The TMR Estimation based on Data Science Technique in Bit Patterned Media Recording System

Atiskit Tanaphatsittha, Anucha Tungkatsathan, and Wiparat Busyatras*

Data and Information Science
Faculty of Science and Technology
Raja Mangala University of Technology Thanyaburi, Thailand
E-mail: wiparat_b@rmutt.ac.th*
Received: February 22, 2025; Revised: April 4, 2025; Accepted: April 23, 2025

Abstract

In large-scale cloud systems, such as data center infrastructure, data storage is still primarily reliant on hard disk drive (HDD) technology. Therefore, advancements in HDD efficiency focus on optimizing storage capacity by improving data density. One of the key technologies that enhances storage efficiency is Bit-Patterned Media Recording (BPMR), which can significantly increase areal density, reaching up to 4 terabits per square inch (Tb/in^2) [1]. However, as track spacing decreases to achieve higher density, the system faces a more severe challenge from two-dimensional interference. This interference can lead to errors in the readback process due to incorrect positioning of the read head, resulting in track misregistration (TMR). Previous research has applied data science techniques, such as the K-Means Algorithm, to predict TMR by comparing the readback signal to the channel signal of each TMR [2]. In this study, we propose a different approach to estimating TMR by preparing a dataset based on 6 various by using 400 record of each TMR level and readback signal features, including the peak amplitude values of read back signal with upper track (Peak Track3), the mean values of read back signal with center track (Mean Track2), the maximum values of read back signal with upper track (Max Track3), the minimum values of read back signal with lower track (Min Track1), the minimum values of read back signal with center track(Min Track2), and maximum of read back signal of upper track(Max Track1). We then applied supervised machine learning 6 techniques, including Random Forest, Decision tree, Support Vector Machines (SVM), General linear model (GLM), Gradient Boosted Tree (GBT) and Deep Learning. The experimental results show that the Decision tree technique provides the best prediction performance, achieving 1.9% Relative Error at TMR level of 20% to 25%.

Keywords: Bit pattern media recording system (BPMR), Track Mis-registration (TMR), Machine learning, Model prediction.

1. Introduction

Nowadays, the demand to store digital data is continuously increasing. Therefore, resulting in the development of new forms of technology. To respond to the need for storage space hard disk drives are the right choice for data storage. Because there is a cheap price per unit of data capacity. Due to the original vertical recording technology (PMR: Perpendicular Magnetic Recording), encounter-red problems with superparamagnetic limitations (Superparamagnetic

Limit) [1-2] which can increase the efficiency of ground density to only 1 - 1.5 terabits per square inch (Tb/in^2)[1-3].

To increase the areal density, Bit pattern media recording (BPMR) data recording technology can increase spatial density [4-5] more than conventional systems. However, the system of BPMR surfaces with a problem with Track-Mis registration into decreased performance of recording. TMR is a situation where the center of the read head and the center of the main track do not match. Previous research has applied data science techniques, such as the K-Means Algorithm, to predict TMR by comparing the readback signal to the channel signal of each TMR [2]. In this study, we propose a different approach to estimating TMR by preparing a dataset based on various readback signal features, including the peak amplitude values of read back signal with upper track (Peak Track3), the mean values of read back signal with center track (Mean Track2), the maximum values of read back signal with upper track (Max Track3), the minimum values of read back signal with lower track (Min Track1), the minimum values of read back signal with center track (Min Track2), and maximum of read back signal of upper track(Max Track1). Then, we applied supervised machine learning techniques, including Random Forest, Decision tree, Support Vector Machines (SVM), General linear model (GLM), Gradient Boosted Tree (GBT) and Deep Learning for finding TMR prediction performance.

This research consists of seven sections. The first section is the introduction. The second section covers the simulation of the bit pattern media recording system using equations. The third section discusses the TMR simulation by equation. The fourth section focuses on data preparation. The fifth section explains the methodology. The sixth section presents the results. Finally, the seventh section provides the conclusion and discussion.

2. Bit pattern media recording

The BPMR system with TMR issues is modeled using mathematical equations, as referenced in study [4]. The signal is generated according to the following equation:

$$r_{l,k} = \sum_{n}\sum_{m} h_{m,n} x_{l-m,k-n} + n_{l,k}, \qquad (1)$$

Position $x_{l,k}$ is the raw data is recorded, where $l \in \{0,-1,+1\}$, that refers to the position of the upper track, center track, and lower track, respectively, that are disrupted by the addition of white noise (AWGN) obtained from averaging. and the variability of σ^2 will be included in the variable $n_{l,k}$. In practice, the readback signal obtained from the simulated channel is in the form of a coefficient. Where $h_{m,n}$, it can be generated by sampling the data from the simple Icelandic pulse response at the position of the product of the integer with the bit period and the track length, represented as in the following equation [5][10]. Where $h_{m,n}$ is the BPMR with TMR represent by

$$h_{m,n} = H(mTz + \Delta_{TMR} + nTx) \qquad (2)$$

where Tz is the track pitch, Tx is the bit length, and Δ_{TMR} is the read-head offset. We set Tz = Tx = 14.5 nm at 3.0 Tb/in^2 .

3. Track mis registration

Track Misregistration (TMR) is a significant challenge in Bit-Patterned Media Recording (BPMR) systems, where the read/write head deviates from the intended track. This misalignment degrades the accuracy of the data retrieval process, increasing bit error rates (BER) and decreasing overall system performance. TMR arises from several factors [6], including mechanical vibrations, thermal fluctuations, and imperfect head positioning, leading to inter-symbol interference (ISI) and inter-track interference (ITI) [7]. These interferences distort the signal, affecting the channel's ability to correctly interpret the data, which is critical in high-density data storage environments like BPMR [2]. As the areal density of storage systems continues to grow, managing TMR becomes even more crucial. Current research emphasizes machine learning techniques to mitigate the effects of TMR by enhancing prediction and correction algorithms, thereby improving the resilience of storage systems against misregistration errors [3]. The TMR simulation by equation as

$$\Delta_{TMR} = (TMRz \times Tz)/100 \quad (3)$$

where $TMRz$ is the percentage of the TMR. In this research, we simulate the coefficients of read back signal by including several of high TMR as 10% 15% 20% and 25% for finding the model prediction. The level of TMR can be used to design an equalizer that approximates each TMR, allowing performance adjustments to achieve higher efficiency [4-8].

4. Data preparation

In the BPMR system, raw data from read channels is often noisy and inconsistent, requiring pre-processing steps such as data scaling and feature selection to improve model performance. We prepared readback signal data for three tracks: the upper track, center track, and lower track. For each track, TMR issues were simulated at various high TMR levels: 10%, 15%, 20%, and 25%. The six data set is divided into the peak amplitude values of read back signal with center track (Peak Track3), the mean values of read back signal with center track (Mean Track2), the maximum values of read back signal with upper track (Max Track3), the minimum values of read back signal with lower track (Min Track1), the minimum values of read back signal with center track (Min Track2), and maximum of read back signal of upper track (Max Track1), as shown in the following Table 1.

Table 1. Data preparation with six examples in each TMR level.

Peak Track3	Mean Track2	Max Track3	Min Track1	Min Track2	Max Track1	TMR (%)
0.287	0.498	0.821	0.846	0.583	0.846	**0**
0.309	0.659	0.593	0.968	0.614	0.968	**0**
0.248	0.425	0.914	0.748	0.529	0.748	**0**
0.335	0.506	0.579	0.689	0.475	0.689	**0**
0.216	0.260	0.411	0.639	0.307	0.639	**5**
0.248	0.559	0.564	0.688	0.485	0.688	**5**
0.323	0.409	0.682	0.649	0.476	0.649	**5**
0.331	0.647	1.000	0.857	0.470	0.857	**5**
0.368	0.585	0.558	0.874	0.528	0.874	**10**
0.453	0.483	0.637	0.672	0.403	0.672	**10**
0.431	0.529	0.674	1.000	0.515	1.000	**10**
0.430	0.445	0.545	0.785	0.539	0.785	**10**
0.456	0.611	0.717	0.822	0.630	0.822	**15**
0.562	0.488	0.583	0.737	0.502	0.737	**15**
0.483	0.782	0.949	0.724	0.575	0.724	**15**
0.482	0.604	0.578	0.919	0.528	0.919	**15**
0.630	0.386	0.602	0.875	0.558	0.875	**20**
0.738	0.757	0.823	0.884	0.521	0.884	**20**
0.710	0.688	0.895	0.566	1.000	0.566	**20**
0.759	0.707	0.747	0.611	0.513	0.611	**20**
0.875	0.821	0.690	0.835	0.692	0.835	**25**

0.986	0.692	0.848	0.825	0.731	0.825	**25**
1.000	0.581	0.809	0.916	0.589	0.916	**25**
0.932	0.441	0.930	0.751	0.678	0.751	**25**

** This data inside the table is scaled by data normalization technique to fit the data in range between 0 and 1, in the label, Track**X** represents 1, 2, 3 as the upper center and lower tracks.

The dataset for this research is prepared by equation (2) allocating coefficient values to different TMR levels at proportions of 10%, 15%, 20%, and 25%. Each TMR level contains more than 400 records. The training and testing datasets are derived from the entire set of prepared records.

5. Method of model prediction

The model development in this study focuses on creating a predictive model capable of estimating the severity of off-track read errors or TMR in a bit-patterned media recording (BPMR) system using machine learning techniques [9-10] as Figure 1. The model creation process involves several key steps:

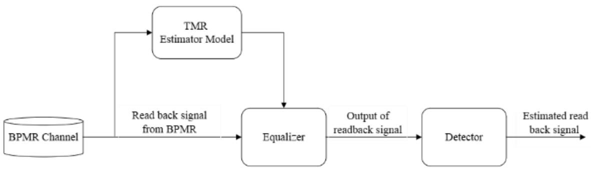


Figure 1. Block diagram of the proposed TMR estimator model in BPMR system.

1)*Data Generation and Feature Selection:* The data used for model development is generated through MATLAB simulations of the BPMR channel. Key parameters that influence the read-back signal, such as pulse width and inter-track interference (ITI), are adjusted to reflect varying levels of off-track read errors. Once the simulation data is collected, feature selection is conducted to identify the most important attributes that significantly impact the TMR problem. This ensures that only the relevant features are used to train the model, improving its performance.

2)*Data Pre-processing:* The simulation data is then pre-processed to prepare it for machine learning model development. This involves data cleaning to handle missing or incorrect values, followed by data normalization to reduce variability and ensure uniformity across all features. Additionally, dimensionality reduction techniques such as Principal Component Analysis (PCA) are applied to minimize the number of features, which helps improve computational efficiency and reduces the risk of overfitting in the model.

3)*Model Development and Training*: With processed data, multiple machine learning algorithms are employed to build the predictive model, including Random Forest, Decision tree, Support Vector Machines (SVM), General linear model (GLM), Gradient Boosted Tree (GBT) and Deep Learning. These algorithms are trained using the training dataset, which is a subset of the entire dataset. After training, the models are evaluated using a testing dataset to determine their accuracy and predictive capability on unseen data. Cross-validation techniques, such as K-Fold Cross-Validation, are employed to ensure the models generalize well to new data and avoid overfitting. The most robust model with the highest accuracy and efficiency is selected for further refinement and application.

6. Result of model prediction

The results of this research highlight the effectiveness of machine learning. All models achieve high performance in predicting Track Mis-registration (TMR) in the BPMR system, with relative errors of less than 10%. Notably, the Decision Tree model consistently achieves a relative error of just 1.9% in distinguishing various levels of TMR severity, as shown in Figure 2 – 4. These significantly lower relative errors suggest a solution for planning the equalizer to filter TMR based on model predictions, thereby improving performance without TMR effects.

Figure 2. Comparison of Relative Error across models.

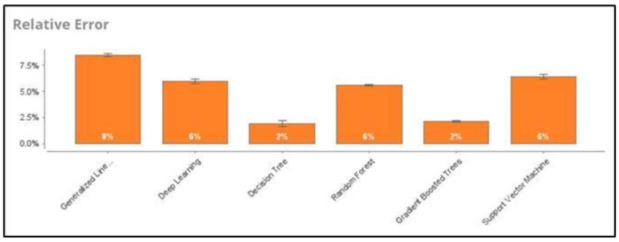


Figure 3. Comparison of processing time for each model.

Figure 4. Overall performance comparison of each model.

A key strength of the model is its robustness in handling noisy and variable data. Even when input data was affected by significant interference, the model maintained stable, accurate predictions, which is crucial for real-world applications where data variability is common. This ability to generalize across different conditions makes it well-suited for high-density data storage environments.

Additionally, the results of all model predictions show high accuracy percentages. Since the accuracy does not differ significantly, selecting the model with the fastest simulation time is a practical approach for future use. This model can be used to simulate performance by designing an equalizer that matches the read-back signal while considering each TMR level.

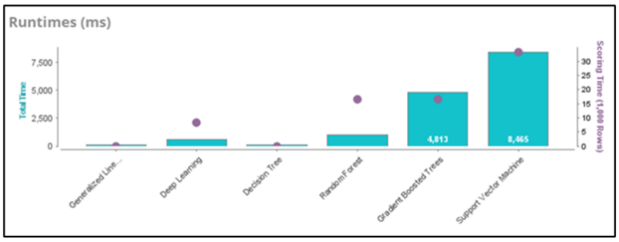
7. Conclusion and Discussion

The experimental results underscore the advantages of using machine learning, particularly deep learning, to address off-track read errors (TMR) in Bit-Patterned Media Recording (BPMR) systems. All the models achieved over 96% accuracy in predicting TMR severity and demonstrated robustness in handling noisy data. This efficiency contributes to overall system reliability, with the model providing quick predictions crucial for high-density storage applications.

However, this research focuses on high TMR levels, which also provide high accuracy percentages. In contrast, low TMR levels, such as 0% and 5%, do not achieve high accuracy.

Model	Relative Error	Standard Deviation	Total Time
Decision Tree	1.9%	± 0.3%	74 ms
Generalized Linear Model	8.5%	± 0.1%	78 ms
Deep Learning	6.0%	± 0.2%	567 ms
Random Forest	5.6%	± 0.1%	1 s
Gradient Boosted Trees	2.1%	± 0.1%	5 s
Support Vector Machine	6.4%	± 0.2%	8 s

To further improve performance, future research should explore hybrid models that combine



various algorithms and implement them with low TMR datasets for data storage technologies in BPMR systems. Additionally, simulating the BER performance of BPMR can help evaluate these models' predictions as criteria for improving system performance and confirming their ability to detect TMR levels.

Acknowledgements

This work received research support from Faculty of Science and Technology Information Technology branch Rajamangala University of Technology Thanyaburi. In addition to the researcher himself There are also many important people who were an important part in the preparation of this thesis, Dr. Wiparat Busayatras and other faculty members. Thank you for your kind advice. and assistance until this research can come to fruition. The researcher would like to use the acknowledgments of this research to express our deepest gratitude.

References

[1] Y. Shiroishi, K. Fukuda, I. Tagawa, and H. Iwasaki, "Future options for HDD storage," *IEEE Trans. Magn.*, vol. 45, no. 10, pp. 3816–3822, Oct. 2009.

[2] S. Jeong and J. Lee, "Track mis-registration estimator based on K-means algorithm for bit-

patterned media recording," *IEEE Trans. Magn.,* vol. 59, no. 3, pp. 1–5, Mar. 2023.

[3] M. H. Kryder, " Magnetic recording beyond the superparamagnetic limit," in *Proc. IEEE Int. Magn. Conf. (INTERMAG)*, Toronto, ON, Canada, Apr. 2000, pp. 14–19.

[4] T. R. Albrecht, H. Arora, and V. Ayanoor-Vitikkate, "Bit-patterned magnetic recording: Theory, media fabrication, and recording performance," *IEEE Trans. Magn.*, vol. 51, no. 5, May. 2015.

[5] S. Nabavi, Signal Processing for Bit-Patterned Media Channels with Inter-Track Interference, Ph.D. dissertation, Carnegie Mellon Univ., Pittsburgh, PA, USA, 2008.

[6] C. Warisarn, W. Busyatras, and L. Myint, " Soft-information flipping approach in multi-head multi-track BPMR systems," *AIP Adv.*, vol. 8, no. 5, pp. 056704-1–056704-7, May. 2018.

[7] S. Nabavi, B. V. K. Vijaya Kumar, and J. A. Bain, " Mitigating the effects of track mis-registration in bit-patterned media," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Beijing, China, May 2008.

[8] S. Karakulak, P. H. Siegel, and J. K. Wolf, " A parametric study of inter-track interference in bit patterned media recording," *IEEE Trans. Magn.,* vol. 46, no. 3, pp. 819-824, Mar. 2010.

[9] P. Kochcha, K. Kankhunthod, and C. Warisarn, "Track misregistration estimation technique based on hybrid K-means and EM algorithm in bit-patterned media recording systems," *IEEE Trans. Magn.,* vol. 61, no. 6, pp. 101-106, Jun. 2025.

[10] K. Kankhunthod and C. Warisarn, " Track misregistration mitigation using CNN-based method on single-reader/two-track reading BPMR systems," *IEEE Access*, vol. 11, pp. 41954-41963, Apr.2023.