BER Analysis of MLC NAND Flash Memories based on an Asymmetric PAM Model

Stelios Korkotsides, Georgios Bikas, Efstratios Eftaxiadis and Theodore Antonakopoulos
University of Patras
Department of Electrical and Computer Engineering
Patras 26504, Greece
e-mails: *stelkork@ece.upatras.gr*, *ece6805@upnet.gr*, *ece6740@upnet.gr* and *antonako@upatras.gr*

***Abstract*—The reliability of multilevel NAND Flash memories, which are used extensively on solid-state drives, is strongly affected by their aging, ie. the number of applied program/erase cycles (P/E). A multilevel memory uses discrete voltage levels to represent the various bit patterns and at the beginning of the life-time of such a device these voltage levels demonstrate distributions with very small variances, resulting to very low bit-error-ratio (BER). As the number of applied P/E cycles increases, the variance of the voltage levels also increases and that results to increased BER. In this paper, we present a general model for four-level NAND Flash memories that can be used to estimate memories' BER as a function of the used NAND technology and the aging process. For that purpose, we use asymmetric Pulse Amplitude Modulation with data-dependent channel noise and we provide analytic expressions for the behavior of such memories, and we associate aging with noise conditions and used technology.**

***Index Terms*—NAND Flash, aging, asymmetric PAM, voltage distributions.**

I. Introduction

The last few years, remarkable technological achievements have been demonstrated in the area of NAND Flash memory. Storage density and capacity have been increased and production cost has been suppressed, resulting to new storage devices that exceed other conventional storage technologies. The main reasons that led to this achievement are the supported high data transfer rates, the adjustment of endurance of various NAND technologies to specific application needs and the low power dissipation that has been achieved, especially during erasing and programming.

In order to satisfy the constantly increasing storage needs, scaling of silicon technology and the use of Multi-level Cell (MLC) was employed [1]. Especially the achievements of MLC increased the storage density and reduced the cost per bit significantly, but also affected the complexity, reliability and endurance of the devices. Except of the interference of adjacent cells due to scaling, the challenge of low probability of erroneous detection between adjacent voltage levels was also introduced.

In this work, we present a model for studying the probability of errors on MLC NAND Flash memories. In this model, we use asymmetric Pulse Amplitude Modulation with data-dependent channel noise. The introduced noise is additive, follows the gaussian distribution but its variance depends on the voltage level used during programming, thus it is data dependent. This noise model is based on measurements in [1] and appears in several experimental studies of non-volatile memories, such as [2], [3], [4], [5]. As it will be shown, there is a strong similarity between this model and the 4-states Pulse Amplitude Modulation (4-PAM) scheme that is used to transmit data over an additive, white gaussian (AWGN) communication channel [6]. The main differences between the typical 4-PAM communication scheme and the proposed model of the NAND Flash memory cell is that the inter-symbol distance between adjacent levels is not fixed, but depends on the voltage levels, and statistical characteristics of the introduced noise depend on the symbol that was written. That is why we use the term Asymmetrical 4-PAM for the analysis that follows in the next sections. The main goal of this work is to study the behavior of critical quantities, such as SNR and BER, when the characteristics of memory cells change as the number of programming cycles increases. In this effort, we also calculate the optimum detection thresholds between adjacent levels and in all measurements these optimum thresholds are used for bit error ratio (BER) calculations.

Section II presents the basic model used in our analysis, while in Section III we present the mathematical analysis of the used model. More specifically, subsection III-A presents the calculation of the optimum threshold voltage in adjacent levels that use different noise distributions, while subsection III-B extends the previous analysis to a 4-levels NAND Flash cell. Subsection III-C presents the analysis of BER and the signal-to-noise ratio (SNR) of such a cell. Section IV highlights the effect of the various parameters of the above model.

II. MLC NAND Flash Model

In a NAND Flash memory cell, several noise sources affect the achieved programming voltages and the sensed voltages during read. These noise sources force the level voltages to fluctuate around their nominal values. These fluctuations can be treated as stochastic signals which can be represented quite accurately with four gaussian distributions as shown in Fig. 1. This has been confirmed with measurements of the initially implemented MLC NAND flash memory [1] and, ever since, it has become the mainstream model of MLC NAND Flash memory cells. Moreover, it has been noticed that the

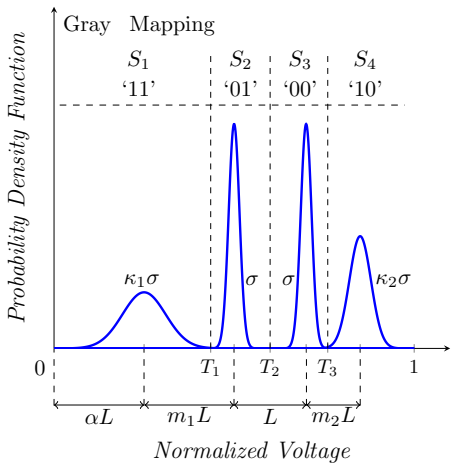

Fig. 1. Voltage distributions of a 4-levels cell.

statistical characteristics of these voltage variations change as the lifetime of the device progresses. More specifically, the standard deviation of the sensed voltages increases and their mean values shift to higher values.

The voltage distributions of Fig. 1 are based on[7], but in order to be able to include in our model the aging effect, as well as to represent as many technologies as possible, we upgraded it to a fully parameterized model, ie. the noises are defined at each voltage level and the inter-symbol distances are not constant. For the following analysis, the level voltages have been normalized to the dynamic range of the used sensing circuit. Therefore, the standard deviations of the four distributions are defined as $\sigma_1 = \kappa_1\sigma$, $\sigma_2 = \sigma_3 = \sigma$ and $\sigma_4 = \kappa_2\sigma$.

This approach can cover all different NAND Flash technologies presented in the literature. For example, based on [8], $\kappa_1 = 2$ and $\kappa_2 = 1$ or in [9], $\kappa_1 = 1.5$ and $\kappa_2 = 1.2$. In other cases, like those found in [10] and [7], other parameters are used: $\kappa_1 = 4$ and $\kappa_2 = 2$. Following the same approach, the nominal voltage levels are expressed as:

$$\begin{aligned} V_1 &= \alpha L \\ V_2 &= (\alpha + m_1)L \\ V_3 &= (\alpha + m_1 + 1)L \\ V_4 &= (\alpha + m_1 + m_2 + 1)L \end{aligned} \tag{1}$$

In order to achieve lower BER, Gray coding is usually used, so that adjacent symbols differ in only one bit [7]. During voltage sensing, voltage thresholds between adjacent levels are specified that result to optimum detection. If V_i and V_{i+1} are two adjacent levels ($V_{i+1} > V_i$), the optimum voltage threshold is the voltage value that minimizes the probability of erroneous level detection, and this is presented in the first part of next section.

III. Theoretical Analysis of Asymmetric PAM

A. Voltage thresholds between adjacent cells

We consider two adjacent voltage levels which are affected by different noise sources, as shown in Fig. 2. The following

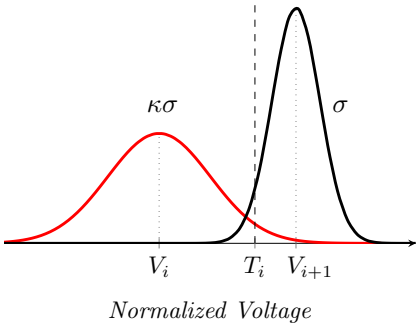

Fig. 2. Voltage distributions and threshold in two adjacent levels.

analysis can be applied in any aging condition of a cell [11].

Let $p_e()$ denotes the probability of erroneous level detection. Using the well-known gaussian distribution we get:

$$\begin{aligned} p_e(T_i) = p_i \frac{1}{\kappa\sigma\sqrt{2\pi}} \int_{T_i}^{+\infty} e^{-\frac{(x-V_i)^2}{2(\kappa\sigma)^2}} dx \quad + \\ + p_{i+1} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{T_i} e^{-\frac{(x-V_{i+1})^2}{2\sigma^2}} dx \end{aligned} \tag{2}$$

where p_i and p_{i+1} are the probabilities to store information at levels i and $i+1$ respectively. In order to determine the threshold that minimizes the probability of error, we have

$$\begin{aligned} &\frac{dp_e}{dT_i} = 0 \Rightarrow \\ &\left(1-\kappa^2\right)T_i^2 + 2\left(\kappa^2 V_{i+1} - V_i\right)T_i + \\ &+ \left[V_i^2 - \kappa^2 V_{i+1} + 2\,\sigma^2 \ln\left(\frac{\kappa p_{i+1}}{p_i}\right)\right] = 0 \end{aligned} \tag{3}$$

B. Multi-level voltage thresholds

In order to analyze the 4-levels cell, we use the previous analysis and we extended it to the three different cases of adjacent levels shown in Fig. 1. Therefore, T_1, T_2 and T_3 are the roots of the following equations:

$$\begin{aligned} &\left(1-\kappa_1^2\right)T_1^2 + 2\left(\kappa_1^2 V_2 - V_1\right)T_1 + \\ &+ \left[(V_1)^2 - \kappa_1^2 V_2^2 + 2\left(\kappa_1\sigma\right)^2 \ln\left(\frac{\kappa_1 p_2}{p_1}\right)\right] = 0 \end{aligned} \tag{4}$$

$$2\left(V_3 - V_2\right)T_2 + \left[2\sigma^2 \ln\left(\frac{p_3}{p_2}\right) + V_2^2 - V_3^2\right] = 0 \tag{5}$$

$$\begin{aligned} &\left(1 - \frac{1}{\kappa_2^2}\right)T_3^2 + 2\left(\frac{1}{\kappa_2^2}V_4 - V_3\right)T_3 + \\ &+ \left[(V_3)^2 - \frac{1}{\kappa_2^2}V_4^2 + 2\left(\kappa_2\sigma\right)^2 \ln\left(\frac{p_4}{\kappa_2 p_3}\right)\right] = 0 \end{aligned} \tag{6}$$

C. BER and SNR in Gray coding MLC

The most important characteristic of Gray codes is that adjacent symbols differ in only one bit. This makes them very attractive to be used inside a NAND Flash cell, since the probability of erroneously detecting a voltage level, that does not belong to adjacent levels during sensing is very low and in our BER analysis we consider the Gray mapping shown in Fig. 1.

The probability of bit error when symbol S_1 is detected is given by:

$$\begin{aligned} P(e_b|S_1) =& \frac{1}{\sqrt{2\pi}\kappa_1\sigma}\left[\frac{1}{2}\int_{T_1}^{T_2} e^{-\frac{(x-V_1)^2}{2\kappa_1^2\sigma^2}}dx + \int_{T_2}^{T_3} e^{-\frac{(x-V_1)^2}{2\kappa_1^2\sigma^2}}dx + \right. \\ &\left. + \frac{1}{2}\int_{T_3}^{+\infty} e^{-\frac{(x-V_1)^2}{2\kappa_1^2\sigma^2}}dx\right] = \\ =& \frac{1}{4}\left[erfc\left(\frac{T_1-V_1}{\sqrt{2}\kappa_1\sigma}\right) + erfc\left(\frac{T_2-V_1}{\sqrt{2}\kappa_1\sigma}\right) - \right. \\ &\left. - erfc\left(\frac{T_3-V_1}{\sqrt{2}\kappa_1\sigma}\right)\right] \end{aligned} \tag{7}$$

where $erfc(x) \equiv \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt$.

Using the same methodology, we calculate the bit error probabilities when symbols S_2, S_3 and S_4 are detected. We now can compute the total probability of bit error by weighting the above probabilities of error with the probability of writing the respective symbols to the memory cell. For example, p_1 is the probability of programming $S_1 = 11$, while p_4 is the probability of programming $S_4 = 10$. Therefore:

$$P(e_b) = p_1P(e_b|S_1)+p_2P(e_b|S_2)+p_3P(e_b|S_3)+p_4P(e_b|S_4) \tag{8}$$

The signal to noise ratio of the sensed voltages is defined as:

$$\begin{aligned} SNR \equiv& \frac{\mathcal{P}_s}{\mathcal{P}_{noise}} = \frac{\mathcal{E}_s}{\mathcal{E}_{noise}} \\ =& \frac{p_1V_1^2 + p_2V_2^2 + p_3V_3^2 + p_4V_4^2}{p_1(\kappa_1\sigma)^2 + p_2\sigma^2 + p_3\sigma^2 + p_4(\kappa_2\sigma)^2} \end{aligned} \tag{9}$$

which is derived by treating symbols as discrete signals and taking into account that the power spectral density of AWGN is equal to its variance.

IV. Model Parameters and Performance Results

In order to study the effect of the various parameters of the above model, we consider various NAND Flash technologies and various probabilities of the programmed data. Table I shows the various NAND Flash technologies and the respective noise distributions, while Table II shows the symbol probabilities for various bit probabilities, where ρ is the probability of bit '1'.

Using the parameters ρ, σ, κ_1, κ_2, α, L, m_1, m_2 and the optimum voltage thresholds determined by using equations (4), (5), (6), we can estimate the expected BER as a function of σ and we can associate σ with SNR.

TABLE I
NAND Flash Technologies

Level	3	2	1	0
Mean	0.2	0.525	0.655	0.85
Deviation (1)	4σ	σ	σ	2σ
Deviation (2)	4σ	σ	σ	σ
Deviation (3)	σ	σ	σ	σ

TABLE II
Symbol Probabilities

	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$
$P(S_1)$	0.0625	0.25	0.5625
$P(S_2)$	0.1875	0.25	0.1875
$P(S_3)$	0.5625	0.25	0.0625
$P(S_4)$	0.1875	0.25	0.1875

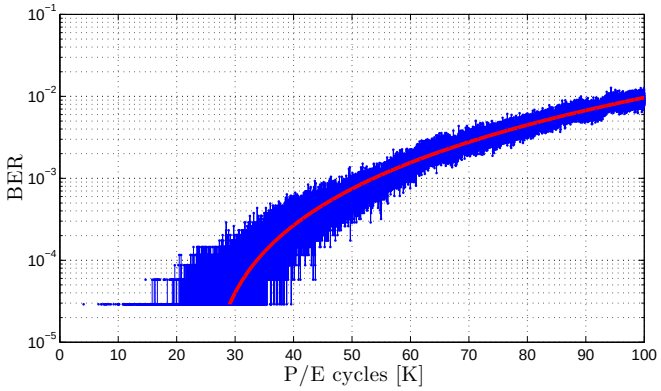

Fig. 3. MLC NAND Flash experimental BER versus P/E cycles.

Fig. 3 shows experimental results from a typical 4-levels NAND Flash memory and how BER is associated with the aging on the memory. Aging is expressed by the number of P/E cycles. The endurance of the experimental memory is 10K P/E cycles and during this experiment we stressed the device an order of magnitude higher than its typical endurance. Depending on the used error correcting codes, the lifetime of such a device is less than 50K P/E cycles.

Using the analysis of Section III, we can study the BER of NAND cells modeled as asymmetric 4-PAM. Fig. 4 shows BER as a function of σ. Comparing the curves of Figs. 3 and 4 we can easily conclude that the presented model describes adequately a MLC NAND Flash memory. Similarly, Fig. 5 shows how SNR is associated with σ for different bit probabilities. Combining the above two figures, we result to Fig. 6 which shows the relation of SNR with BER for various bit and symbol probabilities. As it is shown in this figure, due to the asymmetric nature of the PAM model that accurately represents the NAND Flash cells, the use of line precoding may improve the BER performance at the cost of reduced storage efficiency due to overhead introduced by the line coding that results to non-equal bit probabilities.

Finally, using the NAND experimental results and the theoretical analysis, we can associate aging (P/E cycles) with σ. Fig. 7 highlights this relation for different NAND models.

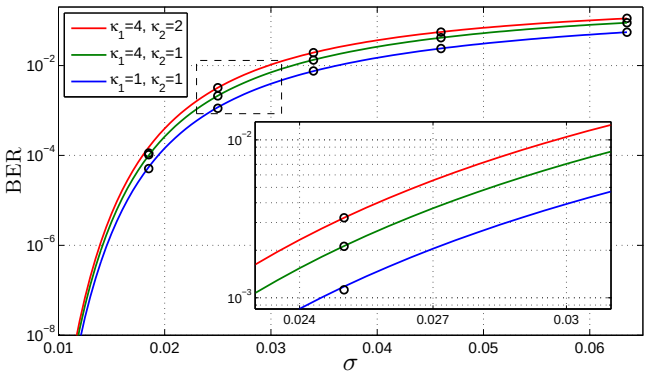

Fig. 4. BER as a function of σ for non-equally probable symbols ($\rho = 0.25$).

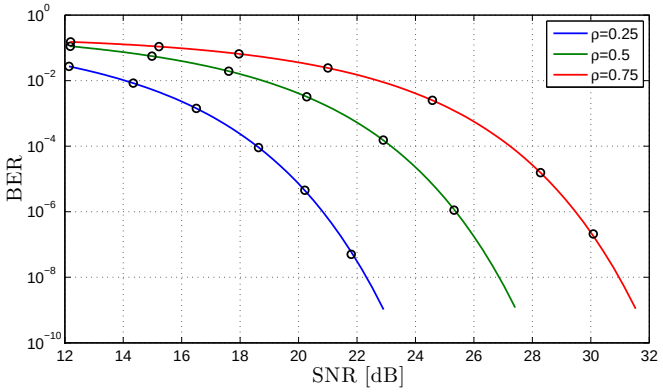

Fig. 6. BER as a function of SNR for non-equally probable symbols.

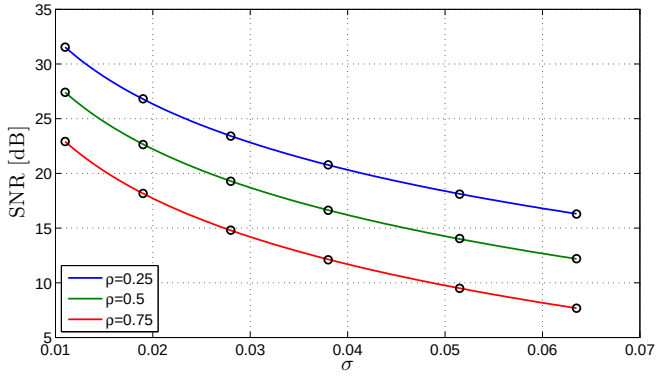

Fig. 5. SNR as a function of σ for non-equally probable symbols.

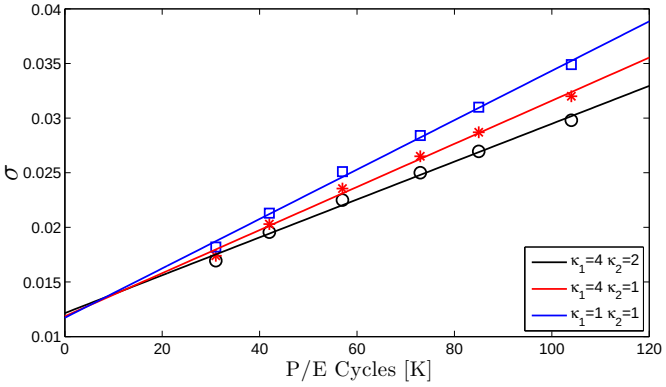

Fig. 7. The relation of introduced noise and the P/E cycles.

Such analysis is very helpful in the design of realistic NAND Flash emulators that can be used for real-time emulation of storage devices and systems.

V. Conclusions

In this paper we presented a model for 4-level NAND Flash memories that is based on asymmetric PAM with data-dependent additive noise. The analysis is independent to a specific NAND Flash technology and is applicable to all known technologies. By using the proper model parameters a specific NAND Flash technology can be accurately modeled and this can be exploited for developing realistic and accurate NAND Flash emulators. The presented analysis can be extended either to model NAND Flashes with more voltage levels, or to describe the behavior of other non-volatile memory technologies.

References

[1] G. Atwood, A. Fazio, D. Mills, and B. Reaves, "Intel strataflash memory technology overview," *Intel Technology Journal*, 1997.

[2] Z. Wang, M. Karpovsky, and A. Joshi, "Reliable mlc nand flash memories based on nonlinear t-error-correcting codes," in *Dependable Systems and Networks (DSN), 2010 IEEE/IFIP International Conference on*, 2010, pp. 41–50.

[3] 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.

[4] J. Brewer and M. Gill, *Nonvolatile Memory Technologies with Emphasis on Flash: A Comprehensive Guide to Understanding and Using Flash Memory Devices*. Wiley. com, 2011, vol. 8.

[5] G. Dong, S. Li, and T. Zhang, "Using data postcompensation and predistortion to tolerate cell-to-cell interference in mlc nand flash memory," *Trans. Cir. Sys. Part I*, vol. 57, no. 10, pp. 2718–2728, Oct. 2010. [Online]. Available: http://dx.doi.org/10.1109/TCSI.2010.2046966

[6] J. Proakis and M. Salehi, *Digital Communications*, ser. McGraw-Hill higher education. McGraw-Hill Education, 2007.

[7] B. Chen, X. Zhang, and Z. Wang, "Error correction for multi-level nand flash memory using reed-solomon codes," in *Signal Processing Systems, 2008. SiPS 2008. IEEE Workshop on*, 2008, pp. 94–99.

[8] S. Li and T. Zhang, "Improving multi-level nand flash memory storage reliability using concatenated bch-tcm coding," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 18, no. 10, pp. 1412–1420, 2010.

[9] Y. Maeda and H. Kaneko, "Error control coding for multilevel cell flash memories using nonbinary low-density parity-check codes," in *Defect and Fault Tolerance in VLSI Systems, 2009. DFT '09. 24th IEEE International Symposium on*, 2009, pp. 367–375.

[10] F. Sun, K. Rose, and T. Zhang, "On the use of strong bch codes for improving multilevel nand flash memory storage capacity," in *IEEE Workshop on Signal Processing Systems (SiPS): Design and Implementation*, 2006.

[11] Y. Cai, E. F. Haratsch, O. Mutlu, and K. Mai, "Threshold voltage distribution in mlc nand flash memory: characterization, analysis, and modeling," in *Proceedings of the Conference on Design, Automation and Test in Europe*. EDA Consortium, 2013, pp. 1285–1290.