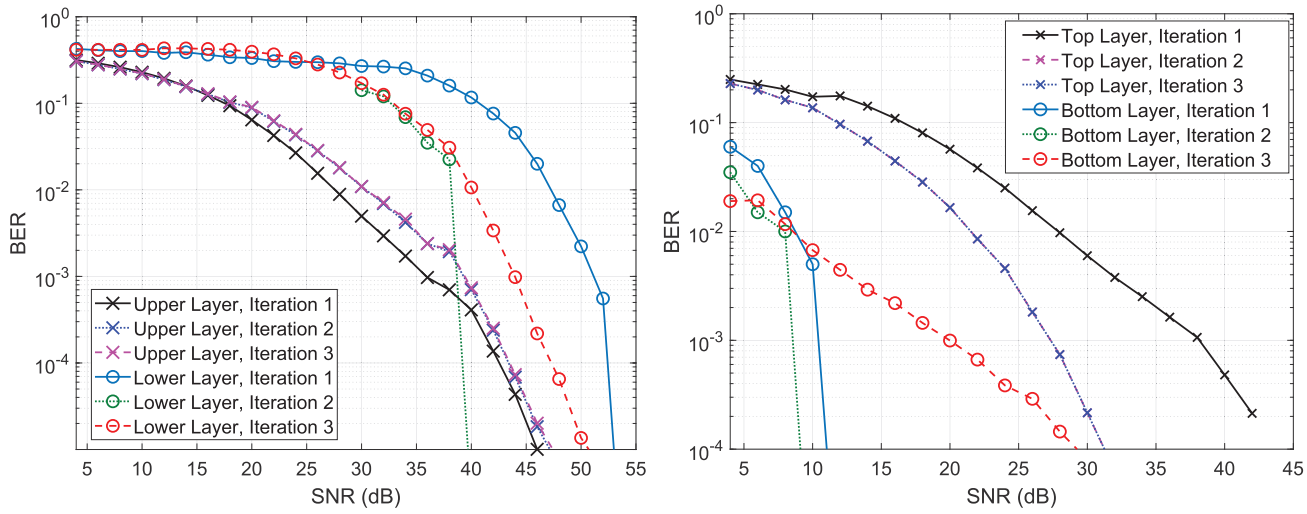

Fig. 11. Demonstrated are the BERs on the upper and lower layers for each iteration of the LS detector with three readers, two samples per upper layer bit, and (a) $U = 1$ and (b) $U = 6$.

layers with high enough SNRs for all $U = 1, \ldots, 6$. For $U = 1, 2$, $U = 3$, and $U = 5, 6$, the upper layer achieves a BER of 10^{-2} by 36, 30, and 24 dB. For $U = 4, 5, 6$, the upper layer is within 2 dB of the one-layer reference for a BER of 10^{-2}. The lower layer achieves a BER of 10^{-2} by 44, 42, 32, 26, 22, and 10 dB for $U = 1, \ldots, 6$, respectively, displaying the sharpest decrease in SNRs for $U = 3$ and $U = 6$.

In Fig. 10(b), the three-reader configuration is assumed, and two samples per upper layer bit are obtained. On the upper layer, a BER of 10^{-2} is achieved by 28, 30, 26, 24, and 20 dB for $U = 1$, $U = 2$, $U = 3, 4$, $U = 5$, and $U = 6$, respectively, which are all within about 3 dB of the reference. The bottom layer reaches a BER of 10^{-2} by 40, 26, 24, and 8 dB for $U = 1, 2$, $U = 3$, $U = 4, 5$, and $U = 6$, respectively. Similar to the two-reader case, the lower layer experiences the largest improvement in SNR for $U = 3$ and $U = 6$. This suggests a tradeoff between SNR and U with the largest SNR gains in return for density happening at $U = 3$ and $U = 6$ using the LS detector.

It is instructive to also demonstrate the BERs of the LS detector per iteration. Fig. 11 illustrates the BER versus SNR performance of the LS detector for different iterations with two samples per upper layer bit and for $U = 1$ and $U = 6$. For $U = 1$, the upper layer's bit estimates are more accurate than the lower layer's bit estimates during the first iteration. Hence, subtracting the effect of the upper layer bits reduces the required SNR for a BER of 10^{-2} on the lower layer by about 10 dB as shown in Fig. 11(a) with iterations 2 and 3 for the lower layer. For $U = 6$, the lower layer's bits are more accurately estimated than the upper layer bits during the first iteration as shown in Fig. 11(b). A second iteration decreases the required SNR on the top layer by about 5 dB and 13 dB for BERs of 10^{-2} and 10^{-3}, respectively, compared with the first iteration.

The singular values removal schedule eliminates all singular values that are less than the $\max(K\sigma_n, \epsilon)$. The lower limit ϵ ensures that a very small number of singular values are discarded regardless of how large the SNR is. In doing so, it prevents the computations of $1/\sigma$ if σ is very close to zero, which could otherwise result in undefined pseudoinverse matrices. Hence, the sharp turns in the BER versus SNR curves of the LS detector happen when the singular values removal scheme switches from removing singular values that are less than $K\sigma_n$ to removing singular values that are less than ϵ. In practice, the operating SNR is known beforehand. Thus, the parameters for singular values removal in Algorithm 1 can be carefully tuned to minimize the BER at the desired SNR value.

B. Viterbi Detector Results

Fig. 12 shows the performance of the Viterbi detector for $N \sim 4$k block sizes. The first two rows hold the results for the situation with two readers and three tracks (#5,#7), turning on and off the ITI, while rows three and four are similar but with three readers. We omit the single-reader/single-track scenarios as the performance with one reader is not good, and the single-track case is not interesting. The graphs also include the configurations where the bottom-layer signal and noise are both removed, as our baseline reference, that we label (#5,#5) when two readers are used and (#7,#5) when three readers are used. The (#5,#5) configuration corresponds to two narrow tracks on the top layer, with a reader centered over each of them, while the (#7,#5) case is similar, but with three evenly spaced readers over the two tracks. These reference configurations would be equivalent to reading in the conventional PMR system and the TDMR system, respectively. These reference curves are unaffected by ITI, as none of the response functions on the top layer fall outside of the two tracks being detected. The remaining curves on each plot correspond to varying values of U from 1 to 4 shown in different colors and, in the cases where it is appropriate, for the top layer and bottom layer with different marker symbols.

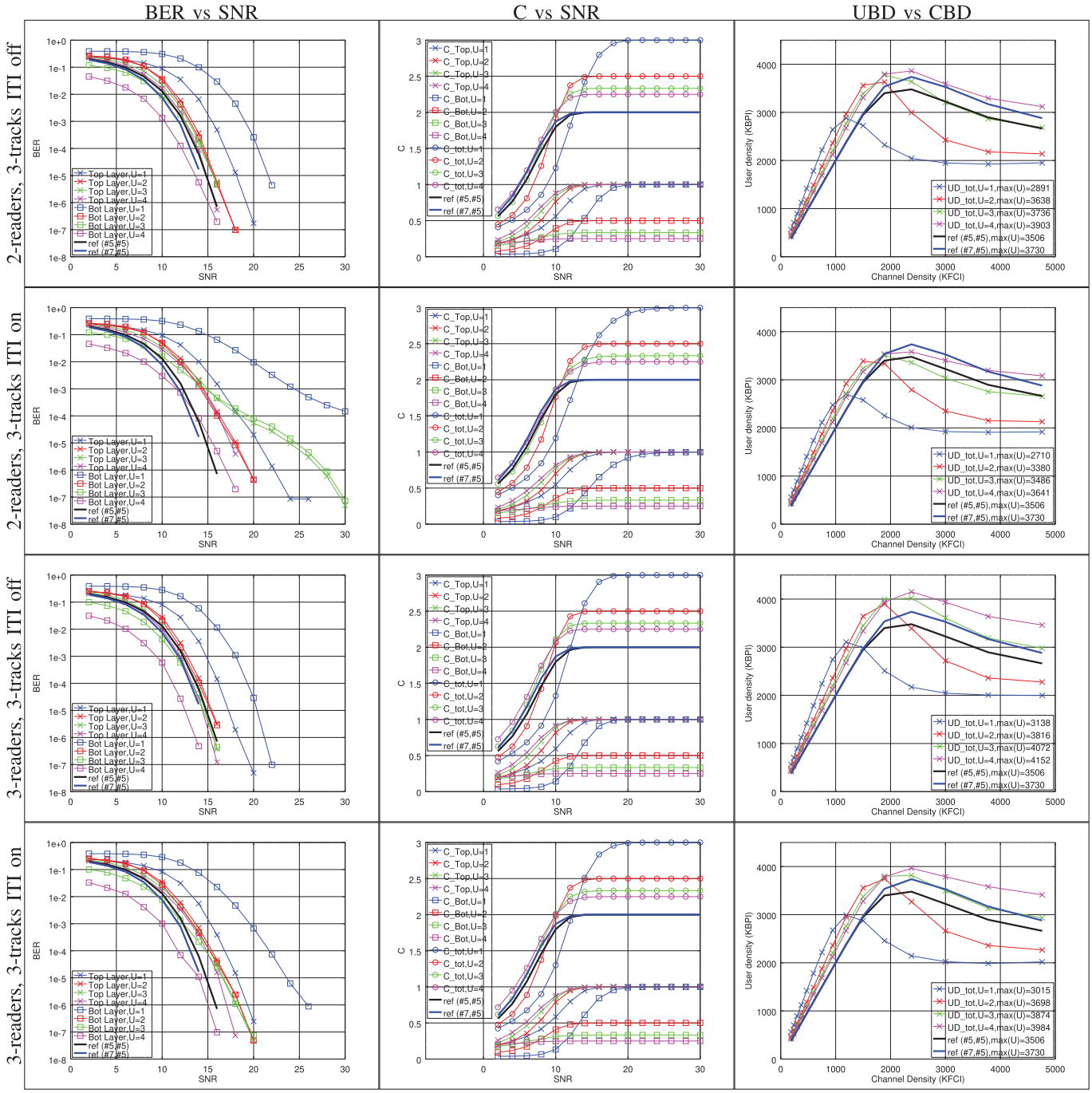

Fig. 12. Performance plots for the Viterbi detector. The leftmost column shows the BER versus SNR plots. In the second column, we convert the BERs into BSC capacities, and in the third column to user densities versus channel densities.

The first column of plots in Fig. 12 show the BER as a function of SNR in each configuration. The general trend observed is that with increasing bit length on the bottom track, U, the BER performance improves for both layers. This is attributed to the increased spectral separation between the signals in the top and bottom layers.

In going from the first to the second row, the ITI is being turned on and we see a marked loss in performance for all U. Compared with the (#7,#7) cases in the third and fourth rows, we see the performance with two readers is more sensitive to ITI than that with three readers. We notice in particular, an unusual second waterfall mode when the ITI is turned on for the (#5,#7) case when $U = 3$ (green curve). We have encountered similar such cases in the past and found it to be attributed to the existence of a low minimum-distance pattern that has a long supporting bit pattern. For the purposes of the discussion, we will call it error pattern EP_A with supporting bit pattern SP_A. The low minimum distance of EP_A means that when SP_A happens, the probability of error is high. However, because SP_A is long, its probability of

occurrence is low. Thus, the Viterbi performance is dominated by another larger minimum-distance error pattern, that has a shorter supporting bit pattern, which we call EP_B and SP_B. At low SNR, the BER is dominated by EP_B/SP_B that occur more frequently because it is shorter, even though it has a larger minimum distance. As the SNR increases, eventually the probability of error of EP_B drops below the probability of occurrence of SP_A revealing the second observed waterfall mode. One of the main observations from these plots is that although there is relatively little loss in going from three readers to two readers when there is no unknown ITI, when it is present, the loss can be considerably larger. We expect this could hold true for TDMR as well.

In the second column of plots in Fig. 12, we transform the BERs into binary-symmetric channel (BSC) capacities using Shannon's BSC capacity formula [18]

$$C_{\text{BSC}} = 1 + p\log_2(p) + (1-p)\log_2(1-p) \tag{30}$$

where we substitute in the BER for the crossover probability p. The BSC capacity computation provides a strict upper bound on the information flow through a channel with random independent bits and errors. In the case of the BER, however, the errors are correlated and the channel has a capacity higher than the BSC bound suggests. We view the BSC capacity as an estimate of the practical information rate that the channel might support when including the options of pattern dependent detection and channel encoding. In the second column of Fig. 12, there are three plots for each value of U: one for the top layer marked with x, one for the bottom layer marked with squares and one for the combination of the two layers marked with circles. The capacity for the combined system is computed as

$$C_{\text{tot}} = 2C_{\text{top}} + \frac{1}{U}C_{\text{bot}} \tag{31}$$

where C_{top} and C_{bot} are the BSC capacities computed from the top and bottom-layer BERs, respectively. There are two tracks on the top layer and the bottom layer only produces $1/U$ bits per clock cycle. This is an estimate at the total number of user bits per clock cycle that the system can detect at a given SNR. We see that when $U = 1$, the system saturates at 3 bits/clock cycle at high SNR: two bits on the top and one on the bottom. As U increases, the curves saturate at lower capacities, because the bottom layer is holding less data. However, it is not the number of bits/clock cycle that the curves saturate at, that ultimately interests us. Our assumption is that the SNR on the horizontal axis of these plots is changing with channel bit density (CBD): the higher the CBD, the lower the SNR, and the poorer the BER. Here, CBD refers to the linear downtrack density on the top layer alone, rather than the density achieved when the bottom layer is included. We use the same assumption as in [13] that the SNR changes for a magnetic recording system by about 6 dB per doubling of CBD. Then choosing an equivalence point allows us to convert the SNRs on the horizontal axis into CBDs in kilo flux change per inch (KFCI): a standard unit for the CBD. Furthermore, we multiply the CBDs by the capacity, which by Shannon's theorem [18] is an estimate of the highest code-rate achievable for "error-free" recording on the BSC, to convert KFCI into kilobits per inch (KBPI), and obtain an estimate of the highest UBD. This is an estimate rather than a bound, as the BSC capacity underestimates the actual capacity of a channel when the errors are correlated. The equivalence that we use to convert SNR into KFCI is shown in Table I, and the plot of UBD versus CBD is shown in the last column of Fig. 12. At low CBDs (high SNRs), the capacity is almost constant at $2+1/U$ and the UBDs increase linearly with CBD. As CBD increases, the SNR drops and at some point, errors begin to occur in the channel at a sufficient rate, such that the capacity begins to drop and therefore, the total achievable UBD decreases again. In between, there is a maximum UBD that can be identified in the charts of the last column of Fig. 12. This should be chosen as the highest operating UBD for the architecture being investigated. We estimate this maximum UBD using parabolic interpolation of the maximum point and two adjacent points, and display its value in the legend of each curve. We see that the $U = 4$ has the highest UBD in all situations. For each of the four scenarios depicted in Fig. 12, the percentage gain over the reference curves are shown in Table II.

TABLE I
TABLE USED TO CONVERT SNR INTO KFCI

SNR	4	6	8	10	12
KFCI	3779.76	3000.00	2381.10	1889.88	1500.00
SNR	14	16	18	20	22
KFCI	1190.55	944.94	750.00	595.28	472.47

TABLE II
GAIN OF $U = 4$ OVER THE TWO REFERENCES

Scheme	(#5,#7) no ITI	(#5,#7) with ITI	(#7,#7) no ITI	(#7,#7) with ITI
% Gain over (#5,#5)	11.3%	3.9%	18.4%	13.6%
% Gain over (#7,#5)	4.6%	-2.4%	11.3%	6.8%

We note here the manifestation of the previous observation that the two-reader case is more sensitive to ITI, losing about $\sim 7\%$ to ITI compared with only $\sim 4.8\%$ for the three-reader case.

C. Optimizing the Reader Position

In the simulations of Fig. 12, the L and R readers were held fixed at positions of $\pm 0.5t$ and the C reader at 0 cross-track offset. In this section, we vary p_R while keeping the left reader symmetrically placed at $p_L = -p_R$, and leaving $p_C = 0$. The results shown in Fig. 13 focus on the (#7,#7) scenario as it is the most interesting, investigating the situations with ITI and ND both turned off and on. In a similar fashion to the case of maximizing the UBD, we have fit parabolas to the BER versus p_R curves to locate the minimum values which are subsequently also listed in the legends of Fig. 13. The SNR is held fixed at 10 dB in these simulations and the reference curves at 10 dB are plotted as a baseline for comparison.

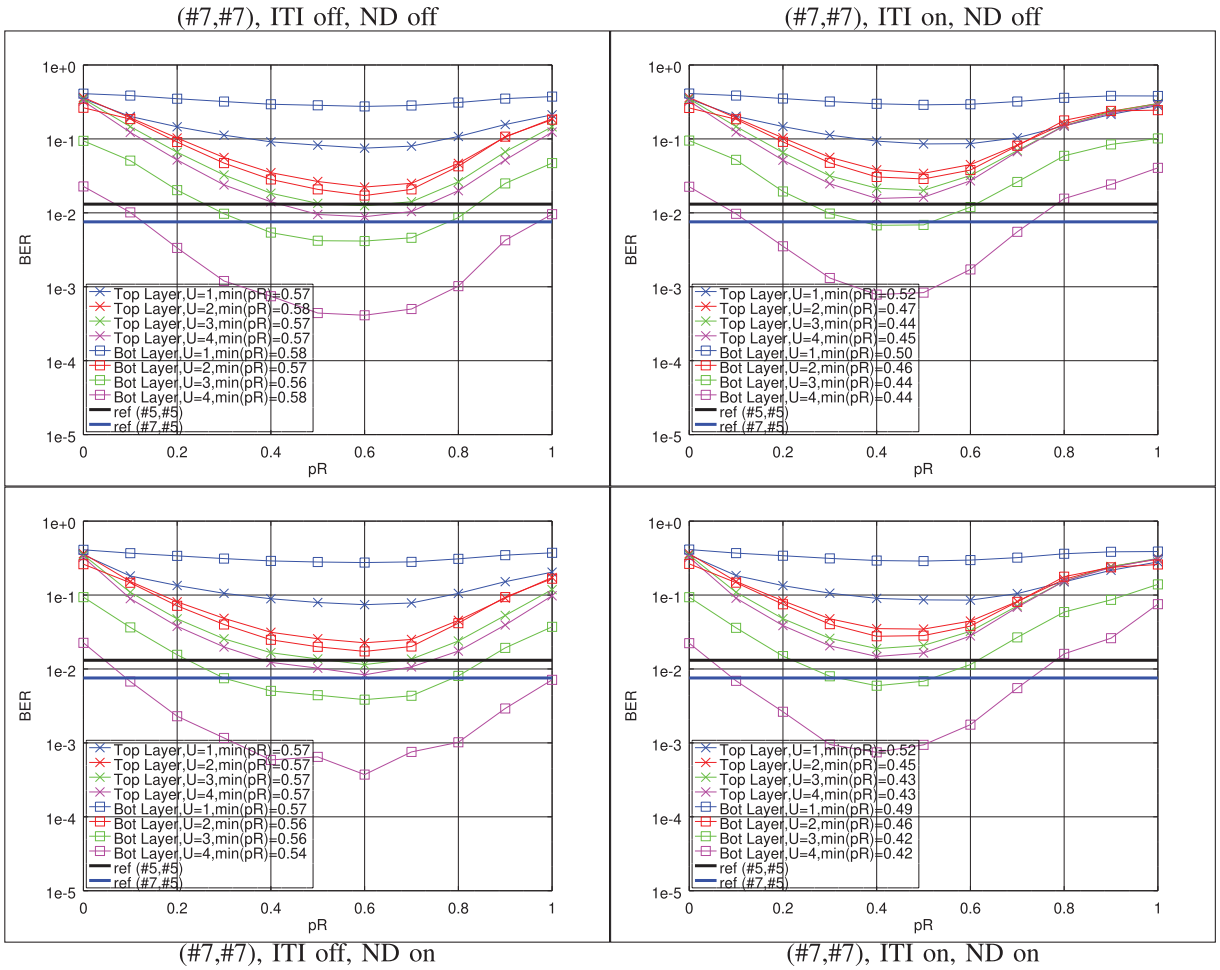

Fig. 13. Error rates as a function of reader cross-track location p_R at SNR = 10 dB.

Comparing the plots in the left and right columns of Fig. 13, we can see that introducing ITI causes higher error rates toward $p_R = 1$, where the readers are pushed out toward the ITI tracks. However, the ITI has no significant effect near $p_R = 0$, which is where the L and R readers are moved in toward the center. On the other hand, comparing the top row against the bottom row plots of Fig. 13, the effect the ND matrix at the input to the Viterbi detector can be seen. The ND gives a slight improvement in BER at low p_R around 0.1 or 0.2. This is the region where the overlap between the RSF of the three readers is strongest leading to the greatest benefit from the ND. However, the benefit from the ND predicted in these simulations is not significant.

We observe a distinct minimum in the BER versus p_R curves of Fig. 13, the location of which depends on whether ITI and ND are off or on. Except for the case $U = 1$, the minima of the curves all tend to agree with less than 10% discrepancy. For the $U = 1$ case at 10 dB, the BERs are quite high, and close to 0.5 for the bottom layer, making the data less accurate for the purposes of estimating the optimum p_R. Therefore, we average together the values of the optimum p_R for the cases excluding $U = 1$ and show them in Table III.

After obtaining the best value of p_R in each case, we reran the simulations and recomputed the BER versus SNR curves, followed by C versus SNR and UBD versus CBD as we did in Fig. 12. We have omitted plotting the BER and C versus SNR plots in this case and jumped directly to the UBD versus CBD plots from which the density gains can be estimated in Fig. 14.

TABLE III
VALUE OF p_R THAT MINIMIZES BER CURVES IN FIG. 13

Scheme	(#7,#7) no ITI, no ND	(#7,#7) with ITI, no ND	(#7,#7) no ITI, with ND	(#7,#7) with ITI, with ND
optimum p_R	0.572	0.450	0.562	0.435

Previously, we stopped our simulations at $U = 4$, but it was found that $U = 4$ was the highest performing, so it was not clear whether this was the optimum value of U. Hence, in Fig. 14, we extended the simulations to also include $U = 5$ and 6 and we see that $U = 4$ does give the maximum UBD that we tabulate in Table IV. We observe a 2% gain in the case without ITI, and a 0.9% gain in the case with ITI due to our optimization of p_R.

A curious phenomenon is observed in that the gain for $U = 4$ is able to exceed 20% in certain conditions. This does not seem feasible at first, as $U = 4$ leads to a quarter of the downtrack linear density and all our simulations have half the track density giving 1/8 AD additional gain on the bottom

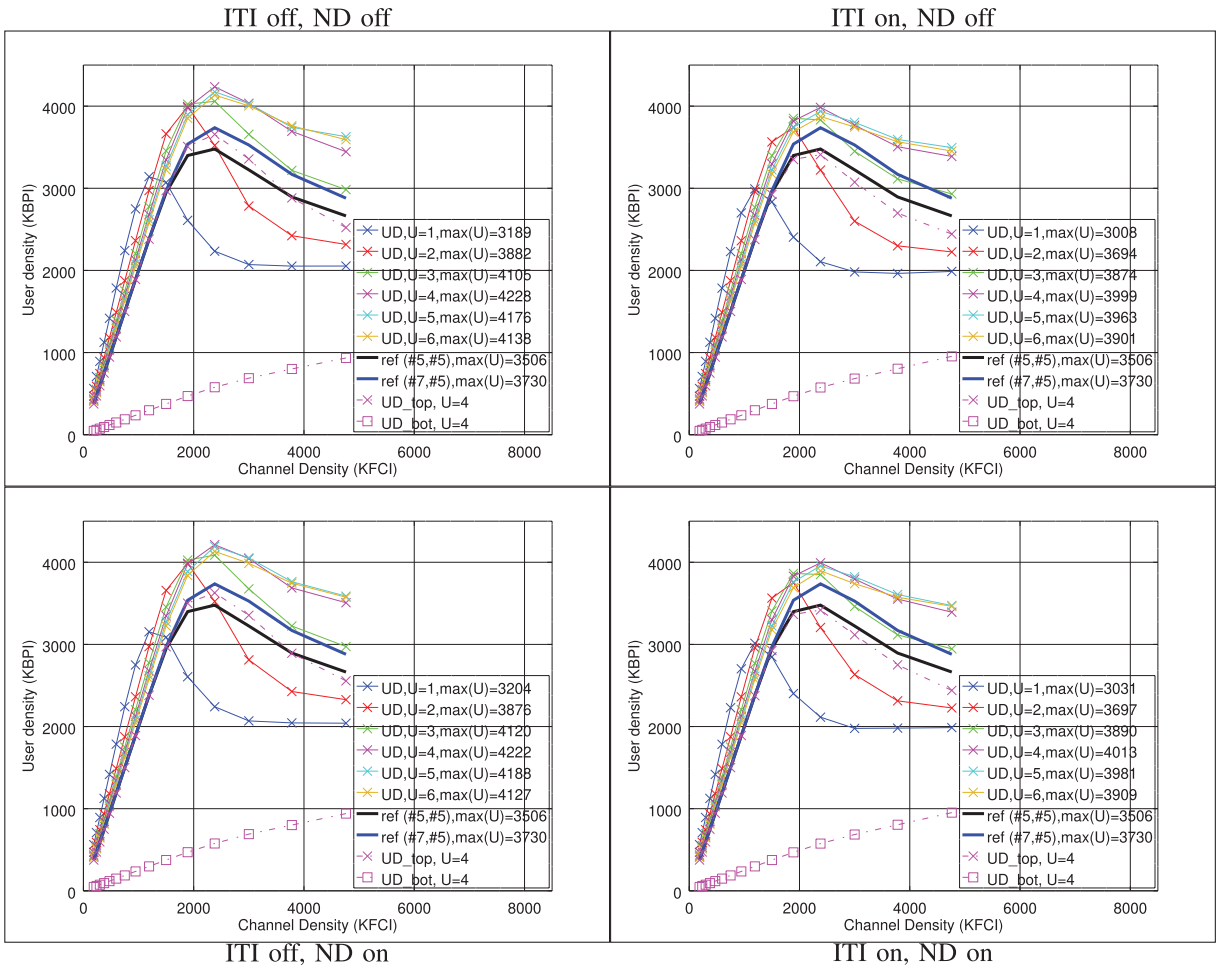

Fig. 14. UBD versus CBD plots for the 2L3T3R Viterbi detector.

TABLE IV
GAIN OF $U = 4$ SCHEME OVER THE REFERENCE, AFTER OPTIMIZING p_R

Scheme	(#7,#7) no ITI no ND	(#7,#7) with ITI no ND	(#7,#7) no ITI with ND	(#7,#7) with ITI with ND
% Gain over (#5,#5)	20.6%	14%	20.4%	14.5%
% Gain over (#7,#5)	13.4%	7.2%	13.2%	7.6%

layer or an expected maximum addition of 12.5%. In order to better understand how 20% was achieved, we have plotted the UBD for the top and bottom layers separately with dashed curves in Fig. 14. We see that the top-layer UBD grows in similar fashion toward its peak, as the (#5,#5) and (#7,#5) reference curves. The bottom-layer UBD on the other hand, grows 1/4 as quickly with CBD and so has not reached its maximum in the range of CBDs in Fig. 14. Even when the top-layer UBD has peaked and started to decline again, the bottom-layer UBD is still increasing faster than the top-layer UBD is declining. This has the effect of shifting the peak of the combined maximum UBD out to a higher CBD and allows it to peak at a higher UBD, resulting in a larger total UBD gain of both layers together.

V. CONCLUSION

In this work, we have proposed a new model for MLMR, that supports up to three readers over three tracks, with two narrower tracks in the top-layer straddling a single double-width track in the bottom layer. The model is based on the ISI contribution of each track to each reader weighted by a coefficient that depends on the cross-track displacement of the reader to the track being sensed. We have also proposed, written, and tested with our model two detection schemes: an LSs detector with a simple iteration scheme and a full joint Viterbi detector.

The LS detector shows promising error rates when the system is overdetermined. Hence, adding the second sample sequence per read head substantially improves the error rates of the LS detector. Expectedly, the detectors exhibit a tradeoff between the density on the lower layer and the SNR required for a particular error rate. The LS detector achieves the largest reductions in SNR for a BER of 10^{-2} at $U = 3$ and $U = 6$.

The Viterbi detector returns the ML bit sequence given the readback waveform. We have evaluated the BER performances for the Viterbi detector and used them to form an estimate of the maximum achievable UBD for our proposed MLMR system, comparing them to reference TDMR and conventional PMR systems. The result is a 13.6% gain over PMR and

6.8% gain over TDMR, respectively. It was also found from the optimization of the cross-track reader locations, that they were close to, but not optimal in, their home positions. After optimizing p_R, the gains went up to 14.5% and 7.6% for TDMR and PMR, respectively.

Furthermore, optimization of p_R where we do not constrain the reader positions to be symmetrical could lead to additional gains. In addition, a simple U-repetition code on the bottom layer is easy to implement, but not necessarily the best choice. These are the candidate topics for us to continue to investigate in our future work.

REFERENCES

[1] R. Wood, "The feasibility of magnetic recording at 1 Terabit per square inch," *IEEE Trans. Magn.*, vol. 36, no. 1, pp. 36–42, Jan. 2000.

[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] X. Sun, B. J. Belzer, and K. Sivakumar, "Dynamic grain state estimation for high-density TDMR: Progress and future directions," *IEEE Trans. Magn.*, vol. 52, no. 2, Feb. 2016, Art. no. 9400107.

[4] R. E. Rottmayer *et al.*, "Heat-assisted magnetic recording," *IEEE Trans. Magn.*, vol. 42, no. 10, pp. 2417–2421, Oct. 2006.

[5] T. Rausch *et al.*, "HAMR drive performance and integration challenges," *IEEE Trans. Magn.*, vol. 49, no. 2, pp. 730–733, Feb. 2013.

[6] G. Woltersdorf and C. H. Back, "Microwave assisted switching of single domain $Ni_{80}Fe_{20}$ elements," *Phys. Rev. Lett.*, vol. 99, Nov. 2007, Art. no. 227207.

[7] J.-G. Zhu, X. Zhu, and Y. Tang, "Microwave assisted magnetic recording," *IEEE Trans. Magn.*, vol. 44, no. 1, pp. 125–131, Jan. 2008.

[8] K. K. Teo, K. S. Chan, S. J. Greaves, and Y. Kanai, "Areal density prediction for microwave-assisted magnetic recording," *IEEE Trans. Magn.*, vol. 51, no. 11, Nov. 2015, Art. no. 7100904.

[9] H. J. Richter *et al.*, "Recording on bit-patterned media at densities of 1 Tb/in^2 and beyond," *IEEE Trans. Magn.*, vol. 42, no. 10, pp. 2255–2260, Oct. 2006.

[10] J. Hu, T. M. Duman, E. M. Kurtas, and M. F. Erden, "Bit-patterned media with written-in errors: Modeling, detection, and theoretical limits," *IEEE Trans. Magn.*, vol. 43, no. 8, pp. 3517–3524, Aug. 2007.

[11] P.-Y. Du, H.-T. Lue, Y.-H. Shih, K.-Y. Hsieh, and C.-Y. Lu, "Overview of 3D NAND Flash and progress of split-page 3D vertical gate (3DVG) NAND architecture," in *Proc. 12th IEEE Int. Conf. Solid-State Integr. Circuit Technol. (ICSICT)*, Oct. 2014, pp. 1–4.

[12] R. Micheloni, L. Crippa, C. Zambelli, and P. Olivo, "Architectural and integration options for 3D NAND flash memories," *Computers*, vol. 6, no. 3, p. 27, 2017.

[13] K. S. Chan, R. Wood, and S. Rahardja, "Maximum likelihood detection for 3D-MAMR," *IEEE Trans. Magn.*, to be published.

[14] K. S. Chan *et al.*, "User areal density optimization for conventional and 2-D detectors/decoders," *IEEE Trans. Magn.*, vol. 54, no. 2, Feb. 2018, Art. no. 3100412.

[15] S. J. Greaves, K. S. Chan, and Y. Kanai, "Optimization of dual-structure recording media for microwave-assisted magnetic recording," *IEEE Trans. Magn.*, vol. 55, no. 7, Jan. 2019, Art. no. 3001305.

[16] S. Greaves, Y. Kanai, and H. Muraoka, "Multiple layer microwave-assisted magnetic recording," *IEEE Trans. Magn.*, vol. 53, no. 2, Feb. 2017, Art. no. 3000510.

[17] J. M. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications, and Control*. London, U.K.: Pearson, 1995.

[18] C. E. Shannon, "A mathematical theory of communication," *The Bell System Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.