Article

Soft-Output Detector Using Multi-Layer Perceptron for Bit-Patterned Media Recording

Seongkwon Jeong and Jaejin Lee *

School of Electronic Engineering, Soongsil University, 369 Sangdo-ro, Dongjak-gu, Seoul 06978, Korea; skwon917@gmail.com

* Correspondence: zlee@ssu.ac.kr; Tel.: +82-2-820-0901

Abstract: As conventional data storage systems are faced with critical problems such as the superparamagnetic limit, bit-patterned media recording (BPMR) has received significant attention as a promising next-generation magnetic data storage system. However, the reduced spacing between islands at increased areal density causes severe intersymbol and intertrack interference, which degrade BPMR system performance. In this study, we introduce a soft-output detector using multi-layer perceptron to predict reliable information. A received signal is equalized and detected by the MLP detector. The MLP detector provides a well-estimated value by using the binary-cross entropy function as a loss function and the identity function as an activation function for the output layer of the MLP detector. This study investigates the received probability distributions out of the detectors and compares the performance of various versions against a conventional detector. Compared with the conventional detection, the proposed MLP detectors provide a small variance and better BER performance than the conventional detection. Simulations of MLP designs show an advantage over conventional detection. Moreover, the proposed MLP detectors with the demodulator exhibit better BER performance than the conventional detector with the demodulator.

Keywords: bit-patterned media recording; multilayer perceptron; signal detection; soft-output detector

Citation: Jeong, S.; Lee, J. Soft-Output Detector Using Multi-Layer Perceptron for Bit-Patterned Media Recording. *Appl. Sci.* **2022**, *12*, 620. <https://doi.org/10.3390/app12020620>

Academic Editor: Amerigo Capria

Received: 7 December 2021

Accepted: 7 January 2022

Published: 10 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

To match data growth, the capacity of hard disk drives (HDDs) is required to increase. However, shrinking the magnetic grains further and cramming them closer to increase the capacity of the HDD results in the superparamagnetic effects [1]. To overcome these problems and increase areal density (AD) beyond one terabit per square inch (Tb/in²), bit-patterned media recording (BPMR) is considered as one of the promising future magnetic storage technologies [2]. Each data bit is recorded on a nano-sized grain or island that is arranged separately. Thus, BPMR can offer advantages such as extending AD to 4 Tb/in², simplified tracking, and reduced track-edge noise, transition noise, and nonlinear bit shift [3].

To achieve a high AD in BPMR, the bit period and track pitch in the down and cross-track direction, respectively, must be reduced. However, the decrease in distances between the islands encourages both intersymbol interference (ISI) and intertrack interference (ITI), which disturb signal detection and degrade the BPMR system performance [4]. To alleviate the effects of ISI and ITI, a two-dimensional (2D) equalizer with a one-dimensional (1D) generalized partial response (GPR) target based on a minimum mean square error (MMSE) criterion for a BPMR system has been introduced [5,6]. Additionally, as ITI is generally more severe than ISI in a BPMR system, modulation coding schemes have been proposed to prevent the occurrence of cross-track patterns such as $[-1, +1, -1]^T$ and $[+1, -1, +1]^T$ [7,8].

In recent years, neural networks (NNs) have been used for equalization, signal detection, and modulation coding for data storage systems. Sayyafan et al. proposed a turbo decoder scheme for a 1D HDD, which consists of a Bahl–Cocke–Jelinek–Raviv (BCJR) detector and a deep neural network (DNN) media noise predictor [9]. The proposed BCJR–DNN

turbo detector exhibited better performance than a pattern-dependent noise prediction detector. In [10], a detection scheme with a track misregistration (TMR) estimator based on multi-layer perceptrons (MLPs) was proposed for BPMP systems. Because the estimated TMR by the MLP-based TMR estimator helps the MLP-based data detector to detect the received signal, the proposed scheme exhibits a better bit error ratio (BER) performance than a conventional partial response maximum likelihood (PRML) detector. For a shingled magnetic recording system, a log-likelihood ratio (LLR) modulator based on an NN was developed to improve the reliability of the decoder output and recalculate the LLR value [11]. The LLR modulator, which was designed to be used in conjunction with a low-density parity check code, improves iterative decoding performance. In [12], a convolutional NN (CNN) modulation decoding scheme was investigated for holographic data storage systems. It was shown that the CNN-based demodulator decreases the number of demodulation errors compared to a hard decision method.

Herein, we introduce a soft-output detector using multi-layer perceptron for a BPMP system to predict reliable information. For the design of the MLP detector, an identity (linear) function and binary cross-entropy are used as an activation function in the output layer and a loss function, respectively, to predict information reliability similar to the soft decision value. This study investigates the received probability distributions out of the detectors and compares the performance of various versions against a conventional detector.

The rest of the paper is organized as follows. In Section 2, we introduce the BPMP channel model. Section 3 explains the design of a soft output MLP detector. Section 4 presents and discuss the simulation results. Finally, Section 5 concludes the paper.

2. BPMP Channel Model

Figure 1 shows the BPMP system with soft output MLP detector. In this channel, to fit numerical pulse without media noise, we use a 2D Gaussian pulse response which is considered as a square island length of 11 nm, a thickness of 10 nm, an element thickness of a magneto-resistive (MR) read-head of 4 nm, an element width of the MR read-head of 15 nm, a gap distance of 6 nm, and a fly height of 10 nm [13]. The 2D Gaussian pulse response is given as

$$H(z, x) = A \cdot \exp \left\{ -\frac{1}{2c^2} \left[\left(\frac{z}{PW_z} \right)^2 + \left(\frac{x}{PW_x} \right)^2 \right] \right\} \quad (1)$$

where A is the normalized peak amplitude, the constant c is the relationship between the standard deviation of a Gaussian function and PW_{50} , z and x are the cross- and down-track directions, and PW_z and PW_x are the PW_{50} of the cross- and down-track pulses, respectively. Here, the parameters related to the 2D Gaussian pulse response are as follows: $A = 1$, $c = 1/2.3548$, $PW_z = 24.8$ nm, $PW_x = 19.4$ nm. The 2D discrete time readback signal is given by

$$r_{j,k} = \sum_{m=-M}^M \sum_{n=-N}^N a_{j-m,k-n} \cdot h_{m,n} + n_{j,k} \quad (2)$$

where M and N are the length of the interference from neighboring islands in the cross- and down-track directions, respectively, $a_{j,k} \in \{-1, 1\}$ is the k -th recorded user data bit of the j -th track, $h_{m,n}$ is the channel response coefficient, and $n_{j,k}$ is the electronic noise modeled as an additive white Gaussian noise with zero mean and variance σ^2 . We set $M = 1$ and $N = 1$. The channel response coefficient $h_{m,n}$ can be written as

$$h_{m,n} = H(m \cdot T_z + \Delta_{TMR}, n \cdot T_x) \quad (3)$$

where T_z is the track pitch, T_x is the bit length, and Δ_{TMR} is a read-head offset. We set $T_z = T_x = 18$ nm at 2.0 Tb/in² and $T_z = T_x = 14.5$ nm at 3.0 Tb/in², respectively, and $\Delta_{TMR} = (TMR_z \times T_z)/100$, where TMR_z is the percentage of the TMR. When the read-head

is not able to continue at the center of the main data track, the system performance is degraded due to TMR.

Figure 1. Block diagram of BPMR system with the soft output MLP detector.

3. Proposed Soft-Output Detector Using Multi-Layer Perceptron

In this section, we describe a design of a soft-output MLP detector. The artificial NN (ANN) has been receiving increased attention as a signal processing tool in communication problems. The MLP, which is a class of ANN, can be applied to solve problems such as two-class classification, multi-class classification, and regression. For regression, the identity function and mean square error are mainly exploited as activation functions in the output layer and loss function to predict a numerical target value. As the input data $\{-1, 1\}$ are used as target values in BPMR, the prediction value of regression is approximately $\{-1, 1\}$, as shown in Figure 2a. Moreover, as the sigmoid function and binary cross-entropy (or categorical cross-entropy) are used as activation functions in the output layer and loss function to solve the binary-class (or multiple-class) classification problem (in which the goal is to predict a discrete label of input data), the classifier that implements binary-class classification classifies input into two classes such as $\{0, 1\}$, as shown in Figure 2b. Thus, as the prediction values are similar to the hard decision value, it is difficult to achieve some decoding gain when conducting error correction coding and modulation coding of the data.

Figure 2. The prediction output value: (a) regression MLP; (b) classification MLP.

To predict reliable information similar to the soft decision value and provide it to the decoder, we design the MLP detector. Figure 3 shows the configuration of the MLP detector, which consists of an input layer, hidden layers, and an output layer. The input layer consists of $(2M_m + 1) \times (2N_m + 1)$ neurons. For each hidden layer, all neurons include a rectified linear unit (ReLU) as an activation function. The output layer has one neuron with an identity function and the binary cross-entropy is used as the loss function to estimate reliable value. For the MLP detector, we use the user data $a_{j,k}$ as the target value and the readback signal $r_{j,k}$ as the input sequence of the MLP detector is given as follows:

$$\mathbf{r}_{j,k} = \begin{bmatrix} r_{j+M_m,k+N_m} & r_{j+M_m,k+N_m-1} & \cdots & r_{j,k} & \cdots & r_{j-M_m,k-(N_m-1)} & r_{j-M_m,k-N_m} \end{bmatrix}^T. \quad (4)$$

Figure 3. Configuration of the MLP detector.

We set $M_m = 1$ and $N_m = 1$ for the MLP detector input sequence. We use batch normalization (BN), which can deal with gradient problems, normalize the input, and scale the result [14]. The BN can be added before or after the activation function. We add the technique after the activation function of all layers. For the initialization of the weights, we use a He initialization, which initializes the weights of all the layers and mitigates gradient problems [15]. Furthermore, to increase the accuracy, we use an ensemble technique, which provides a new prediction by aggregating the predictions from all predictors. The ensemble will help us achieve better results than the best individual predictor by aggregating the predictions of a group of predictors. Adaptive moment estimation (Adam) as an optimizer is used to update the weights of the MLP network and boost the training speed [16]. The MLP detector is designed by Keras, which helps to develop deep-learning models [17].

4. Discussion

To compare the BER performance depending on the detection schemes, we utilize a conventional detection technique based on PRML. A PRML detector employed in data storage systems consists of an equalizer and a channel decoder to mitigate ITI and ISI. In this simulation, we use a 2D equalizer and a 1D GPR target based on an MMSE criterion [5,6]. The size of the 2D equalizer is 3×11 and the length of the 1D GPR target is 3. After the readback signal is processed by the 2D MMSE equalizer, the equalizer output is fed to a four-state soft-output Viterbi algorithm (SOVA). In this simulation, we assume the regular BPMP channel model without media noise. The signal-to-noise ratio (SNR) is defined as $10\log_{10}(1/\sigma^2)$.

Figure 4 shows the BER performance comparison of the models at TMR of 0% and 30%. In Figure 4, each MLP detector is different only in the number of hidden layers and neurons per hidden layer, and random initialization by the He initialization, but has the same structure such as the activation functions, the locations of batch normalization, and so on. To compare the performance for the different number of hidden layers and neurons, two MLPs are designed. One has one hidden layer with sixteen neurons (blue solid line with triangle marker). The other has two hidden layers with 64 neurons per hidden layer (blue dotted line with triangle marker). Additionally, to investigate the performance of exploiting ensemble techniques, we apply the ensemble of five MLP detectors with 1 hidden layer and 16 neurons in each MLP detector (red solid line with circle marker) and the ensemble of 5 MLP detectors with 2 hidden layers and 64 neurons per hidden layer in each MLP detector (red dotted line with circle marker). Among various learning methods of the ensemble learning, the average is used in this simulation. Thus, the estimated values from five MLP detectors is shown in Figure 5 are averaged.

Figure 4. BER performance comparison of the models: (a) TMR of 0%; (b) TMR of 30%.

Figure 5. Block diagram of BPMR system with ensemble of MLP detectors.

As shown in Figure 4a, when SNR = 16 dB and TMR = 0%, the performances of the MLP (16) and (64, 64) detector alone (blue lines with triangle marker) are worse than that of the PRML detector. Furthermore, it shows unstable performance due to the overfitting problem. However, the ensemble schemes of MLP detectors (red lines with circle marker) show better BER performances compared to the PRML detector and provide stable BER performances compared to MLP alone. Additionally, the ensemble of MLP (64, 64) detectors performs better by approximately 0.3 dB compared to the ensemble of MLP (16) detectors at the BER of 10^{-4} . Thus, using the ensemble technique can provide more stable performance than using the MLP alone. As shown in Figure 4b, when TMR = 30%, the MLP detector (16) and (64, 64) alone show the inconsistent performance. However, the BER curves of the ensemble schemes of MLP (16) and (64, 64) detectors exhibit better performance than the MLP detector (16) and (64, 64) alone and the PRML detector. Again, the ensemble schemes of MLP detectors provide a consistent BER performance compared to the schemes of MLP detector alone.

We investigate the probability density function of the output of the PRML detector, MLP (64, 64) detector alone, and an ensemble of the MLP (64, 64) detector at 3 Tb/in² and SNR = 14 dB, as shown in Figure 6. The ensemble of MLP detectors can reduce the overlapping area more than the PRML detector and MLP detector alone. Table 1 presents the mean and variance of the models. Compared with the PRML detector and MLP detector alone, the ensemble of MLP detectors provides a small variance. Thus, it is clear that the ensemble of MLP detectors provides more reliable information.

Figure 6. Probability density function of the output of PRML, MLP detector alone, and ensemble of MLP detectors at SNR of 14 dB.

Table 1. Mean and variance of the models.

	Mean		Variance	
	−1 bit	+1 bit	−1 bit	+1 bit
PRML	−3.59	3.59	2.07	2.07
MLP detector alone	−3.52	3.38	1.24	1.32
Ensemble of MLP detectors	−3.37	3.26	1.18	1.20

Figure 7 shows the BER performance comparisons according to the models using a 4/6 modulation coding scheme. For pair comparison, the user density, which is equal to $AD \times$ code rate (R), should be considered. Because we investigate the performance using a 4/6 ($=R$) modulation code at an AD of 3 Tb/in^2 [18], the PRML detector alone (uncoded system) must be tested at an AD of 2 Tb/in^2 . All models using modulation code at 3 Tb/in^2 exhibit better BERs performance after decoding than the PRML at 2 Tb/in^2 . At the BER of 10^{-4} and TMR of 0%, ensemble of the MLP (64, 64) detectors with the demodulator performs better by approximately 1.3 and 1.7 dB, compared with the PRML with the demodulator and the PRML detector alone, respectively. Even with increasing TMR, the ensemble of the MLP (64, 64) detectors with the demodulator exhibit better BER performance.

Figure 8 presents the BER performance of detectors and demodulators according to MLP detectors. All of the MLP detectors consist of an input layer, 2 hidden layers, and an output layer and have 64 neurons in each hidden layer. Firstly, we have compared the performances of the MLP detectors (dotted lines). When the MLP detectors are used alone, the performances of the regression and proposed MLP detector are better than that of the classification MLP detector. The performance of the regression MLP detector is similar to that of the proposed MLP detector. Secondly, we have verified the performance of a Euclidean distance demodulator depending on MLP detectors (solid lines). The classification MLP detector with a demodulator shows the worst performance because the prediction values are similar to the hard decision value $\{0, 1\}$ and it is difficult to achieve some decoding gain when conducting error correction coding and modulation coding of the data. The decoding performance of the proposed MLP detector with demodulator gives the best performance since the proposed MLP detector provides more reliable information to the demodulator. However, to use the proposed MLP detector, the trade-off between complexity and BER performance should be considered in the training process.

Figure 7. BER performance comparison of the models using the 4/6 modulation coding scheme at TMR of 0 and 30%.

Figure 8. BER performance of detectors and demodulators according to MLP detectors.

5. Conclusions

Herein, we introduce a soft-output detector using multi-layer perceptron for a BPMR system to predict reliable information. A multi-layer perceptron (MLP) can be commonly used for regression or classification tasks. Regression MLP output is the predicted value and classification MLP output can be interpreted as the estimated probability. However, the output of the regression MLP converges toward the target value, and the range of the output of the classification MLP is restricted between 0 and 1. Thus, it is difficult to achieve some decoding gain when conducting error correction coding and modulation coding of the data. To provide reliable information similar to the soft decision value, we design the MLP detector that uses the identity function as an activation function in the output layer and binary cross-entropy as the loss function. Additionally, techniques such as BN and ensemble are applied to the MLPs to increase the accuracy. An ensemble of MLP detectors reduces the overlapping area and provides a smaller variance than the PRML detector and MLP detectors alone. It is clear that the ensemble of the MLP decoders estimates reliable information. As the MLP detector provides reliable information similar to the soft decision

value, the ensemble of the MLP detectors with and without the demodulator improves the BER performance and is more robust to TMR compared to PRML.

Author Contributions: S.J. contributed to this work in experiment planning, experiment measurements, data analysis and manuscript preparation; J.L. contributed to experiment planning, data analysis, and manuscript preparation. 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.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. White, R.L.; New, R.M.H.; Pease, R.F.W. Patterned media: A viable route to 50 Gbit/in² and up for magnetic recording. *IEEE Trans. Magn.* **1997**, *33*, 990–995. [\[CrossRef\]](#)
2. Richter, H.J.; Dobin, A.Y.; Heinonen, O.; Gao, K.Z.; Veerdonk, R.J.M.V.D.; Lynch, R.T.; Xue, J.; Weller, D.; Asselin, P.; Erden, M.F.; et al. Recording on bit-patterned media at densities of 1 Tb/in² and beyond. *IEEE Trans. Magn.* **2006**, *42*, 2255–2260. [\[CrossRef\]](#)
3. Zhu, J.; Lin, Z.; Guan, L.; Messner, W. Recording, noise, and servo characteristics of patterned thin film media. *IEEE Trans. Magn.* **2000**, *36*, 23–29.
4. Nutter, P.W.; Ntokas, I.T.; Middleton, B.K.; Wilton, D.T. Effect of island distribution on error rate performance in patterned media. *IEEE Trans. Magn.* **2005**, *41*, 3214–3216. [\[CrossRef\]](#)
5. Moon, J.; Zeng, W. Equalization for maximum likelihood detectors. *IEEE Trans. Magn.* **1995**, *31*, 1083–1088. [\[CrossRef\]](#)
6. Ng, Y.; Cai, K.; Kumar, B.V.K.V.; Zhang, S.; Chong, T.C. Modeling and two-dimensional equalization for bit-patterned media channels with media noise. *IEEE Trans. Magn.* **2009**, *45*, 3535–3538.
7. Nguyen, C.D.; Lee, J. Elimination of two-dimensional intersymbol interference through the use of a 9/12 two-dimensional modulation code. *IET Commun.* **2016**, *10*, 1730–1735. [\[CrossRef\]](#)
8. Jeong, S.; Lee, J. Modulation code for reducing intertrack interference on staggered bit-patterned media recording. *Appl. Sci.* **2020**, *10*, 5295. [\[CrossRef\]](#)
9. Sayyafan, A.; Belzer, B.J.; Sivakumar, K.; Shen, J.; Chan, K.S.; James, A. Deep neural network based media noise predictors for use in high-density magnetic recording turbo-detectors. *IEEE Trans. Magn.* **2019**, *55*, 6701406. [\[CrossRef\]](#)
10. Han, S.; Kong, G.; Choi, S. A detection scheme with TMR estimation based on multi-layer perceptrons for bit patterned media recording. *IEEE Trans. Magn.* **2019**, *55*, 3100704. [\[CrossRef\]](#)
11. Nishikawa, M.; Nakamura, Y.; Osawa, H.; Okamoto, Y.; Kanai, Y. A study on iterative decoding with LLR modulator using neural network in SMR system. *IEEE Trans. Magn.* **2019**, *55*, 3100604. [\[CrossRef\]](#)
12. Katano, Y.; Muroi, T.; Kinoshita, N.; Ishii, N.; Hayashi, N. Data demodulation using convolutional neural networks for holographic data storage. *Jpn. J. Appl. Phys.* **2018**, *57*, 09SC01. [\[CrossRef\]](#)
13. Nabavi, S.; Kumar, B.V.K.V.; Bain, J.A.; Hogg, C.; Majetich, S.A. Application of image processing to characterize patterning noise in self-assembled nano-masks for bit-patterned media. *IEEE Trans. Magn.* **2009**, *45*, 3523–3526. [\[CrossRef\]](#)
14. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning (PMLR), Lille, France, 6–11 July 2015.
15. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
16. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the International Conference Learn. Represent. (ICLR), San Diego, CA, USA, 7–9 May 2015.
17. Chollet, F. *Deep Learning with Python*, 1st ed.; Manning Publications: Shelter Island, NY, USA, 2018; pp. 61–62.
18. Jeong, S.; Lee, J. Modulation code and multilayer perceptron decoding for bit-patterned media recording. *IEEE Magn. Lett.* **2020**, *11*, 6502705. [\[CrossRef\]](#)