Serial Maximum a Posteriori Detection of Two-Dimensional Generalized Partial Response Target for Holographic Data Storage Systems

Thien An Nguyen and Jaejin Lee *

Abstract: Holographic data storage (HDS) is an emerging technology that promises to revolutionize digital data storage and access. Unlike traditional storage media such as hard drives and flash memory, HDS uses light to write and read page-oriented two-dimensional (2D) data from volume media. This allows for significantly higher densities and faster data transfer rates in HDS systems. However, 2D interference is a serious issue in HDS due to hologram dispersion during the reading process. Therefore, we present a novel detection algorithm based on maximum a posteriori (MAP) detection to mitigate 2D interference. In our proposed model, we inherited the structure of the serial generalized partial response target to design the serial structure for MAP detection. The simulation results show that the proposed model can achieve a high bit error rate performance.

Keywords: detection; holographic data storage; estimate channel; generalized partial response; maximum a posterior

1. Introduction

Holographic data storage (HDS) has emerged as a promising technology for achieving high-density data storage with short read and write times [1]. In an HDS system, the information is stored as interference patterns or holograms created by the intersection of two laser beams. These page-oriented two-dimensional (2D) holograms can be stored and retrieved using various media, including photorefractive crystals and polymers. A key advantage of HDS over traditional data storage technologies is its potential for high storage density [2]. HDS can achieve storage densities of hundreds of terabytes per cubic inch owing to its ability to store multiple holograms in a single volume of storage medium. In addition, HDS has the potential for fast read and write times, making it suitable for high-performance computing applications.

However, the development of HDS systems involves 2D inter-symbol interference (ISI), misalignments, and noise [3]. In addition, diffusion of the refractive index pattern over time can certainly have an impact on the detection of data in holographic data storage. When the refractive index pattern diffuses, it can cause the hologram to lose its sharpness and result in a decreased signal-to-noise ratio (SNR) in the detected signal. This can lead to errors in the data detection and potentially a loss of information. Researchers have attempted to overcome these problems in two ways: First, by enhancing the structure of the system. A pixel matching method was introduced by Shelby et al. [4]. Bernal et al. proposed an optimization of the aperture size [5]. Wang et al. introduced aperture and medium defocusing [6]. The second category is the postprocessing and coding methods. Modulation methods for converting data bits into pixels in a hologram were introduced in some studies [7–9]. A combination of modulation and error-correcting codes was proposed to avoid isolated patterns and reduce the effect of 2D ISI [10–13]. The isolated patterns are presented in Figure 1.
In signal processing, an equalizer can be used to mitigate ISI. However, using a 2D equalizer directly to recover the original data results in a low bit error rate (BER) performance. Therefore, a generalized partial response (GPR) target was used with a 2D equalizer to estimate the channel parameters. GPR is a type of equalization technique that helps to reduce inter-symbol interference (ISI) caused by the channel response. These parameters are then used to detect the received signal in the original signal while maintaining a high BER performance. Initially, Cideciyan introduced the partial response maximum likelihood (PRML) method to balance and remove ISI in digital magnetic recording [14], which is a technology used for storing digital data on magnetic tape or other magnetic storage devices. In the PRML method, the received signal is converted into the desired target signal, which includes the known interference, by using an equalizer. Then, the output of the equalizer is supplied to the detection to decide the original data. Nabavi and Kumar proposed a GPR target to improve the detection performance of the PRML [15]. Regularly, researchers have used the one-dimensional (1D) form for PR and GPR targets because calculating all Viterbi paths is very complicated when using the 2D form for PR and GPR targets. Therefore, many studies have used and modified PRML and GPR to design and improve 2D detection for storage systems [16–19]. Recently, Nguyen and Lee proposed a serial target for bit-patterned media recording (BPMR) systems based on the 2D GPR target and the serial structure of detection [20–22]. Serial detection is based on the serial form of the 2D target. After estimating the 2D target, it was decomposed into a series of two 1D targets. In the serial detection structure, the first detector is used to estimate the values of six possible target symbols. Then, the output of the first detector, which contains 1D interference, is supplied to the second detector to detect the 0/1 bits and recover the original data. In the serial detection, because the 2D target is analyzed into two 1D targets, we can easily use the conventional Viterbi algorithm (VA).

In addition, to use the 1D VA for the 2D target, we use the intertrack–interference (ITI)/ISI estimator to remove the interference from neighboring tracks. The ITI and ISI are the parts of 2D interference in BPMR systems, as in Figure 2.

| 1 1 1 | 0 0 0 |
| 1 0 1 | 0 1 0 |
| 1 1 1 | 0 0 0 |

Figure 1. Isolated patterns.

In addition, to use the 1D VA for the 2D target, we use the intertrack–interference (ITI)/ISI estimator to remove the interference from neighboring tracks. The ITI and ISI are the parts of 2D interference in BPMR systems, as in Figure 2.

Figure 2. ITI and ISI of 2D interference.

When removing ITI/ISI, the received signal rests the 1D interference on the main track. Therefore, we can apply the 1D VA on the main track to restore the original data. In [23,24], Nguyen and Lee proposed the ITI/ISI estimator for BPMR and HDS systems, respectively. The authors used the extended trellis to predict the ITI/ISI. The ITI/ISI estimator based on extended trellis achieved high BER performance for both BPMR and HDS systems. However, their model obtained very high complexity. To avoid using the extended trellis, Nguyen and Lee introduced a multilayered 2D GPR target [25]. In [25], the authors used...
three equalizers and targets. The first target was designed following the 1D form. With a 1D target, we can use the 1D VA for detection. In the second and third targets, a 2D form is exploited. Then, the output from 1D VA detection was used to predict ITI/ISI in the second and third targets.

Serial targets and detection methods have also been applied to the HDS system by Nguyen and Lee [26]. As mentioned above, the serial structure in their study used the VA to detect the received signal. Therefore, to develop serial detection, in the present study, we propose a new detection algorithm based on the maximum a posteriori (MAP) algorithm to improve the BER performance in HDS systems. After distortion by the 2D interference in the HDS system, the received signal is converted by an equalizer into the desired target signal. The 2D target is analyzed into two serial targets: a 1D horizontal target and a 1D vertical target. The parameters of the horizontal and vertical targets are supplied to the first and second detectors, respectively. Based on the output of the equalizer, symbol information is estimated by the first detector. Dissimilar to the VA, which creates a hard output signal, the MAP algorithm creates a soft output signal. Therefore, the MAP algorithm can improve the BER performance of serial detection. Then, these values were supplied for the second detection step to restore the original bit. In the second detector, the algorithm is then modified to extract information from these symbols and supply it to the trellis. Symbol information is detected in the bit data during the second detector, and the original signal (bit information) is recovered. The bit information is also used to create a priori information about the symbol (probability of the symbols) and feedback for the first detection. This helps in creating an iterative algorithm for our proposed detection. The simulation results show that the proposed algorithm outperforms existing detection methods and significantly improves the BER performance of HDS systems.

The remainder of this paper is organized as follows: In Section 2, we provide a brief review of the literature on MAP detection. Section 3 presents the proposed model using a MAP-based detection algorithm. Section 4 presents the simulation results and a performance evaluation of the proposed algorithm. Finally, in Section 5, we provide concluding remarks and discuss the implications of our study for the field of HDS.

2. MAP Detection

After applying the method used by Nguyen and Lee [26], we obtain the following target form:

$$G = \begin{bmatrix} r_p & l_2 & r_p \\ r_l & l_2 & r_l \\ r_p & l_2 & r_p \end{bmatrix} = \begin{bmatrix} p \\ l_1 \\ r \end{bmatrix},$$ (1)

where \( [p, l_1, p]^T \) is the vertical interference (VI) or intertrack interference (ITI) and \([r, l_2, r] \) is the horizontal interference (HI) or ISI. To remove the VI and HI, we used outer and inner detectors, as shown in Figure 3.

![Figure 3. Serial MAP detection.](image)

(1) **Horizontal detection**

Signal \( b[j,k] \) estimates the results of \( a[j,k] \otimes [p, l_1, p]^T \). We can observe that, because of the modulated signal \( a[j,k] \in \{-1, 1\} \), signal \( b[j,k] \in \{-2p - l_1; -l_1; 2p - l_1; -2p + l_1; l_1; 2p + l_1\} \). Six values of \( b[j,k] \) are assigned as six symbols \( [R_1 = -2p - l_1; R_2 = -l_1; R_3 = 2p - l_1; \)
For horizontal detection using the MAP algorithm, we calculate the probability of \( P(R_i \mid z) \) as follows:

\[
P(R_i \mid z) = \sum_{R_{(i-1)}} P(s', s \mid z) = \sum_{R_{(i-1)}} P(s', z_{(i-1)}, s \mid z) P(z_{(i-1)} \mid s)
\]

\[
= \sum_{R_{(i-1)}} P(s', z_{(i-1)}) P(z_{(i-1)} \mid s) P(z_{(i-1)} \mid s) P(R_i \mid z)
\]

(2)

where \( s', s, \) and \( z \) are the previous state, current state, and received symbol of MAP detection in the proposed model, respectively. \( P(s', z_{(i-1)}) \) is the joint probability of the previously stated sequence received in the past; \( P(z_{(i-1)} \mid s) \) represents the conditional probability of the current state \( s \) and the current received symbol \( z_{(i-1)} \); \( P(z_{(i-1)} \mid s) \) is the conditional probability of the future received sequence \( z_{(i-1)} \) under the current state \( s \). We assigned \( \alpha_{(i-1)}(s') = P(s', z_{(i-1)}), \gamma_{(i)}(s', s) = P(z_{(i-1)}, s \mid s') \), and \( \beta_{(i)}(s) = P(z_{(i-1)} \mid s) \) for simplification. The branch transition probability of the horizontal trellis \( \gamma_{(i)}(s', s) \) is estimated using the following equation.

\[
\gamma_{(i)}(s', s) = P(z \mid Z_i) P(R_i) = \exp \left( \frac{-(z - Z_i)^2}{2\sigma^2} \right) P_e(R_i)
\]

(3)

where \( z \) is the value of \( z[j,k] \), \( Z_i \) is the discrete value of \( z \) in the noiseless case (216 symbols on the trellis), \( P(z \mid Z_i) \) is expressed by a Gaussian function, and \( P_e(R_i) = P(R_i) \) is the extrinsic information from the vertical detection, which helps create the feedback path. In the forward path, we set \( P_e(R_i) = 1/6 \).

(2) Vertical detection

For vertical detection, we estimate the probability of \( \{-1, 1\} \) in signal \( \alpha[j,k] \). In addition, from the probability of \( \{-1, 1\} \) in \( \alpha[j,k] \), we must determine the probability of \( \{R_i; i = 1, 2, \ldots, 6\} \) in signal \( \beta[j,k] \). Firstly, the conditional probability \( P(R_i \mid z) \) is received from the horizontal detection; therefore, the branch transition probability \( \gamma_{(i)}(s', s) \) in the vertical detection must be modified as follows:

\[
\gamma_{(i)}(s', s) = \frac{P(R_i \mid z)}{\sum_i P(R_i \mid z)}
\]

(4)

The probabilities of \( \{-1, 1\} \) and \( \{R_i; i = 1, 2, \ldots, 6\} \) are calculated as follows:

\[
P(\pm 1) = P(a = \pm 1 \mid R) = \sum_{a = \pm 1} P(s', s, R)
\]

(5)

\[
P_e(R_i) = P(b = R_i \mid R) = \sum_{b = R_i} P(s', s, R)
\]

(6)

where \( a \) is the value of the signal \( \alpha[j,k] \); \( b \) is the value of the signal \( \beta[j,k] \); \( R \) is the set of 6 symbols \( R_i \) with \( i = 1, 2, 3, \ldots, 6 \).

Considering the example in Figure 4, the summation \( P(s', s, R) \) of the dotted lines is \( P(a = -1 \mid R) \). The summation \( P(s', s, R) \) of lines 1/R5 (between states \(-1, 1\) and \(-1, -1\)) and 1/R5 (between states \(-1, -1\) and \(-1, 1\)) is \( P(b = R_5 \mid R) \).
Figure 4. Trellis for the horizontal detection.

(3) Feedback line

We can use \( P_e(R_i) \) from the output of the vertical detection to supply horizontal detection as extrinsic information. Using the feedback line, iterative detection can be implemented as follows:

- **Initialization**: We assign \( P_e(R_i) = 1 \) with \( i = 1, 2, 3, \ldots, 6 \) in the horizontal detection.
- **Step 1**: Calculate \( P(R_i | z) \) using (2) for horizontal detection.
- **Step 2**: Estimate \( P_e(R_i) \) using (4) and (6) for vertical detection.
- **Step 3**: If achieving the stopping condition, we calculate \( P_e(\pm 1) \) using (5) in the vertical detection.

3. Proposed Model

Figure 5 illustrates the diagram of the proposed model.

In a holographic data storage system, data are first converted into a pattern of light and dark areas using a spatial light modulator (SLM). A laser beam is split into two: a reference beam and a data beam. The data beam is modulated by the pattern created by the SLM. The reference beam and modulated data beam intersect at a photosensitive material, such as a crystal or photopolymer, creating an interference pattern that contains the data. This interference pattern is recorded in the material, which changes its refractive index, creating a 3D holographic image of the data. To read the data, a laser beam is shone onto the material at the same angle as the reference beam used to record the data. This creates an identical reference beam that interferes with the holographic image, reconstructing
the data. The reconstructed data can then be captured by a detector, such as a camera or photodiode array.

In the model, an SLM was used to modulate the user data \( u[k] \) into the data page \( a[j,k] \). The signal and reference beams were then combined to create a hologram. Data were stored in the medium as holograms. The readback signal \( y[j,k] \) was affected by 2D interference and noise as below [3,26]:

\[
y[j,k] = a[j,k] \otimes h[j,k] + w[j,k],
\]

where \( \otimes \) is the 2D convolution operator and \( w[j,k] \) is the additive white Gaussian noise (AWGN) with zero mean and variance \( \sigma^2 \). We used the range of the discrete point spread function (PSF), \( h[j,k] \) with \( 5 \times 5 \) array pixels. The discrete PSF \( h[j,k] \) can be calculated as follows:

\[
h[j,k] = \int_{j-\Delta/2}^{j+\Delta/2} \int_{k-\Delta/2}^{k+\Delta/2} h(q,t) dq dt,
\]

where \( \alpha \) denotes the linear fill factor. We choose \( \alpha = 1 \), same as that selected by Kim and Lee [16]. \( h(q,t) \) is the continuous PSF, calculated as follows:

\[
h(q,t) = \frac{1}{\sigma_b^2} \sin^2 \left( \frac{q - m_q}{\sigma_b}, \frac{t - m_t}{\sigma_b} \right),
\]

where \( \sin c(q,t) = \left( \sin(\pi q) / \pi q \right) \left( \sin(\pi t) / \pi t \right) \), \( \sigma_b \) is the blur grade, and \( m_q \) and \( m_t \) are the misalignments in the horizontal and vertical directions, respectively. We define SNR as follows:

\[
\text{SNR} = 10 \log_{10} \left( \frac{1}{\sigma^2} \right)
\]

4. Simulation Results and Discussions

In the simulation, we used hardware with CPU i5-11400 2.6 GHz, 16 GB RAM, and a GTX 1650 graphic card. For the software, we created the program with MATLAB 2022b. To estimate the parameters of the serial target \( G \) and equalizer \( F \), we input a data page with a size of 1200 \( \times \) 1200 into the system and collected the output. The coefficients of the target were calculated and applied to detection using the method described by Nguyen and Lee [26]. We input ten data pages to the model in Figure 5 to estimate the BER performance of the system. To be simpler and shorter, we did not draw modulation and demodulation in Figure 5. In modulation, we arranged the 1D user data \( u[k] \) into the 2D modulated signal \( a[j,k] \). We can see that the signal \( a[j,k] \) represents the light and dark pixels in SLM. Here, to fit the MAP algorithm, we use \(-1 \) and \( 1 \) to present light and dark pixels, respectively. Then, the signal \( a[j,k] \) was inputted into the HDS channel. In the HDS channel, the coefficients of \( h[j,k] \) was created, as in Section 3. The output of HDS channel \( y[j,k] \) was obtained, as in (7). The signal \( z[j,k] \), which is close to the desired signal \( d[j,k] \), was created by the convolution of \( y[j,k] \) and the equalizer matrix \( F \) as below.

\[
z[j,k] = y[j,k] \otimes F,
\]

The signal \( z[j,k] \) was inputted to the serial detection in Section 2 to restore the signal \( \hat{a}[j,k] \), which is the estimation of \( a[j,k] \). Finally, we recovered the user data \( \hat{u}[k] \) by arranging the signal \( \hat{a}[j,k] \). We collected and compared the signal \( \hat{u}[k] \) with \( u[k] \) for calculating the BER performance.

In Table 1, we present the results of the estimation of the matrix target \( G \) and equalizer \( F \). The matrix \( F \) helps convert the received signal \( y[j,k] \) into the signal \( z[j,k] \), which is similar
to the desired signal $d[j,k]$. Therefore, we varied SNR from 10 to 16 dB and compared the signals $z[j,k]$ and $d[j,k]$. We used mean square error as below.

$$MSE = E[(z[j,k] - d[j,k])^2], \quad (12)$$

where $E$ denotes the expectation.

<table>
<thead>
<tr>
<th>SNR</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.4036</td>
<td>0.39</td>
<td>0.3776</td>
<td>0.3677</td>
<td>0.3580</td>
<td>0.3497</td>
<td>0.3425</td>
</tr>
</tbody>
</table>

From Table 1, we can see that when the SNR increases, the MSE decreases, and the accuracy of the estimation is improved. The estimator was designed to predict the 2D interference in the channel. However, the AWGN disturbs the estimator too much to exactly predict the 2D interference. Therefore, the equalizer $F$ just converts the signal $z[j,k]$ into an approximation of the desired signal $d[j,k]$.

In the first experiment, the signal $z[j,k]$ was inputted to the horizontal detection to extract the probability of the symbols $b[j,k]$. Then, the signal $b[j,k]$ is supplied to the vertical detection to estimate the probability of bits 0 and 1 in $d[j,k]$. We did not use the feedback in this experiment, and the signal $d[j,k]$ is restored into the user data $u[k]$. We fixed the blur $\sigma_b$ at 1.85 and 0% misalignment ($m_q = m_t = 0$) during the first experiment.

As shown in Figure 6, our proposed model can improve the BER performance compared with previous studies. We implemented the GPR target and 1D detection for the 1D GPR target reported by Nabavi and Kumar [15]. Serial detection using Viterbi is a serial target detection method [26]. Serial MAP detection achieved a gain of 3.5 dB at a BER of $10^{-3}$. Because the soft output signal was created in the inner part of the serial structure when applying the MAP algorithm, the proposed model can improve the BER performance compared with the VA.

![Figure 6. BER performance of the proposed model without feedback.](image-url)
In the subsequent experiment, we used feedback to create an iterative algorithm. The model is similar to the model in the first experiment. However, in vertical detection, the signal $b[j,k]$ was exploited to calculate $P_e(R_i)$ in (6). Then, $P_e(R_i)$ was supplied to the horizontal detection, as in (3). To optimize the number of iterations, we implemented the experiments and show the results in Figure 7. We fixed the blur $\sigma_b$ at 1.85 and 0% misalignment ($m_q = m_t = 0$) during the second experiment.

![Figure 7. BER performance of the proposed model with the iterative algorithm.](image)

In Figure 7, we assume that the inter-track interference (ITI) is known when we can completely remove the ITI and remind the ISI. We can detect the 1D Viterbi with the ISI to achieve the ideal case with the best performance. By applying feedback, the proposed model can be improved to the ideal case and optimized after two iterations. In addition, when applying the feedback line, a gain of 1 dB can be achieved at a BER of $10^{-3}$ compared with the proposed model without feedback. In the serial structure, the 2D interference can be analyzed into two 1D interferences. Therefore, the horizontal detector (the first detector) lacks vertical interference information and vice versa. When the symbol information from the vertical detector (the second detector) is feedbacked to the first detector, the vertical interference is supplied to the first detector. This helps the first detector have information about both the horizontal and vertical interferences. Therefore, the feedback line can improve the BER performance of the system. However, in the second iteration, because the feedback information did not change, the performance was not improved compared with the first iteration.

In this study, we investigated the effects of blurring. When the level of blur increases, the desired pixel gets a significant 2D interference from neighboring pixels. We varied the blur $\sigma_b$ from 1.8 to 3 with SNR = 13 dB. The results are shown in Figure 8. When the blur increases to three, our proposed model achieves the same serial detection performance as Viterbi. Therefore, the proposed model can resist the blur effect of less than 3.
Figure 8. BER performance of the proposed model according to blur. As shown in Figure 10, the proposed model achieved a gain of 3 dB at a BER of $10^{-2}$.

We set a 0% misalignment for the HDS channel in the above experiments. In the next experiment, we added a 10% misalignment ($m_q = m_t = 0.1$) to the HDS channel to close the reality. This effect caused detection to miss the position of the data. Therefore, the misalignment effect decreased the detection ability. We set SNR = 13 dB, added a 10% misalignment, and varied the blur from 1.8 to 3, as shown in Figure 9. When misalignment appears, the GPR target can estimate the level of misalignment and compensate for that difference with the equalizer. The output of the equalizer was approximately the desired signal and only left the 2D interference, which is suitable for our proposed model to detect the signal. However, when the level of misalignment was too large, the GPR target could not estimate the misalignment which led to the proposed model not working.

Figure 9. BER performance of the proposed model according to blur with 10% misalignment.

In addition, we set blur $\sigma_b = 1.85$, 10% misalignment, and varied the SNR from 10 to 16, as shown in Figure 10. Our proposed model achieved the best performance in blurred and misaligned environments. As shown in Figure 10, the proposed model achieved a gain of 3 dB at a BER of $10^{-2}$.
Finally, we show the complexity of our proposed model and that of previous studies in Table 2. The operators per the detected bit are counted from the equalizer to the detection in each method. In Table 2, the operators of our proposed model are counted without iteration. We can see the trade-off between the BER performance and the complexity of the models.

Table 2. Complexity of the proposed model and previous studies.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Mul/Div</th>
<th>Add/Sub</th>
<th>Log/Exp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our model</td>
<td>1201</td>
<td>733</td>
<td>336</td>
</tr>
<tr>
<td>Serial detection in [26]</td>
<td>252</td>
<td>700</td>
<td>0</td>
</tr>
<tr>
<td>1D GPR [15]</td>
<td>33</td>
<td>48</td>
<td>0</td>
</tr>
</tbody>
</table>

5. Conclusions

In this study, we proposed a method for improving serial detection using the MAP algorithm. First, the 2D interference from the HDS channel was estimated using the 2D GPR target used by Nguyen and Lee [26]. These interferences are then supplied to the serial MAP detection. The first detection (horizontal detection) estimates the conditional probabilities of the six symbols. These probabilities are then input to the second detection (vertical detection) to exploit the probabilities of bits 0 or 1. Simultaneously, vertical detection extracts extrinsic information to provide feedback on horizontal detection. Based on this feedback, we designed an iterative algorithm for serial MAP detection. The results showed that the MAP algorithm improved the gain by 3.5 dB at a BER of $10^{-3}$. Furthermore, the iterative algorithm improved performance by up to 1 dB over the proposed model without the feedback line. The proposed model can withstand up to 3% of the blur effect. Finally, in the case of misalignment, the proposed model achieved the best performance.

Author Contributions: Conceptualization, T.A.N. and J.L.; methodology, T.A.N. and J.L.; software, T.A.N.; validation, T.A.N. and J.L.; formal analysis, T.A.N.; investigation, T.A.N. and J.L.; writing—original draft preparation, T.A.N.; writing—review and editing, T.A.N. and J.L.; supervision, J.L.; project administration, J.L.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government. (MSIT) (2021R1A2C1011154).

Institutional Review Board Statement: Not applicable.