Computer and Information Science; Vol. 14, No. 2; 2021
ISSN 1913-8989 E-ISSN 1913-8997
Published by Canadian Center of Science and Education

A Nonuniform Reference Voltage Optimization Based on Relative-Precision-Loss Ratios in MLC NAND Flash Memory

Caifeng Lv[1], Xiujie Huang[1], & Shancheng Zhao[1]

[1] College of Information Science and Technology, Jinan University, China

Correspondence: Caifeng Lv, College of Information Science and Technology, Jinan University, China.

Received: March 1, 2021 Accepted: March 31, 2021 Online Published: April 13, 2021

doi:10.5539/cis.v14n2p75 URL: https://doi.org/10.5539/cis.v14n2p75

Abstract

In Multi-Level-Cell (MLC) NAND flash memory, cell-to-cell interference (CCI) and retention time have become the main noise that degrades the data storage reliability. To mitigate such noise, a relative precision loss (RPL) nonuniform reference voltage sensing strategy is proposed in this paper. First, based on the NAND flash channel model with CCI and retention noise, we simulate the data storage process of MLC NAND flash by Monte Carlo method, and find that the threshold-voltage of each disturbed storage state shows approximately to be Gaussian distributed. Then, by Gaussian approximation, the distribution of threshold voltage can be estimated easily in mathematics with a little loss. Second, we introduce a concept of log-likelihood ratio (LLR)-based RPL ratio to determine the dominating overlap regions, and then propose a new nonuniform reference voltage sensing strategy. This strategy does not only reduce the memory sensing precision (i.e., the number of reference voltages), but also maintains the reliability of the soft information of NAND flash memory channel output for soft decoding. Third, we implement extensive simulations to verify the performance of the new nonuniform sensing strategy. The BER performances of LDPC codes for different sensing strategies are provided to show that the proposed LLR-based RPL-nonuniform sensing strategy can make a good compromise between memory sensing latency and error-correction performance.

Keywords: MLC-NAND flash memory, LDPC code, nonuniform sensing, cell-to-cell interference, retention time

1. Introduction

Today, in order to provide high-quality services to end users, data centers need fast and highly reliable storage. Multi-Level-Cell (MLC) NAND Flash memory has become a mainstream storage device due to its high performance and nonvolatile nature while reducing the over space and power. With the increased storage density of flash memory, memory blocks is increasingly susceptible to a variety of channel noises, including data retention, Cell-to-Cell interference (CCI), program/erase (P/E) cycles and read disturb (Cai, 2017; Cai, Luo, Ghose & Mutlu, 2015; Wang, Dong, Pan, Zhou & Stievano, 2011), which can shift the state threshold voltage. Consequently, the raw bit error rate of data stored in the memory increases. To improve the integrity and reliability of information stored in memory, error-correction codes (ECCs) are used, such as Bose-Chaudhuri-Hocquenghem (BCH), Reed-Solomon (RS) codes (Liu, Rho, & Sung, 2006; Micheloni, Ravasio, Marelli, Alice, Altieri, Bovino & Won, 2006) have been introduced to MLC NAND flash memory system. However, the traditional hard-decision ECCs are not enough to meet the reliability requirements of the high-density flash memory. Compared with hard-decision ECCs, soft decision based ECCs have a higher error-correcting performance, especially the low-density parity check (LDPC) codes (Gallager, 1962) which is a class of powerful iteratively decodable ECC. The Belief Propagation (BP) decoding algorithm of LDPC code is a typical iterative decoding algorithm that works iteratively. During iterations, the log-likelihood-ratio (LLR) soft information is iteratively updated to have excellent error performance in NAND flash memory. Therefore, the accuracy of the LLR value output by the NAND flash memory channel determines the decoding performance of the LDPC code.

For high-quality and accurate LLRs, NAND flash memory chips demand perform small range of soft-decision memory-cell sensing. However, fine-grained memory sensing can lead to an increase in the sensing time consumed by the chip and incur access latency penalty (Zhao, Dong, Sun, Zheng, & Zhang, 2012). Hence, it is necessary to make a trade-off between the memory sensing levels and decoding performance of flash memory

chip. In addition to higher sensing precision, accurate channel initial LLRs also need to estimate the distribution of storage states. However, the threshold voltage distribution of storage states are severely shifted as the NAND flash chip scaling down and retention time increase (Prall, 2007). Literature (Lee, & Sung, 2013; Peng, Wang, Fu & Huo, 2017) study the distribution of varies noise and demonstrated that the distribution of each storage states can be approximately expressed as a Gaussian mixture function. Nonuniform sensing strategy (Dong, Xie & Zhang, 2011) is proposed to reduce memory sensing precision, yet gives a high precision within three dominating overlap region. However, only the cell-to-cell interference is regarded as the main noise in that work.

Our work attempts to overcome the shortcomings mentioned above. First, based on the channel model of NAND flash memory system, we use Monte Carlo to simulate flash memory with LDPC codes as ECC, which suffer from various noises, including cell-to-cell interference and retention time. We find the PDF of each state approximately obeys Gaussian distribution. Further, we derive mathematical formulations to estimate the parameters of state distributions. Second, combining the calculation of LLRs, we propose a new nonuniform sensing strategy based on the Relative-Precision-Loss (RPL) which can determine the boundary of dominating overlap regions. Final, we estimated the PDF of threshold voltage over a wide range of coupling strength factors s and different retention times T and calculated the LLRs of channel. Simulation results show that the proposed new nonuniform sensing strategy exhibits better error-correcting performance with respect to other existing counterparts while considering more noise interference.

2. NAND Flash Memory Channel

In the multi-level cell (MLC) NAND Flash memory, each cell is supposed to fall into one of the four non-overlapping threshold voltage windows and store 2 bits. Each threshold voltage window specifies one storage state, and the first state is the erased state storing 11, the other three states are the programmed states storing 10, 00 and 01, respectively. The smallest unit of erasure is a block, and all cells in a block should be erased before starting programming. Due to the inevitable process variability, the threshold voltage of the erased-state cell shifts, which is usually considered to be Gaussian distributed (Takeuchi, Tanaka & Nakamura, 1996; Wang, Dong, Pan, Zhou & Stievano, 2011). The PDF of the erased state threshold voltage is

$$P_e(x) = \frac{1}{\sigma_e \sqrt{2\pi}} e^{-\frac{(x-u_e)^2}{2\sigma_e^2}} \tag{1}$$

where u_e and σ_e are the mean and the standard deviation, respectively. The incremental step pulse programming (ISPP) procedure is one common technique used to program memory cell (Suh, Lim, Kim, Choi, Koh & Lim, 1995). By the ISPP technique, the threshold-voltage distribution of the *k*-programmed state is uniform (Dong, Pan, & Zhang, 2014; Xu, Gong, Chen, Michael & Li, 2015), whose PDF is given by

$$p_p^{(k)}(x) = \begin{cases} \dfrac{1}{\Delta V_{pp}}, & \text{if } V_p < x < V_p + \Delta V_{pp} \\ 0, & \text{else} \end{cases} \tag{2}$$

where V_p is the verify voltage of the k-th programmed state and ΔV_{pp} is the incremental programming voltage step size.

In the MLC NAND Flash channel, there are many noise factors such as CCI, read disturb errors, random telegraph errors, retention time and P/E cycling errors, and various analytical methods are designed to work out with different noises. Among these factors, CCI is one sever noise, because the MLC technology reduces the width of threshold voltage for each storage state and narrows the gaps between adjacent states, which degrades the reliability of the memory. Retention noise is a predominant noise, especially when the flash device is powered-off for a long time or has been used with many P/E cycling operations. Hence, in our work, we mainly consider the flash memory with CCI and retention noise.

2.1 Cell-to-Cell Interference

For a memory chip, the Cell-to-Cell interference occurs due to parasitic capacitive coupling between adjacent memory cells. As the threshold voltage of a flash cell (interfering cell) increases, the threshold voltage of its adjacent (victim) cells also shift (Lee, Hur, & Choi, 2002). The unintended threshold voltage shifts may cause the victim cell moving into a different storage state, which brings data distortions. The charges change of the

victim cell caused by CCI can be accurately modeled as a linear combination of the threshold voltage shift of the adjacent cells which are programmed after the victim cell (Cai, Mutlu, Haratsch & Mai, 2013). It can be expressed as

$$\Delta F = \sum_{n} (\Delta V_t^{(n)} \cdot \gamma^{(n)}) \tag{3}$$

Where $\gamma^{(n)}$ is the coupling-capacitive ratio between the n-th interfering cell and the victim cell, and $\Delta V_t^{(n)}$ is the threshold voltage change of the n-th interfering cell during programming. In addition, the computation of ΔF depends on the architecture of NAND Flash memory (Park, Kang, Kim, Hwang, Choi, Lee, Kim & Kim, 2008). In our work, we focused on NAND flash memories with all-bit-line structure where a victim cell is mainly disturbed by three neighboring cells on the next word-line, as shown in figure 1. Here, γ_y and γ_{xy} represent the vertical and diagonal coupling-capacitive ratios, respectively, which are assumed to be Gaussian According to the literature (Shibata, Maejima, 2008), the relation between of means of γ_y and γ_{xy} is set to be approximately 0.08:0.006, which shows that the vertical coupling ratio plays a major role in CCI. Therefore, we ignore the diagonal coupling ratio to simplify the following mathematics. To study a wide range of CCI, the parameter of cell-to-cell coupling strength factor s is introduced and the means of the verticality-coupling ratio and the diagonal-coupling ratio can be expressed as 0.08s and 0.006s, respectively.

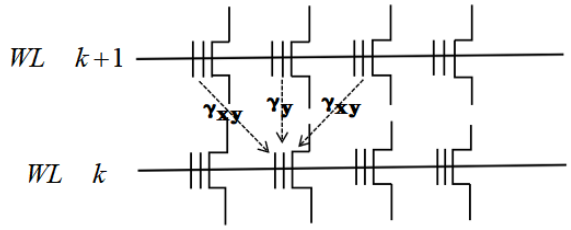


Figure 1. Illustration of the CCI interference in all-bit-line structure and parasitic coupling Capacitance among neighboring cells

Let $p_c(x)$ be the PDF of CCI. Supposing that the interfering cells be written to each state with the same probability (i.e., each state has the same probability of $1/K$ and $K = 4$ for MLC NAND flash memory), the PDF of CCI can be modeled as (Dong, Xie, Zhang, 2011)

$$p_c(x) = \frac{1}{2K\gamma_y \Delta V_{pp}} \times \sum_{v=1}^{K-1} \left(erf\left(\frac{V_r^{(v)} - x/\gamma_y - u_e}{\sqrt{2}\sigma_e} \right) - erf\left(\frac{V_l^{v(\ )} x/\gamma_y - u_e}{\sqrt{2}\sigma_e} \right) \right) + \frac{\delta(x}{K} \tag{4}$$

where

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{5}$$

is the error-function and $\delta(x)$ represents the Dirac delta function. When the interfering cell remains in the erased state, it will not induce CCI to the victim cell. Hence, the corresponding interference is represented by Dirac delta function. Based on the above analysis, the PDF of the threshold-voltage for the victim cell that was programmed to the k-th programmed state and has been interfered by CCI can be expressed by

$$p_v^{(k)}(x) = p_p^{(k)}(x) \otimes p_c(x) \tag{6}$$

where the sign $\otimes$ is the convolution operation. Supposing the victim cell is in the 0-th state (erased state), then

the PDF $p_v^{(k)}(x)$ in (6) will be computed by using $p_e(x)$ since $p_p^{(k)}(x) = p_e(x)$.

2.2 Date Retention Noise

Retention noise (RN) is caused by charge leakage over time after a flash memory is programmed and repeated P/E operations. According to reference (Lee, Choi, Park & Kim, 2003; Cai, Yalcin, Mutlu, Haratsch, Cristal, Unsal & Mai, 2012; Cai, Yalcin, Mutlu, Haratsch, Crista, Unsal & Mai, 2013; Dong, Xie & Zhang, 2013), the retention noise can be approximated with a Gaussian distribution, whose PDF is given as

$$p_r^{(k)}(x) = \frac{1}{\sigma_{rk}\sqrt{2\pi}} e^{\frac{-(x-u_{rk})^2}{2\sigma_{rk}^2}} \tag{7}$$

where the mean u_{rk} and the standard variance σ_{rk} are data-dependent and defined as (Dong, Xie & Zhang, 2013), as given by

$$u_{rk} = K_k(u_c^{(k)} - x_0)\cdot(A_t \cdot N_c^{a_i} + B_t \cdot N_c^{a_o})\cdot\log(1+t) \tag{8}$$

$$\sigma_{rk} = 0.3\cdot|u_{rk}| \tag{9}$$

where K_k, x_0, A_t, B_t, a_i, a_o are constants, N_c and t represents the P/E cycles and the retention time respectively. For the threshold-voltage distribution of the k-th storage state after CCI, we denote $u_c^{(k)}$ to represent its mean.

Combining the analysis of CCI and RN model, we finally estimate the PDF of the overall threshold-voltage distribution for the victim cell disturbed by CCI and RN as

$$p^{(k)}(x) = \int_t p_v^{(k)}(x) p_r^{(k)}(x-t)dt \tag{10}$$

2.3 The Mathematical Formula of LLR

Since the charges in a storage cell can be detected by finite number of sensing voltage, when probability density function of each storage state is available, it is easy to compute the initial LLR information for the flash memory channel. For MLC NAND flash memory, storing two bits in a cell, a physical page referring to all cells in one word-line consists of a least significant bit (LSB) page and a most significant bit (MSB) page. The left bit of the storing two bits is the MSB and the right one is the LSB. Recall that the four storage states 11, 10, 00 and 01 are denoted as the state $k = 0,1,2,\text{and } 3$, respectively. If a sensed threshold-voltage V_{th} of memory cells falls into the reference voltage interval $(R_l, R_r]$, then the LLR information of the LSB and MSB are respectively calculated as

$$L_{MSB} = \log\frac{\int_{R_l}^{R_r}[p^{(0)}(x) + p^{(1)}(x)]dx}{\int_{R_l}^{R_r}[p^{(2)}(x) + p^{(3)}(x)]dx} \tag{11}$$

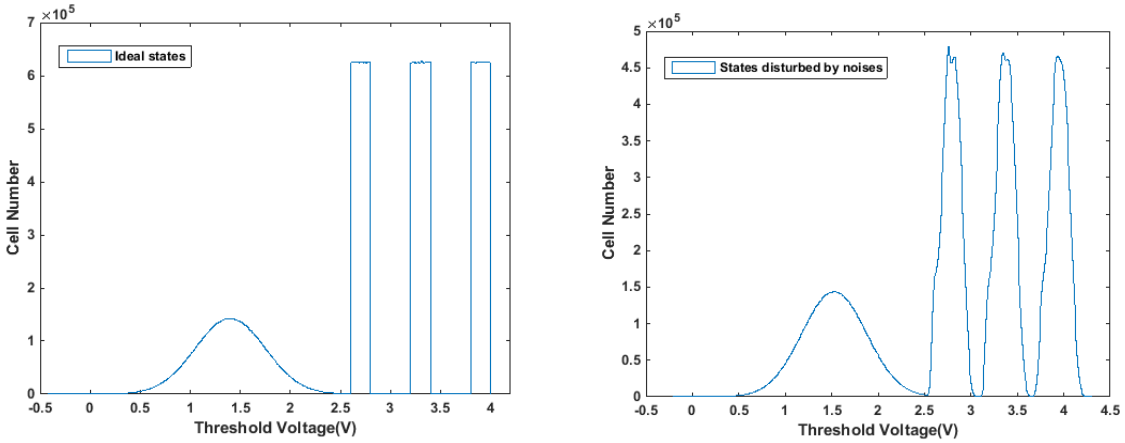
$$L_{LSB} = \log\frac{\int_{R_l}^{R_r}[p^{(0)}(x) + p^{(3)}(x)]dx}{\int_{R_l}^{R_r}[p^{(1)}(x) + p^{(2)}(x)]dx} \tag{12}$$

If a series of reference voltages is provided, then the above LLR information can be precalculated. If a soft decoding is performed, the calculated LLRs of channel are inputted to the LDPC decoder as initial soft-decision information. The decoding performance is affected by the choice of reference voltages. Therefore, it is critical to design a method to find (sub-) optimal reference voltages for soft decoding.

3. New Nonuniform Memory Sensing

3.1 Gaussian Approximation of Threshold-Voltage Distribution

To compute the LLR informations in (11) and (12), it is necessary to know the exact threshold voltage distributions of cells in four storage states that are specified by $p^{(k)}(x)$. To this end, the Monte-Carlo simulation is conducted and shows that there are three overlapping regions for MLC NAND flash memory, and that the shapes of threshold-voltage PDFs for four storage states are Gaussian-like in the presence of CCI and RN, as shown in Figure 2(b).



(a) (b)

Figure 2. Threshold-voltage distribution of MLC NAND flash memory before disturbed (a) and after disturbed by CCI and retention noises (b)

Assuming that $u_I^{(k)}$ and $\sigma_I^{(k)}$ are respectively the mean and the standard deviation of the targeted distribution for the storage state k without CCI and RN. In presence of CCI and RN, the threshold voltage of the storage state k after disturbed is approximated by the Normal distribution with mean $u_f^{(k)}$ and standard deviation $\sigma_f^{(k)}$. The parameters of mean and standard deviation are defined as

$$u_f^{(k)} = u_I^{(k)} + \delta_c - u_{rk} \tag{13}$$

$$\sigma_f^{(k)} = [(\sigma_I^{(k)})^2 + (\sigma_c)^2 + (\sigma_{rk})^2]^{1/2} \tag{14}$$

where δ_c is the mean of CCI distribution and calculated by

$$\delta_c = \int x \cdot p_c(x) dx \tag{15}$$

and σ_c is the standard deviation of CCI distribution, given as

$$\sigma_c = (\int x^2 \cdot p_c(x)\, dx - \delta_c^2)^{1/2} \tag{16}$$

It should be noted that the values of $u_f^{(k)}$ and $\sigma_f^{(k)}$ vary with the coupling strength factor s. Nevertheless, we can precalculate the means and the variances of the threshold voltage distributions for four states under a wide range of coupling strength factors s and store them in a look-up table.

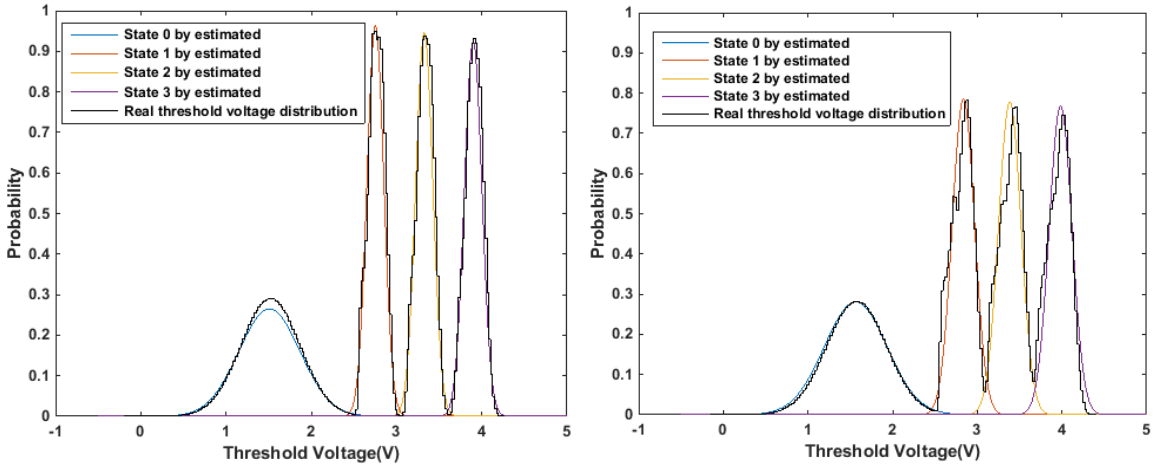


Figure 3. PDFs of the actual threshold-voltage simulated by Monte-Carlo and the derived Gaussian distribution, where the strength factor $s = 1.0$ is 1.0 (a) and $s = 1.5$ (b), $PE = 10\text{k}$ and retention time equal 10^4.

To verify the validity of the above Gaussian approximation, we fit the distribution of storage states obtained by Monte-Carlo simulation with the derived Gaussian distribution. We find that the coincidence is high for a wide range of coupling strength factors. Figure 3 shows the similarity of the PDFs of the two kind distributions for two different coupling strength factor, i.e., $s = 1.0 \text{ and } 1.5$.

3.2 RPL-Nonuniform Memory Sensing Strategy

As mentioned in subsection 2.3, it is a key topic to find (sub-) optimal sensing voltages for soft decoding, because the error-correcting performance of soft decoding for LDPC codes is affected by the value of the LLR information as defined in formulas (11) and (12) that depends on the choice of sensing voltages $(R_l, R_r]$. In the literature of Dong et al. (2013), a nonuniform sensing strategy was designed to sense the dominating overlap area with a higher precision while the rest region with a lower precision. The dominating overlap region was determined by the borders that were selected by adjusting a probability ratio R of the adjacent two storage states based on entropy. The basic idea of the nonuniform sensing strategy is shown in Figure 4.

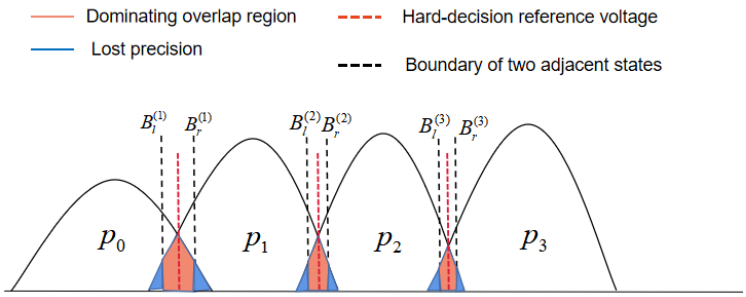


Figure 4. Illustration of the nonuniform sensing strategy, lost precision and borders of a dominating overlap region

The key of nonuniform sensing scheme is how to determine the boundary of the dominant overlapping regions of adjacent states. If each size of the dominant overlapping region is large, more sensing levels are introduced for better decoding performance, but heavier implementation and latency overhead is required. If the size of the dominant overlapping region is reduced, the number of sensing levels is reduced, but the performance of soft decoding may degrades. From computations of the LLR information defined by formulas (11) and (12), we find

that each LLR is mainly dependent on two largest probability items in the overlapping region. In this work, by this observation, we introduce a concept of LLR-based relative-precision-loss (RPL) ratio θ. Then the boundary $B_l^{(k)}$ and $B_r^{(k)}$ are obtained by adjusting the value of the LLR-based RPL ratio θ. The LLR-based RPL ratio is defined as follows

$$\log \frac{\int_{-\infty}^{B_l^{(k)}} p^{(k-1)}(x)dx}{\int_{-\infty}^{B_l^{(k)}} p^{(k)}(x)dx} = \theta_l^{(k)} \tag{17}$$

$$\log \frac{\int_{B_r^{(k)}}^{+\infty} p^{(k)}(x)dx}{\int_{B_r^{(k)}}^{+\infty} p^{(k-1)}(x)dx} = \theta_r^{(k)} \tag{18}$$

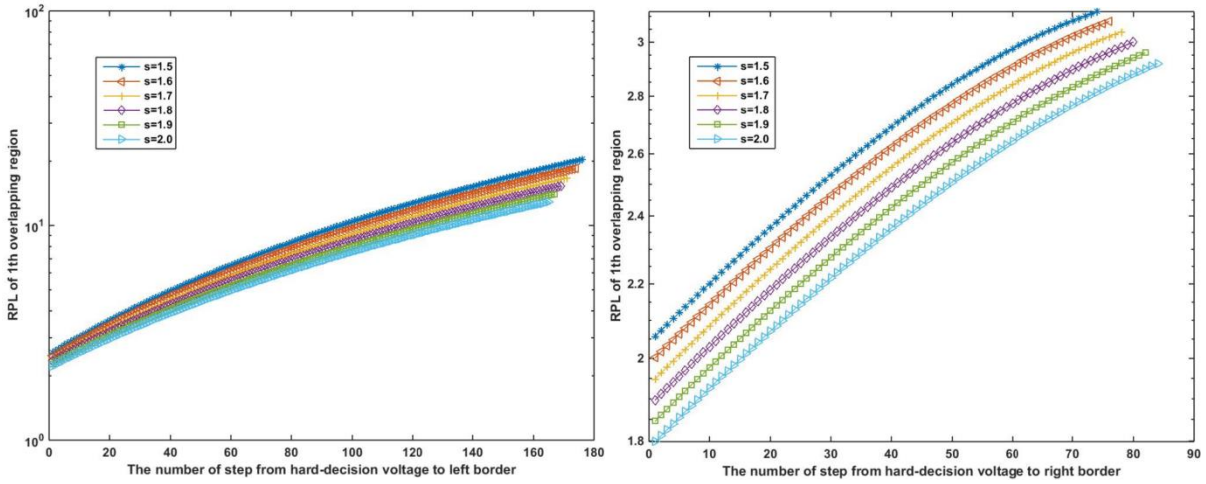
where $\theta_l^{(k)}$ and $\theta_r^{(k)}$ presents the Relative-Precision-Loss of LLR in the k-th(k=1,2,3) overlapping region.

From the definition given by (17) and (18), we can see that the values of $\theta_l^{(k)}$ and $\theta_r^{(k)}$ change with $B_l^{(k)}$ and $B_r^{(k)}$. Such changes are depicted in Figure 5 for a wide range of different coupling strength factors. From Figure 5, we can find that $\theta_l^{(k)}$ and $\theta_r^{(k)}$ in the 2-th and 3-th overlapping regions almost have the same trends for different coupling strength factors. After a lot of data processing and experimentations, we choose the LLR-based RPL ratios in three dominating overlap regions as

$$\theta_l^{(k)} = \begin{cases} 4.40, & k=1 \\ 2.55, & k=2,3 \end{cases} \tag{19}$$

$$\theta_r^{(k)} = \begin{cases} 2.45, & k=1 \\ 2.85, & k=2,3 \end{cases} \tag{20}$$

And the corresponding boundaries of the dominant overlapping regions of every two adjacent storage states can be obtained at the same time. In each dominant overlapping region, a more fine-grained sensing with more voltage levels is adopted; but in the remainder region, a less precision sensing with less voltage levels is utilized. In this work, we apply the equal division to separate the sensing range between the hard-decision reference voltage and the neighbor boundary of the adjacent storage states. And each remainder region has no further division.



(a) (b)

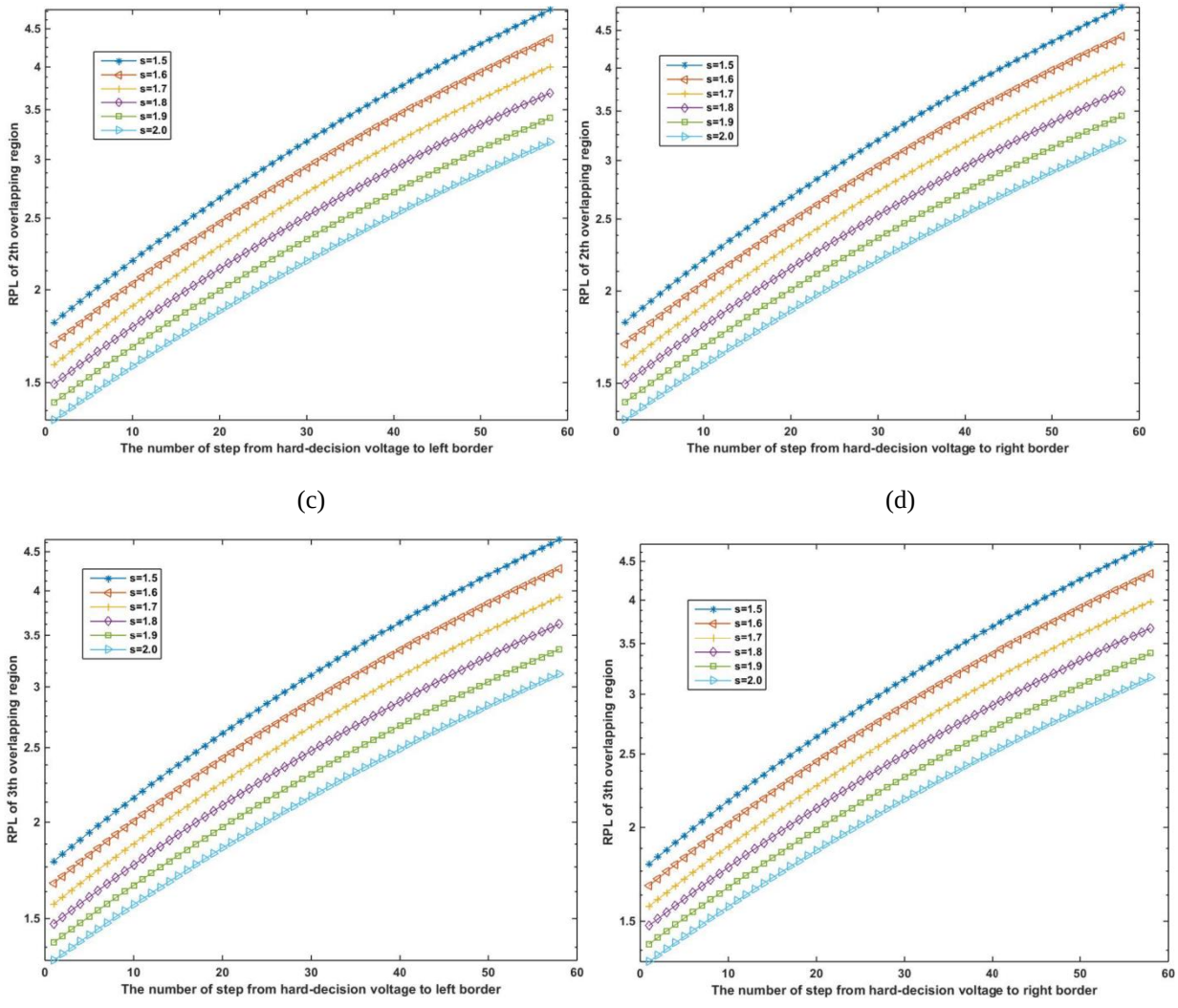


Figure 5. The values of $\theta_l^{(k)}$ and $\theta_r^{(k)}$ with different borders of $B_l^{(k)}$ and $B_r^{(k)}$ for different coupling strength factors s, where k=1, 2, 3

3.3 Simulation Results

In this subsection, we employ a rate-0.85(750, 5000) LDPC code in MLC NAND flash memory to evaluate the error correcting performance of soft decoding when the new nonuniform sensing strategy based on the LLR-based RPL ratios proposed in previous subsection is applied. For comparison, we also evaluate the soft decoding performances for the uniform sensing strategy presented by Alrod *et. al* in (2013) and the nonuniform sensing strategy based on entropy presented by Dong *et.al* in (2013) where the probability ratio R is set as 512 to determine the boundaries. In simulations, the parameters value of MLC NAND flash memory system are set as follows: $u_e = 1.4,\ \sigma_e = 0.35,\ \Delta V_{pp} = 0.2,\ K_k = 0.333,\ A_t = 3.5\times10^{-5},\ B_t = 2.35\times10^{-4},\ a_i = 0.62,\ a_o = 0.3,$ $V_p = 2.6, 3.2, \text{and } 3.8.$ The CCI coupling strengthen factor s varies from 1.5 to 2.0 with the step size as 0.1. In simulations, we plot the BER performances of the proposed RPL-nonuniform sensing strategy, Dong's nonuniform sensing strategy and the traditional uniform sensing strategy under different sensing precisions.

Figure 6 shows the BER versus the CCI. It can be observed that the proposed RPL-nonuniform sensing strategy

outperforms the traditional uniform sensing strategy, and that the BER performance of the proposed RPL-nonuniform sensing strategy enhances as the number of sensing levels increases. It can be seen that the proposed RPL-nonuniform sensing strategy outperforms Dong's nonuniform sensing strategy. For a given number (e.g., 15 and 31) of sensing levels, the performance advantage usually increases as the CCI coupling strength factor s becomes larger. This performance advantage may decrease as the number of sensing levels increases because of the more precise sensing for more levels. It can be seen also that, the performance of the proposed 15-level RPL-nonuniform sensing is close to that of 31-level uniform sensing, which shows that our proposed strategy reduces the sensing precision latency by about 45%.

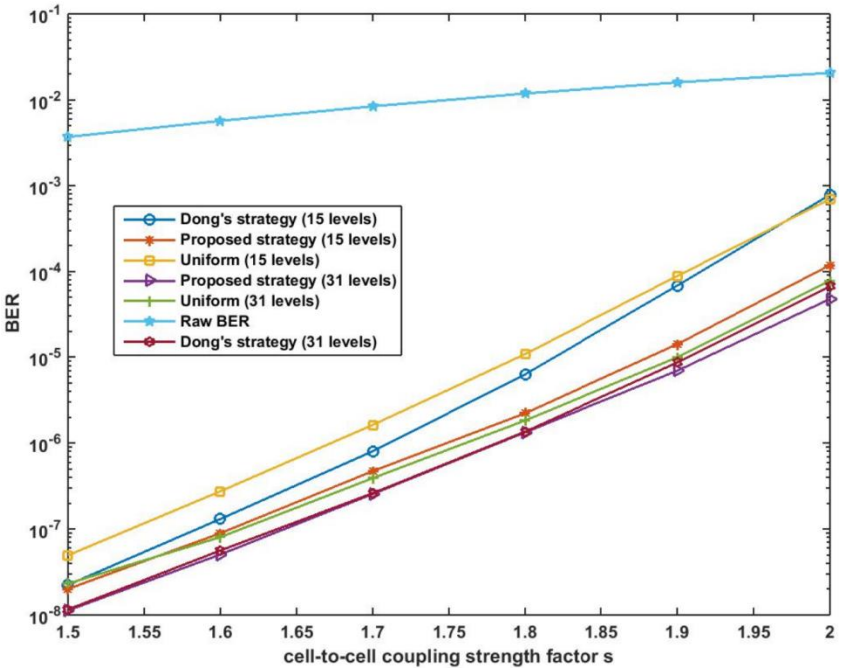


Figure 6. BER performance of LDPC code with cell-to-cell coupling strength factor s for different sensing level strategies, where the P/E cycles equal 10000 and retention time equal 10000

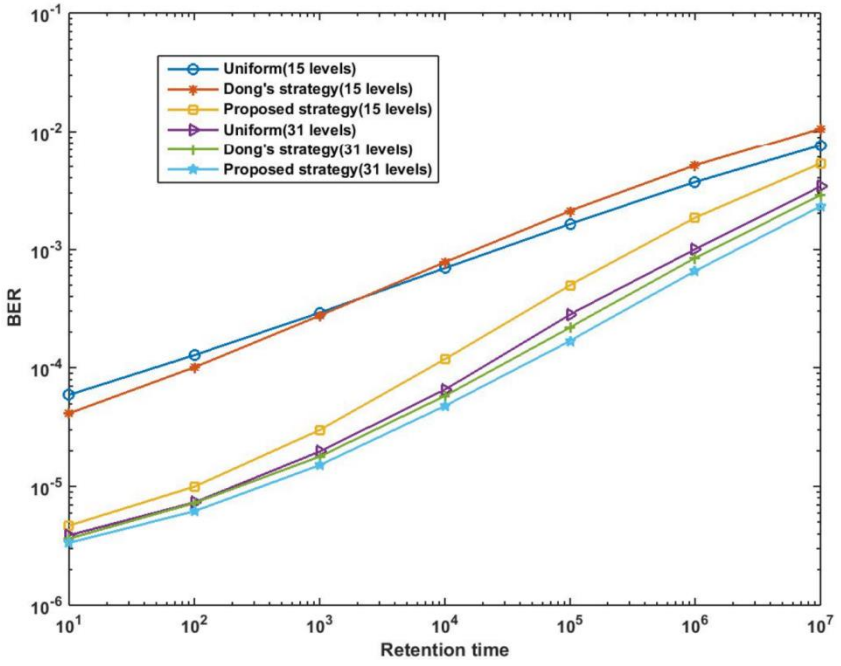


Figure 7. BER performance of LDPC code curves versus retention time for different sensing level strategies, where the coupling strengthen factor $s = 2.0$ and PE cycles equal 10000

Figure 7 shows the BER versus the retention time. It can be observed that, compared with the Dong's nonuniform sensing strategy and the traditional uniform sensing method, the proposed RPL-nonuniform sensing scheme can compensate the impact of retention time on the stability of flash memory cells to a greater extent.

4. Conclusion

In this paper, through Monte Carlo simulation, we discover that the threshold voltage disturbed by various noises (including CCI and RN) approximately to be Gaussian distributed, and then derive an easily-computable mathematical formulation to characterize the distribution of storage states. We further analyze the calculation of LLRs to introduce the parameter of relative-precision-loss ratio, and then propose a new nonuniform memory sensing strategy. With this sensing strategy, a series of reference voltages is obtained, and the soft-decoding LLRs for LDPC codes are evaluated. The proposed RPL-nonuniform sensing scheme still maintains the good anti-noises performance in MLC NAND flash memory while reducing the memory sensing level, which is also verified by a comprehensive simulation and comparison.

Acknowledgments

This work was supported by the Guangdong Provincial NSF under Grant 2021A1515011906. The authors would like to thank Dr. Suihua Cai and Mr. Qianfan Wang for their helps.

References

Cai, Y., Ghose, S., Haratsch, E. F., Luo, Y., & Mutlu, O. (2017). Error characterization, mitigation, and recovery in flash-memory-based solid-state drives[J]. *Proceedings of the IEEE*, *105*(9), 1666-1704. https://doi.org/10.1109/JPROC.2017.2713127

Cai, Y., Luo, Y., Ghose, S., & Mutlu, O. (2015, June). Read disturb errors in MLC NAND flash memory: Characterization, mitigation, and recovery[C]. In 2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (pp. 438-449). IEEE. https://doi.org/10.1109/DSN.2015.49

Cai, Y., Luo, Y., Haratsch, E., Mai, K., & Mutlu, O. (2015). Data retention in MLC NAND flash memory: Characterization, optimization, and recovery. *2015 IEEE 21st International Symposium on High Performance Computer Architecture (HPCA)*, 551-563. https://doi.org/10.1109/HPCA.2015.7056062

Cai, Y., Mutlu, O., Haratsch, E., & Mai, K. (2013). Program interference in MLC NAND flash memory: Characterization, modeling, and mitigation [J]. *2013 IEEE 31st International Conference on Computer Design (ICCD)*, 123-130. https://doi.org/10.1109/ICCD.2013.6657034

Cai, Y., Yalcin, G., Mutlu, O., Haratsch, E. F., Crista, A., Unsal, O. S., & Mai, K. (2013). ERROR ANALYSIS AND RETENTION-AWARE ERROR MANAGEMENT FOR NAND FLASH MEMORY [J]. *Intel Technology Journal*, *17*(1).

Cai, Y., Yalcin, G., Mutlu, O., Haratsch, E., Cristal, A., Unsal, O., & Mai, K. (2012). Flash correct-and-refresh: Retention-aware error management for increased flash memory lifetime [C]. *2012 IEEE 30th International Conference on Computer Design (ICCD)*, 94-101. https://doi.org/10.1109/ICCD.2012.6378623

Clark Jr, G. C., & Cain, J. B. (2013). *Error-correction coding for digital communications*. Springer Science & Business Media. Retrieved from https://dl.acm.org/doi/book/10.5555/1088886

Dong, G., Pan, Y., & Zhang, T. (2014). Using Lifetime-Aware Progressive Programming to Improve SLC NAND Flash Memory Write Endurance [J]. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, *22*, 1270-1280. https://doi.org/10.1109/TVLSI.2013.2267753

Dong, G., Xie, N., & Zhang, T. (2011). On the Use of Soft-Decision Error-Correction Codes in nand Flash Memory [J]. *IEEE Transactions on Circuits and Systems I: Regular Papers*, *58*, 429-439. https://doi.org/10.1109/TCSI.2010.2071990

Dong, G., Xie, N., & Zhang, T. (2013). Enabling NAND Flash Memory Use Soft-Decision Error Correction Codes at Minimal Read Latency Overhead [J]. *IEEE Transactions on Circuits and Systems I: Regular Papers*, *60*, 2412-2421. https://doi.org/10.1109/TCSI.2013.2244361

Gallager, R. (1962). Low-density parity-check codes [J]. *IRE Transactions on information theory*, *8*(1), 21-28. https://doi.org/10.1109/TIT.1962.1057683

Lee, D., & Sung, W. (2013). Estimation of NAND Flash Memory Threshold Voltage Distribution for Optimum Soft-Decision Error Correction [J]. *IEEE Transactions on Signal Processing*, *61*, 440-449. https://doi.org/10.1109/TSP.2012.2222399

Lee, J., Choi, J., Park, D., & Kim, K. (2003). Degradation of tunnel oxide by FN current stress and its effects on data retention characteristics of 90 nm NAND flash memory cells [J]. *2003 IEEE International Reliability Physics Symposium Proceedings, 2003. 41st Annual.*, 497-501.

Lee, J., Hur, S., & Choi, J. (2002). Effects of floating-gate interference on NAND flash memory cell operation [J]. *IEEE Electron Device Letters, 23*, 264-266. https://doi.org/10.1109/55.998871

Lee, S., Kim, H., Bae, J., Yoo, H., Choi, N.Y., Kwon, D., Lim, S., Park, B., & Lee, J. (2019). High-Density and Highly-Reliable Binary Neural Networks Using NAND Flash Memory Cells as Synaptic Devices. *2019 IEEE International Electron Devices Meeting (IEDM)*, 38.4.1-38.4.4. https://doi.org/10.1109/IEDM19573.2019.8993478

Liu, W., Rho, J., & Sung, W. (2006, October). Low-power high-throughput BCH error correction VLSI design for multi-level cell NAND flash memories [J]. *In 2006 IEEE Workshop on Signal Processing Systems Design and Implementation* (pp. 303-308). IEEE. https://doi.org/10.1109/SIPS.2006.352599

Micheloni, R. (2013). *Inside solid state drives (SSDs)* (p. 381). A. Marelli, & K. Eshghi (Eds.). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-5146-0

Micheloni, R., Crippa, L., & Marelli, A. (2010). *Inside NAND flash memories*. Springer Science & Business Media. https://doi.org/10.1007/978-90-481-9431-5

Micheloni, R., Ravasio, R., Marelli, A., Alice, E., Altieri, V., Bovino, A., ... Won, S. (2006, February). A 4Gb 2b/cell NAND flash memory with embedded 5b BCH ECC for 36MB/s system read throughput[J]. In *2006 IEEE International Solid State Circuits Conference-Digest of Technical Papers* (pp. 497-506). IEEE. https://doi.org/10.1109/ISSCC.2006.1696082

Parat, K., & Goda, A. (2018). Scaling Trends in NAND Flash. *2018 IEEE International Electron Devices Meeting (IEDM)*, 2.1.1-2.1.4. https://doi.org/10.1109/IEDM.2018.8614694

Park, K., Kang, M., Kim, D., Hwang, S., Choi, B., Lee, Y., Kim, C., & Kim, K. (2008). A Zeroing Cell-to-Cell Interference Page Architecture With Temporary LSB Storing and Parallel MSB Program Scheme for MLC NAND Flash Memories. *IEEE Journal of Solid-State Circuits, 43*, 919-928. https://doi.org/10.1109/JSSC.2008.917558

Peng, J., Wang, Q., Fu, X., & Huo, Z. (2017). Dynamic LLR scheme based on EM algorithm for LDPC decoding in NAND flash memory [J]. *IEICE Electron. Express, 14*, 20170820. https://doi.org/10.1587/elex.14.20170820

Prall, K. (2007). Scaling Non-Volatile Memory Below 30nm. *2007 22nd IEEE Non-Volatile Semiconductor Memory Workshop*, 5-10. https://doi.org/10.1109/NVSMW.2007.4290561

Shibata, N., Maejima, H., Isobe, K., Iwasa, K., Nakagawa, M., Fujiu, M., Shimizu, T., Honma, M., Hoshi, S., Kawaai, T., Kanebako, K., Yoshikawa, S., Tabata, H., Inoue, A., Takahashi, T., Shano, T., Komatsu, Y., Nagaba, K., Kosakai, M., Motohashi, N., Kanazawa, K., Imamiya, K., Nakai, H., Lasser, M., Murin, M., Meir, A., Eyal, A., & Shlick, M. (2008). A 70 nm 16 Gb 16-Level-Cell NAND flash Memory [J]. *IEEE Journal of Solid-State Circuits, 43*, 929-937. https://doi.org/10.1109/JSSC.2008.917559

Shiga, H., Tanzawa, T., Umezawa, A., Taura, T., Miyaba, T., Saito, M., Kitamura, S., Mori, S., & Atsumi, S. (1999). A sampling weak-program method to tighten Vth-distribution of 0.5 V for low-voltage flash memories [J]. *1999 Symposium on VLSI Circuits. Digest of Papers (IEEE Cat. No.99CH36326)*, 33-36. https://doi.org/10.1109/VLSIC.1999.797226

Suh, K. D., Suh, B. H., Lim, Y. H., Kim, J. K., Choi, Y. J., Koh, Y. N., ... Lim, H. K. (1995). A 3.3 V 32 Mb NAND flash memory with incremental step pulse programming scheme [J]. *IEEE Journal of Solid-State Circuits, 30*(11), 1149-1156. https://doi.org/10.1109/4.475701

Takeuchi, K., Tanaka, T., & Nakamura, H. (1996). A double-level-V/sub th/select gate array architecture for multilevel NAND flash memories [J]. *IEEE Journal of Solid-State Circuits, 31*(4), 602-609. https://doi.org/10.1109/4.499738

Wang, X., Dong, G., Pan, L., Zhou, R., & Stievano, P. I. (2011). Error correction codes and signal processing in flash memory[J]. *Flash Memories*, 57-82. https://doi.org/10.5772/19083

Xu, Q., Gong, P., Chen, T., Michael, J., & Li, S. (2015). Modelling and characterization of NAND flash memory channels [J]. *Measurement, 70*, 225-231. https://doi.org/10.1016/j.measurement.2015.04.003

Zhao, W., Dong, G., Sun, H., Zheng, N., & Zhang, T. (2012). Reducing latency overhead caused by using LDPC codes in NAND flash memory [J]. *EURASIP Journal on Advances in Signal Processing, 2012,* 1-9. https://doi.org/10.1186/1687-6180-2012-203