This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2011 proceedings.

Statistical Analysis of Flash Memory Read Data

Jaekyun Moon, Jaehyeong No
Department of Electrical Engineering
KAIST
Daejeon, Korea
Email: jmoon@kaist.edu, eee4u@kaist.ac.kr

Sangchul Lee, Sangsik Kim and Joongseop Yang
Department of Flash Solution Development
Hynix Semiconductor
Icheon, Korea
Email: {sangchul.lee, sangsik1.kim, joongseob.yang}@hynix.com

***Abstract*—This paper discusses a technique for analyzing real data from flash memory cells. The goal is to identify and isolate various sources that cause the shifts and variations of the read values with respect to the intended write values. The analysis reveals how the neighboring cells interfere with the victim cell. Using the proposed analysis technique, the contribution of a specified set of neighboring cells towards the random read variation of the victim cell can also be quantified.**

I. Introduction

Characterization of the flash memory cell's input-output behavior is of great interest in the flash memory industry. As the flash memory cell density increases, the effect of interference among neighboring cells also increases, causing a shift of as well as a variation around the written cell value (threshold voltage) [1][2]. For example, the threshold voltage level of a cell (victim cell) may change with a change in the threshold voltage of adjacent cells (interfering cells) through floating-gate voltage coupling of neighboring cells [1]. Isolating the source of interference and noise would provide valuable insights in the flash memory manufacturing process and in the design of advanced signal processing schemes. See, for example, the pre- and post-compensation techniques described in [3] based on a cell-to-cell interference model.

In this paper, we discuss a statistical analysis method applied to a set of experimental read data corresponding to a block of commercial flash memory cells, with a goal in mind to isolate and quantify the sources of the mean-shift and the variation observed on the read value of the cell. Our approach is based on specifying a mask pattern that captures a certain number of cells, typically including the victim cell. Assuming a particular set of input (write) values for the cells positioned within the mask, the 2-dimensional (2-D) read data array is scanned over with the mask and the read (output) value of the victim cell is collected every time the write values of the cells captured under the scanning mask match the assumed input values. This process is repeated for each combination of input values for the cells within the mask pattern. For each such set of cell input values, averaging the collected read values for the victim cell provides the conditional sample mean, which is an unbiased estimate of the victim cell's read value conditioned on the particular input values for the cells within the mask, given the uncertainties introduced due to random noise and to the effect of changing input values for the cells outside the assumed mask.

This work was supported by Hynix Semiconductor and the National Research Foundation of Korea under Grant 2010-0029205.

When the data set is limited, however, the least square fit gives a highly reliable estimate for the conditional mean; in this paper we opt to use the least square fit by running a random-access-memory (RAM)-based training algorithm that keeps adjusting the RAM contents whenever the local input cell data matches the given RAM address. Each RAM address corresponds to some specific write value combination for the cells positioned within the mask. When the training is done, the RAM contents contain valuable statistical information based on which the effect of the specified cells on the victim cell can be accurately analyzed.

Existing data analysis methods are based on simply fixing the write values of the victim cell and the suspected interfering cells in the neighbor, and then scanning over the data collecting the victim cell's read value whenever the local input pattern matches the assumed write values. This method can identify the interfering cells causing the mean-shift in the victim cell's read value, but fails to isolate the sources of the random variation in it. In contrast, the technique discussed in this paper can quantify the effects of different sources on the random variation of the victim cell's read value.

II. Data Analysis Method

A. Problem Statement

Let x_{jk} and r_{jk} denote the input (write) and output (read) values of the jk-th memory cell in the 2-D array of cells. We can write

$$r_{jk} = x_{jk} + f_{jk}(\mathbf{x}) + n_{jk}(\mathbf{x}) \quad (1)$$

where $f_{jk}(\mathbf{x})$ and $n_{jk}(\mathbf{x})$ represent the deterministic and random portions, respectively, of interference on the jk-th cell that in general depend on the input values for all cells, $\mathbf{x}$. Given the pairs of write and read values for a finite-size 2-D array of cells, we wish to characterize $n_{jk}(\mathbf{x})$ conditioned on a specified set of write values on an arbitrary selection of potentially interfering cells. The interference signal of $f_{jk}(\mathbf{x})$ can be viewed as the shift in the threshold voltage of the victim cell due to interfering cells' taking particular input values $\mathbf{x}$. The random interference $n_{jk}(\mathbf{x})$, on the other hand, is the noise in r_{jk} due to the collective contribution of the random deviations around the nominal written charge levels of individual interfering cells.

978-1-4244-9268-8/11/$26.00 ©2011 IEEE

In particular, given that the error signal defined as $e_{jk} = r_{jk} - x_{jk} = f_{jk}(\mathbf{x}) + n_{jk}(\mathbf{x})$ can be measured for all (j, k) for some random input pattern $\mathbf{x}$, we are interested in answering the following questions. What is the probability density function (pdf) of e_{jk} conditioned on a specified local pattern of write values? Can we separate the noise from the interference? Where are the interfering cells and can we quantify their impact on the victim cell at position (j, k)? Specifically, can we accurately estimate $f_{jk}(\mathbf{x})$ and the variance of the entire noise for different input values for the interfering cells?

An existing, straightforward data analysis method consists of collecting the read values r_{jk} corresponding to some fixed write value x_{jk} and a specified write value for a suspected interfering cell, say, the cell at position $(j + 1, k)$, as the pass is taken over all victim cell positions (j, k). Fig. 1 shows typical histogram plots for the conditional pdfs of r_{jk} obtained using this method. The suspecting interfering cell is the one below the victim cell in this case and different conditional pdfs correspond to different input levels assumed for the interfering cell. The cells can take four different write levels in our experiment: the erasure (ER) level and three different program voltage levels, ranging from PV1 to PV3. The victim cell was fixed at PV2 in the figure. Notice that the conditional pdfs have different means and variances, signifying that the amount of interference varies depending on the input value of the interfering cell.

While this method predicts with reasonable accuracy the mean shift in the read value of the victim cell due to the interfering cell, it does not allow isolation of the sources for the random variations around the means.

As will be shown below, our proposed analysis method overcomes this shortcoming, allowing identification and quantitative assessment of different sources for the random variations in the victim cell's read value. In particular, our method can isolate the effect of arbitrarily chosen cells.

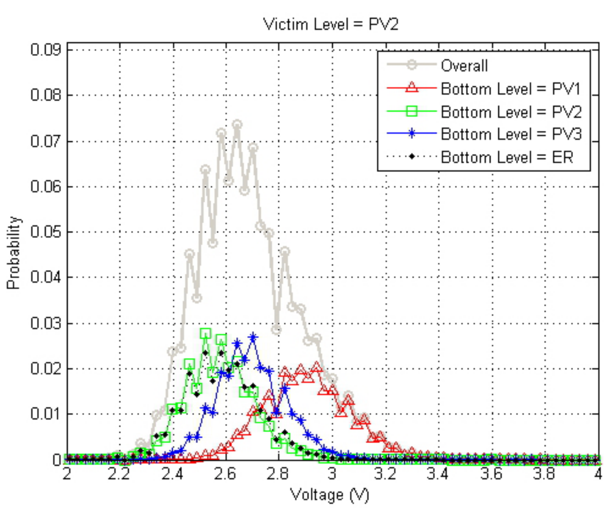


Fig. 1. Probability distributions of victim cell read values conditioned on different write values for the interfering cell

B. Determining the Mask Pattern

Our method is based on dividing the cells into a group of potentially interfering nearby cells and the rest of cells that are distant and thus are not likely to interfere. We first define a mask that captures the potentially interfering cells. We then attempt to average out the effect of the cells outside of this mask, so a focus can be made only on likely interfering cells. The size and the shape of the mask must be chosen judiciously. Let us call the cells captured within the mask internal cells and the ones outside the external cells. Since we are dealing with a finite set of data, if the mask size is too big, there would not be enough data samples to average out the effect of the external cells. If the mask size is too small, there would not be enough samples to reliably estimate the random distribution that arises due to internal cells. In this paper, the size of the mask is limited to cover 5$\sim$9 cells. For example, the mask may be of a 3 by 3 square pattern consisting of 8 neighboring cells surrounding a victim cell in the center.

C. RAM Training

Once the mask pattern is determined and a set of write values are assumed for the internal cells, a pass is taken over the given 2-D data set and the read value is collected whenever the local internal input values match the assumed values. Eventually an average value is to be obtained for each set of assumed write values for the mask. This average represents the read value specific to the given local internal write pattern, averaged over all possible combinations of the external cell write values as well as over all random deviations around the nominal input values of the internal cells.

Given the limited data set, though, we opt to use the least square fit rather than direct averaging, in removing the effect of external cells and internal random variations. For this, we employ RAM-based training where each RAM address corresponds to the local write pattern and the RAM content is adjusted in the direction of minimizing the squared error between the up-to-date content and the read data whenever the local write pattern matches the address, as the pass is taken over the whole data set. See Fig. 2.

For a given address, the update is done according to the well-know rule [4]:

$$RAM_{l+1}(address) = RAM_l(address) + \mu e_l \qquad (2)$$

where e_l represents the error between the actual read value r and $\hat{r}$, and $0 < \mu < 1$ is the update step size which controls the RAM update speed. RAM-based training is frequently used in the application of the decision feedback equalizer [4][5].

D. Analysis

Let P represent the mask pattern. We also need to define a subpattern P' as well as $Q = P \cap \bar{P}'$. See Table I for some example patterns. We shall use lower case letters p, p' and q to denote the specific combination of nominal write or input values for the cells under the corresponding patterns. Assuming that all signal and noise processes are stationary (i.e., statistical parameters are position-invariant), we write the read or output value of the victim cell as

$$r(p) = x(p) + n(p) + f(p) + E(P) \tag{3}$$

where $x(p)$ denotes the victim cell's nominal write value corresponding to p, a deterministic array of input values; the noise term $n(p)$ is a zero-mean random variable representing the various noise sources induced during writing and random noise during reading as well as random variations around the nominal write values in all internal cells in P; $f(p)$ represents interference caused by the nominal write values of the internal cells of the mask and is non-random once the internal write pattern p is fixed; and another interference term $E(P)$ denotes potential interference coming from all cells outside the chosen mask P and is a random variable if the input values of the external cells are not specified. In writing (3), we have assumed that the victim cell's noise due to random variation around the nominal write level in each external cell is negligible.

We are ready to write the mathematical expression for the trained RAM content of an address corresponding to p:

$$\begin{aligned}\hat{r}(p) &= \bar{x}(p) + \bar{n}(p) + \bar{f}(p) + \bar{E}(P)\\ &= x(p) + f(p) + \bar{E}(P)\end{aligned} \tag{4}$$

where the over-bar denotes average over successive hits on the given RAM address (or the given p value). Notice that once the training is done with enough hits on each address, we should have $\bar{n}(p) = 0$ for any p. Also, we have $\bar{x}(p) = x(p)$, $\bar{f}(p) = f(p)$ since they are deterministic once p is given. At

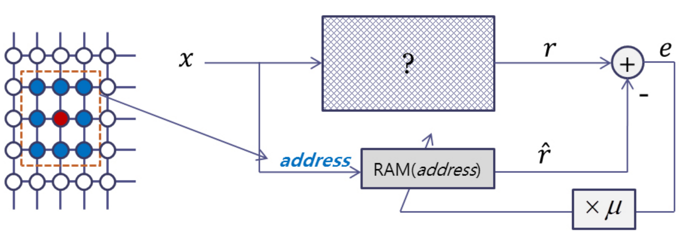
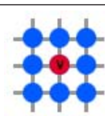
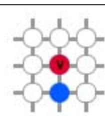
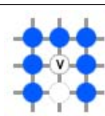


Fig. 2. Averaging out the effect of external cells using RAM training

TABLE I
EXAMPLE MASK PATTERN AND SUBPATTERNS

Pattern	Definition	Example
P	A mask pattern covering the victim cell (v) and selected, potentially interfering neighboring cells (all cells shown)	
P'	A subset pattern of P including the victim cell (red+blue)	
$Q = P \cap \bar{P}'$	A pattern covering the cells in P but not those in P' (blue)	

this point, we take a close look at the existing pdf extraction method described in II.A. We rewrite (3) with p replaced by p' to describe the victim cell's read value conditioned on a specific p', the input values for cells in some submask P' :

$$r(p') = x(p') + n(p') + f(p') + \underbrace{E(P')}_{f(Q)+E(P)} . \tag{5}$$

Notice that in (5) we are assuming $n(Q) = 0$. Now the average (taken over all possible input values for cells outside the mask P') is:

$$\begin{aligned}\bar{r}(p') &= \bar{x}(p') + \bar{n}(p') + \bar{f}(p') + \bar{f}(Q) + \bar{E}(P)\\ &= x(p') + f(p') + \bar{f}(Q) + \bar{E}(P).\end{aligned} \tag{6}$$

Let us compute the variance of $r(p')$:

$$\begin{aligned}\sigma^2_{r(p')} &= E\left[\left\{r(p') - \bar{r}(p')\right\}^2\right]\\ &= \sigma^2_{n(p')} + \sigma^2_{f(Q)} + \sigma^2_{E(P)}\end{aligned} \tag{7}$$

where $\sigma^2_{f(Q)}$ is the victim cell's read noise contribution due to variations in the input values for the cells in Q while $\sigma^2_{E(P)}$ is due to similarly introduced interference from cells outside the mask P. It is now clear that the noise variance computed in this way is a mixture of effects from multiple sources: the random write and read variation around the nominal value in the victim cell depending on the write level of the interfering cell in mask P', the input pattern variations in Q and the input pattern variations in cells outside P. The existing analysis does not tell us how to separate the effects of these sources.

We now come back to our approach and show how noise sources can be isolated and quantitatively analyzed. As we discussed above, the RAM value $\hat{r}$ does no longer contain the effect of cells outside the mask P. Let us fix p' (the cell input values within the submask P') and collect those $\hat{r}$ values corresponding to the common p'. The histogram plot of these values reflects the distribution generated by input pattern variations for cells under Q. Define the corresponding random variable as

$$\hat{r}(P|p') = x(p') + f(p') + f(Q) + \bar{E}(P) \tag{8}$$

where only $f(Q)$ is a random variable causing a distribution around some mean in the right hand side. The mean of $\hat{r}(P|p')$ is

$$\bar{\hat{r}}(P|p') = x(p') + f(p') + \bar{f}(Q) + \bar{E}(P). \tag{9}$$

The variance of $\hat{r}(P|p')$ can be computed using (8) and (9) as

$$\sigma^2_{\hat{r}(P|p')} = E\left[\left\{\hat{r}(P|p') - \bar{\hat{r}}(P|p')\right\}^2\right] = \sigma^2_{f(Q)} \tag{10}$$

which is precisely the contribution due to the input pattern variations in Q. Thus, by carefully designing the masks P and P' (and thus Q), we are able to compute the victim's read noise variance due to the input pattern variations in specific interfering cells in Q.

Putting (7) and (10) together, we subtract out the effect $\sigma^2_{f(Q)}$:

$$\sigma^2_{r(p')} - \sigma^2_{\hat{r}(P|p')} = \sigma^2_{n(p')} + \sigma^2_{E(P)} \tag{11}$$

the left side of which is obtained from the direct conditional averaging of the data as well as RAM training results. The right hand side shows two separate noise sources. If we assume $\sigma^2_{E(P)} \approx 0$, then (11) allows us to quantify the noise variance $\sigma^2_{n(p')}$ due to intrinsic noise and random variation in the specified interfering cell's write level which induces a random variation in the victim cell's read value, whereas (10) lets us to extract the read noise contribution due to input pattern variations in the remaining interfering cells in the chosen mask. Thus, by appropriately choosing P and P', we are able to compute the entire noise contributions of cells in P'.

We can also obtain $\sigma^2_{n(p)}$ with P instead of P' by using mean-shifted read data. First write

$$r(p) - \bar{r}(p) = n(p) + \big[E(P) - \bar{E}(P)\big] \tag{12}$$

which leads to the relation

$$\sigma^2_{r(p)} = \sigma^2_{n(p)} + \sigma^2_{E(P)}. \tag{13}$$

If there is one dominant interfering cell and it is captured within the submask P', then we would expect $\sigma^2_{n(p)} \approx \sigma^2_{n(p')}$. In this case, comparing (11) and (13) gives

$$\sigma^2_{r(p)} \approx \sigma^2_{r(p')} - \sigma^2_{\hat{r}(P|p')}. \tag{14}$$

The approximate equality (14) would be true if

$$\big[r(p) - \bar{r}(p)\big] + \hat{r}(P|p') \approx r(p'). \tag{15}$$

Notice that the means of $\hat{r}(P|p')$ and $r(p')$ are the same. In words, (15) means that we should be able to generate the probability distribution of $r(p')$ by convolving the distribution of $\hat{r}(P|p')$ with that of the mean-shifted $r(p)$. In the next section, this will indeed be shown to be true, using real data.

III. Numerical Results and Interpretation

Real data written into and read from a block of multi-level cells (MLC) were used in the experiment and analysis. Fig. 3 shows the distribution of $\hat{r}(P|p')$, the trained RAM contents for different p' assuming the submask P' contains the victim cell and the cell right below it. Only the victim cells that reside on the odd-bit lines are considered here.

Figs. 1 and 3 consistently show that the victim cell's read value contains the most bias to the right side when the input level of the cell below is PV1. This is due to the fact that the change in the applied charge level is the largest when the written cell level transitions from ER to PV1, inducing the worst program disturbance effect in the victim cell. The second biggest transition is made when the target write level is PV3, as reflected in the second largest bias when the interfering cell's write level is PV3. We note that for the MLC cells under investigation, the two bits written into each cell originate from two separate logical pages. The first bit is written by either retaining the original ER level of the cell or programming the cell value to a level somewhat below PV2. At the writing of the second bit, if the first bit corresponded to the ER level, either the same level is retained or the the charge level is raised to PV1, depending on the second bit value. If the first bit was written as the level somewhat below PV2, then the second bit would drive the new charge level to either PV2 or PV3. Raising the charge level at each programming time involves the well-known incremental step pulse programming (ISPP) method based on repeated program-verify steps with incremental increase in the programming voltage level [6]. ISPP is used to prevent cells being programmed to a charge level higher than intended, in which case the entire block must be erased again. In ISPP, the random variation around the nominal written value depends on the incremental step size [3].

Fig. 4 shows the distribution of mean-shifted $r(p)$ for different p' values. With $\sigma^2_{E(P)} \approx 0$, (13) indicates this distribution arises mainly due to the randomness in the write level of the victim cell as well as of neighboring cells, rather than due to the varying input patterns of the interfering cells.

Fig. 5 shows the convolution of the RAM content distribution and the $r(p)$ distribution, displaying mean and variance characteristics very similar to that of the directly obtained distributions in Fig. 1, as the analysis of the previous section predicted.

We also mention that analysis based on RAM training using new masks patterns that capture cells outside of the original 3 by 3 mask reveals that the interference coming from these distance cells is negligible, i.e., $\sigma^2_{E(P)} \approx 0$ if P is as shown in Table I.

Table II shows the mean-shift values in the victim cell output value caused by its bottom neighbor for different combinations of the victim cell and the bottom cell input values. For this, the RAM content values are utilized; $\bar{\hat{r}}(P|p') - \bar{\hat{r}}(P|p' = v) = f'(p')$ is shown with P' as given in Table I.

Table III shows the variance of mean-shifted $r(p)$ distribution: $\sigma^2_{r(p)}$. The same table also shows in the right-most column the calculated variance according to the left-hand side of (11). The fact that the numbers in these two columns are

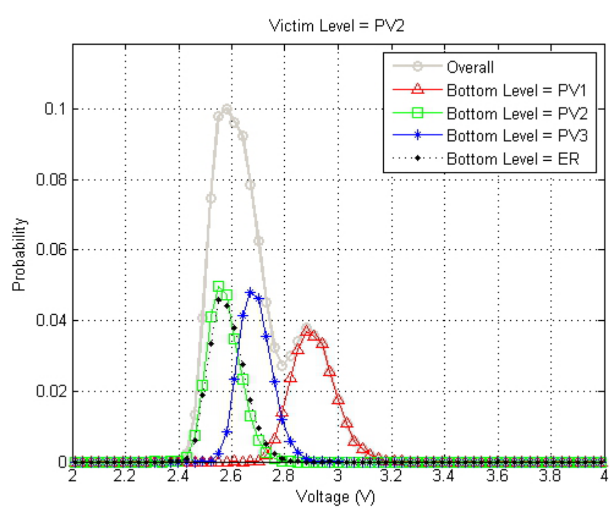


Fig. 3. Distribution of RAM contents $\hat{r}(P|p')$

very similar validate (14). The noise associated with PV3 on

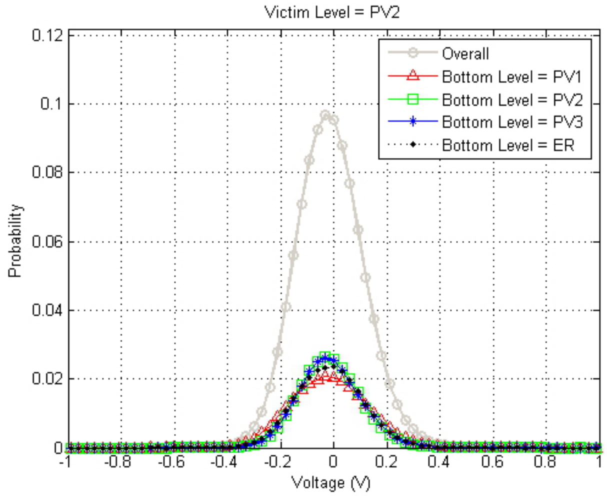


Fig. 4. Distribution of mean-shifted $r(p)$

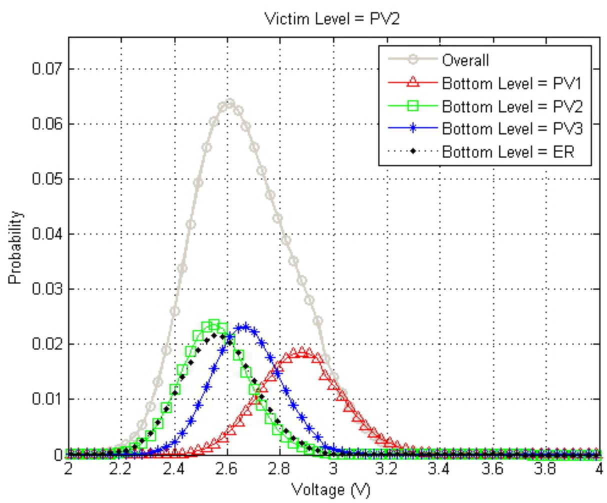


Fig. 5. Estimated $r(p')$ distribution (convolution of Fig. 3 and Fig. 4)

TABLE II
AMOUNT OF INTERFERENCE FROM BOTTOM NEIGHBOR: MEAN SHIFT

Victim level	Bottom level	$f'(p')$
PV1	PV1	0.226
	PV2	-0.126
	PV3	-0.007
	ER	-0.093
PV2	PV1	0.219
	PV2	-0.116
	PV3	0.001
	ER	-0.104
PV3	PV1	0.245
	PV2	-0.134
	PV3	-0.017
	ER	-0.094

TABLE III
VARIANCE OF MEAN-SHIFTED $r(p)$

Victim level	Bottom level	$\sigma^2_{r(p)}$	$\sigma^2_{r(p')} - \sigma^2_{\hat{r}(P\mid p')} \approx \sigma^2_{n(p')}$
PV1	PV1	0.0173	0.0168
	PV2	0.0100	0.0097
	PV3	0.0105	0.0101
	ER	0.0127	0.0124
PV2	PV1	0.0203	0.0196
	PV2	0.0138	0.0133
	PV3	0.0145	0.0138
	ER	0.0156	0.0153
PV3	PV1	0.0459	0.0448
	PV2	0.0364	0.0355
	PV3	0.0367	0.0355
	ER	0.0400	0.0392

victim is the largest while PV1 on victim yields the least noise. If the bottom cell level is PV1, the noise variance is the largest for the same victim level.

To get the results summarized in the tables shown above, the victim cells lying on the odd-bit lines and even-bit lines were analyzed separately. The flash memory considered in this paper has the alternate odd/even bit line structure where the cells belonging to odd (even) bits lines are programmed together. In this structure, for each odd cell, three cells below it are programmed afterwards, thus creating potential program disturbance for the odd victim cell [3]. For each even cell, two cells on both sides as well as three bottom cells can interfere. The interference coming from the single lower cell tend to dominate for the victim cells on the odd-bit lines, whereas there are some contributions from the side cells for the even-lined victim cells. The results summarized in Tables II and III reflect the victim cells on odd-bit lines only.

IV. CONCLUSION

A technique was discussed for statistically analyzing read data from flash memory cells. RAM training is utilized based on a specified mask pattern. Once the RAM training was done, the RAM contents were used to quantitatively analyze the impact of the interfering cells within the mask on the victim cell's read value. Using the method discussed in this paper, different noise sources inducing random variations around the nominal read value of a cell can be extracted. The proposed technique can be used to aid the flash memory design process and to develop a communication-type channel model which, in turn, would be useful in the design of advanced coding and signal processing algorithms that require accurate statistical characterization of the noise and interference. Various useful interpretations were made on the results of our analysis applied to the data read off a commercial NAND flash memory chip.

REFERENCES

[1] J. Lee, S. Hur and J. Choi, "Effects of floating-gate interference on NAND flash memory cell operation," *IEEE Electron Device Letters*, vol. 23, pp. 264-266, May 2003.

[2] K. Kim, "Future memory technology: Challenges and opportunities," in *Proc. Int. Symp. VLSI Technol.*, pp. 5-9, Apr. 2008.

[3] Guiqiang Dong, Shu Li and Tong Zhang, "Using Data Postcompensation and Predistortion to Tolerate Cell-to-Cell Interference in MLC NAND Flash Memory," *IEEE Trans. Circuits and Systems*, vol. 57, no. 10, pp. 2718-2728, Oct. 2010.

[4] B. Rothenberg, J. Brown, P. Hurst, and S. Lewis, "A Mixed-Signal RAM Decision-Feedback Equalizer for Disk Drives," *IEEE Journal on Solid-State Circuits*, vol. 32, no. 5, pp. 713-721, May 1997.

[5] K. D. Fisher, J. M. Cioffi, W. L. Abbott, P. S. Bednarz, and C. M. Melas, "An adaptive RAM-DFE for storage channels," *IEEE Trans. Comm.*, vol. 40, pp. 1559-1568, Nov. 1991.

[6] K. Suh, B. Suh, Y. Lim, J. Kim, Y. Choi, Y. Koh, S. Lee, S. Kwon, B. Choi, J. Yum, J. Choi, J. Lim and H. Lim, "3.3 V 32 Mb NAND flash memory with incremental step pulse programming scheme," *IEEE Journal on Solid-State Circuits*, vol. 30, no. 11, pp. 1149-1156, Nov. 1995.