Dynamic Write-Level and Read-Level Signal Design for MLC NAND Flash Memory

Chaudhry Adnan Aslam
School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore 639801
Email: ad0001ry@e.ntu.edu.sg

Yong Liang Guan
School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore 639801
Email: EYLGuan@ntu.edu.sg

Kui Cai
Data Storage Institute (DSI)
Singapore 117608
Email: Cai_Kui@dsi.astar.edu.sg

***Abstract*—In this paper, we propose dynamic write-level and read-level voltage scheme for MLC NAND flash memory. We study the characteristics of flash channel which is considered as mixture of Uniform and Exponential distribution. Since this channel shows non-stationary behavior and the distribution of programmed cells varies over Program/Erase(PE) operations, we present the probability of error analysis and introduce the concept of dynamically adjusting the verify-level (write-level) and quantization-level (read-level) voltage values over varying channel conditions. The proposed dynamic voltage based method outperforms fixed verify-level voltage scheme. We also demonstrate the improvements in bit-error-rate (BER) performance and cell storage capacity for the proposed dynamic signal design scheme.**

***Index Terms*—MLC NAND Flash, PE, verify-level, quantization-level, BER**

I. Introduction

The Multi-level/Cell (MLC) NAND Flash memory has been widely used as non-volatile storage medium in many digital electronic products. The impressive technology scaling and advancements in chip manufacturing processes have tremendously reduced the cost per bit and size of NAND flash memory and simultaneously increased the storage capacity leading to its dominance into semiconductor storage market. The NAND flash memory cells are organized into an array of rows and columns connected through word-line and bit-line connections respectively. As shown in Fig. 1, each memory cell is made-up of floating-gate transistor which comprises of source, drain, channel and two transistor gates; floating and control gate. There is an insulating medium, known as oxide-layer, between the floating-gate and transistor channel. We store electrons on floating gate through programming the flash cells in which we shift the threshold voltage of MOS transistor. For example, to store 2-bits/cell, we program the flash cell into 4 (2^2) distinct states, also called verify-level or write-levels voltages V_i, each representing a particular stored data symbol $(00, 01, 10, 11)$.

As the feature size of flash cell has been scaled-down, it has also raised concerns over its data reliability [1]. The smaller size of transistor floating-gate results into less number of electrons that can be stored on it and hence the gap between adjacent cell states has also reduced making it more susceptible to errors. Besides, we also observe noise and interference effect on flash memory cells which further reduce the data reliability. Random Telegraph Noise (RTN) [2] is considered as one of the noise factors related with cell PE cycles which affects the cell threshold voltage. As the flash memory is cycled through PE operations, the voltage applied across the transistor terminals, gradually cause damage to the oxide layer. Consequently, multiple traps are generated at the oxide layer which can gain or loose electrons causing fluctuations in cell threshold voltage and resulting into more errors. This RTN effect becomes more severe as the flash experience higher PE cycles and this limits the operational lifetime of flash memory.

To improve the flash reliability, we employ different signal processing and coding schemes [3], [4]. One of the approach is through the Incremental Step Pulse Programming (ISPP) [5] technique which has been widely used to tightly limit the range of threshold voltage values. In ISPP, we apply voltage across the word-line and then compare with the desired verify-level voltage. If the desired voltage is achieved, we stop the programming operation, otherwise, we increase the word-line voltage by step size ΔV_{pp} and continue this process until the cell is completely programmed. With this iterative programming approach, we can program the flash cell very close to the desired verify-level voltage. However, after flash cell is ideally programmed, it is disturbed by RTN and its effect become more adverse as device approaches more PE cycles. Since the channel flash exhibit non-stationary behavior over PE count, is it very important to adaptively tune flash device parameters (write-level, read-level voltages) such that the it can endure more PE cycles.

The rest of the paper is organized as follows. Section II, reviews the probabilistic model of flash channel. Section III presents the probability of error analysis for the flash channel. Based on this analysis, in Section IV, we present the proposed signal design technique which is optimized for non-stationary flash channel. For the proposed scheme, section V estimates cell storage capacity and computes threshold voltage quantization-levels. Conclusions are presented in Section VI.

II. NAND Flash Channel Model

According to [6], the erased cells of MLC NAND flash are modeled with Gaussian distribution with mean voltage

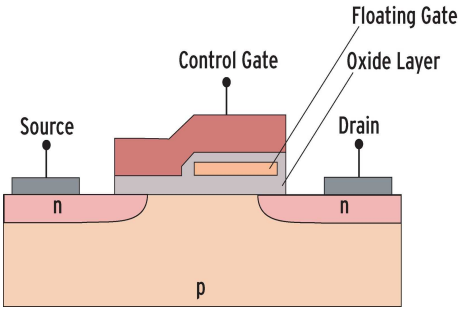

Fig. 1: Flash Memory Cell (Floating Gate Transistor).

V_{min} and variance ${\sigma_e}^2$, whereas, the programmed cells tend to follow uniform distribution between V_i and ΔV_{pp} due to ISPP technique and have symmetric exponential tail below V_i and above $V_i + \Delta V_{pp}$ owing to the effect of RTN [7]. Considering m-bit/cell flash memory, we use P_{s0} and P_{si} for $i = 1$ to 2^{m-1} to represent the distribution of erased and programmed cells respectively as given by (1) and (2).

$$P_{s0}\left(V_{th}\right) = \frac{1}{\sqrt{2\pi{\sigma_e}^2}} exp^{-\frac{\left(V_{th}-V_{min}\right)^2}{2{\sigma_e}^2}} \tag{1}$$

$$P_{si}\left(V_{th}\right) = \begin{cases} \frac{c}{\Delta V_{pp}}\left[1 - exp^{-\frac{\Delta V_{pp}}{\lambda}}\right] exp^{-\frac{\left(V_{si}-V_{th}\right)}{\lambda}}, \\ \qquad\qquad \text{for } V_{th} < V_{si} \\ \frac{c}{\Delta V_{pp}}, \\ \qquad\qquad \text{for } V_{si} \leqslant V_{th} \leqslant V_{si} + \Delta V_{pp} \\ \frac{c}{\Delta V_{pp}}\left[exp^{\frac{\Delta V_{pp}}{\lambda}} - 1\right] exp^{-\frac{\left(V_{th}-V_{si}\right)}{\lambda}}, \\ \qquad\qquad \text{for } V_{th} > V_{si} + \Delta V_{pp} \end{cases} \tag{2}$$

Here λ is RTN distribution parameter ptoportional to PE count in power law fashion [8] and c is normalizing factor given by:

$$c = \frac{1}{1 + \frac{2\lambda}{\Delta V_{pp}}\left(1 - exp^{-\frac{\Delta V_{pp}}{\lambda}}\right)}$$

In Fig. 2, we have plotted the the distribution of 2-bit/cell flash memory. Here $s0$ is the distribution of erased state and $s1, s2, s3$ with V_1, V_2, V_{max} as verify-level threshold voltages for the distribution of programmed states.

III. Probability of Error Analysis

Previously the probability of error for flash memory is reported in [9], however it is based on simplified Gaussian channel. In this paper, we refer to channel model described in preceding section and compute the expression for probability of error. The objective is to determine the optimized values of verify-level voltages which minimize the probability of error. With the help of Fig. 3, we can write the probability of error for 2-bit/cell flash memory with equi-probable symbols as:

$$P_e = \frac{1}{4}\{P\left(e|s0\right) + P\left(e|s1\right) + P\left(e|s2\right) + P\left(e|s3\right)\} \tag{3}$$

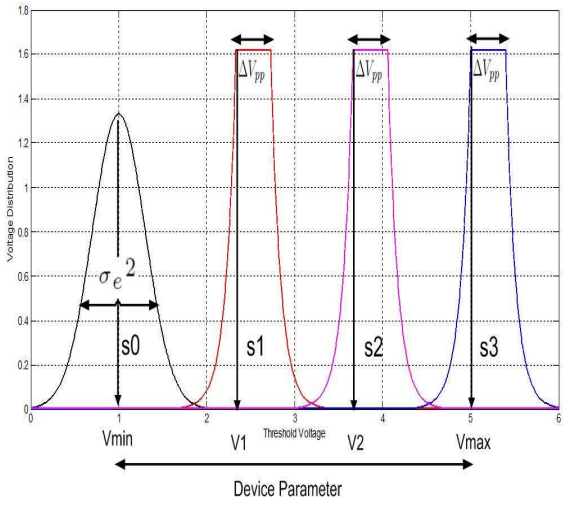

Fig. 2: Illustration of NAND flash channel distribution.

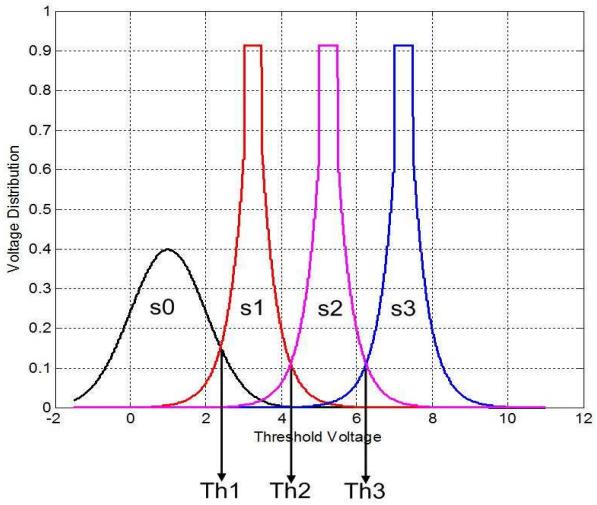

Fig. 3: Decision boundaries between adjacent cell states (symbols).

In the expression, $P(e|s = s0, s1, s2, s3)$ represent the pair-wise probability of error for individual symbols (states) given by:

$$\begin{aligned} P\left(e|s0\right) &= P_{s0}\left(V_{th} > Th_1\right) \\ P\left(e|s1\right) &= P_{s1}\left(V_{th} < Th_1\right) + P_{s1}\left(V_{th} > Th_2\right) \\ P\left(e|s2\right) &= P_{s2}\left(V_{th} < Th_2\right) + P_{s2}\left(V_{th} > Th_3\right) \\ P\left(e|s3\right) &= P_{s3}\left(V_{th} < Th_3\right) \end{aligned}$$

To compute these pair-wise error probabilities, we first require the decision boundaries between adjacent symbols (Th_1, Th_2, Th_3) as shown in Fig. 3. We use the Maximum-Aposteriori-Probability (MAP) detection criterion to compute the decision boundaries and according to that, we get Th_1

by:

$$\begin{aligned} P_{s0}\left(V_{th}=Th_1\right) &= P_{s1}\left(V_{th}=Th_1\right) \\ \frac{1}{\sqrt{2\pi\sigma_e{}^2}}exp^{-\frac{\left(Th_1-V_{min}\right)^2}{2\sigma_e{}^2}} &= \lambda f\left(\lambda\right)exp^{\frac{V1-Th_1}{\lambda}} \\ \text{where } f\left(\lambda\right)=\frac{c}{\Delta V_{pp}}\left[1-exp^{-\frac{\Delta V_{pp}}{\lambda}}\right] & \end{aligned}$$

Solving the above relation for Th_1, we get:

$$Th_1=\frac{-\left(2\sigma_e{}^2-2\lambda V_{min}\right)+\sqrt{\left(2\sigma_e{}^2-2\lambda V_{min}\right)^2-4\lambda h\left(V_1\right)}}{2\lambda}$$

where $h=\left(2\lambda\sigma_e{}^2 log\{f\left(\lambda\right)\sqrt{2\pi\sigma_e{}^2}\}+\lambda V_{min}{}^2-2V_1\sigma_e{}^2\right.$

Similarly, we compute Th_2 and Th_3 using MAP criterion as:

$$\begin{aligned} P_{s1}\left(V_{th}=Th_2\right) &= P_{s2}\left(V_{th}=Th_2\right) \\ \lambda g\left(\lambda\right)exp^{\frac{Th_2-V_1}{\lambda}} &= \lambda f\left(\lambda\right)exp^{\frac{V_2-Th_2}{\lambda}} \\ Th_2 &= \left(\frac{V_1+V_2}{\lambda}-log\{\frac{f(\lambda)}{g(\lambda)}\}\right)\frac{\lambda}{2} \end{aligned}$$

$$\begin{aligned} P_{s2}\left(V_{th}=Th_3\right) &= P_{s3}\left(V_{th}=Th_3\right) \\ \lambda g\left(\lambda\right)exp^{\frac{Th_3-V2}{\lambda}} &= \lambda f\left(\lambda\right)exp^{\frac{V_{max}-Th_3}{\lambda}} \\ Th_3 &= \left(\frac{V_2+V_{max}}{\lambda}-log\{\frac{f(\lambda)}{g(\lambda)}\}\right)\frac{\lambda}{2} \end{aligned}$$

where $g\left(\lambda\right)=\frac{c}{\Delta V_{pp}}\left[exp^{\frac{\Delta V_{pp}}{\lambda}}-1\right]$

Once we have computed the decision boundaries, the pairwise error probabilities are given by:

$$\begin{aligned} P\left(e|s0\right) &= P_{s0}\left(V_{th}>Th_1\right) \\ P\left(e|s0\right) &= Q\left(\frac{Th_1-V_{min}}{\sigma_e}\right) \end{aligned}$$

$$\begin{aligned} P\left(e|s1\right) &= P_{s1}\left(V_{th}<Th_1\right)+P_{s1}\left(V_{th}>Th_2\right) \\ P\left(e|s1\right) &= \lambda f\left(\lambda\right)exp^{\frac{Th_1-V_1}{\lambda}}+\lambda g\left(\lambda\right)exp^{\frac{V_1-Th_2}{\lambda}} \end{aligned}$$

$$\begin{aligned} P\left(e|s2\right) &= P_{s2}\left(V_{th}<Th_2\right)+P_{s2}\left(V_{th}>Th_3\right) \\ P\left(e|s2\right) &= \lambda f\left(\lambda\right)exp^{\frac{Th_2-V_2}{\lambda}}+\lambda g\left(\lambda\right)exp^{\frac{V_2-Th_3}{\lambda}} \end{aligned}$$

$$\begin{aligned} P\left(e|s3\right) &= P_{s3}\left(V_{th}<Th_3\right) \\ P\left(e|s3\right) &= \lambda f\left(\lambda\right)exp^{\frac{Th_3-V_{max}}{\lambda}} \end{aligned}$$

where $Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^{\infty} exp^{-\frac{1}{z^2}}dz$

Using pairwise error probabilities, (3) becomes:

$$\begin{aligned} P_e\left(V_1,V_2\right) &= \frac{1}{4}\{P\left(e|s0\right)+P\left(e|s1\right)+P\left(e|s2\right)+P\left(e|s3\right)\} \\ &= Q\left(\frac{Th_1-V_{min}}{\sigma_e}\right) \\ &+\lambda f\left(\lambda\right)exp^{\frac{Th_1-V_1}{\lambda}}+\lambda g\left(\lambda\right)exp^{\frac{V_1-Th_2}{\lambda}} \\ &+\lambda f\left(\lambda\right)exp^{\frac{Th_2-V_2}{\lambda}}+\lambda g\left(\lambda\right)exp^{\frac{V_2-Th_3}{\lambda}} \\ &+\lambda f\left(\lambda\right)exp^{\frac{Th_3-V_{max}}{\lambda}} \end{aligned} \quad (4)$$

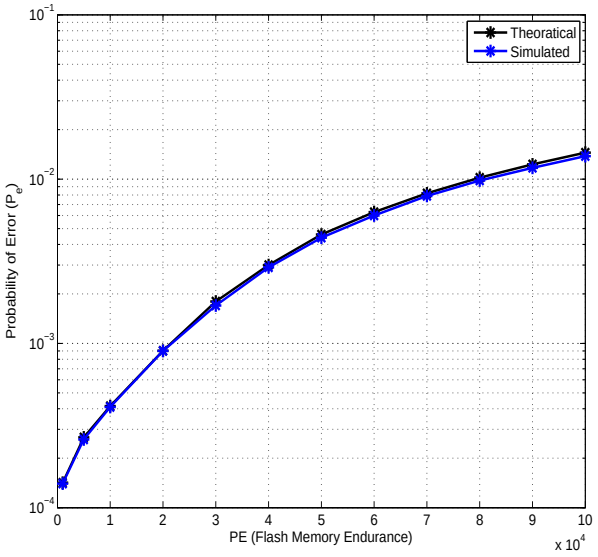

Fig. 4: Comparison of Theoretical and Simulated P_e with flash parameter taken from [10]: $V_{min}=1.4, V_1=2.6, V_2=3.2, V_{max}=3.93, V_p=0.2, \sigma_e=0.35, \lambda=0.00025\left(PE\right)^{0.5}$.

To verify the analysis, we have plotted P_e for simulated NAND flash and compared with theoretical P_e given by (4) as shown in Fig. 4. We can see that both curves are very close to each other validating the accuracy of (4).

IV. Dynamic Verify-level Voltage Scheme

In this paper, we dynamically adjust the verify-level threshold voltage values based on the channel condition. Since the flash channel is non-stationary and varies over PE cycles, we would like to modify the verify-level voltages such that the overall probability of error is minimized. Until now, most of the research has been done on estimating the desired quantization-levels (read-level) based on the flash channel model [9], [11], [12], however, we propose to compute and adjust both verify-level (write-level) and quantization-level (read-level) voltage values according to channel condition. Given the channel parameters σ_e and λ and mean threshold voltage of erased V_{min} and highest programmed state V_{max} which is device dependent parameter, we would like to optimize V_1 and V_2 by minimizing probability of error (4). We write the cost function as:

$$\left(V_1^*,V_2^*\right)=\max_{\left(V_1,V_2\right)} P_e(V_1,V_2) \quad (5)$$

In Fig. 5, we have plotted P_e as function of V_1 and V_2. We can observe that P_e is convex function. Therefore, we can use any convex optimization technique to minimize this function. We partially differentiate (4) w.r.t. V_1 and V_2, equate to 0 and simplify for V_1^* and V_2^* as:

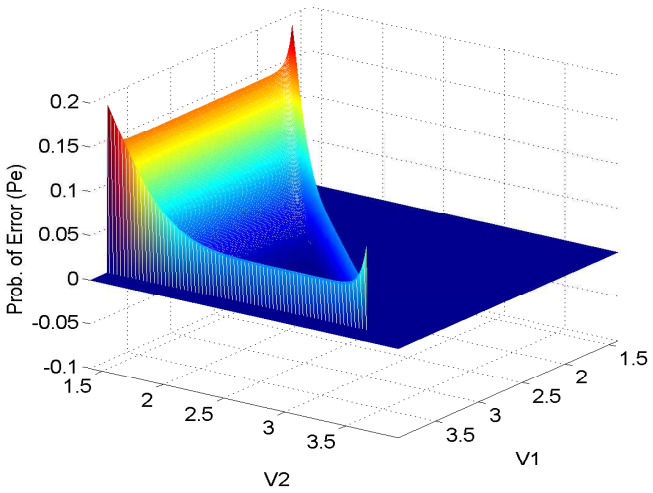

Fig. 5: Plot of P_e as function of V_1 and V_2 with simulation parameters: $V_{min} = 1.4, V_1 = 2.6, V_2 = 3.2, V_{max} = 3.93, V_p = 0.2, \sigma_e = 0.35, PE = 10K, \lambda = 0.00025\,(PE)^{0.5}$.

$$\frac{\partial P_e(V1,V2)}{\partial V1} = 0$$

$$\frac{1}{2}\{g(\lambda)\,exp^{0.5log\left(\frac{f(\lambda)}{g(\lambda)}\right)} + f(\lambda)\,exp^{0.5log\left(\frac{g(\lambda)}{f(\lambda)}\right)}\}exp^{\frac{V_1-V_{max}}{4\lambda}} = \sqrt{\frac{2}{\pi\,(b^2-4ac)}}exp^{\frac{(Th_1-V_{min})^2}{2{\sigma_e}^2}} + f(\lambda)\,exp^{\frac{Th_1-V_1}{\lambda}}\left(1-\frac{2{\sigma_e}^2}{\sqrt{b^2-4ac}}\right) \quad (6)$$

Similarly,

$$\frac{\partial P_e(V1,V2)}{\partial V2} = 0$$

$$V_2^* = \frac{V_1^* + V_{max}}{2} \quad (7)$$

With (6) and (7), we can numerically evaluate the optimized verify-level values for non-stationary channel over PE cycles as gives in Table I. In Fig. 6, we have plotted the probability of error using the optimized verify-level voltages for different PE cycles. For comparison, we have also shown error plot for fixed verify-level scheme. We can clearly observe the potential performance gain with the proposed dynamic voltage scheme over the fixed one. Although, for fixed voltage scheme, we have adjusted the decision boundaries given Th_1, Th_2, Th_3 for different PE count, still it has higher probality of error. Even if we select any other fixed verify-level voltage values, its performance will turn out to be inferior to dynamically optimized threshold voltage scheme.

V. Effect on Cell Storage Efficiency and Quantization-Levels

In this section, we investigate the proposed signal design technique from information theoretic perspective. We compute the cell storage capacity/efficiency of 2-bit/cell flash memory

TABLE I: Optimized verify-level voltage using (6) and (7). Simulation parameter: $V_{min} = 1.0, V_{max} = 5.0, V_p = 0.2, \sigma_e = 0.75$

PE Cycles	V_1^*	V_2^*	P_e
1000	3.1383	3.5342	5.38×10^{-7}
5000	2.9085	3.4193	2.81×10^{-5}
10000	2.7998	3.3649	1.47×10^{-4}
20000	2.6903	3.3102	6.82×10^{-4}
30000	2.6276	3.2788	0.0015
40000	2.5843	3.2572	0.0027
50000	2.5517	3.2409	0.0040

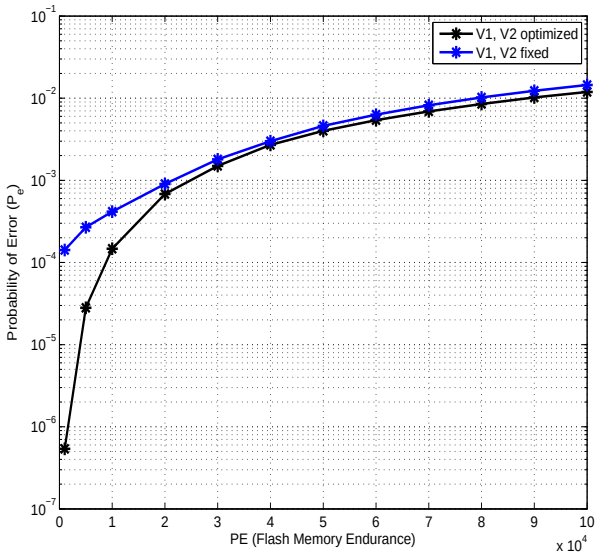

Fig. 6: Comparison of dynamic vs. fixed verify-level voltage scheme. Simulation parameter: $V_{min} = 1.4, V_{max} = 3.93, V_p = 0.2, \sigma_e = 0.35$, black curve($V_1^*, V_2^*$ from TABLE I), blue curve($V_1 = 2.6, V_2 = 3.2$).

for both dynamic and fixed verify-level voltage schemes. We use mutual information to represent the capacity of flash memory cell as:

$$C = I(X;Y) = H(Y) - H(Y|X)$$

Here $I(X,Y)$ is the mutual information between input symbol $X \in \{s_0, s_1, s_2, s_3\}$ and output threshold voltage $Y \in (-\infty, \infty)$. The terms $H(Y)$ and $H(Y|X)$ represent the entropy and conditional entropy of output Y respectively and given as:

$$H(Y) = \int_{-\infty}^{\infty} p(y) log(\tfrac{1}{p(y)})$$
$$H(Y|X) = \sum_{x\in\{s_0,s_1,s_2,s_3\}} p(x) \int_{-\infty}^{\infty} p(y|x) log(\tfrac{1}{p(y|x)})$$

Using these relations, the capacity can be written as:

$$C = \sum_{x\in\{s_0,\ldots,s_3\}} p(x) \int_{-\infty}^{\infty} p(y|x) log(\frac{p(y|x)}{\sum\limits_{x\in\{s_0,\ldots,s_3\}} \{p(x)p(y|x)\}})dy \tag{8}$$

In (8), the quantity $p(y|s_0)$ represents the conditional probability of erased cell given by (1) and $p(y|s_1, s_2, s_3)$ is the probability of programmed cell given by (2). This capacity equation is used for soft information (infinite-level quantization), however, we also want to compute the capacity based on finite-level quantization since we can only sense (read) the flash memory at some finite discrete voltage levels. For this purpose, we adopt the idea presented in [13] where author has chosen the optimized discrete quantization levels that maximize the capacity expression. However, our work is distinct from [13] as we are using a different flash channel model. To find the quantization levels, we divide the output threshold voltage values into discrete regions as depicted in Fig. 7. In this figure, we have shown 6-levels of quantization $(y_1, y_2, y_3, y_4, y_5, y_6)$ with 7 non-overlapping regions $(R_1, R_2, R_3, R_4, R_5, R_6, R_7)$ such that:

$$R1 = (-\infty, y_1], R2 = (y_1, y_2], R3 = (y_2, y_3], R4 = (y_3, y_4],$$
$$R5 = (y_4, y_5], R6 = (y_6, y_7], R7 = (y_7, \infty)$$

After splitting the threshold voltage values into discrete regions, we can re-write (8) as:

$$C = \sum_{x\in\{s_0,\ldots,s_3\}} p(x)$$
$$\{\sum_{R\in\{R_1,\ldots,R_7\}} \int_R p(y|x) log(\frac{p(y|x)}{\sum\limits_{x\in\{s_0,\ldots,s_3\}} \{p(x)p(y|x)\}})dy\}$$

Now the objective is to maximize C over $(y_1, y_2, y_3, y_4, y_5, y_6)$:

$$(y_1^*, y_2^*, y_3^*, y_4^*, y_5^*, y_6^*) = \max_{(y_1,y_2,y_3,y_4,y_5,y_6)} C \tag{9}$$

We can use any optimization technique for (9), however, we have used particle swarm optimization (PSO) for this purpose. With similar approach we can modify (9) and optimize for any specific number of quantization levels. Using (8) and (9), we have plotted the capacity expression over different PE cycles with soft information (infinite-level) and PSO optimized 6-level, 15-level quantization for both proposed and fixed voltage schemes as shown in Fig. 8. In fixed voltage scheme, we have optimized quantization levels but verify-level voltage values are fixed. The proposed scheme has better cell storage capacity over the fixed voltage scheme for different quantization levels.

In flash memory system, we always use some error correction coding (ECC) to improve the bit-error-rate (BER) performance. The modern ECC schemes require high-precision quantization levels to yield better performance, but higher quantization levels incur more memory sensing latency. Therefore, we desire to obtain the benefits of ECC scheme with minimum quantization levels. In this paper, we have used Low-Density Parity Check (LDPC) code over 2-bit/cell flash memory whereby the quantization levels are obtained as stated previously. In Fig. 9 and Fig. 10, we have shown the probability of bit/frame error for LDPC coded system with code rate 0.91 and block length $4544bits$. For the fixed voltage scheme, the best possible performance is obtained with infinite-level quantization (black dotted curve) whereas we can perform closer to that with proposed scheme and only 15-level of quantization (red solid curve). Hence, we can reduce the memory sensing latency by using less quantization levels and at the same time improve the data reliability with the implementation of proposed signal design scheme.

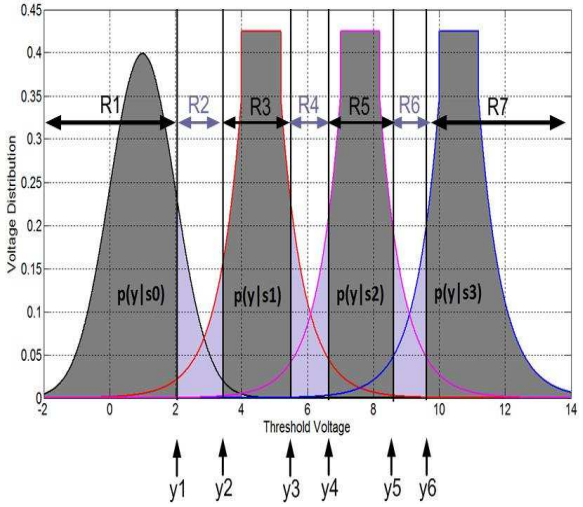

Fig. 7: Threshold Voltage spectrum is split into discrete regions.

VI. Conclusion

In this paper, we have proposed the concept of adjusting the verify-level voltage values for MLC NAND flash memory. As the flash channel changes with PE count, it is desirable to tune the flash parameters accordingly. In previous works, the research has been carried over simplified Gaussian channel, however, we have analyzed the mixture of Uniform and symmetric exponential model. Based on the error analysis, we have proposed verify-level (write-level) values that are optimized for non-stationary flash channel. We have also found the optimized threshold-level (read-level) values by maximizing mutual information. With cell storage capacity and bit-error-rate simulations, we have shown that the proposed scheme performs better than the fixed verify-level voltage scheme. Using this system level signal design technique, we expect that flash memory performance can be further improved.

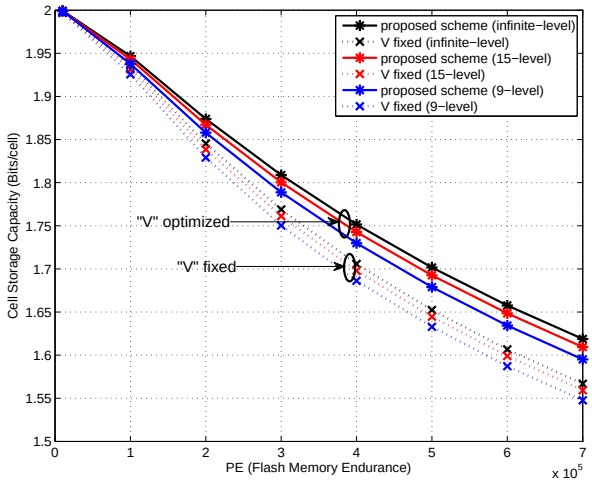

Fig. 8: Plot of cell storage capacity using dynamic and fixed verify-level voltage schemes. Simulation parameters: $V_{min} = 1.4, V_{max} = 3.93, V_p = 0.2, \sigma_e = 0.35$, doted curves (fixed voltage $V_1 = 2.6, V_2 = 3.2$).

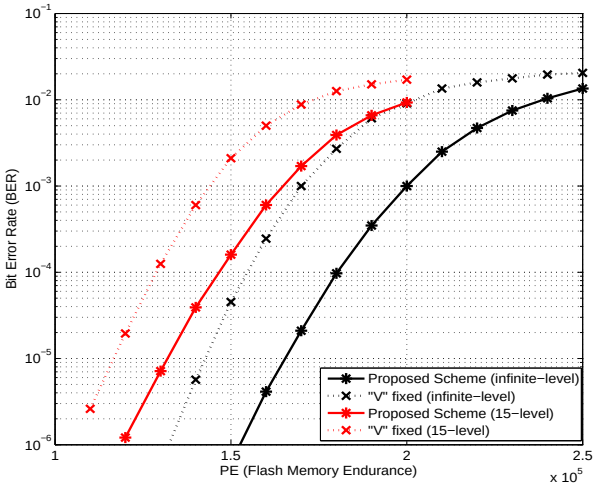

Fig. 9: Illustration of Bit-Error-Rate Performance of proposed and fixed voltage scheme. Simulation parameters: $V_{min} = 1.4, V_{max} = 3.93, V_p = 0.2, \sigma_e = 0.35$, doted curves (fixed voltage $V_1 = 2.6, V_2 = 3.2$).

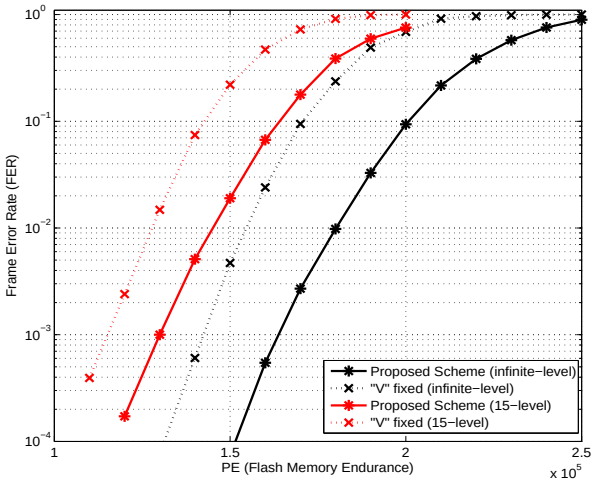

Fig. 10: Illustration of Frame-Error-Rate Performance of proposed and fixed voltage scheme. Simulation parameters: $V_{min} = 1.4, V_{max} = 3.93, V_p = 0.2, \sigma_e = 0.35$, doted curves (fixed voltage $V_1 = 2.6, V_2 = 3.2$).

REFERENCES

[1] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, "Introduction to flash memory," *Proceedings of the IEEE*, vol. 91, no. 4, pp. 489–502, 2003.

[2] S.-M. Joe, J.-H. Yi, S.-K. Park, H. Shin, B.-G. Park, Y.-J. Park, and J.-H. Lee, "Threshold voltage fluctuation by random telegraph noise in floating gate nand flash memory string," *Electron Devices, IEEE Transactions on*, vol. 58, no. 1, pp. 67–73, 2011.

[3] G. Dong, N. Xie, and T. Zhang, "Enabling nand flash memory use soft-decision error correction codes at minimal read latency overhead," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 60, no. 9, pp. 2412–2421, 2013.

[4] Q. Huang, S. Lin, and K. A. S. Abdel-Ghaffar, "Error-correcting codes for flash coding," *Information Theory, IEEE Transactions on*, vol. 57, no. 9, pp. 6097–6108, 2011.

[5] K.-D. Suh, B.-H. Suh, Y.-H. Lim, J.-K. Kim, Y.-J. Choi, Y.-N. Koh, S.-S. Lee, S.-C. Kwon, B.-S. Choi, J.-S. Yum, J.-H. Choi, J.-R. Kim, and H.-K. Lim, "A 3.3 v 32 mb nand flash memory with incremental step pulse programming scheme," *Solid-State Circuits, IEEE Journal of*, vol. 30, no. 11, pp. 1149–1156, 1995.

[6] K. Takeuchi, T. Tanaka, and H. Nakamura, "A double-level-vth select gate array architecture for multilevel nand flash memories," *Solid-State Circuits, IEEE Journal of*, vol. 31, no. 4, pp. 602–609, 1996.

[7] C. Monzio Compagnoni, M. Ghidotti, A. Lacaita, A. Spinelli, and A. Visconti, "Random telegraph noise effect on the programmed threshold-voltage distribution of flash memories," *Electron Device Letters, IEEE*, vol. 30, no. 9, pp. 984–986, 2009.

[8] G. Dong, Y. Pan, N. Xie, C. Varanasi, and T. Zhang, "Estimating information-theoretical nand flash memory storage capacity and its implication to memory system design space exploration," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 20, no. 9, pp. 1705–1714, 2012.

[9] Y. Kim, J. Kim, J. J. Kong, B. K Vijaya Kumar, and X. Li, "Verify level control criteria for multi-level cell flash memories and their applications," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, p. 196, 2012. [Online]. Available: http://asp.eurasipjournals.com/content/2012/1/196

[10] G. Dong, Y. Pan, N. Xie, C. Varanasi, and T. Zhang, "Estimating information-theoretical nand flash memory storage capacity and its implication to memory system design space exploration," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 20, no. 9, pp. 1705–1714, Sept 2012.

[11] G. Dong, N. Xie, and T. Zhang, "On the use of soft-decision error-correction codes in nand flash memory," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 58, no. 2, pp. 429–439, 2011.

[12] W. Kang, Y. Zhang, M. Wang, and G. Li, "Improving flash memory reliability with dynamic thresholds: Signal processing and coding schemes," in *Communications and Networking in China (CHINACOM), 2012 7th International ICST Conference on*, 2012, pp. 161–166.

[13] J. Wang, T. Courtade, H. Shankar, and R. Wesel, "Soft information for ldpc decoding in flash: Mutual-information optimized quantization," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, 2011, pp. 1–6.