

Cascaded Channel Model and Analysis for STT-MRAM

Kui Cai*, Kees A. Schouhamer Immink†, and Zhen Mei*

*Singapore University of Technology and Design, Singapore; †Turing Machines Inc., The Netherlands
Email: cai_kui@sutd.edu.sg, immink@turing-machines.com, mei_zhen@sutd.edu.sg

I. INTRODUCTION

Spin-torque transfer magnetic random access memory (STT-MRAM) is a promising emerging non-volatile memory (NVM) technology which shows high potential for replacing the dynamic random access memory (DRAM), as well as for various embedded applications. However, STT-MRAM suffers from process variation induced variations of the access transistor size, the magnetic tunneling junction (MTJ) geometry and resistance, as well as random thermal fluctuations [1], leading to the write error, the read disturb error, and the read decision error. Therefore, advanced channel coding and signal processing techniques are expected to improve system's reliability in the presence of both write errors and read errors. Channel modeling is the prerequisite for the design of channel codes and detectors. In the literature, many works have been done to model the MTJ based on transport physics and magnetization dynamics [2]. Other works have also been reported which model the STT-MRAM cell from circuit and system perspective including the compact models [3]. However, none of the above models can be considered as a communication type channel model. Moreover, they are all too slow to support the error rate simulations of the STT-MRAM array.

In this work [4], we first propose a novel cascaded channel model, which incorporates both the write errors and read errors for STT-MRAM. We then derive analytically the channel raw bit error rate (BER), based on which we obtain the optimum hard memory sensing threshold. We further derive the bit log-likelihood-ratio (LLR) which enables soft-decision decoding (SDD) of error correction codes (ECCs). We also derive the maximum likelihood (ML) decision criterion for the cascaded channel. These theoretical works serve as basis for the design of advanced channel coding schemes for STT-MRAM. We finally present a hybrid decoding algorithm for extended Hamming codes for STT-MRAM, to illustrate the application of the proposed channel model and its analysis.

II. CASCADED CHANNEL MODEL AND ANALYSIS

In STT-MRAM, each memory cell has an MTJ as the storage element and an nMOS transistor as the access control device. The MTJ has a low resistance state and a high resistance state that can be switched by applying a write current with different polarizations, and hence can represent an input information bit of '0' and '1', respectively. Due to process variation and thermal fluctuation, the write error occurs when the MTJ fails in the switching from one resistance state to the other. It has been found that the write error

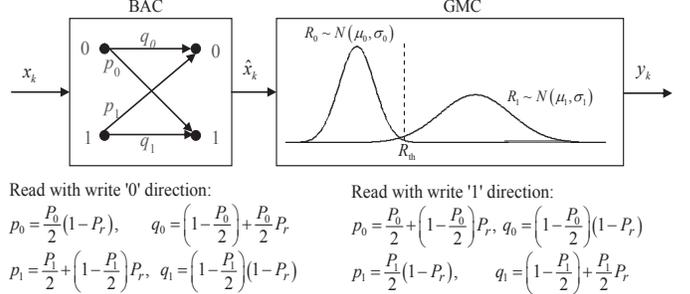


Fig. 1. The cascaded BAC-GMC model.

rate for '0'→'1' switching, denoted by P_1 , is much higher than that for '1'→'0' switching (denoted by P_0) [1], [2]. Since the switching of STT-MRAM cells is asymmetric and independent with each other, we propose to use a binary asymmetric channel (BAC) to model the write error. The read error consists of the read disturb error and the read decision error. The read disturb error, with an error rate of P_r , is caused by the accidental flipping of MTJ during the read operation, due to a large read current caused by process variation or thermal fluctuation [1]. As it can only occur in one direction, it can be modeled as a Z channel. Moreover, since the read disturb error occurs immediately after the write error, we can combine the BAC for the write error and the Z channel for the read disturb error into a combined BAC. The read decision error occurs when the memory cell resistance is incorrectly sensed/detected during reading. It is mainly caused by the process variation induced widened distributions of the low and high resistances of the STT-MRAM cell, which are found to approximately follow Gaussian distributions [5]. Therefore, we propose to use a Gaussian mixture channel (GMC) to model the read decision error, where the distributions of two Gaussian variables, R_0 and R_1 , follow those of the two cell resistances. The mean and variance of R_0 and R_1 are given by μ_0 , σ_0 , μ_1 , and σ_1 , respectively. Note that the Gaussian distributions can be replaced if the cell resistances follow other distributions. As the writing and reading of memory cells are independent processes, we concatenate the combined BAC with the GMC, and obtain a cascaded channel model, the BAC-GMC model, as shown by Fig. 1. Based on the cascaded BAC-GMC channel, we derive the following theorems [4].

Theorem 1: The LLR of bit x_k input into the cascaded BAC-GMC channel is given by

$$L(x_k) = \log \frac{\frac{e^{L(x_k)q_1}}{p_0 + e^{L(x_k)q_1}} e^{L(\hat{x}_k)} + \frac{e^{L(x_k)p_1}}{q_0 + e^{L(x_k)p_1}}}{\frac{p_0}{p_0 + e^{L(x_k)q_1}} e^{L(\hat{x}_k)} + \frac{q_0}{q_0 + e^{L(x_k)p_1}}}, \quad (1)$$

where

$$L(\hat{x}_k) = -\log \sigma_1 + \log \sigma_0 - \frac{(y_k - \mu_1)^2}{2\sigma_1^2} + \frac{(y_k - \mu_0)^2}{2\sigma_0^2}$$

This work has previously been published [4]; publicly available at <http://ieeexplore.ieee.org/document/7938399/>. This work is supported by SUTD-ZJU grant ZJURP1500102, Singapore Agency of Science and Technology (A*Star) PSF research grant, and SUTD SRG grant SRLS15095.

$$+ \log \frac{p_0 + q_1 e^{l(x_k)}}{q_0 + p_1 e^{l(x_k)}}, \quad (2)$$

and $l(x_k) = \log \frac{\Pr(x_k=1)}{\Pr(x_k=0)}$.

Theorem 2: For the cascaded BAC-GMC channel, assume a codeword $\mathbf{x} = [x_0, x_1, \dots, x_{n-1}]$ of an arbitrary code C , the ML decision criterion is given by

$$\hat{\mathbf{x}}_{\text{ML}} = \arg \max_{\mathbf{x} \in C} \sum_{k=0}^{n-1} \max^*(\log(q_k) - \log(\sqrt{2\pi}\sigma_k) - \frac{(y_k - s_k)^2}{2\sigma_k^2}, \log(p_k) - \log(\sqrt{2\pi}\bar{\sigma}_k) - \frac{(y_k - \bar{s}_k)^2}{2\bar{\sigma}_k^2}), \quad (3)$$

with $s_k = (1 - x_k)\mu_0 + x_k\mu_1$, $\sigma_k = (1 - x_k)\sigma_0 + x_k\sigma_1$, $\bar{s}_k = x_k\mu_0 + (1 - x_k)\mu_1$, $\bar{\sigma}_k = x_k\sigma_0 + (1 - x_k)\sigma_1$, $q_k = (1 - x_k)q_0 + x_kq_1$, $p_k = (1 - x_k)p_0 + x_kp_1$, and $\max^*(a, b) \triangleq \log(e^a + e^b)$.

III. AN APPLICATION EXAMPLE

Based on the above presented theoretical analysis, various SDD algorithms of ECCs can be derived. However, the choice of the algorithms is confined by the potential applications. For example, in order for STT-MRAM to replace DRAM, its design needs to meet the requirement of DRAM-like read latency of 10ns to 30ns. The state of the art ECC used by Everspin's 16Mb toggle MRAM is a (71, 64) Hamming code [6]. In this work, we consider the extended Hamming codes, and propose a two-stage hybrid decoding algorithm. That is, we first carry out hard-decision decoding during the first-stage of decoding. If the hard-decision decoder (HDD) declares a successful error correction, output the decoded codeword. Otherwise, continue with the second-stage decoding, using a modified Chase II decoder [4], whose metric is based on the ML decision rule for the cascaded channel given by (3). Moreover, the cascaded channel LLR given by (1) is used to identify the least reliable Q -bit positions. While computing the LLR, we quantize the GMC by L quantization bits by maximizing the capacity of the cascaded channel [4].

Simulations use experimental data taken from [1]. Fig. 2 shows that there is a high error floor at the frame error rate (FER) of 4×10^{-4} , for HDDs of both the (71, 64) Hamming code and the (72, 64) extended Hamming code. It is caused by the read decision errors occur at $\sigma_0/\mu_0 = 9.5\%$. The hybrid decoder of the (71, 64) code only slightly lower this error floor. This is due to the mis-correction of 2-bit errors into 3-bit errors made by the HDD of the (71, 64) code during the first-stage decoding, which disables the second-stage Chase decoding for correcting the 2-bit errors. However, with the (72, 64) code whose HDD can detect all the 2-bit errors, we observe that the hybrid decoder overcomes the high error floor, and improves the maximum affordable write error rate to $P_1 = 1.2 \times 10^{-4}$, at FER = 10^{-6} . Note that the hybrid decoder of the (72, 64) code approaches the performance of the full Chase decoder with ML metric for the cascaded BAC-GMC channel, while the latter brings substantial performance gain over both the squared Euclidean distance (SED) metric (*i.e.* ML metric for AWGN channel) and the ML metric for the GMC only. This demonstrates that the proposed hybrid

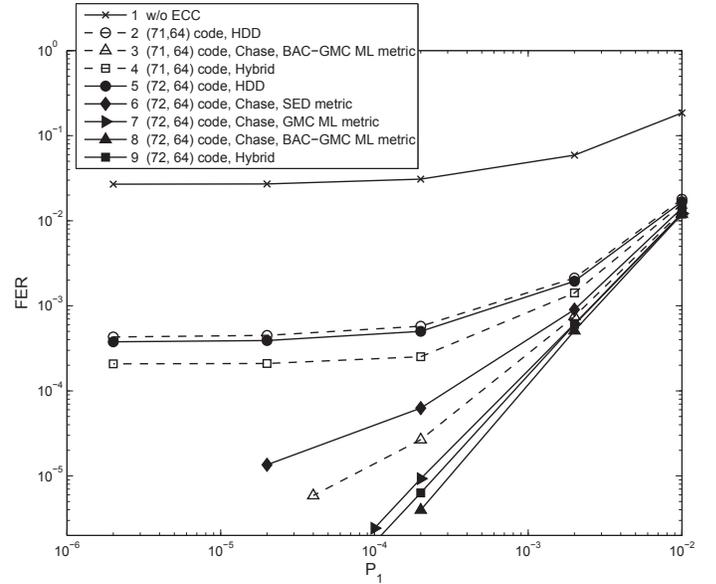


Fig. 2. FER of the (72, 64) extended Hamming code and (71, 64) Hamming code over different P_1 , with different decoders and $\sigma_0/\mu_0 = 9.5\%$. For a 45nm \times 90nm in plane MTJ under a PTM 45nm technology node, with $\mu_0 = 1 \text{ k}\Omega$, $\mu_1 = 2 \text{ k}\Omega$, and $\sigma_0/\mu_0 = \sigma_1/\mu_1$ [1]. We follow [2] and assume P_0 and P_r to be two orders of magnitude lower than P_1 . The hybrid decoders take a $Q = 2$ in Chase decoding, with $L = 3$ bits quantization.

decoder for extended Hamming codes can greatly improve the system tolerance to the write errors and read disturb errors, irrespective of the process variation induced resistance spread. This will facilitate a reduction of the write current and write energy [1]. Meanwhile, although the complexity of the Chase decoder is much higher than the HDD, it is only activated when HDD fails during the first-stage decoding. Our analysis shows that the decoding latency of the hybrid decoder for the (72, 64) code is around 2% higher than that of the HDD [4] at FER = 10^{-6} . If the error rate is further lower, the increase in decoding latency will be even less.

REFERENCES

- [1] Y. Zhang, X. Wang, and Y. Chen, "STT-MRAM cell design optimization for persistent and non-persistent error rate reduction: a statistical design view," in *Tech. Dig. IEEE/ACM International Conference on Computer Aided Design*, pp. 471-477, Nov. 2011.
- [2] Y. Zhang, X. Wang, Y. Li, A.K. Jones, and Y. Chen, "Asymmetry of MTJ switching and its implication to STT-RAM design," in *Tech. Dig. Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Dresden, Germany, pp. 1313-1318, Mar. 2012.
- [3] L.B. Faber, W. Zhao, J.O. Klein, T. Devolder, and C. Chappert, "Dynamic compact model of spin-transfer torque based magnetic tunnel junction (MTJ)," in *International Conference on Design & Technology of Integrated Systems in Nanoscale Era (DTIS)*, pp. 130135, Apr. 2009.
- [4] K. Cai and K. Immink, "Cascaded channel modeling, analysis, and hybrid decoding for spin-torque transfer magnetic random access memory (STT-MRAM)" *IEEE Trans. Magn.*, vol. 53, no. 11, # 8204311, Nov. 2017.
- [5] E. Chen, D. Apalkov, Z. Diao, A. Driskill-Smith, D. Druist, D. Lottis, V. Nikitin, X. Tang, S. Watts, S.A. Wolf, A.W. Ghosh, S.J. Poon, M. Stan, W.H. Butler, S. Gupta, C.K.A. Mewes, T. Mewes, and P.B. Visscher, "Advances and future prospects of spin-transfer torque random access memory," *IEEE Trans. Magn.*, vol. 46, no. 6, pp. 1873-1878, Jun. 2010.
- [6] <https://www.everspin.com/file/162/download>