Ensemble Learning for STT-MRAM Channel Detection

Chi Dinh Nguyen[1] and Thao Thu Nguyen[2]

[1]Faculty of Computing Fundamentals, FPT University, Ha Noi 13113, Viet Nam, chind3@fe.edu.vn*
[2]Faculty of Mathematics, Hanoi National University of Education, Ha Noi 13113, Viet Nam, thaont_k681028@stu.hnue.edu.vn

We propose a stacking-based ensemble detector for STT-MRAM that combines deep neural networks (base learners) with a meta-model to fuse their outputs. Simulation results demonstrate that, even without error-correction coding, the ensemble detector achieves up to two orders of magnitude lower bit error rate (BER) compared to a conventional threshold detector and maintains graceful degradation as noise and offset severity increase. When combined with an error-correcting modulation code, the proposed approach further reduces BER under the severe effects of unknown offsets. These findings highlight the potential of stacking-based ensemble learning to enhance the reliability of next-generation nonvolatile memories substantially.

***Index Terms*—Nonvolatile RAM, ensemble learning, spin-torque transfer magnetic random access memory (STT-MRAM), neural network, asymmetric write error rate.**

I. Introduction

SPIN-TRANSFER torque magnetic random-access memory (STT-MRAM), which utilizes spin-polarized currents to manipulate the magnetization states of magnetic tunnel junctions (MTJs), is widely regarded as a leading candidate for next-generation nonvolatile memory (NVM) technologies [1]. Despite its significant advantages, the path toward widespread adoption of STT-MRAM is accompanied by several technical challenges. Notably, STT-MRAM is subject to two critical technical challenges: asymmetric write error (AWE) and unknown offsets [1, 2]. To mitigate the impact of such offsets, conventional approaches include the design of offset-cancellation sensing circuits that autonomously compensate for systematic deviations during readout [3]. In addition, advanced modulation codes and channel detection techniques have been explored to suppress offset-induced errors and overcome AWE [4 – 7]. More recently, neural network-based models have been proposed to estimate and correct readout offsets dynamically, demonstrating improvements in both the accuracy and robustness of STT-MRAM systems [5].

II. STT-MRAM Cascaded Channel Model

An STT-MRAM cell comprises two fundamental components: a magnetic tunnel junction (MTJ) and an nMOS transistor. The MTJ is composed of a reference layer, a free layer, and an ultrathin tunneling oxide barrier. The bit is encoded by the magnetization direction of the free layer; that is, if the free layer is parallel to the reference layer, the MTJ is in a low resistance state (representing bit '0'), and conversely (representing bit '1').

In this study, we adopt the STT-RAM cascaded channel model proposed by Cai and Immink [2] to simulate the error behavior of the STT-MRAM channel. Let P_1, P_0, and P_r denote the write error rate for 0 → 1 switching, the write error rate for 1 → 0 switching, and the read disturb error rate, respectively. The writing process is modeled as a binary asymmetric channel (BAC). The read disturb phenomenon is modeled as a **Z**-channel, while the read decision errors are modeled using a Gaussian Mixture Channel (GMC). In this model, the low resistance state R_0 is represented by a Gaussian distribution with its mean and standard derivation of μ_0 and σ_0, while the high resistance state R_1 follows a Gaussian distribution characterized by μ_1 and σ_1. For the effect of system offsets, we assume a resistance offset arising due to elevated operating temperatures. The offset, which only occurs with the high resistance state R_1, is regarded as a Gaussian distribution with $\mathcal{N}(\mu_{ofs}, \sigma_{ofs}^2)$.

III. Proposed Stacking-Based Ensemble Learning

Ensemble learning [6] is a machine learning paradigm in which multiple base models are trained, and their predictions are aggregated to enhance generalization and robustness compared to any individual model. In this study, we propose a stacking ensemble framework with three distinct base learners; Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM). MLP learns complex nonlinear mappings from noisy voltage samples to bit probabilities; CNN (with Conv1D) captures localized spatial error patterns; and LSTM models temporal dependencies in readout sequences to detect time-dependent faults. The stacking procedure in this study is structured as follows.

The MLP, CNN, and LSTM models are independently trained on raw input sequences derived from the STT-MRAM channel. The outputs (predicted probabilities) from all three base models are concatenated to form a combined feature vector. The meta-model is then trained using the stacked predictions from the base models as input, learning to combine them to yield the final decision on each bit. Fig. 1 illustrates the proposed model for STT-MRAM channel. Accordingly, the stacking ensemble learning model can be represented as follows. Each base model receives the input as a vector $\boldsymbol{r} \in \mathbb{R}^n$ (where n is the bit sequence length) and outputs a prediction $\hat{\boldsymbol{y}}_m \in \mathbb{R}^{\boldsymbol{n}}$ where $m \in \{MLP, CNN, LSTM\}$. Predictions from base models can be represented as follows,

$$\hat{\boldsymbol{y}}_{MLP} = f_{MLP}(\boldsymbol{r}, \theta_{MLP})$$
$$\hat{\boldsymbol{y}}_{CNN} = f_{CNN}(\boldsymbol{r}, \theta_{CNN})$$
$$\hat{\boldsymbol{y}}_{LSTM} = f_{LSTM}(\boldsymbol{x}, \theta_{LSTM})$$

where f_{MLP}, f_{CNN}, and f_{LSTM} are functions learned from the training process of MLP, CNN, and LSTM models; θ_{MLP}, θ_{CNN},

θ_{LSTM} are the corresponding parameters of each model. The predictions from the base models are concatenated to form a new (combined) feature vector $\boldsymbol{z} = [\widehat{\boldsymbol{y}}_{MLP};\ \widehat{\boldsymbol{y}}_{CNN};\ \widehat{\boldsymbol{y}}_{LSTM}]^T \in \mathbb{R}^{3n}$. The vector $\boldsymbol{z}$ contains aggregated information from the predictions of the MLP, CNN, and LSTM. The meta model is a neural network that receives $\boldsymbol{z}$ as input and learns how to combine the predictions from the base models to produce the final prediction. The meta model can be expressed by,

$$\hat{\boldsymbol{c}} = f_{meta}\left(\begin{bmatrix} f_{MLP}(\boldsymbol{r}, \theta_{MLP}) \\ f_{CNN}(\boldsymbol{r}, \theta_{CNN}) \\ f_{LSTM}(\boldsymbol{x}, \theta_{LSTM}) \end{bmatrix}, \theta_{meta}\right) \quad (1)$$

where f_{meta} is function learned from the training process, and θ_{meta} is the corresponding parameters of the meta model.

IV. Simulation Results and Discussion

The experimental parameters utilized in this study for simulating the STT-MRAM cell are adopted from [2]. Specifically, the mean resistance values of the two states are set as $\mu_0 = 1\ k\Omega$ and $\mu_1 = 2\ k\Omega$, respectively. The severity of read decision errors can be systematically controlled by adjusting the ratio σ_0/μ_0 (hence σ_1/μ_1). For write errors, we assume a fixed write error probability of $P_1 = 2 \times 10^{-4}$. To model offset effects, a Gaussian distribution with $\mu_{ofs} = -0.2\ k\Omega$ and $\sigma_{ofs}/\mu_1 = 7\%$ is setup for the simulation.

We first evaluate the performance of the proposed model on the raw data without coding. In Fig. 2a, it can be easily observed that the ensemble learning (EL) detector already achieves a BER on the order of 10^{-4}, whereas the threshold detector is near 10^{-2}, a two-order-of-magnitude improvement, at the read errors $\sigma_0/\mu_0 = 3\%$. This demonstrates that the EL detector not only reduces baseline error but also degrades much more gracefully as the read-decision errors raise.

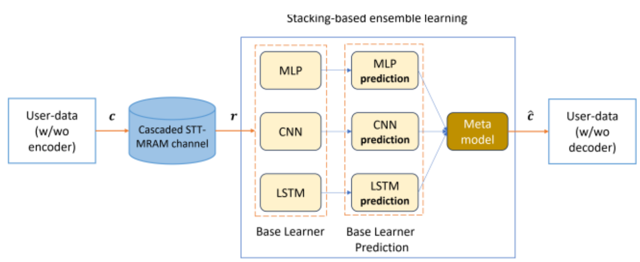

Fig. 1. Proposed model.

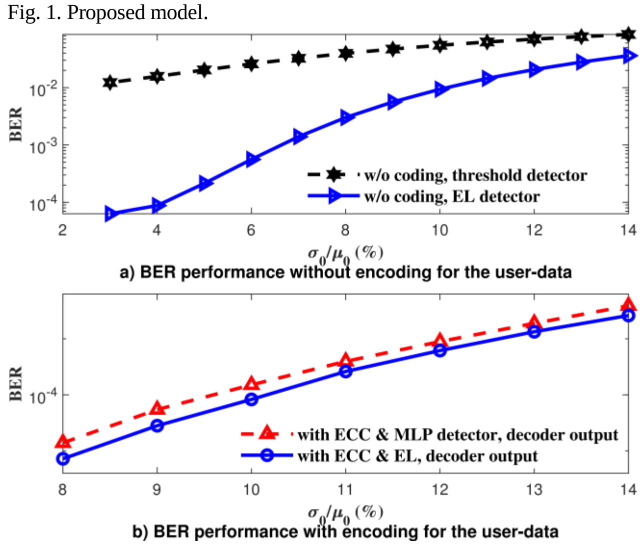

Fig. 2. BER performance for the proposed model.

In addition, we validated the proposed detection scheme under a scenario where the user signal is encoded using an error correction modulation code designed with a minimum Hamming distance (d_{min}) of 3, enabling single-bit error correction. The 4/9-rate code used in this study is constructed following the theoretical framework outlined in [7], where the selected codewords satisfy the $d_{min} = 3$. As illustrated in Fig. 2b, the inclusion of this code notably enhances system performance, especially under high-offset and noisy conditions, further reinforcing the effectiveness of the proposed stacking-based ensemble detection scheme. Both curves in Fig. 2b show significantly improved BER compared to Fig. 2a. At $\sigma_0/\mu_0 = 8\%$, the EL detector achieves a BER of roughly 10^{-5} versus 3×10^{-5} for the MLP detector. As the level of the read errors increases to 14%, both schemes converge toward BER of 10^{-3}, but the EL detector maintains a consistent gap of approximately 2×10^{-4} in absolute BER. As a result, when combined with ECC, the EL detector further enhances decoding performance over an MLP-only approach, yielding lower BER across the entire noise range.

V. Conclusion

We have shown that a stacking ensemble of MLP, CNN, and LSTM base learners, guided by a meta-model, can effectively mitigate both write-asymmetry and unknown offset challenges in STT-MRAM. The proposed detector outperforms single-model and threshold-based approaches by up to two orders of magnitude in uncoded BER and demonstrates superior robustness under high noise and offset conditions. Integration with a lightweight error-correcting modulation code further enhances performance, achieving consistently low BER values around 10^{-5} at 8% read error levels.

References

[1] C. Worledge *et al.*, “STT-MRAM - Status and Outlook,” in *Proc. IEEE 33rd Magn. Recording Conf. (TMRC)*, Aug. 2022, doi: 10.1109/TMRC56419.2022.9918607.

[2] K. Cai and K. A. Schouhamer Immink, “Cascaded channel model, analysis, and hybrid decoding for spin-torque transfer magnetic random access memory,” *IEEE Trans. Magn.*, vol. 53, no. 11, Nov. 2017, Art. no. 8204311.

[3] T. Na, S. H. Kang, and S.-O. Jung, “STT-MRAM sensing: A review,” *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 1, pp. 12–18, Jan. 2021.

[4] C. D. Nguyen and J. Lee, “Improving SOVA output using extrinsic informations for bit patterned media recording,” in *Proc. IEEE Int. Conf. Consum. Electron.*, Jan. 2015.

[5] Z. Mei, K. Cai, and X. Zhong, “Neural network-based dynamic threshold detection for non-volatile memories,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019.

[6] Gautam Kunapuli, *Ensemble Methods for Machine Learning*, Manning, 2023.

[7] C. D. Nguyen, “On the design of 7/9-rate sparse code for spin-torque transfer magnetic random access memory,” *IEEE Access*, vol. 9, pp. 164562–164569, 2021.