SCIENCE CHINA
Information Sciences

• LETTER •

December 2024, Vol. 67, Iss. 12, 229304:1–229304:2
https://doi.org/10.1007/s11432-024-4228-3

Unreliability normalization weighted bit-flipping algorithms of LDPC decoding for ReRAM systems

Qike PANG, Zheng MA* & Xiaohu TANG

Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 610031*, China*

Received 18 June 2024/Revised 25 September 2024/Accepted 18 November 2024/Published online 3 December 2024

As the emerging non-volatile storage technology, resistive random-access memory (ReRAM) [1] has garnered considerable interest from diverse industries. It is also one of the most promising candidates for the in-memory computing (IMC) paradigm [2], which is an effective scheme for overcoming the "memory wall". ReRAM comprises a multilayer crossbar array, with memristor positioned on each row-column intersection (Appendix A). Nevertheless, the performance of ReRAM is limited by the "sneak paths" [3] interference (Appendix B) and random resistance interference. Thus, the application of error-correcting code (ECC) technology for data protection is essential, with the exceptional performance of low-density parity-check (LDPC) codes being among the most widely adopted. Unlike traditional communication systems, decoding with low-complexity and low-latency is paramount for storage devices. Therefore, the performance-optimized soft decoding algorithms may not be the practical choice. Instead, the hard decoding algorithm, known as bit-flipping (BF), offers lightweight computation complexity while sacrificing some decoding capability, making it more suitable for storage applications.

In this study, for a better complexity/performance tradeoff, we first propose the unreliability normalization weighted bit-flipping (UNWBF) algorithm for the ReRAM system. Particularly, the proposed algorithms take into account the influence of "sneak paths", providing crucial side information that can enhance decoding performance. Then, we introduce the quantified-UNWBF (Q-UNWBF) algorithms, which are based on the quantitative principle of maximum mutual information (MMI) [4]. The Q-UNWBF algorithms incorporate multi-bit flipping capabilities to optimize decoding performance while minimizing decoding latency. Furthermore, a Q-UNWBF with flip-counting-bias (Q-UNWBF-FCB) is further proposed to alleviate the error-floor problem.

Channel model of ReRAM. Following [5], the sneak paths interference and random noise are characterized comprehensively by a cascaded channel model, where ϵ_{ij} is the sneak paths occurrence probability for cell (i, j). We adopt additive Gaussian noise to model the random noise of resistances. The details of the model and the relevant symbols can be found in Appendix B.

Proposed UNWBF algorithm. It should be stated that the symbols related to the algorithm are defined in Appendix C. The UNWBF exploits "resistance difference" ($\delta R_i = |R_i - R_{\text{ref}}|$) as the reliability metric of cell. And the core steps of the UNWBF algorithm are as follows (the details in Appendix C.2):

• Step 1. Acquiring the number of unreliable cells, $T_{m,n}$, contained in the set $\mathcal{N}(m)\backslash n$, as in (1). According to the channel model, threshold R_{th} can be employed to distinguish between reliable and unreliable cells. Especially, when $T_{m,n} = 0$, it is evidence that whether or not the check node (CN) is satisfied depends only on the current variable node (VN). The significant effect on the flipping function should be imposed, i.e., the parameter is set to $T_{m,n} = 0.5$ in the case.

$$T_{m,n} = |\{i \in \mathcal{N}(m)\backslash n | R_i \leqslant R_{\text{th}}\}|, \tag{1}$$

where $|\cdot|$ denotes the cardinality of the set.

• Step 2. The calculation of weights $w^0_{m,n}$ and $w^1_{m,n}$ as (2). In order to locate the position of the erroneous bits more precisely, the weights ought to consider both the information provided by the check equations and their own reliability.

$$\begin{aligned} w^0_{m,n} &= \min_{i\in\mathcal{N}(m)\backslash n}\{\delta R_i\} + \frac{\delta R_n}{T_{m,n}}, \\ w^1_{m,n} &= \min_{i\in\mathcal{N}(m)\backslash n}\{\delta R_i\} + \frac{\max_{i\in\mathcal{N}(m)}\{\delta R_i\} - \delta R_n}{T_{m,n}}. \end{aligned} \tag{2}$$

• Step 3. The calculation of flipping function as (3) and each iteration flips the bit in the set $\boldsymbol{n}' = \arg\max_n E_n$.

$$E_n = \sum_{m\in\mathcal{M}(n)} ((s_m - 1)w^0_{m,n} + s_m w^1_{m,n}). \tag{3}$$

Based on the fact that the failure of the check equation may imply that more than one of the relevant symbols is wrong, we consider the number of unreliable nodes contained in the check equation as the normalizing factor for the weights. More specifically, we explain the motivation for "unreliability normalization" from three aspects in detail in Appendix C.2. In addition, in order to minimize the raw bit error rate (RBER), the channel parameter estimation is necessary to obtain the optimal R_{ref}, we mainly consider

* Corresponding author (email: zma@swjtu.edu.cn)

© Science China Press 2024

two of them: (1) random electronic noise ($\hat{\sigma}$); (2) the sneak paths interference noise ($\hat{\epsilon}$). The estimation process is also shown in Appendix C.2.

Q-UNWBF algorithm. Subject to the requirements of the fast read/write speed of the storage device, the UNWBF decoder with soft-value weighting, which solely flips the bit with the highest functional value, may not be feasible. Thus, the quantified-UNWBF is presented in this section, which enables the multi-bit flipping and accelerates the convergence of the decoding process. Naturally, determining the appropriate quantitative decision resistance (QDR) becomes a pivotal task in harnessing the full potential of the WBF decoder. The quantization criterion based on MMI [4] is the classical solution towards this problem. As an example, Eq. (4) is the formula for calculating mutual information with 5 levels of quantization and the corresponding discrete memoryless channel model can be found in Appendix D. Based on the MMI criterion, the quantization scheme is obtained by searching for the combination of thresholds that maximizes $I(r, r_q)$. It is worth mentioning that R_{th} is used as a fixed threshold due to the large difference between high and low resistance values.

$$\begin{aligned} I(r, r_q) =& H\left(\frac{p_{1,-2}+\hat{\epsilon}\times p_{0',-2}}{1+\hat{\epsilon}}, \frac{p_{1,-1}+\hat{\epsilon}\times p_{0',-1}}{1+\hat{\epsilon}},\right. \\ & \left.\frac{p_{1,1}+\hat{\epsilon}\times p_{0',1}}{1+\hat{\epsilon}}, \frac{p_{1,2}+\hat{\epsilon}\times p_{0',2}}{1+\hat{\epsilon}}\right) \\ & - \frac{H\left(p_{1,-2}, p_{1,-1}, p_{1,1}, p_{1,2}\right)}{1+\hat{\epsilon}} \\ & + \frac{\hat{\epsilon}\times H(p_{0',-2}, p_{0',-1}, p_{0',1}, p_{0',2})}{1+\hat{\epsilon}}, \end{aligned} \tag{4}$$

where H is the information entropy function, $H = \sum_i -p_i \cdot \log(p_i)$.

Q-UNWBF-FCB algorithm. The Q-UNWBF decoder, which relies on integer processing, significantly enhances computational efficiency. However, the multi-bit flipping of Q-UNWBF will prematurely lead to "error-floor". In Appendix E, we will analyze the reasons for this phenomenon. And we present a compensating method, i.e., a re-formulating of the flipping-function to include the counting of flipping as (5). The flipping-loops will lead to inefficient utilization of iterative resources until the decoding fails. The Q-UNWBF-FCB decoder that incorporates the effect of the number of flipping in the flipping function possesses the ability to automatically escape from the flipping loops.

$$E_n = \sum_{m\in\mathcal{M}(n)} ((s_m - 1)w^0_{m,n} + s_m w^1_{m,n}) - F_c, \tag{5}$$

where F_c is the number of flipping.

Simulation results and comparisons. We illustrate the superiority of the algorithm based on different LDPC codes with varying rates, row-weights, column-weights and code-lengths, whose parameters are shown in Appendix F. For a more explicit demonstration of the performance of the proposed algorithms, there are three points that we explain in the following. (1) Figure 1 indicates that the proposed algorithm outperforms other WBF algorithms under the ReRAM channel. It should be noted that although the RRWBF algorithm performs better than UNWBF under Code 2 in the presence of high random noise, it will suffer from error-floor prematurely. Thus, it is not suitable for the ReRAM system. (2) The 7-level quantization endows the Q-UNWBF with multi-bit flipping properties while guaranteeing the resistance difference. Therefore, Q-UNWBF-7-level not only realizes rapid decoding but also achieves performance improvement when there are more error bits. (3) The Q-UNWBF-7-level suffers an obvious error floor, which can be mitigated by the Q-UNWBF-7-level-FCB decoder. As displayed in Figure 1, the Q-UNWBF-7-level-FCB decoder also yields decoding-gain in low-noise regions. In terms of implementation, the Q-UNWBF-7-level-FCB decoder sacrifices negligible register resources for a large performance gain. The counting bias does not involve any quantization operation and only a simple addition operation is attached, which is very convenient for hardware implementation. Additional performance comparisons and complexity analyses of proposed algorithms are shown in Appendix F.

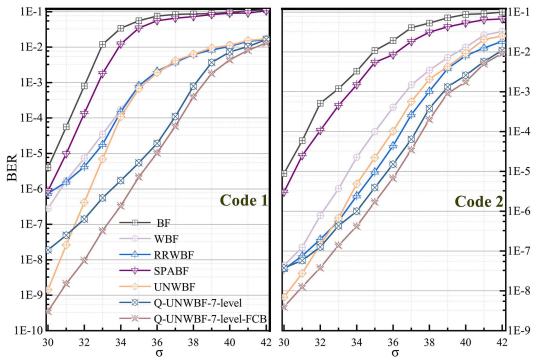

Figure 1 (Color online) Performance comparison of several BF and WBF algorithms based on different regular codes, where $\epsilon = 0.5$ and the maximum of iteration is 50.

Conclusion. This study proposes a novel weighted bit-flipping algorithm, UNWBF, tailored specifically for ReRAM systems. The proposed algorithm mitigates the effects of sneak paths interference, where the flipping metric takes into account several factors to more accurately localize the error bits. To reduce decoding delay, we further propose Q-UNWBF decoder and Q-UNWBF with the flip-counting-bias decoder. These are equipped with multi-bit flipping capabilities and are highly suitable for hardware implementations.

Acknowledgements This work was supported in part by Key Project of Sichuan Provincial Natural Science Foundation (Grant No. 2022NSFSC0043).

Supporting information Appendixes A–F. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Strukov D B, Snider G S, Stewart D R, et al. The missing memristor found. Nature, 2008, 453: 80–83
2 Ling Y T, Wang Z W, Yang Y H, et al. An isolated symmetrical 2T2R cell enabling high precision and high density for RRAM-based in-memory computing. Sci China Inf Sci, 2024, 67: 152402
3 Zidan M A, Fahmy H A H, Hussain M M, et al. Memristor-based memory: the sneak paths problem and solutions. MicroElectron J, 2013, 44: 176–183
4 Wang J, Vakilinia K, Chen T Y, et al. Enhanced precision through multiple reads for LDPC decoding in flash memories. IEEE J Sel Areas Commun, 2014, 32: 880–891
5 Song G, Cai K, Zhong X, et al. Performance limit and coding schemes for resistive random-access memory channels. IEEE Trans Commun, 2021, 69: 2093–2106