Channel Modeling and Target Design for Two-Dimensional Optical Storage Systems

Li Huang1,2, Student Member, IEEE, George Mathew1, Member, IEEE, and Tow Chong Chong1,2, Member, IEEE
1Department of Electrical and Computer Engineering, National University of Singapore, 117576 Singapore
2Data Storage Institute, A*STAR, 117608 Singapore

In this paper, we first show that the symbol response of the two-dimensional (2-D) optical storage (TwoDOS) channel can be computed by a one-dimensional (1-D) Hankel transform, instead of the 2-D Fourier transform. This results in a computationally efficient approach for generating readback signals in the presence of pit-size variations. We also show how to design a 2-D minimum mean-square error (MMSE) equalizer for channels with these distortions. Second, we present a novel way to jointly design equalizer and target under linear for generating readback signals in the presence of pit-size variations. We also show how to design a 2-D minimum mean-square error by a one-dimensional (1-D) Hankel transform, instead of the 2-D Fourier transform. This results in a computationally efficient approach for generating readback signals in the presence of pit-size variations.

Index Terms—Hankel transform, monic constraint, target design, two-dimensional optical storage.

I. INTRODUCTION

In response to the continuously increasing demand for storage capacity, optical recording research has undergone three generations [namely, compact disc (CD; 650 MB), digital versatile disc (DVD; 4.7 GB), blu-ray disc (BD; 25 GB)] of optical media, with BD being the most recent [1]–[4]. Even though each generation offers significant improvement in storage capacity, the growth rate of recording density in optical storage lags behind that of magnetic storage, largely due to the slow pace at which the wavelength of laser diode has improved. The relatively slow pace of physical improvements in optical storage motivates the use of advanced signal processing techniques to achieve further increase in recording density. One promising example of this is the recently introduced two-dimensional optical storage (TwoDOS) approach [5]. Compared with the conventional one-dimensional (1-D) optical storage, this technology noticeably reduces the track pitch and thus makes it possible to record at much higher track density that is comparable to the linear density. This higher track density is realized by grouping a number of adjacent tracks with shorter track pitch between them, and using a larger track pitch as the group boundaries. Further, TwoDOS maximizes the capacity of the disc by adopting hexagonal bit cells instead of traditional square/rectangular cells (which was used in [6], [7]). Moreover, a higher data rate can be achieved with the use of parallel writing and read-out. Even though the presence of a guard-band between groups prevents the interference from adjacent groups, the use of shorter track pitch between the tracks within the group results in severe two-dimensional (2-D) intersymbol interference (ISI) from bits across as well as along the tracks. Consequently, it becomes very important to develop powerful 2-D signal processing and coding techniques to ensure reliable data recovery. In this paper, we address the issues of channel modeling, and the design of equalizer and partial response (PR) target for a 2-D Viterbi detector (VD) based receiver.

The linear 2-D channel model for TwoDOS was first introduced in [8] by taking the 2-D inverse Fourier transform of the 2-D modulation transfer function (MTF), resulting in 2-D impulse response of the channel. However, this channel model may not be accurate enough unless the density is sufficiently high since this model does not account for the presence of the linear pulse modulator in the write circuit. In this paper, we present the linear channel model which incorporates the linear pulse modulator. More importantly, we show that by exploiting the radial symmetry property of the symbol response, the 2-D symbol response can be efficiently calculated by a 1-D Hankel transform approach [9]. Though our linear channel model is constructed solely from a signal processing point of view, it reflects physical reality since it is consistent with the linear part of the physical channel model constructed by Coene [10] based on scalar diffraction theory. We also develop the 2-D symbol response model in the presence of pit-size variations, and show that the Hankel transform approach can be used to develop a computationally simple approach for generating readback signals in the presence of pit-size variations. Because of the presence of serious 2-D ISI in TwoDOS readback signals, the 2-D VD has been used for bit detection [11]. As the complexity of 2-D VD grows exponentially with both channel length and number of tracks per group, 2-D PR equalization has been used to shorten the channel memory [12]. In our work, instead of choosing a fixed target to design the equalizer, we optimize both the equalizer and target responses based on minimum mean-square error (MMSE) approach to improve the performance of 2-D VD. However, to avoid the trivial solution, we need to impose a suitable constraint on the 2-D PR target. Since doing constrained minimization in 2-D is quite difficult, we propose a novel method which converts the 2-D target design problem into a 1-D problem. We also develop an appropriate target constraint that results in good bit detection performance from 2-D VD.

Section II describes the TwoDOS channel model and the Hankel transform approach to the computation of symbol response. In Section III, the 2-D MMSE equalizer design approach is presented. In Section IV, we investigate the theoretical
performance of 2-D VD in the presence of pit-size variations. A novel 2-D target design method is proposed in Section V. The investigation of different target constraints along with performance comparison through bit-error-rate (BER) simulations are presented in Section VI. Finally, Section VII concludes the paper.

II. CHANNEL MODEL

TwoDOS uses a multipoint light beam for parallel read-out. For linear read-out, each spot can be characterized by the MTF given by the Braat–Hopkins formula [14]

$$F_{\phi,p}(\phi, p) = \left\{ \begin{array}{ll}
\frac{2}{\pi} \arccos \left( \frac{p}{\rho_c} \right) \sqrt{1 - \left( \frac{p}{\rho_c} \right)^2} & , \quad \text{if } \rho \leq \rho_c \\
0 & , \quad \text{else}
\end{array} \right. \quad (1)$$

where \( \rho \) is the spatial angular frequency, \( \rho_c \) is the angular cutoff frequency, and \( \phi \) is the azimuth angle in the 2-D spatial frequency plane. The 2-D linear pulse modulator corresponds to the pit region used in the writing process (i.e., pit hole), and is defined as

$$C_{\phi} (\theta, r) = \left\{ \begin{array}{ll}
1, & \text{when } r \leq R \\
0, & \text{else}
\end{array} \right. \quad (2)$$

where \( r \) and \( \theta \) are radial and angular coordinates, respectively, in spatial domain and \( R \) is the radius of the pit hole.

We use polar coordinates in deriving the channel symbol response. Let \( C_{\phi,p} (\phi, p) \) represent the linear pulse modulator in spatial frequency domain. Then, the symbol response in spatial frequency domain is given by

$$H_{\phi,p}(\phi, p) = C_{\phi,p}(\phi, p)F_{\phi,p}(\phi, p). \quad (3)$$

By taking 2-D Fourier transform on both sides of (2), we obtain

$$C_{\phi,p}(\phi, p) = 2\pi R J_0(pR/p) \quad (4)$$

where \( J_0(x) \) is the Bessel function of the first kind and first order. Then the symbol response in spatial domain, \( H_{\phi}(\theta, r) \), can be obtained by taking the 2-D inverse Fourier transform of \( H_{\phi,p}(\phi, p) \), i.e.,

$$H_{\phi}(\theta, r) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho J_0(p\rho) e^{ij\rho \cos(\theta - \phi)} \ d\rho \ d\phi. \quad (5)$$

It should be noted that the above symbol response considers only linear ISI and it is developed purely from a signal processing point of view. In the latter part of this section, we will modify this to include nonlinear interferences also. A rigorous derivation of the channel model based on physical generation of the signal, including linear and nonlinear ISI, can be found in [10]. However, as we show below, our model in (3) is consistent with the model in [10] when only linear ISI is considered.

Based on the general formulation in [10], we can write the linear symbol response \( H_{\phi}(\theta, r) \) in spatial domain as \((x, y)\) being Cartesian coordinates

$$H_{\phi}(x, y) = \int \int |p(x', y')|^2 C_s(x' - x, y' - y) \ dx' \ dy' \quad (6)$$

where \( p(x', y') \) is the spot function and \( C_s(x' - x, y' - y) \) is the window function centered at \((x, y)\). Because \( C_s(x, y) \) is circularly symmetric in TwoDOS, \( C_s(x' - x, y' - y) = C_s(x' - x, y' - y) \). Therefore, (6) implies that \( H_{\phi}(x, y) \) is the 2-D convolution between \( |p(x, y)|^2 \) and \( C_s(x, y) \). Since \( F_{\phi,p}(\phi, \rho) \) is the 2-D Fourier transform of \(|p(x, y)|^2\), the signal processing based model given in (3) is consistent with the physical model given in (6).

While studying TwoDOS channels with time-varying characteristics (e.g., channels with pit-size variations), it may be necessary to recompute the channel symbol response very often. Therefore, in such situations, directly using the 2-D Fourier transform approach may not be desirable in view of its computational complexity. We now show that the 2-D Fourier transform can be replaced by Hankel transform in TwoDOS, thus resulting in significant computational saving in the calculation of \( H_{\phi}(\theta, r) \). From (1) and (4) we find that since \( F_{\phi,p}(\phi, \rho) \) and \( C_{\phi,p}(\phi, \rho) \) are both radially symmetric, the symbol response in spatial frequency domain \( H_{\phi,p}(\phi, \rho) \) is also radially symmetric, i.e., \( H_{\phi,p}(\phi, \rho) = H_{\phi,p}(\rho) \). Then, (5) can be simplified as

$$H_{\phi}(\theta, r) = \frac{1}{2\pi} \int_0^{\rho_c} \rho J_0(\rho R) \ d\rho \quad (7)$$

where \( J_0(x) \) is the Bessel function of the first kind and zero order, and is independent of \( \theta \). Thus, \( H_{\phi}(\theta, r) \) is also radially symmetric, i.e., \( H_{\phi}(\theta, r) = H_{\phi}(\rho) \). Equation (7) shows that \( H_{\phi}(\theta, r) \) is the inverse Hankel transform of \( H_{\phi,p}(\rho) \) [9]. Thus, by taking advantage of the circular symmetry of the TwoDOS channel, the 2-D Fourier transform can be reduced to a 1-D Hankel transform. This is important since, as we show below, the Hankel transform approach is a computationally efficient means to compute \( H_{\phi}(\theta, r) \), compared with 2-D Fourier transform. The fast Hankel transform can be computed with a complexity of about \( O(N_s \log_2 N_s) \), where \( N_s \) denotes the number of data points in each dimension [15], [16]. Compared to this, the complexities of traditional 2-D discrete Fourier transform (DFT) and 2-D fast Fourier transform (FFT) are \( O(N_s^2 \log_2 N_s) \) and \( O(N_s^2 \log_2 N_s) \), respectively [17]. Thus, the fast Hankel means of computing the channel symbol response is indeed a computationally efficient approach when the radial symmetry property is maintained in TwoDOS systems.

Note that \( H_{\phi}(r) \) characterizes a single spot only. The channel response of the 2-D system is given by \( H = [H_{\phi,0}, H_{\phi,1}, \ldots, H_{\phi,N_{\phi}-1}]^T \), where \( N_{\phi} \) is the channel length, \( H_{\phi} \) is a \( N_{\phi} \times N_{\phi} \) matrix, and \( N_{\phi} \) denotes the number of tracks forming one group. Fig. 1 shows the arrangement of bit cells in TwoDOS recording system. The dashed slanted rectangular boxes indicate the array of bit cells onto which the parallel read-out laser beam is focused at each time instant.
Considering the hexagonal structure of each bit cell, the \((i, j)\)th element of \(H_k\) can be obtained as

\[
H_k(i, j) = \hat{H}_k \left( \sqrt{x_{i,j}^2 + y_{i,j}^2} T \right), \quad i, j = 1, 2, \ldots, N_r
\] (8)

where \(x_{i,j} = (i - j) \cos 60^\circ + t_k, y_{i,j} = (j - i) \sin 60^\circ\), \(t_k = (k - (N_b + 1)/2) T\) is the spatial difference along the track (for odd \(N_b\)), and \(T\) is the center-to-center distance between adjacent bits. In fact, \(H_k(i, j)\) can be considered as the interference of \(j\)th track to \(i\)th track. The readback signal resulting from the parallel read-out at time index \(n\) is given by

\[
z(n) = \sum_{k=0}^{N_b-1} H_k a(n-k) + \theta(n)
\] (9)

where \(z(n) = [z_1(n), z_2(n), \ldots, z_{N_b}(n)]^T\), \(z_k(n)\) denotes the readback signal component from the \(k\)th track, \(a(n) = [a_1(n), a_2(n), \ldots, a_{N_b}(n)]^T\), \(a_k(n) \in \{-1, 1\}\) denotes the data bit written on the \(k\)th track, \(\theta(n) = [\theta_1(n), \theta_2(n), \ldots, \theta_{N_b}(n)]^T\) and \(\theta_k(n)\) denotes the noise picked up from the \(k\)th track, for \(k = 1, 2, \ldots, N_b\).

It has been shown that the optimum radius of the pit hole is

\[R = (1/2) \sqrt{3} T\] [11]. In practice, however, it is difficult to accurately control the size of the recorded domain. For example, the domain bloom, which is due to under- or over-etching during the mastering process, causes pits to be systematically larger or smaller and it is one of the major write imperfections. In this case, (4) becomes

\[
C_{\text{st}}(\rho, \Delta_b) = 2\pi (1 + \Delta_b) R \frac{J_1(\rho(1 + \Delta_b) R)}{\rho}
\] (10)

where \(\Delta_b\) reflects the degree of bloom. A positive value of \(\Delta_b\) means that the recorded pits are systematically larger than the nominal pits and vice versa when \(\Delta_b\) is negative. The channel symbol response \(H_k\) \((k = 0, 1, \ldots, N_b - 1)\) should be computed by taking this into account. Unlike the 1-D system, where bloom manifests only when pit-to-land or land-to-pit transitions occur, the transitions in TwoDOS occur whenever bits corresponding to “+1” are written. This is because every “+1” is recorded as a pit of radius \(R\) smaller than \(T/2\) and thus there are two transition edges between consecutive pits, instead of none in the 1-D system.

In this section, it has been implicitly assumed that the recorded domains has infinitely steep vertical edges. In reality, however, the edges of the recorded domains are not always sharply defined and tend to exhibit random pit-size variations. Accounting for this random pit-size variation in (10), we get

\[
C_{\text{st}}(\rho, \Delta_b) = 2\pi (1 + \Delta_b + \Delta_e) R \frac{J_1(\rho(1 + \Delta_b + \Delta_e) R)}{\rho}
\] (11)

where the normalized parameter \(\Delta_e\) is a zero-mean random variable that is statistically independent of the input data vector \(a(n)\). Due to the same reason as explained above, this random pit-size variation occurs whenever the input bit is “+1.” Thus, the random variable \(\Delta_e = \Delta_b + \Delta_e\) may be used to represent the effective pit-size variation. The bloom component \(\Delta_b\), which remains constant for all pits, is the mean of \(\Delta_e\), and the component \(\Delta_e\) varies randomly for each pit.

In the presence of pit-size variations, it is easy to see that the channel becomes a time-varying linear system since the symbol response tends to differ from one bit to another. As a result, the readback signal takes the form (see Fig. 2)

\[
z(n) = \sum_{k=0}^{N_b-1} H_k a(n-k) + \theta(n)
\] (12)

where the channel matrices \(H_k\) are now also dependent on the time reference “\(n\).” The elements \(H_k(i, j)\) \((i, j = 1, 2, \ldots, N_r)\) of \(H_k\) can be computed by (8), (7), and (3) after replacing the \(C_{\text{st}}(\rho, \Delta_b)\) in (3) with the time-varying form given in (10) or (11). Thus, to compute \(z(n)\) for each “\(n\)” we need to recompute \(H_k\) for all \(k\), which is a very computationally demanding task. A simple table lookup can be used to save computational complexity by storing precomputed channel responses for quantized values of pit sizes. On the other hand, since the time-varying symbol response still satisfies the radial symmetry condition in spatial frequency domain, we can use the fast Hankel transform approach to compute the individual symbol responses. This approach results in significant savings in computational complexity and memory requirements compared with the 2-D Fourier transform approach and the table lookup approach, respectively. This is very important for doing simulation studies over a very large length of data bits.

III. 2-D MMSE EQUALIZER DESIGN

Fig. 2 shows the discrete-time channel model of TwoDOS system with PR equalization and VD. In the figure, \(G_k\) \((k = 0, 1, \ldots, N_g - 1)\) and \(W_k\) \((k = 0, 1, \ldots, N_w - 1)\) represent
$N_r \times N_r$ coefficient matrices of PR target and equalizer, respectively, $m_0$ and $\Delta$ denote the delays (in number of bits) from channel input to equalizer output and detector output, respectively. In this section, we assume that the channel symbol response $H_k$ ($k = 0, 1, \ldots, N_b - 1$) is linear and time-invariant, and the PR target $G_k$ is known.

The 2-D MMSE equalizer design for TwoDOS was first proposed in [13]. However, the derivation in [13] assumes the input data $a(n)$ as uncoded and the channel noise $\theta(n)$ as 2-D white. In this section, we present a somewhat generalized approach, which can also be used when the channel contains pit-size variations.

From Fig. 2, we get the error vector $e(n)$ as

$$e(n) = x(n) - d(n)$$

$$= \sum_{k=0}^{N_b-1} W_k z(n-k) - \sum_{k=0}^{N_b-1} G_k a(n-k)$$

$$= W^T z(n) - G^T a(n - m_0)$$

(13)

where $z(n) = [z^T(n), z^T(n-1), \ldots, z^T(n - N_w + 1), z^T(n - N_g + 1)]^T$, $a(n) = [a^T(n), a^T(n-1), \ldots, a^T(n - N_g + 1)]^T$, $W = [W_0, W_1, \ldots, W_{N_N-1}], G = [G_0, G_1, \ldots, G_{N_G-1}]$.

Then, the total mean square error (MSE) of the equalizer output can be expressed as

$$J(W) = E\{e^T(n)e(n)\}$$

$$= \text{trace}(W^T R_{ee} W - 2 W^T R_{ez} G + G^T R_{za} G)$$

(14)

where $R_{ee} = E\{z(n)z^T(n)\}$, $R_{ez} = E\{z(n)a^T(n - m_0)\}$, and $R_{za} = E\{a(n - m_0)a^T(n - m_0)\}$. By minimizing the MSE, the optimum taps of the equalizer are obtained as

$$R_{opt} = R_{ee}^{-1}R_{ez}G$$

(15)

and the MMSE becomes

$$J_{\text{min}}(W_{opt}) = \text{trace} \left[ G^T R_{za} G - (R_{ee}^{-1}R_{ez}G)^T (R_{ee}^{-1}R_{ez}G) \right].$$

(16)

We now derive the expressions for $R_{ee}$, $R_{ez}$, and $R_{za}$. Here, $R_{ee}$ is a $N_b N_r \times N_b N_r$ autocorrelation matrix of the equalizer input and its $(k,l)$th submatrix $(k,l = 0, 1, \ldots, N_b - 1)$ is given by $R_{ee}^{kl} = E\{z(n-k)z^T(n-l)\}$. Since $a(n)$ and $\theta(n)$ are mutually uncorrelated and are of zero mean, using (9) we get

$$R_{ee}^{kl} = H^T R_{k\theta}^{kl} H + R_{kl}$$

(17)

where $R_{k\theta}^{kl} = E\{a(n-k)a^T(n-l)\}$ is a $N_b N_r \times N_b N_r$ autocorrelation matrix of the input data, $\theta(n) = [\theta^T(n), \theta^T(n-1), \ldots, \theta^T(n - N_g + 1)]^T$, and $R_{kl} = E\{\theta(n-k)\theta^T(n-l)\}$ is a $N_r \times N_r$ autocorrelation matrix of the noise. The $(k',l')$th submatrix of $R_{ee}^{kl}$ is given by

$$R_{ee}^{kl}(i, j) = E[a(n-k-k')a(n-l-l')]$$

(18)

for $k', l' = 0, \ldots, N_b - 1$, and $(i, j)$th elements of $R_{k\theta}^{kl}$ and $R_{kl}^{kl}$ are given by

$$R_{k\theta}^{kl}(i, j) = E[a(n-k-k')a(n-l-l')]$$

(19)

$$R_{kl}^{kl}(i, j) = E[\theta(n-k)\theta^T(n-l)]$$

(20)

respectively, for $i, j = 1, 2, \ldots, N_r$.

The matrix $R_{zza}$ is a $N_w N_r \times N_g N_r$ cross-correlation matrix between the channel output and input data, and its $(k',l')$th submatrix $(k, l = 0, 1, \ldots, N_w - 1; l = 0, 1, \ldots, N_g - 1)$ is given by

$$R_{za}^{kl} = E[z(n-k)a^T(n - m_0 - l)]$$

Then, as before, we get

$$R_{za}^{kl} = H^T R_{k\theta}^{kl} H$$

(21)

where $R_{za}^{kl} = E[a(n-k)a^T(n - m_0 - l)]$ is a $N_b N_r \times N_g N_r$ autocorrelation matrix of the input data. Its $(k',l')$th submatrix is given by

$$R_{za}^{kl}(i, j) = E[a(n-k-k')a(n-l-l')]$$

(22)

for $k', l' = 0, 1, \ldots, N_b - 1$, and $(i, j)$th element of $R_{za}^{kl}$ is given by

$$R_{za}^{kl}(i, j) = E[a(n-k-k')a(n-l-l')]$$

(23)

for $i, j = 1, 2, \ldots, N_r$.

Finally, $R_{za}$ is a $N_g N_r \times N_g N_r$ autocorrelation matrix of input data. As before, its $(k,l)$th submatrix $(k, l = 0, 1, \ldots, N_g - 1)$ is given by

$$R_{za}^{kl} = E[a(n-k-m_0)a^T(n-m_0-l)]$$

(24)

for $i, j = 1, 2, \ldots, N_r$.

Special Cases:

1) Input data and noise are 2-D wide-sense stationary.

In this case, (19), (20), (23), and (24) can be simplified as

$$R_{k\theta}^{kl}(i, j) = r_{a}(l-k-l') \delta_{i-j}$$

(25)

$$R_{kl}^{kl}(i, j) = r_{\theta}(l-k-l') \delta_{i-j}$$

(26)

$$R_{za}^{kl}(i, j) = r_{a}(l-k-m_0-l) \delta_{i-j}$$

(27)

$$R_{za}^{kl}(i, j) = r_{\theta}(l-k-l') \delta_{i-j}$$

(28)

respectively, where $r_{a}(k)(i)$ and $r_{\theta}(k)(i)$ are defined as $E[a(n-k)a^T(n-k)]$ and $E[\theta(n-k)\theta^T(n-k)]$, respectively, for any $n$ and $j$.

2) The noise is 2-D white with variance $\sigma^2$ and the input data bits are uncoded.

This is the simplest case. In this case, (19), (20), (23), and (24) can be simplified as

$$R_{k\theta}^{kl}(i, j) = r_{\delta}(l-k-l') \delta_{i-j}$$

(29)

$$R_{kl}^{kl}(i, j) = \sigma^2 \delta_{i-j}$$

(30)

$$R_{za}^{kl}(i, j) = \delta_{i-j}$$

(31)

$$R_{za}^{kl}(i, j) = \delta_{i-j}$$

(32)
respectively, where \( \delta_k = 1 \) when \( k = 0 \) and \( \delta_k = 0 \) when \( k \neq 0 \). Substituting (29) and (30) in (17), we get

\[
\mathbf{R}_{\mathbf{d}} = \begin{cases} \mathbf{H}_l \mathbf{H}_l^T & \text{if } k = l, \\ \sigma^2 \mathbf{I}_{N_p}, & \text{else} \end{cases}
\]

where \( \mathbf{H}_l \) is the mirror filter of \( \mathbf{H} \), i.e., \( \mathbf{H}_l(k) = \mathbf{H}(-k) \), \( \mathbf{I}_{N_p} \) is the \( N_p \times N_p \) identity matrix and \( \ast \) refers to 2-D convolution given by \( \mathbf{A} * \mathbf{B} = \sum_{n} \mathbf{A}_n \mathbf{B}_{n-k} \). Using (21), we get the \( k \)th submatrix of \( \mathbf{R}_{\mathbf{d}_{\mathbf{a}} \mathbf{G}} \) as

\[
(\mathbf{R}_{\mathbf{d}_{\mathbf{a}} \mathbf{G}})_k = \sum_{k'=0}^{N_k-1} \sum_{\ell=0}^{N_{k'}-1} \mathbf{H}_k \tilde{\mathbf{R}}_{d_{\mathbf{a}} \mathbf{G}}^{a} \mathbf{G}_l^T.
\]

Substituting (31) in (34) and letting \( l = k + k' = m_0 \), we obtain

\[
\mathbf{R}_{\mathbf{d}_{\mathbf{a}} \mathbf{G}} = \sum_{k'}^{N_k-1} \begin{bmatrix} \mathbf{H}_{k'} \mathbf{G}_l^T & \mathbf{H}_{k'} \mathbf{G}_l^T & \cdots & \mathbf{H}_{k'} \mathbf{G}_l^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{k'} \mathbf{G}_l^T & \mathbf{H}_{k'} \mathbf{G}_l^T & \cdots & \mathbf{H}_{k'} \mathbf{G}_l^T \\ \mathbf{H}_k \mathbf{G}_l^T & \mathbf{H}_k \mathbf{G}_l^T & \cdots & \mathbf{H}_k \mathbf{G}_l^T \\ \mathbf{H}_k \mathbf{G}_l^T & \mathbf{H}_k \mathbf{G}_l^T & \cdots & \mathbf{H}_k \mathbf{G}_l^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_k \mathbf{G}_l^T & \mathbf{H}_k \mathbf{G}_l^T & \cdots & \mathbf{H}_k \mathbf{G}_l^T \end{bmatrix}.
\]

We can compute the optimum equalizer taps by using (33) and (35) in (15). Since \( \mathbf{R}_{\mathbf{d}_{\mathbf{a}} \mathbf{G}} \) is a \( N_g N_p \) \( N_g N_p \) identity matrix in this simplest case, by using (16) we get the corresponding MMSE as

\[
J_{\text{min}}(\mathbf{W}) = \text{trace} \left[ \mathbf{G}_l^T \mathbf{G} - (\mathbf{R}_{\mathbf{d}_{\mathbf{a}} \mathbf{G}})^T \mathbf{R}_{\mathbf{d}_{\mathbf{a}} \mathbf{G}} \right].
\]

The expressions (33)–(36) corresponding to the simplest case can be found in [13].

3) Channel contains pit-size variations.

The Fourier transform of the channel response in the presence of pit-size variations is given by [using (11) and (3)]

\[
\mathbf{H}_{\text{st}}(\phi, \rho) = 2\pi (1 + \Delta_e) R J_1(\rho + \Delta_e) R \mathbf{F}_{\text{st}}(\phi, \rho)
\]

where \( \Delta_e = \Delta_b + \Delta_t \) with \( \Delta_b \) and \( \Delta_t \) are the normalized domain bloom and random pit-size variation for the bit under consideration, respectively. If \( \Delta_b \) and \( \Delta_t \) are sufficiently small and the recording density is sufficiently high, the above transfer function can be approximated as [18]

\[
\mathbf{H}_{\text{st}}(\phi, \rho) \approx 2\pi (1 + \Delta_e) R J_1(\rho R) \mathbf{F}_{\text{st}}(\phi, \rho)
\]

where \( \mathbf{H}_{\text{st}}(\phi, \rho) \) is the time-invariant response given in (3). Using this approximation, we can write (12) as

\[
\mathbf{z}(n) = \sum_{k=0}^{N_k-1} \mathbf{H}_k \tilde{\mathbf{a}}(n-k) + \mathbf{\theta}(n)
\]

where \( \mathbf{H}_k(k = 0, 1, \ldots, N_k - 1) \) corresponds to the time-invariant symbol response matrices and \( \tilde{\mathbf{a}}(n) = [\tilde{a}_1(n), \tilde{a}_2(n), \ldots, \tilde{a}_{N_k}(n)]^T \) is the modified input data vector where

\[
\tilde{a}_i(n) = \begin{cases} (1 + \Delta_i) a_i(n), & \text{if } a_i(n) = +1 \\ a_i(n), & \text{if } a_i(n) = -1 \end{cases}
\]

for \( i = 1, 2, \ldots, N_k \), where \( \Delta_i(n) \) corresponds to the effective pit-size variation experienced by the bit \( a_i(n) \).

It is now easy to see that Case 1 discussed above (i.e., wide-sense stationary data and noise) can be used to design the optimum equalizer for the case with random pit-size variation since the modified data \( \tilde{\mathbf{a}}(n) \) and white electronics noise \( \mathbf{\theta}(n) \) are mutually uncorrelated and wide-sense stationary. In this case, (19), (20), (23), and (24) can be simplified as

\[
\mathbf{R}_{\mathbf{d}_{\mathbf{a}} \mathbf{G}}^{(l,k)(m_0)(m_0)} = r_{\mathbf{a}}(l - k - l' - k')(i - j)
\]

\[
\mathbf{R}_{\mathbf{d}}^{(l,k)(l,k + m_0)(l,k + m_0)} = r_{\mathbf{a}}(l - k - l' - k')(i - j)
\]

\[
\tilde{\mathbf{R}}_{\mathbf{d}_{\mathbf{a}} \mathbf{G}}^{(l,k)(m_0)(m_0)} = r_{\mathbf{a}}(l - k - l' + m_0)(i - j)
\]

\[
\tilde{\mathbf{R}}_{\mathbf{a}}^{(l,k)(l,k + m_0)(l,k + m_0)} = r_{\mathbf{a}}(l - k - l' + m_0)(i - j)
\]

respectively, where \( r_{\mathbf{a}}(k)(i) \) and \( r_{\mathbf{d}}(k)(i) \) are defined as \( E[\tilde{a}_{i+j}(n+k)\tilde{a}_j(n)] \) and \( E[\mathbf{\theta}_{i+j}(n+k)\theta_j(n)] \), respectively, for any \( n \) and \( j \). Let \( \Delta_i(n) \) represent the random pit-size variation experienced by the bit \( a_i(n) \). Noting that \( \Delta_i(n) \) is statistically independent of \( a_j(n) \) and \( a_i(n) \) is of zero mean, it can be shown that

\[
r_{\mathbf{a}}(k)(i) = \left[ 1 + r_{\mathbf{a}}(k)(i) [\Delta_i^2 + r_{\mathbf{d}}(k)(i)] \right]^{1/2} + (1 + \Delta_i)r_{\mathbf{a}}(k)(i)
\]

where \( r_{\mathbf{a}}(k)(i) \) is defined as \( E[\tilde{a}_{i+j}(n+k)\tilde{a}_j(n)] \) for any \( n \) and \( j \). If \( \Delta_e = 0 \), the noise is 2-D white with variance \( \sigma^2 \) and the input data bits are uncoded, then the expressions of correlations given in (19), (20), (23), and (24) can be simplified for this case as

\[
\mathbf{R}_{\mathbf{d}_{\mathbf{a}} \mathbf{G}}^{(l,k)(l,k + m_0)(l,k + m_0)} = (1 + \sigma^2/2) \delta_{l-k-l'-k'} \delta_{i-j}
\]

\[
\mathbf{R}_{\mathbf{d}}^{(l,k)(l,k + m_0)(l,k + m_0)} = (1 + \sigma^2/2) \delta_{l-k-l'-k'} \delta_{i-j}
\]

\[
\tilde{\mathbf{R}}_{\mathbf{d}_{\mathbf{a}} \mathbf{G}}^{(l,k)(m_0)(m_0)} = (1 + \sigma^2/2) \delta_{l-k-l'+m_0} \delta_{i-j}
\]

\[
\tilde{\mathbf{R}}_{\mathbf{a}}^{(l,k)(l,k + m_0)(l,k + m_0)} = (1 + \sigma^2/2) \delta_{l-k-l'+m_0} \delta_{i-j}
\]

respectively, where \( \sigma^2 \) is the variance of random pit-size variation \( \Delta_e \). The factor \( 1/2 \) accounts for the the probability of pit bit “+1” where random pit-size variation may occur.

IV. PERFORMANCE OF VITERBI DETECTOR FOR TwoDOS

To deal with the serious 2-D ISI present in TwoDOS readback signal and to ensure reliable data recovery, powerful detection approaches such as the Viterbi algorithm become necessary [11]. The BER performance of 2-D VD may be used as a criterion to evaluate different PR targets. Even though a BER expression for 2-D VD has been reported in [13], the variables used in the expression correspond to 1-D VD. In this section,
we briefly derive the derivation of BER of 2-D VD using the 2-D definitions and notations of error event, target, etc.

As a starting point for determining the BER, we define a 2-D error event vector \( \mathbf{e} \). Let \( \mathbf{a}(n) \) denote the detected data vector corresponding to \( \mathbf{a}(n) \). Further, define \( \mathbf{a} = [\mathbf{a}^T(n), \mathbf{a}^T(n+1), \ldots, \mathbf{a}^T(n+N_e-1)]^T, \mathbf{a} = [\mathbf{a}^T(n), \mathbf{a}^T(n+1), \ldots, \mathbf{a}^T(n+N_e-1)]^T, e = \frac{1}{2} = [e^T(n), e^T(n+1), \ldots, e^T(n+N_e-1)]^T \) with \( e(n) = (a(n) - \hat{a}(n))/2 \). Then, \( \mathbf{e} \) is a 2-D error event of length \( N_e \) if:

1. \( e(n) \neq 0 \) and \( e(n+N_e-1) \neq 0 \);
2. the length of strings of zero vectors in \( \mathbf{e} \) does not exceed \( N_f \);
3. \( \hat{a}(n+i) = a(n+i) \) for \( -N_f \leq i < 0 \) and \( N_e \leq i \leq N_e + N_f - 1 \).

Here, \( N_f \) is called the “error-free interval” and \( N_f \geq N_g \).

Let \( W(\mathbf{e}) \) be the number of error bits in the error event \( \mathbf{e} \). Then, the BER of 2-D VD is given by

\[
P_e = \sum_{\mathbf{a} \in E_e} \sum_{\mathbf{b} \in S_e} W(\mathbf{e})P(\mathbf{a})P(\text{error } | \mathbf{a})
\]

where \( E_e \) is the set of all possible error events, \( S_e \) is the set of all possible patterns that support the error event \( \mathbf{e} \), and \( P(\text{error } | \mathbf{a}) \) is the conditional probability that the VD detects \( \hat{a} \) as the recorded data when the true recorded data is \( a \).

According to the principle of VD, \( P(\text{error } | \mathbf{a}) \) can be upper-bounded by the probability that the metric corresponding to path \( \hat{a} \) is smaller than that corresponding to path \( a \). That is

\[
\sum_i \left[ (x(n+i) - (G \times \hat{a})_{n+i})^T \right] (x(n+i) - (G \times a)_{n+i}) < \sum_i \left[ (x(n+i) - (G \times a)_{n+i})^T \right] (x(n+i) - (G \times a)_{n+i})
\]

where \( x(n) \) is the equalizer output which can be expressed as

\[
x(n) = \sum_i W_i x(n-i)
\]

where \( M_0, M_1, \ldots, M_{N_f+N_g-2} \) represent the taps of the 2-D residual ISI channel, i.e., \( M_n = (W \times H_n - G_n) \), and \( \theta(n) \) is a 2-D white Gaussian noise vector with variance \( \sigma^2 \). Using (52), we can express (51) as

\[
d^2(\mathbf{e}) + \sum_{i=0}^{N_e+N_g-2} \mathbf{e}^T(n+i)(M \times a)_{n+i} + (W \times \theta)_{n+i} < 0
\]

where \( d^2(\mathbf{e}) = \sum_{i=0}^{N_e+N_g-2} (G \times e)_{n+i}^T(G \times e)_{n+i} \) and \( e(n) = (G \times e)_{n+i} \). For given \( \mathbf{a} \) and \( \mathbf{e} \), the left-hand side of the inequality in (53) is a Gaussian random variable with mean

\[
m_u = d^2(\mathbf{e}) + \sum_{i=-N_e}^{N_g} (M^T \times \mathbf{e})_{n+i} a(n+i)
\]

and variance

\[
\sigma^2_n = \sum_{i=-N_e}^{N_g} (W^T \times \mathbf{e})_{n+i}^T(W^T \times \mathbf{e})_{n+i}
\]

Thus, we get

\[
P(\text{error } | \mathbf{a}) \leq Q \left( \frac{m_u}{\sigma_n} \right)
\]

where \( Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt \) is the tail probability of Gaussian distribution. By substituting (56) into (50), we get the BER of 2-D VD as

\[
P_e \leq \sum_{\mathbf{a} \in E_e} \sum_{\mathbf{b} \in S_e} W(\mathbf{e})P(\mathbf{a})Q \left( \frac{m_u}{\sigma_n} \right)
\]

Based on the BER expression given above, we may define an effective detection signal-to-noise ratio (SNR\( \text{eff} \)) for a given error event \( \mathbf{e} \) and target \( \mathbf{G} \) as

\[
\text{SNR}_{\text{eff}} = \left( \frac{d^2(\mathbf{e}) + \sum_{i=0}^{N_e+N_g-2} (M^T \times \mathbf{e})_{n+i}^T a(n+i)}{\sigma_n \sqrt{\sum_{i=0}^{N_e+N_g-2} (W^T \times \mathbf{e})_{n+i}^T (W^T \times \mathbf{e})_{n+i}}} \right)^2
\]

The error event \( \mathbf{e}_t \) that minimizes \( \text{SNR}_{\text{eff}} \) for a given target \( \mathbf{G} \) is called the dominant error event since this error event has the highest probability of occurrence. Consequently, we could use \( \text{SNR}_{\text{eff}} \) as a criterion to determine the target since it is closely related to BER and can be computed very easily compared to BER. The optimum target should maximize the \( \text{SNR}_{\text{eff}} \) corresponding to the dominant error event. In other words, the design of optimum target is a max-min optimization problem

\[
\mathbf{G}_t = \arg\{ \max_{\mathbf{G}} \min_{\mathbf{e}} \text{SNR}_{\text{eff}} \}
\]

As seen from (58), the presence of the residual ISI makes this optimization data-dependent in nature. Therefore, the usual practice is to assume that the residual is very small and to add it as part of the Gaussian noise \( (W \times \theta)_n \). Further, the effect of noise correlation on the detection performance can be inferred by examining the denominator of \( \text{SNR}_{\text{eff}} \).

For the case when the channel contains pit-size variations, (53) is revised as

\[
\sum_{i=0}^{N_e+N_g-2} e^T(n+i)(M \times a)_{n+i} + (W \times \theta)_{n+i} < 0
\]

where \( \mathbf{a}(n) \) is the modified input data vector defined in (40) and \( \mathbf{G}_n = \mathbf{G}_n + \mathbf{M}_n \). Under the assumption that random pit-size variation and thermal noise \( \theta(n) \) are uncorrelated and jointly Gaussian distributed, the left-hand side of the inequality in (60) is still Gaussian with mean

\[
m_{\text{tar}} = d^2(\mathbf{e}) + \sum_{i=0}^{N_e+N_g-2} (M^T \times \mathbf{e})_{n+i} a(n+i)
\]

and variance

\[
\sigma^2_{\text{tar}} = \sum_{i=0}^{N_e+N_g-2} (W^T \times \mathbf{e})_{n+i}^T(W^T \times \mathbf{e})_{n+i}
\]

where \( Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt \) is the tail probability of Gaussian distribution. By substituting (56) into (50), we get the BER of 2-D VD as
and variance
\[
\sigma_{ur}^2 = \sigma^2 \sum_i (\mathbf{W}_c^T \mathbf{e})_{n+i} (\mathbf{W}_c^T \mathbf{e})_{n+i} + \sigma^2 \sum_i [\mathbf{p}(n+i)]^T [\mathbf{p}(n+i)] , \tag{62}
\]
where \( \mathbf{p}(n) = (\mathbf{C}_c^T \mathbf{e}_h \cdot \mathbf{u}(n)) \), \( \Delta_k \) reflects the degree of domain bloom, \( \sigma^2 \) is the variance of random bit-size variation, \( \mathbf{u}(n) = (1 + \mathbf{a}(n))/2 \), and "/" indicates element-wise multiplication of vectors. Then, (58) becomes
\[
\text{SNR}_{\text{eff}} = \left( \frac{m_{ur}}{\sigma_{ur}} \right)^2 . \tag{63}
\]
From (63), it can be seen that the presence of random bit-size variation will degrade the performance since it increases the denominator of (63). However, its destructive effect may be reduced if we consider it in the target design that will be covered in the next section.

V. TARGET DESIGN FOR TWDOS

Several techniques have been reported in the literature for designing the PR target for 1-D data storage systems [19]–[21]. For the sake of convenience, the cost function used for the design is mostly the MSE (at equalizer output) rather than the SNR_{eff} given in (58). The design of equalizer and target by minimizing MSE subject to monic constraint (i.e., first tap of the target is set to unity) has been reported to result in near-optimum performance in 1-D system [20]. The use of a constraint is necessary to avoid the trivial all-zero solutions for equalizer and target. In this section, we investigate the type of target constraint required to achieve good performance from 2-D VD. Since the PR target is in the form of a sequence of matrices in the 2-D case (as compared to a vector target in 1-D case), solving the constrained minimization is quite difficult. Therefore, we introduce a novel method which transforms the 2-D target design problem from 2-D into 1-D form. Using this method, targets with different constraints can be designed with relatively less effort as in the 1-D case.

Without any loss of generality, we define a vector \( \mathbf{g} \) which transforms the target \( \mathbf{G} = [\mathbf{G}_{00}, \mathbf{G}_{11}, \ldots, \mathbf{G}_{N_0 N_1}]^T \) into a 1-D vector target \( \mathbf{g} = [g_0, g_1, \ldots, g_{N_0 N_1}]^T \) where \( g_i \) (\( i = 0, 1, \ldots, N_0 N_1 - 1 \)) are the distinct nonzero elements in \( \mathbf{G} \). It should be noted that imposing some symmetry constraints on the taps of the target can decrease the number of parameters to optimize and thus reduce the complexity of target design process. Replacing \( \mathbf{G} \) with \( \mathbf{g} \) in (14), the MSE for the 2-D case becomes
\[
J(\mathbf{W}, \mathbf{g}) = \text{trace} \{ \mathbf{W}_c^T \mathbf{R}_c \mathbf{W} - 2 (\mathbf{W}_c^T \mathbf{R}_c)^T \mathbf{g} + \mathbf{g}^T (\mathbf{R}_a^r \mathbf{g}) \} \tag{64}
\]
where \( (\mathbf{R}_c^r)^T \) and \( (\mathbf{R}_a^r)^T \) represent rearranged forms of \( \mathbf{R}_c^r \) and \( \mathbf{R}_a^r \), respectively. More specifically, \( (\mathbf{W}_c^T)^T \) is a vector given by
\[
(\mathbf{W}_c^T)^T = [\text{1st row of } \mathbf{W}_c^T \quad \cdots \quad \text{(N_r)th row of } \mathbf{W}_c^T] . \tag{65}
\]
Then, we can obtain \( (\mathbf{R}_c^r)^T \) as
\[
(\mathbf{R}_c^r)^T = \begin{bmatrix}
\text{1st column of } \mathbf{R}_c^r & 
\vdots & 
N_r\text{-th column of } \mathbf{R}_c^r \\
\vdots & 
\ddots & 
\vdots \\
N_r\text{-th column of } \mathbf{R}_c^r & 
\vdots & 
\text{1st column of } \mathbf{R}_c^r \\
\end{bmatrix}^T . \tag{66}
\]
where \( \text{trace}(\mathbf{W}_c^T \mathbf{R}_c^r \mathbf{G}) = (\mathbf{W}_c^T \mathbf{R}_c)^T \mathbf{g} \) and \( \mathbf{R}_c^r \) is the matrix associated with the coefficient \( g_k \) such that
\[
\mathbf{R}_c^r \mathbf{G} = \mathbf{R}_c^r \mathbf{g}_0 + \mathbf{R}_c^r \mathbf{g}_1 + \cdots + \mathbf{R}_c^r \mathbf{g}_{N_0 N_1 - 1} . \tag{67}
\]
Similarly, \( (\mathbf{R}_a^r)^T \) is a \( N_r \times N_r \) matrix that can be obtained such that \( \text{trace}(\mathbf{G} \mathbf{R}_a^r \mathbf{g}) = \mathbf{g}^T \mathbf{R}_a^r \mathbf{g} \).

To obtain the solutions subject to different constraints, we formulate a cost function using Lagrange multiplier method [23]:
\[
J(\mathbf{W}, \mathbf{g}) = \text{trace}(\mathbf{W}_c^T \mathbf{R}_c \mathbf{W} - 2 (\mathbf{W}_c^T \mathbf{R}_c)^T \mathbf{g} + \mathbf{g}^T (\mathbf{R}_a^r \mathbf{g}) - \lambda \mathbf{c}(\mathbf{g}) \tag{68}
\]
where \( \mathbf{c}(\mathbf{g}) \) represents the constraint function and \( \lambda \) is the Lagrange multiplier. Taking the gradient of \( J(\mathbf{W}, \mathbf{g}) \) with respect to \( \mathbf{g} \), we get
\[
\nabla_{\mathbf{g}} J(\mathbf{W}, \mathbf{g}) = -2[(\mathbf{R}_c^r)^T \mathbf{g}]^T \mathbf{W}_c^T \mathbf{R}_c \mathbf{W} + 2 (\mathbf{R}_a^r \mathbf{g}) - \lambda \nabla_{\mathbf{g}} \mathbf{c}(\mathbf{g}) . \tag{69}
\]
From (15) and (67), the optimum equalizer matrix can be expressed as
\[
\mathbf{W}_{\text{opt}} = \mathbf{R}_z^{-1} \left( \mathbf{R}_c^r \mathbf{g}_0 + \mathbf{R}_c^r \mathbf{g}_1 + \cdots + \mathbf{R}_c^r \mathbf{g}_{N_0 N_1 - 1} \right) . \tag{70}
\]
Then the transpose of the rearranged transposed optimum equalizer matrix can be written as
\[
(\mathbf{R}_c^r)^T = (\mathbf{R}_z^{-1} \mathbf{R}_a^r \mathbf{g}) \tag{71}
\]
where \( (\mathbf{R}_z^{-1} \mathbf{R}_a^r)^T \) can be obtained as
\[
(\mathbf{R}_c^r)^T = \begin{bmatrix}
\text{1st column of } \mathbf{R}_z^{-1} \mathbf{R}_a^r & 
\vdots & 
\text{N_r-th column of } \mathbf{R}_z^{-1} \mathbf{R}_a^r \\
\vdots & 
\ddots & 
\vdots \\
\text{N_r-th column of } \mathbf{R}_z^{-1} \mathbf{R}_a^r & 
\vdots & 
\text{1st column of } \mathbf{R}_z^{-1} \mathbf{R}_a^r \\
\end{bmatrix}^T \tag{72}
\]
Substituting (71) in (69), we get
\[
\nabla g^T f(W, g) = -2((R_{ca})^T (R_c^{-1} R_{ca})^T g + 2(R_{na})^T g - \lambda \nabla g^T (g).
\]
(73)

Thus, the crux of this novel method is appropriate rearrangement of the matrices. An example is given in [22] to illustrate the method for \(N_g = 3\) and \(N_r = 3\).

In view of the hexagonal bit-cells and the radially symmetric symbol response, we may choose a target that is also symmetric, i.e., we assume that all the values of the cross interference to the nearest spot are the same. Further, we assume the interference from beyond adjacent bits to be zero. Then, the target can be written as \((N_g = 3, N_r = 5)\)

\[
G_0 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix},
\]

\[
G_1 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix},
\]

\[
G_2 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix},
\]

and the 1-D form of the target matrix becomes \(g = [g_0 \ g_1]^T\), where \(g_0\) is value of the cross interference to the nearest spot, and \(g_1\) is value of the self interference of the central spot. Even though we are considering symmetric targets here for illustration, the target design approach presented here is applicable for any general target.

We design the targets for three different constraints. The first constraint used is \(g_0 = 1\), which we call “1-D monic constraint.” The constraint is widely used in 1-D systems and is known to have good noise-whitening properties. Secondly, we use the energy constraint. The rationale behind this constraint is as follows. For a single-bit error event \(I_e = [1 \ 0 \ \cdots \ 0]^T\) of length \(N_e = 1\), neglecting noise colorization and channel misequalization, the resulting minimum SNR is given by

\[
\text{SNR}_{\text{eff}} = \frac{\sum_{k} (G_k I_e)^T (G_k I_e)}{\sigma_{\text{min}}^2} = \frac{4g_0^2 + g_1^2}{\sigma_{\text{min}}^2}
\]

(75)

where \(\sigma_{\text{min}}^2\) is the MMSE at the equalizer output. The “energy constraint” mentioned above refers to setting \(4g_0^2 + g_1^2 = 1\) so that minimizing the MSE results in maximizing the SNR_{\text{eff}}. A similar argument for choosing a similar energy constraint in 1-D data storage system was made in [21]. The third constraint used is \(g_1 = 1\), which we call “2-D monic constraint.” The rationale behind this target constraint is as follows. Let the target response for a single spot in the spatial frequency domain be radially symmetric, i.e., \(G_{st}(\phi, \rho) = \hat{G}_{st}(\rho)\). Then for the noise to be 2-D white, \(\hat{G}_{st}(\rho)\) should be ideally of the form \(\mu |\hat{H}_{st}(\rho)|^2\), where \(\mu\) is some constant of proportionality. Of all the targets with the same magnitude characteristics of \(\hat{G}_{st}(\rho)\), the one with the energy optimally concentrated near \(g_1\) has the largest possible amplitude \(g_1\). Equivalently, for a constrained value \(g_1 = 1\), the target with the energy optimally concentrated near \(g_1\) will yield the smallest possible \(\mu\) and therefore the smallest possible noise enhancement. Since we put symmetric constraint that appears in (74), the target has its energy almost concentrated near \(g_1\). Consequently, making the target with the energy optimally concentrated near \(g_1\) will not burden the finite length equalizer much. So setting \(g_1 = 1\) is a reasonable target constraint.

For the 1-D monic constraint, the constraint function is \(c(g) = 2(i^T g - 1)\), where \(i = [1 \ 0]^T\). Then, the optimum target can be obtained as [using (73)]

\[
g = \lambda \left( (R_{na})^T \left( (R_c^{-1} R_{ca}) \right)^{-1} \right)^{-1} I
\]

(76)

where

\[
\lambda = \frac{1}{i^T (R_{na})^T - (R_{ca})^T (R_c^{-1} R_{ca})^{-1} i}.
\]

(77)

And it can be shown that MMSE is equal to the Lagrange multiplier \(\lambda\).

For the energy constraint, we have \(c(g) = g^T (R_{ca})^T g - 1\), where \((R_{ca})^T g = \sum_k (G_k I_e)^T (G_k I_e) g\). Using this in (73) and setting the gradients zero, we obtain

\[
\left\{(R_{ca})^{-1/2} (R_{ca})^{-1/2}\right\} \left\{(R_{ca})^{-1/2}\right\}^{-1} g = \lambda (R_{ca})^{-1/2} g
\]

(78)

where \((R_{ca})^{-1/2}\) is such that \((R_{ca})^{-1/2} (R_{ca})^{-1/2}\) is the matrix [(Eq.)] and \((R_{ca})^T g = (R_{ca})^{-1} (R_{ca})^T g\). It can be shown that the MMSE is given by the minimum eigenvalue of the matrix \((R_{ca})^{-1/2} (R_{ca})^{-1/2}\) and \((R_{ca})^{-1/2} g\) is the corresponding eigenvector.

For the 2-D monic constraint, the solutions are identical to the 1-D monic constraint case except for \(i = [0 \ 1]^T\).

VI. SIMULATION RESULTS

Computer simulations are carried out for the TwoDOS system shown in Fig. 2 with channel length \(N_b = 41\), equalizer length \(N_w = 3\), target length \(N_g = 3\), and \(N_r = 5\) tracks per group. The target matrix \(G\) is set up using the symmetric constraint, and the coefficients are given by (74). The normalized optical cutoff frequency is set to 1/√3 to avoid aliasing [5], [11]. The signal-to-noise ratio (SNR) is defined as

\[
\text{SNR} = 10 \log_{10} \left( \frac{\sum_{x,y} \hat{H}_{st}^2 (x,y)}{\sigma^2} \right)
\]

(79)

where \(\sum_{x,y} \hat{H}_{st}^2 (x,y)\) is the energy for a single spot with \(\hat{H}_{st}(x,y)\) being the symbol response in Cartesian coordinates, and \(\sigma^2\) is the variance of the additive white
Gaussian noise (AWGN) on each individual track (i.e., $E[|h_{ij}^2(z)|] = \sigma^2, i = 1, 2, \ldots, N_r$). In addition to AWGN, we also consider pit size variations for performance evaluation. We use four different PR targets for performance comparison: fixed PR target ($g_0 = 1, g_1 = 2$), 1-D monic constrained target ($g_0 = 1$), energy constrained target ($4g_0^2 + g_1^2 = 1$), and 2-D monic constrained target ($g_1 = 1$). BER is used as the performance measure. The equalizer and target are designed with the approaches described in Sections III and V for each SNR.

Fig. 3 shows the performance of 2-D VD for the four different targets. Observe that the 1-D monic constrained target performs worse than all the other targets when SNR is smaller than 33 dB, whereas the fixed PR [1, 2] target becomes the worst for SNR > 33 dB. At high SNRs, residual ISI dominates the MSE. Hence, jointly designing the target and equalizer, rather than fixed target approach, is preferable at higher SNRs since noise is not significant at these SNRs. For this reason, for SNR > 33 dB, the 1-D monic constrained target outperforms the fixed PR target. Fig. 3 also shows that the energy constraint is a reasonable target constraint. More importantly, the target with 2-D monic constraint performs best among the four targets at all the SNRs, thus indicating that concentrating the energy around $g_1$ is a good choice for symmetric targets [e.g., see (74)].

To further understand the performance trends observed in Fig. 3, we examine the noise correlation at the VD input for the different target designs, since the performance of VD is known to be influenced significantly by noise correlation. Moreover, the 2-D equalization could result in significant noise correlation. This problem becomes even more severe for shorter targets (e.g., $N_g = 3$). Furthermore, we know that the MMSE design approach focuses on minimizing the total noise power, and not noise correlation. The normalized noise correlation along the third track is shown in Table I for the four different targets at SNR = 32 dB and SNR = 34 dB. Similar trends in the correlation are observed for other tracks as well as in the cross-track direction. Of all the targets, we see that the 2-D monic constrained target results in least noise correlation at both SNRs, which is consistent with its best BER performance among the four targets. Further, comparing the noise correlation produced by the fixed PR target with the 1-D monic constrained target at SNR = 32 dB and SNR = 34 dB, we can roughly understand why the BER curves for these two targets cross each other around SNR = 33 dB in Fig. 3. Finally, the relative noise correlation obtained for the energy constraint target with respect to the other targets is consistent with the BER performance observed in Fig. 3.

Table II shows the values of $g_1$ (normalized by $g_0$) for the four targets at different SNRs. Comparing Table II with Fig. 3, it can be concluded that the target that has the smallest normalized $g_1$ results in the worst BER performance and vice versa. This observation reconfirms that the 2-D monic constraint is a reasonable target constraint since it causes the energy to be concentrated near $g_1$, thus resulting in the largest normalized $g_1$ compared with the other targets. The 1-D monic constraint can be considered as causing the target with the energy to be concentrated near $g_0$ and as a result it minimizes the normalized $g_1$. For this reason, the 1-D monic constraint is not a good target constraint for 2-D channels.

We did further investigations to see if the BER performance can be further improved (beyond that in Fig. 3) by removing the symmetry constraint on the target. This addresses the situa-

### Table I

<table>
<thead>
<tr>
<th>SNR</th>
<th>Spatial difference along the track</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>32dB</td>
<td>Noise correlation for fixed PR target</td>
<td>1</td>
<td>0.2811</td>
<td>-0.1390</td>
<td>0.0548</td>
<td>-0.0083</td>
</tr>
<tr>
<td></td>
<td>Noise correlation for 1D monic constrained target</td>
<td>1</td>
<td>0.2903</td>
<td>-0.1446</td>
<td>0.0492</td>
<td>-0.0038</td>
</tr>
<tr>
<td></td>
<td>Noise correlation for 2D monic constrained target</td>
<td>1</td>
<td>0.2320</td>
<td>-0.1059</td>
<td>0.0714</td>
<td>-0.0266</td>
</tr>
<tr>
<td></td>
<td>Noise correlation for energy constrained target</td>
<td>1</td>
<td>0.2607</td>
<td>-0.1259</td>
<td>0.0639</td>
<td>-0.0160</td>
</tr>
<tr>
<td>34dB</td>
<td>Noise correlation for fixed PR target</td>
<td>1</td>
<td>0.2595</td>
<td>-0.1365</td>
<td>0.0655</td>
<td>-0.0061</td>
</tr>
<tr>
<td></td>
<td>Noise correlation for 1D monic constrained target</td>
<td>1</td>
<td>0.2545</td>
<td>-0.1306</td>
<td>0.0684</td>
<td>-0.0100</td>
</tr>
<tr>
<td></td>
<td>Noise correlation for 2D monic constrained target</td>
<td>1</td>
<td>0.2094</td>
<td>-0.0811</td>
<td>0.0773</td>
<td>-0.0329</td>
</tr>
<tr>
<td></td>
<td>Noise correlation for energy constrained target</td>
<td>1</td>
<td>0.2312</td>
<td>-0.1043</td>
<td>0.0761</td>
<td>-0.0239</td>
</tr>
</tbody>
</table>

### Table II

<table>
<thead>
<tr>
<th>SNR</th>
<th>$g_1$ for Different Target Constraints</th>
<th>32</th>
<th>33</th>
<th>34</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>32dB</td>
<td>Fixed PR Target</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1D monic constrained target</td>
<td>1.9468</td>
<td>1.9901</td>
<td>2.0310</td>
<td>2.0696</td>
</tr>
<tr>
<td></td>
<td>2D monic constrained target</td>
<td>2.2784</td>
<td>2.2828</td>
<td>2.2901</td>
<td>2.2997</td>
</tr>
<tr>
<td></td>
<td>Energy constrained target</td>
<td>2.1150</td>
<td>2.1411</td>
<td>2.1668</td>
<td>2.1919</td>
</tr>
</tbody>
</table>

![Fig. 3. BER performance of 2-D VD for different target constraints.](image-url)
Fig. 4. BER performance for different target constraints with \(-3\%\) domain bloom.

Fig. 5. BER performance for different target constraints with \(3\%\) domain bloom.

Fig. 6. BER performance for different target constraints in the presence of random pit-size variation when SNR = 31 dB. The pit-size variation is normalized with respect to the bit duration \(T\).

Fig. 7. BER performance for different target constraints in the presence of random pit-size variation with \(\Delta_b = 0\) whose probability density function is Gaussian with zero mean and variance \(\sigma_T^2\), where \(\sigma_T\) is specified as a percentage of \(T\). In this case, the equalizer and target are designed using the modified correlation matrices given by (46)–(49). Observe that the 2-D monic constrained target still outperforms other targets. In conclusion, the 2-D monic constraint is a suitable target constraint in TwoDOS with or without pit-size variations.

VII. CONCLUSION

In this paper, Hankel transform is used to construct the symbol response of TwoDOS in the presence and absence of pit-size variations. Based on this channel, we introduce a novel method which transforms the 2-D target design problem from 2-D into 1-D form. We also propose a novel 2-D monic constraint for 2-D partial response targets. By evaluating the BER performance of 2-D Viterbi detector for targets designed using different constraints, the proposed 2-D monic constraint is shown to be a reasonable target constraint for TwoDOS systems. The noise correlation resulting from the different target constraints is examined to support the observed BER performances.
REFERENCES


Manuscript received May 27, 2004; revised May 13, 2005.