Evaluation of Design for Reliability Techniques in Embedded Flash Memories

Benoît Godard, Jean-Michel Daga, Lionel Torres, Gilles Sassatelli

To cite this version:

Benoît Godard, Jean-Michel Daga, Lionel Torres, Gilles Sassatelli. Evaluation of Design for Reliability Techniques in Embedded Flash Memories. DATE: Design, Automation and Test in Europe, Apr 2007, Nice, France. pp.1593-1598, 10.1109/DATE.2007.364529 . lirmm-00179951

HAL Id: lirmm-00179951

https://hal-lirmm.ccsd.cnrs.fr/lirmm-00179951

Submitted on 11 Dec 2007

Evaluation of Design for Reliability Techniques in Embedded Flash Memories

Benoît Godard [1,2], Jean-Michel Daga [1], Lionel Torres [2], Gilles Sassatelli [2]

[1] Libraries and Design Tools Department – Embedded Non-Volatile Memory Group
ATMEL Rousset - 13106 Rousset Cedex, France
benoit.godard, jean-michel.daga@rfo.atmel.com

[2] Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier – LIRMM
Université de Montpellier II / UMR5506 CNRS
161, rue Ada – 34392 Montpellier Cedex 5, France
torres, sassatelli@lirmm.fr

Abstract

Non-volatile Flash memories are becoming more and more popular in Systems-on-Chip (SoC). Embedded Flash (eFlash) memories are based on the well-known floating-gate transistor concept. The reliability of such type of technology is a growing up issue for embedded systems; endurance and retention are of course the main features to analyze. To enhance memory reliability current eFlash memories designs use techniques such as Error Correction Code (ECC), Redundancy or Threshold Voltage (V_T) Analysis. In this paper, a memory model to evaluate the reliability of eFlash memory arrays under distinct enhancement schemes is developed.

1 Introduction

Different types of memory can be embedded in a SoC as SRAM, DRAM, EEPROM and Flash. The increased use of portable electronic devices produces a high demand for eFlash memories. eFlash memories exhibit low power characteristics and some security features (lock bits). In addition, these memories allow In Situ Programming (ISP), resulting in very flexible solutions for code development and updates. In parallel with the recent market evolution, SoCs with embedded memories are facing technological issues due to reliability and chip yield. An increasing silicon area is dedicated to memories and storage elements. The Semiconductor Industry Association (SIA) confirms this trend [1] forecasting that memory content will approach 94% of a SoC silicon area by 2015. As a result, memory reliability will be the main detractor of the SoC reliability. Additional constraints and reliability objectives may be added to SoCs in order to cover applications such as automotive, aeronautic or biomedical. Manufacturers should borrow specific methods and design solutions to certify defect-free devices.

Error Correcting Codes (ECC) are the most popular method to prevent memories from online random errors. Some parity bits are stored with information bits. Depending on the adopted ECC scheme, a certain number of errors can be detected and corrected. To enhance yield, designers choose row and/or column redundancy. During the test production phase, defective memory elements are disconnected and replaced with error-free redundancy elements. In SoC context, Built-in Self Repair (BISR) has been realized successfully [2]. Throughout the years, these methods have been mixed. Architectures combining ECC and redundancy for yield enhancement [3] [4], and/or reliability enhancement [5] [6] have been developed.

Additionally, Flash memories could be considered as analog memories that allow specific reliability enhancement methods. The mainstream operation is based on the floating gate concept on which charges can be stored or removed by high voltage biasing. It results in a shift of the memory cell threshold voltage (V_T). During a read operation, the modulation of the biasing conditions allows V_T level analysis. Bits whose charge levels are weak can be detected [7]. A cell refreshing scheme [8] and an error detection/correction scheme [9] based on the V_T level analysis have already been proposed.

In literature, architecture reliability evaluations are usually performed with a constant failure rate reflecting the SRAM cell reliability. Charge loss and cycling degradations are not taken into account even if multiple models for Flash reliability prediction exist [10], [11]. In this paper, we compare different methods to enhance Flash reliability using ECC, online redundancy, and V_T analysis. For this purpose, a memory array model using the compact model exposed in [10] has been developed. The aim of this work is to help designers to choose the most efficient

978-3-9810801-2-4/DATE07 © 2007 EDAA

reliability scheme to implement depending on the technology, memory architecture and reliability objective.

The rest of the paper is organized as follow. In section 2, the cell reliability model is exposed. Next, in section 3, we describe the concept of cell-V_T repartition in a word and we introduce reliability enhancement methodologies in section 4. Results and discussion on these reliability techniques are presented in section 5. Finally, in section 6, we conclude this paper and introduce our future work.

2 Cell reliability modeling

When electrons are stored in the floating gate, the memory cell has a high V_T. The cell is erased. When electrons are removed from the floating gate, the memory cell has a low V_T. The cell is written. The threshold voltage value of the cell is directly related to the quantity of electrons stored in the floating gate referred as Q_{FG} by:

$$V_T = V_{T0} + \frac{Q_{FG}}{C} \quad (1)$$

where, C, V_{T0} are the equivalent capacitance of the floating gate and the virgin threshold voltage, respectively. Q_{FG}/C is also the floating gate potential V_{FG}. By convention, erased and written cells correspond to the logic value “1” and the logic value “0”.

Floating gate reliability is covered by two aspects: endurance and retention. Endurance is the memory cell ability to keep good physical characteristics so as to be usable even after multiple write/erase cycles. Retention is the cell ability to retain information throughout time since the last writing operation. Manufacturers targets are typically 10^5 cycles in endurance and 10 years in retention.

From a functional point of view, during the memory life, V_T are distributed over two populations, one for cells with high V_T (logic value “1”), and another for cells with low V_T (logic value “0”). Because of charge leakage mechanism, distributions drift with time, “1” and “0” tends to opposite values and becomes weak or erroneous. In the rest of the paper, only one distribution that goes from high V_T values (erased cells) to low V_T values is considered. As shown in the figure 1, the charge leakage mechanism can be modeled by a capacitor being discharged by a source through a thin oxide. The fundamental expression linking threshold voltage with leakage current is derived from (1) and expressed by:

$$\frac{dV_T}{dt} = -\frac{I_L}{C} \quad (2)$$

where, I_L is the charge leakage current.

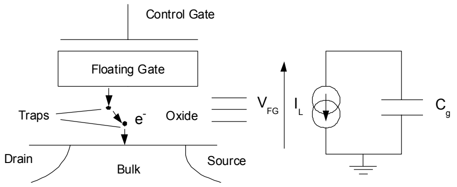

Figure 1 – Defective Flash cell and equivalent cell model

The modeling of the leakage current I_L is still a big issue. When low electrical fields are applied to the oxide, conduction mechanisms are not well known. The current leakage is extremely low and difficult to measure accurately. So, the conduction mechanism is usually chosen by using empirical assumptions depending on experimental observations and not on an accurate knowledge of the conduction mechanism. The main retention issues are related to the *Stress-Induced Leakage Current* (SILC) that becomes predominant when tunnel oxide is thinned. With high field stress used for programming, properties of the insulating layers are degraded. SILC seems to be due to some *Trap-Assisted Tunneling* effect (TAT) [12]. Multiple models have been developed to describe the SILC phenomenon. The percolation model [12] seems a good way to explain the underlying physical phenomenon.

In our case, the compact model developed in [10] based on an exponential I-V characteristic is used. In this model, V_T variation between cells is explained by parameters modulation that acts as V_T-shift of the cell threshold distribution. As in [10], the following assumptions on the V_T evolution are made:

- V_T drift is independent of the initial V_T value,
- V_T drift is linear with the logarithm of the time,
- V_T drift is linear with the logarithm of the number of write/erase cycles,
- V_T drift is linear with the inverse of the temperature,
- The ratio of cells below a given V_T is exponentially distributed.

The resulting reliability modeling determines the probability that cell threshold is higher than a voltage limit V_{Limit} depending on constants c_0, c_1, c_2, c_3, c_4 and variables such as time t, number of cycles n_{cycles} and temperature T:

$$R_{Normal} = p(V_T > V_{Limit}) = f(t, n_{cycles}, T, V_{Limit}) \quad (3)$$

$$\ln(-\ln(1-R_{Normal}) = c_0 + c_1 \cdot V_{Limit} + c_2 \cdot \ln(t) + c_3 \ln(n_{cycles}) + c_4 \cdot \frac{1}{T} \quad (4)$$

where, constant c0, c1, c2, c3, c4 are determined from experimental results.

Moreover, erratic bits are considered. The erratic behavior is a floating gate specific issue that usually affects a ratio of cells randomly distributed in an array. For these bits, the threshold voltage may evolve by steps throughout time. These bits are at the root of reliability loss because they cannot be detected during the test production phase: even if a memory has successfully passed tests, it may be defective due to erratic cells. The underlying phenomenon is not known but some explanations have been proposed. Bi-stable traps in the oxide would create a TAT effect. For a ratio of the memory life α_{ON}, these traps would be in an ON state, adding an additional current leakage I_{ON} in the model. The rest of the time, traps would be in an OFF state. This effect is taken into account adding the constant term c_0' to (4):

$$\ln(-\ln(1-R_{Erratic}) = c_0 + c_0' + c_1 \cdot V_{Limit} + c_2 \cdot \ln(t) + c_3 \ln(n_{cycles}) + c_4 \cdot \frac{1}{T} \quad (5)$$

Erratic bits and normal bits are thus part of two independent distributions. The combination of (4) and (5) gives the reliability for one cell:

$$R_{cell} = \alpha \cdot R_{Erratic} + (1-\alpha) \cdot R_{Normal} \tag{6}$$

where, α represents the ratio of erratic bits in a population. By the way, this expression depends on t, n_{cycles}, T and V_{Limit}. In our model, the ageing of normal bits is responsible for the normal memory wear-out, whereas, erratic bits are abnormally increasing the in-line failure rate at the beginning of the memory life.

3 Cell-V_T repartition in a word

The figure 2 is composed of three V_T limits: a low voltage value V_L, a nominal voltage value V_N and a high voltage value V_H. When a read operation is performed, a V_T limit is chosen. Bits with V_T higher than this limit will correspond to logic value “1”. In the same way, bits with V_T lower than this limit will correspond to logic value “0”. We can note that selecting a V_T limit is equivalent to perform a read with a particular biasing of the cell control gate.

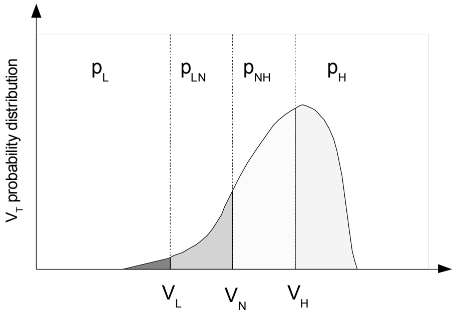

Figure 2 – V_T probability distribution with V_T limits

As mentioned in the previous section, only erase operations are considered here. After a write/erase cycle, all the V_T distribution is shifted towards high V_T values. But, with time, the distribution drifts to lower values as illustrated in the figure 2. Then, the V_T of each bit in a word is located in one of four V_T slices with a certain probability as shown in table 1. In this table, notions of error bits and weak bits are also defined.

Slice	Type	Corresponding probability
$V_T > V_H$	good bits	$p_H = R_{cell}(V_H)$
$V_T \in \bullet[V_N, V_H]$	weak good bits	$p_{NH} = R_{cell}(V_N) - R_{cell}(V_H)$
$V_T \in \bullet[V_L, V_N]$	weak failing bits	$p_{LN} = R_{cell}(V_L) - R_{cell}(V_N)$
$V_T < V_L$	hard failing bits	$p_L = 1 - R_{cell}(V_L)$

Table 1 – Bits convention and associated probabilities

To know the efficiency of a reliability scheme, the cells V_T repartition in a word must be considered. In a word composed of N_{Cells}, the probability of having respectively N_L, N_{LN}, N_{NH}, N_H bits in slices 1, 2, 3 and 4 is described by a multinomial repartition [14]:

$$p_w(N_L, N_{LN}, N_{NH}, N_H) = \frac{Ncells}{N_L! \cdot N_{LN}! \cdot N_{NH}! \cdot N_H!} \cdot p_L^{N_L} \cdot p_{LN}^{N_{LN}} \cdot p_{NH}^{N_{NH}} \cdot p_H^{N_H} \tag{7}$$

where, $N_{Cells} = N_L + N_{LN} + N_{NH} + N_H$.

In the rest of the paper, the V_T limit is always V_N when a read is performed. However, this V_T limit is changed during a V_T analysis. Indeed, the V_T analysis is the process that detects and locates weak bits in a word i.e. bits that have their V_T in the slice $[V_L, V_H]$. This operation is carried out by two read operations choosing successively V_L and V_H as V_T limits. Then, weak bits locations are found making a bitwise comparison of read operation’s results. Logic values “1” will reveal weak bits.

4 Reliability enhancement methods

4.1 Array modeling

Figure 3 represents a memory array. It implements additional bits for an error detection/correction system and spare rows for an on-line repair system. Here, column redundancy has not been considered. Indeed, Flash are page-oriented during write/erase operations. These operations are time-consuming (few ms). To replace an entire column with redundancy, all pages of the array should be erased and written back in order to modify only one bit position. This on-line repair process is not realistic because the memory will not be available for a few seconds depending on the depth of the array. For instance, if the replacement of a bit position takes 4 ms and the array has 1024 pages, the column repair process will take more than 4 seconds. On the contrary, in case of row redundancy, only one page has to be programmed during the repair process.

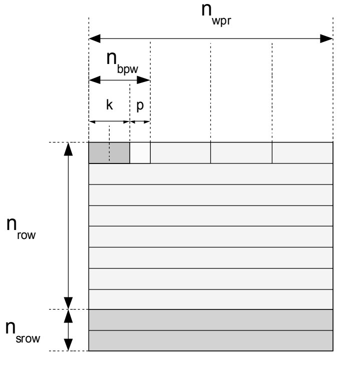

Figure 3 – Flash memory array modeling

In the rest of the paper, the following notations are used: each word is composed of k information bits and p parity bits with $n_{bpw} = k + p$ bits per word. cc and dc are defined as the error correction capacity and the error detection capacity associated to the error correcting code respectively. There are n_{wpr} words per row (or page) and the array has n_{row} normal rows and n_{srow} spare rows.

Reliability enhancement procedures always incorporate three steps:

- Error Detection (ED) – An error state is detected in a memory word. This process is usually performed by a control of the likelihood based on an error detection/correction code.
- Error Localization (EL) – Locations of the error bits in a memory word are determined. This step is performed using the capacity correction of an error correction code and/or a V_T analysis.
- Retrieval Mechanism (RM) – When a bit is detected to be weak or in error, the information sent to the user must be corrected. Additionally, operation can be performed on the memory array to physically repair it (using redundancy), to refresh the data stored (using a refresh process) or to correct the information on the fly (using on-line detection/correction).

4.2 Array Reliability with detection/localization procedures

Three detection/localization procedures A, B and C are studied. As shown in table 2, procedures depend on the error correcting code implemented and so, on the number of parity bits added per word. Potentially, procedures A and B permit one error per word correction whereas the procedure C allows two errors correction.

		Procedure A	Procedure B	Procedure C
Error correction code implemented		Parity Code	Hamming Standard	Extended Hamming
p		1	$\log_2(k)+1$	$\log_2(k)+2$
dc		1	1	2
cc		0	1	1
Detection		Error correcting code (detection capacity)		
Localization	**One error**	V_T analysis	Error correction code	Error correction code
	Double error	–	–	V_T analysis
Total number of correctable errors		1	1	2

Table 2 – Detection/localization procedures

When a word is read, the online ECC mechanism is used to detect errors. If error correction capacity of the ECC is sufficient, errors are automatically corrected and the result is sent to the user. If the error correction capacity is exceeded but error detection capacity is still sufficient, a V_T analysis will determine weak bits in the slice $[V_L, V_H]$. Then, the following assumption is made: the weak bits discovered during the V_T analysis are the failing bits that have not drifted enough to be hard errors. Consequently, if the number of weak bits in the word equals the number of error detected, the inversion of weak bits permits to recover the correct word. To illustrate that purpose, the procedure C is considered. In this case, the Extended Hamming Code is used, so 2 errors can be detected (dc = 2) and only one can be corrected (cc = 1). A read is performed on a word. If the word has a single error, the correction capacity is not exceeded. The ECC mechanism is able to transparently detect and correct the error. Now, if the word has two errors, the correction capacity is exceeded but not the detection capacity. The ECC mechanism is able to analyze the problem: two errors have been detected but not located. Then, a V_T analysis is launched to locate weak bits. If two weak bits are found in the word, their values are inverted and the word is supposed to have been corrected. Table 2 summarizes detection and localization procedures A, B and C.

Reliability enhancements with the three detection/localization procedures are now analyzed. For that purpose, we enumerate reliable V_T repartitions in a word:

- Procedure A: A memory word is correct if, for all cells, $V_T > V_N$ i.e. multiple bits may be weak but there is no error. Or, one cell has a V_T in $[V_L, V_N]$ slice and all the others have $V_T > V_H$ i.e. one error is present due to one weak bit. It corresponds to the probability:

$$p^A_{word} = \sum_{i=0}^{n_{bpw}} p_w(0,0,i,n_{bpw}-i) + p_w(0,1,0,n_{bpw}-1) \tag{8}$$

- Procedure B: A memory word is correct if, for all cells, $V_T > V_N$ i.e. multiple bits may be weak but there is no error. Or, one cell has a $V_T < V_N$ and all the others have $V_T > V_N$ i.e. there is only one error. It corresponds to the probability:

$$p^B_{word} = \sum_{i=0}^{n_{bpw}} p_w(0,0,i,n_{bpw}-i) + \sum_{i=0}^{n_{bpw}-1} p_w(0,1,i,n_{bpw}-1-i) + \sum_{i=0}^{n_{bpw}-1} p_w(1,0,i,n_{bpw}-1-i) \tag{9}$$

- Procedure C: A memory word is correct if, for all cells, $V_T > V_N$ i.e. multiple bits may be weak but there is no error. Or, one cell has $V_T < V_N$ and all the others have $V_T > V_N$ i.e. there is one error. Or, two cells have a V_T in $[V_L, V_N]$ slice and all the others have $V_T > V_H$ i.e. there are two errors due to weak bits. It corresponds to the probability:

$$p^C_{word} = \sum_{i=0}^{n_{bpw}} p_w(0,0,i,n_{bpw}-i) + \sum_{i=0}^{n_{bpw}-1} p_w(0,1,i,n_{bpw}-1-i) + \sum_{i=0}^{n_{bpw}-1} p_w(1,0,i,n_{bpw}-1-i) + p_w(0,2,0,n_{bpw}-2) \tag{10}$$

There are n_{wpr} . n_{row} words per array. Consequently, reliability expressions for pages and arrays without redundancy in procedures A, B and C are obtained from (8), (9), (10):

$$R^i_{page} = (p^i_{word})^{n_{wpr}} \tag{11}$$

$$R^i_{ecc} = (p^i_{word})^{n_{wpr} \cdot n_{row}} \tag{12}$$

where, i is replaced by A, B or C.

4.3 Array Reliability with on-line repair procedure

Online repair with row redundancy can be considered as a retrieval mechanism. As soon as an error is detected, the entire page is replaced with row redundancy if available. The corresponding reliability is given by:

$$R^i_{red} = \sum_{k=0}^{n_{srow}} C^k_{n_{row}+n_{srow}} (R^i_{page})^{n_{row}+n_{srow}-k} \left(1 - R^i_{page}\right)^k \qquad (13)$$

4.4 Array Reliability with mixed repair and ECC procedure

In reality, when all redundancy rows have been used, the error detection/localization system still corrects errors and the memory continues to be reliable. In other words, the reliability expression (13) is used for memories whose number of pages in error is at least equal to the number of spare rows. Assuming $n_{srow} > 0$, the probability that some redundancy is still available is expressed by:

$$p^i_{n_{srow}} = \sum_{k=0}^{n_{srow}-1} C^k_{n_{row}+n_{srow}} (R^i_{page})^{n_{row}+n_{srow}-k} \left(1 - R^i_{page}\right)^k \qquad (14)$$

With this procedure, the reliability of the array is:

$$R^i_{red+ecc} = (1 - p^i_{n_{srow}}) \cdot R^i_{ecc} + p^i_{n_{srow}} \qquad (15)$$

5 Results and Discussion

The figure 4 is a comparison of the reliability between a standard eFlash array and eFlash arrays with detection/localization procedures (12) developed in the section 4.2 after 10^5 program/erase cycles. The V_T limits have been chosen as follow: $V_L = -1$ V, $V_N = 0$ V and $V_H = 1$ V. There are $n_{row} = 1024$ rows and $n_{wpr} = 64$ words per row. All curves are normalized in time by the *Mean Time To Failure* (MTTF) of the standard array noted $MTTF_{standard}$. Model parameters have been calibrated on a 180 nm eFlash technology, based on measurements performed on samples of a 2Mb memory.

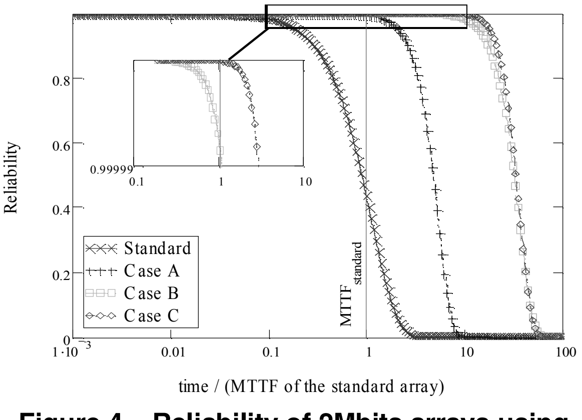

Figure 4 – Reliability of 2Mbits arrays using detection/localization procedures only

With the procedure A, only one error can be corrected using a V_T analysis. This scheme can correct one weak failing bit per word. As illustrated in the table 3, the MTTF is improved by a factor 4.67 with this procedure in comparison with a standard array. However, the hard failing bits are not reachable. Consequently, the reliability improvement is lesser than with the procedure B where one weak failing or one hard failing bit can be corrected thanks to the Hamming Correcting Code. In the procedure B, the MTTF is improved by a factor 25.1 in comparison with a standard array. Namely, if a memory has a MTTF equal to 1 year, then in procedures A and B, the MTTF will be respectively improved to 4.67 years and 25.1 years. In the procedure C, there is no noticeable improvement of the reliability compared to the procedure B even if an Extending Hamming Code and V_T analysis are used to correct up to two errors: the MTTF gain is only 26.3. Nevertheless, the reliability decrease occurs later in the procedure C than in the procedure B. This phenomenon is observed focusing on the beginning of the curve decrease as shown in the figure 4. For instance, at time $MTTF_{standard}$, 56.3% of the standard arrays would have failed. In procedures A, B and C, only 6263, 7.2 and 0.08 parts per million (ppm) would have failed respectively. The two decades difference between procedures B and C may justify the adoption of the procedure C for products needing high reliability rates.

The impact of the erratic bits ratio on the reliability scheme is low. Indeed, if the ratio is increased by a factor 10^2, then, 9.7 ppm, 0.11 ppm would have failed in procedures B and C at time $MTTF_{standard}$.

2Mbits eFlash array with distinct word lengths have been reported in the table 3. Array overheads and MTTF gains are presented. MTTF gains are independent of the number of cycles due to the logarithmic dependence of n_{cycles} in our cell model. Our cost function is defined as the ratio between the array overhead and the logarithm of the MTTF gain. When the word length is increased the MTTF gain is reduced but the overhead impact decreases far more rapidly. In consequence, the cost function decreases. Array requiring moderate reliability improvement for a very low cost may adopt the procedure A. On the contrary, to be fully reliable, the procedures B or C should be adopted.

	Procedure A			Procedure B			Procedure C		
Max. Cor.	1			1			2		
Constant Parameters	n_{row} = 1024, n_{wpr} = 64, V_L = -1 , V_N = 0, V_H = 1								
k	32	64	128	32	64	128	32	64	128
p	1	1	1	6	7	8	7	8	9
n_{wpr}	64	32	16	64	32	16	64	32	16
Array Overhead (%)	3.1	1.6	0.8	18.7	10.9	6.2	21.9	12.5	7.0
MTTF Gain	4.67	3.80	3.16	25.1	23.4	20.9	26.3	24.0	21.4
Cost Overhead/ log(MTTF gain)	4.6	2.8	1.6	13.4	8.0	4.7	15.4	9.1	5.3
# Defective arrays at $MTTF_{standard}$ (ppm)	6263	12663	17637	7.2	12.6	23.4	0.08	0.2	0.59

Table 3 – Summary of 2Mbits arrays using detection/localization procedures only

The figure 5 shows a reliability comparison between a standard array and arrays with mixed detection/localization procedures (A or B) and the online repair developed in the section 4.4 after 10^5 program/erase cycles. The V_T limits have been chosen as follow: $V_L = -1$ V, $V_N = 0$ V and $V_H = 1$ V. There are n_{row} = 1024 rows and n_{wpr} = 64 words per row. The number of row redundancy is a parameter.

At first look, the online repair makes the reliability slope sharper. In table 4, we have reported the number of defective arrays at time $MTTF_{standard}$. This number becomes zero as soon as some redundancy is added. This observation is independent of the detection/localization procedures used. Thanks to table 4, we can note that the procedure A with 2 rows results in less defective arrays after $MTTF_{standard}$ than the procedure B with 0 rows. Consequently, the array becomes very reliable at time $MTTF_{standard}$ adding the online repair. The figure 5 shows also that the MTTF gain is increased adding few rows. But, the improvement reduces slowly with each new row. In the table 4, this is traduced by a decrease slow down of the cost function. As a result, only a low number of rows are useful. The procedure A associated with online repair would be a very good choice to manage reliability at a reduced array overhead cost.

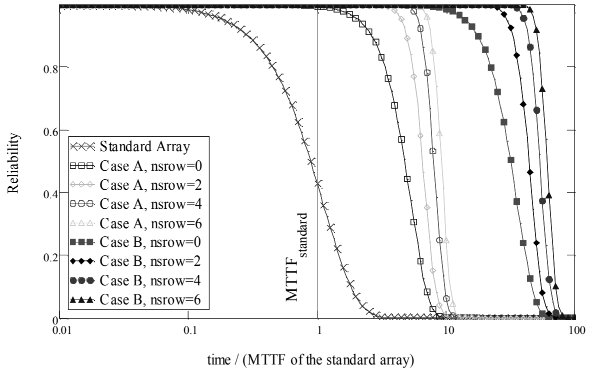

Figure 5 – Reliability of 2Mbits arrays using mixed detection/localization and online repair procedures

	Procedure A				Procedure B			
Max. Cor	1				1			
Constant parameters	k=32, n_{row}= 1024, n_{wpr}= 64, V_L= -1 , V_N= 0, V_H= 1							
p	1				6			
n_{srow}	0	2	4	6	0	2	4	6
Array Overhead (%)	3.1	3.3	3.5	3.7	18.7	19.0	19.2	19.4
MTTF Gain	4.67	6.31	7.76	8.91	25.1	41.7	50.1	57.5
Cost Overhead/ Log(MTTF Gain)	4.6	4.1	3.9	3.9	13.4	11.7	11.3	11
# Defective arrays at $MTTF_{standard}$ (ppm)	6263	0.1	~0	~0	7.2	~0	~0	~0

Table 4 – Summary of 2Mbits arrays reliability using mixed detection/localization and online repair procedures.

6 Conclusion

In this paper, an eFlash array reliability model has been developed. We have clearly shown that by mixing different reliability techniques (ECC, redundancy, V_T analysis), high reliability improvement can be reached. Our work is based on the fact that eFlash memories are analog devices; analog information can be extracted from cells for reliability purpose. The V_T analysis was proven to be a powerful method with a low additional cost in order to localize errors when ECC can only detect it. The online redundancy has also been studied. This method allows reducing the number of defective chips after $MTTF_{standard}$. For a given technology and a given eFlash memory architecture, this work helps designer to adopt the most adapted scheme in order to reach a defined ppm and MTTF objective. In our future work, we will focus on the implementation of such schemes. A modified eFlash memory architecture and a logical memory wrapper to perform online repair, ECC and V_T analysis will be presented.

References

[1] Semiconductor Industry Association (SIA), "International technology roadmap for semiconductors (ITRS)", http://www.sia-online.org/home.cfm, 2005.

[2] C. T. Huang, J. C. Yeh, R. F. Huang, C. W. Wu, P. Y. Tsai, A. Hsu and E. Chow, "A built-in self-repair scheme for semiconductor memories with 2-D redundancy", *in Proc. Int. Test Conf.*, p.393, 2003.

[3] M. Choi, N. Park, F. Lombardi, Y. B. Kim and V. Piuri, "Optimal Spare Utilization in Repairable and Reliable Memory Cores", *in Proc. Of Int. Work. On Mem. Tech. and Design*, p.64, 2003.

[4] C-H. Chen, Y-Y. Hsaio, C-W. Wu, "A Built-In Self-Repair Scheme for NOR-type Flash Memory", *in Proc. of the VLSI Test Symp.*, p.114-119, 2006.

[5] C-L. Su, Y-T. Yeh, C-W. Wu ,"An Integrated ECC and Redundancy Repair Scheme for Memory Relibability Enhancement", *in Proc. of the Int. Symp. Defect and Fault Tolerance in VLSI Systems*, p.81-89, 2005.

[6] C. Wickman, D. G. Elliott, and B. F. Cockburn, "Cost model for large file memory DRAMs with ECC and bad block marking", *in Proc. of the Int. Symp. Defect and Fault Tolerance in VLSI Systems*, p.319, 1999.

[7] J.M. Portal, H. Aziza, D. Née, "Eeprom memory: threshold voltage built in self diagnosis", *in Proc. of the Int. Test Conf.*, p.23-28, 2003.

[8] Y. Furuta, T. Okumura, "Non-volatile memory refresh control circuit", *US Patent n°4 218 764*, 1980.

[9] La Rosa, "One bit error correction in a chain of bits", *EP Patent n° 1 109 321 A1*, 2001.

[10] H. P. Belgal, N. Righos and al., "A new reliability model for post-cycling charge retention of Flash memories", *in Proc. of the Annual Int. Rel. Phys. Symp.*, p.7-20, 2002.

[11] A. Hoefler, J. M. Higman and al., "Statistical modeling of the program/erase cycling acceleration of low temperature data retention in floating gate non-volatile memories", *in Proc. of the Annual Int. Rel. Phys. Symp.*, p.60-66, 2002.

[12] R. Degreave, F. Schuler and al., "Analytical percolation model for predicting anomalous charge loss in Flash memories", *in Trans. On Elect. Dev.*, p.1392-1400, 2004.

[13] J. Harthong, "Probabilities and statistics", Ed. Diderot, 1980.