Design and Implementation of an Adaptive Proactive Reconfiguration Technique for SRAM Caches

Peyman Pouyan,Esteve Amat,Francesc Moll,Antonio Rubio
Department of Electronic Engineering
UPC Barcelona Tech, Barcelona, Spain
peyman.pouyan,esteve.amat,francesc.moll, antonio.rubio@upc.edu

***Abstract*—Scaling of device dimensions toward nano-scale regime has made it essential to innovate novel design techniques for improving the circuit robustness. This work proposes an implementation of adaptive proactive reconfiguration methodology that can first monitor process variability and BTI aging among 6T SRAM memory cells and then apply a recovery mechanism to extend the SRAM lifetime. Our proposed technique can extend the memory lifetime between 2X to 4.5X times with a silicon area overhead of around 10% for the monitoring units, in a 1kB 6T SRAM memory chip.**

I. INTRODUCTION

One of the most important challenges in design of reliable nano-scale VLSI circuits is the Bias Temperature Instability (BTI) phenomenon, known as Positive Bias Temperature Instability (PBTI) in NMOS transistors and Negative Bias Temperature Instability (NBTI) in PMOS transistors [1][2]. The BTI aging mechanism mainly leads to threshold voltage shifts in the electronic devices and its effect gets even worse in scaled technology (sub-45nm) and consequently reduces the circuit performance [3]. It has been demonstrated that the BTI is composed of two damage components a recoverable and a permanent one [4]. The important characteristic of BTI aging mechanism, that some wear-out recovery is possible when the stress condition is removed, motivates the design of dynamic reliability enhancement techniques, which can utilize this feature and result in a system lifetime enhancement.

Process variation is another reliability concern that is becoming more important in the nano-scale design and is divided into two groups of interdie and intra die variations [5]. Interdie variations are mainly the result of fluctuations in length (L), width (W) and oxide thickness (TOX) parameter, while the intradie variations occur because of random dopant fluctuations (RDF) and line edge roughness (LER) [5]. With continuation of device scaling toward nanometer regime, large statistical process variations in some of the transistor parameters such as the threshold voltage (V_T), emerge and have added another constrain in the design of reliable circuits.

Static Random Access Memory (SRAM) 6T cell is one of the most used circuit structures in the cache design. The 6T SRAM cell might need to store a bit for a long time causing its transistors to experience stress for a long period of time, and this kind of operation along with the fact that SRAM is built in near to minimum size dimensions make them susceptible to process variation and BTI aging mechanism.

978-3-9815370-0-0/DATE13/©2013 EDAA

There are some techniques to enhance the SRAM reliability such as the periodic cell flipping technique, standby supply voltage (Vdd) scaling and adaptive body biasing [5]. However, these techniques such as the adaptive body bias technique, could not mitigate any part of the BTI aging occurring in the transistors since the memory cells does not experience any recovery time during their operation.

The adaptive proactive reconfiguration is a dynamic approach, which can benefit from BTI recovery mechanism and mitigate some part of BTI wearout. It also distributes the activity between the cells according to their time zero process variation value and the amount of V_T shift during BTI stress and balances the lifetime among the 6T SRAM cells. This technique results in significant memory lifetime extension [6].

In this work, we present an implementation methodology of the adaptive proactive technique in a 1kB SRAM memory and propose a circuit design technique to be implemented in SRAM memories, which can monitor and mitigate part of process variation impact, together with the recoverable BTI aging. Section II reviews our proposed methodology and shows the benefit of using it. Section III describes the implementation and shows the simulation results, and finally section IV concludes the paper.

II. ADAPTIVE PROACTIVE CONCEPT AND METHODOLOGY

A. Concept

Conventionally, one of the techniques to improve the yield in memory is to utilize spare units, rows/columns to substitute the faulty components. This kind of reconfiguration scheme is named reactive reconfiguration [7] and though it is effective and important, it cannot mitigate the BTI aging and process variation effect.

The proactive reconfiguration concept, introduced by IBM [7], is based on the utilization of the spare elements in normal operation of the system instead of reserving them in the system up-to the time that a failure happens. This would allow active elements to operate in activated or deactivated modes for example in a rotating basis, based on a recovery schedule, and therefore recovers some part of the BTI aging effects.

Proactive reconfiguration in memories can be implemented in different granularity levels, however in the following, memory columns are used as the reconfiguration elements, as

they are more effective in repairing faults [8]. The columns can have two modes, activated or deactivated, based on a recovery schedule. The recovery periods are equal in the proactive reconfiguration approach and can result in memory lifetime extension, however equal recovery periods of units cannot mitigate the time zero process variation of transistors in SRAM memories.

Our adaptive proactive reconfiguration is an improved version of proactive reconfiguration [6] in which not only it extends the memory lifetime but also it adapts the aging of the devices according to their time zero process variation from the nominal design value and BTI stress time aging value.

B. *Our Methodology*

First the SRAM cells in the columns should be monitored by applying a test procedure in order to measure their V_T values (which is affected by process variation and BTI aging), and each column, will be recognized by its cell with the highest V_T value. Based on these values the appropriate recovery period for each column is assigned with respect to the time zero V_T value and aging status of the columns. The memory columns will go to recovery mode on a round robin schedule and experience recovery periods that adapts their aging speed with respect to their time zero process variation and the previous round BTI stress. This reconfiguration cycle (test, recovery period calculation, and each column in recovery mode) will repeat up to the end of memory columns lifetime. To show the benefits of the adaptive proactive reconfiguration approach, as a matter of example, some randomly V_T values are generated for a set of fresh 6T SRAM cells in memory columns under normal distribution with a given mean and standard deviation (300 mV and 30 mV, respectively). These arbitrary time-zero V_Ts are assumed as: $V_{T1}=330mV, V_{T2}=300mV, V_{T3}=290mV, V_{T4}=310mV, V_{T5}=320mV$ The maximum acceptable V_T aging value (H) before cell failure is assumed at 400mV and the aging slope is considered to be 10mV V_T shift per year. We assume that this system is composed of 4 working columns and 1 spare column and wear-out recovery is assumed at 30%. Figure 1 illustrates the system behavior among the memory columns for the proposed technique, where the presented example results to a lifetime extension of 2.3X times.

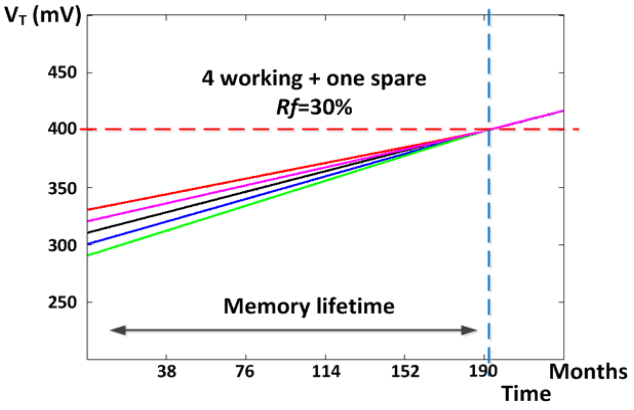


Figure 1. Aging of columns with adaptive proactive

The adaptive proactive approach makes the V_T values of the memory columns to converge toward a common value during their lifetime and balances their activity distribution.

C. *Multi spare adaptive proactive reconfiguration*

Our adaptive reconfiguration can be extended to more than one spare column in one memory column set. When having R spare columns in a group of memory columns, the columns that have V_T values close to each other are organized in one class. Then during the memory lifetime the R columns in each one of classes go to recovery mode together, which means all the columns in one class will experience the same recovery period. The recovery time of each class is adapted with process variability and degradation status of the column in the class with the highest V_T value. Figure 2 presents the results of the Monte Carlo analysis showing the lifetime extension of memory columns in multi spare adaptive proactive technique when having a set of 16 operational columns. This figure presents that; our technique could result to lifetime extensions between 2X to 4.5X times depending on the amount of BTI mitigation (30% or 50% varied according to different technologies), when having one spare column (case B).

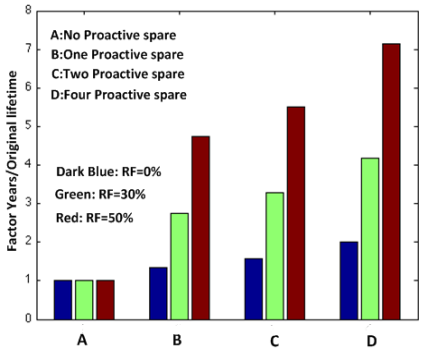


Figure 2. The lifetime extensions for a memory system based on 16 columns,with different number of proactive spare columns

The lifetime extension is increased as the number of proactive spare columns increases, for instance when the recovery factor is 50%, using one spare column can increase the lifetime about 4.5X times while having 4 spare columns can improve the lifetime of the column set around 7.2X times.

In the next section we present the implementation of our technique having only one spare column in a group of columns, however the proposed implementation can be adapted to the multi spare adaptive proactive by a small modification in the design.

III. IMPLEMENTATION OF 1KB SRAM MEMORY ARRAY WITH ADAPTIVE PROACTIVE RECONFIGURATION MECHANISM

In this section we describe a hardware implementation approach of a 1kB SRAM memory with adaptive proactive reconfiguration methodology that we proposed previously. The 1kB SRAM memory consists of 128 columns divided into 8 groups of 16 columns and each memory column contains 64 6T SRAM cells. It is assumed that the 1kB memory contains 8 spare columns and each one of the spare columns belongs to

one set of the 16 columns. Figure 3 shows the memory architecture proposed in this work with its proactive circuitry for a set of 17 columns (16+1 spare).

The memory columns start their normal operation and during the memory lifetime they go to recovery mode one by one. When a column goes to the recovery mode a test is applied to the corresponding column in order to measure the degradation of each one of the 6T SRAM cells in the column. It is assumed that the 6T SRAM cells are built up with high-k transistors, and consequently the effect of PBTI is more significant than the NBTI [3], therefore we evaluate the aging of each SRAM cell by only measuring the PBTI aging in the two NMOS pull down transistors of the SRAM (see figure 4).

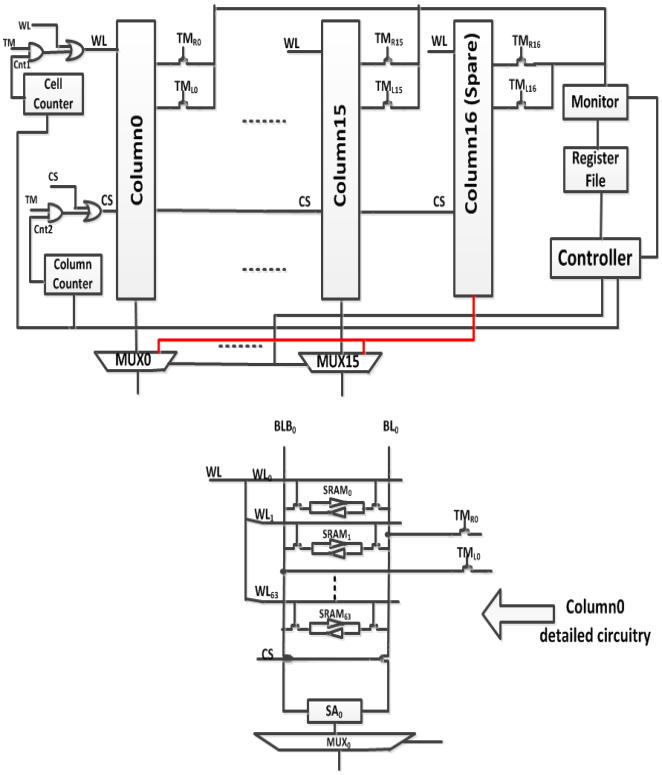


Figure 3. A group of 16+1 columns with adaptive proactive methodology and column0 with detailed circuitry for the reconfiguration

The test process has two steps as following: First a logic value one is written to all the cells in the column being tested (the column that is in recovery mode), then a counter enables each wordline from 0 to 63 and each time that one wordline is enabled the monitoring circuit measures the degradation in one of the pull down NMOS transistors of a specific SRAM cell. In the second phase a zero is written to all cells of the column, again the counter turns on each wordline from 0 to 63 and this time the monitoring circuit measures the degradation in the other pull down NMOS transistor of the SRAM cell. Observe that, since the monitoring is applied when a column is in recovery mode it does not cause any time loss in the normal memory operation. When the test of all columns finishes, the digital values representing the aging status of each column are recorded and used by reconfiguration controller in order to let the columns be in the recovery mode with a specified time adapted to their BTI aging and time zero process variability. The monitoring circuit is explained in the next section.

A. SRAM BTI monitoring circuit

In order to evaluate the time zero variation and aging status (V_T value) of the SRAM cells in a cache memory array, an on chip monitoring circuit is needed. There are already some existing approaches to measure the degradation of SRAM cells [9], [10], [11] but each one them has some drawbacks [12].

In this work, we propose a monitoring approach that can measure the BTI wearout of the individual SRAM memory cells in each memory column, which also considers the process variation among the SRAM cells and has a small physical overhead. Our approach monitors the SRAM cells degradation in a column-by-column sequence and has no effect on the normal memory operation since it is applied when the specified memory column is in recovery mode and is disconnected during the normal operation of the column. Our proposed circuit adds a small capacitance to the overall bitline capacitance but this effect can be neglected during the normal memory operation.

We have used a commercial 45nm PTM (predictive technology model) transistor model [13] to implement our circuit design in HSPICE. To model the aging behavior of NMOS transistors in the SRAM cells in the spice, a voltage source is added at the gate of the NMOS transistors in series with the gate input signal.

To implement the monitoring circuit for the 6T SRAM cells, a current mirror is connected to the bitlines of a memory column, which measures the aging in the two SRAM NMOS transistors (the pull down NMOS transistors) by tracking their current during the cell lifetime. Figure 4 shows our proposed degradation monitoring circuit of the SRAM cells and depicts the current path for this monitoring approach (dotted line).

The transistors (AC, TM) on the circuit path do not experience significant aging as they are only switched on for a short time when a specific SRAM is accessed, therefore the value of the mirrored current will be an indication of the corresponding SRAMs pull down NMOS transistor strength.

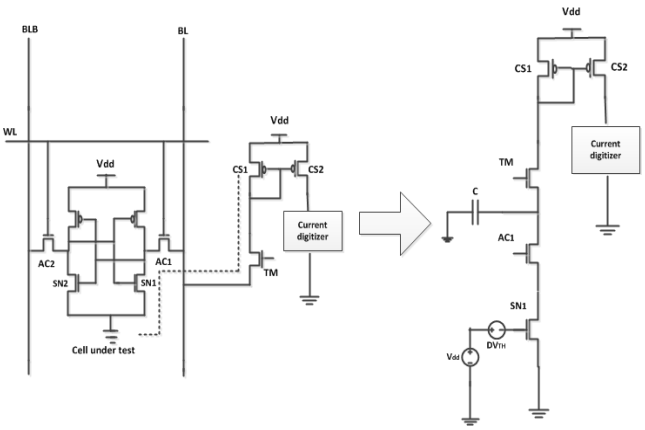
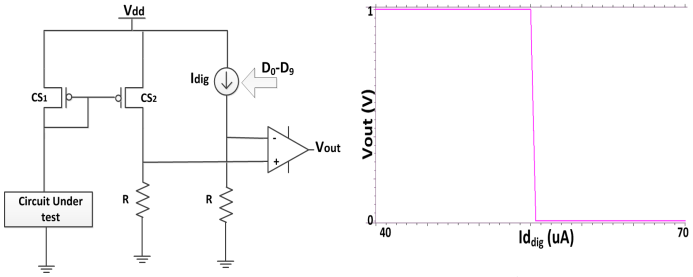


Figure 4. SRAM degradation monitoring scheme,the dotted line exhibits the path for monitoring the aging in one of the pull down NMOS transistors that is under test, the circuit on the right side depicts the aging measurement devices

We use one monitoring circuit per set of 17 columns and this avoids the impact of process variation and aging mismatch among the sensor devices as the same measurement condition is applied among all the cells and columns in a reconfiguring set. Note that the TM transistor and the current mirror transistors are designed as long and wide channel transistors in order to avoid short channel effects in the devices and assure a perfect match of mirrored current. Figure 5 shows the approach we use to digitalize the current. The current from the analyzed SRAM`s NMOS transistor (SN1) is first mirrored by the current mirror, and then it would be compared with a digital current which is adjusted by the controller through a comparator. The point where these currents get equal the output of the comparator changes the state (Figure 5b).



5a. 5b.

Figure 5.Circuit for measuring the degradation current (5a), capturing the device current in the state change of the comparator output (5b).

The controller monitors the state change of the comparator and the current number at this point, and then the digital value of the V_T in respect to it is recorded in the register file.

B. Evaluation results

We have evaluated our implementation in terms of silicon area overhead for the monitoring units. The digital units (the register file and the counters) are coded in VHDL and synthesized with RTL compiler toward CMOS 45nm lp (low power) technology library and the area overhead is indicated in the Table I. It also shows the area of our monitoring circuit implementation and the overall proactive reconfiguration monitoring design. The overall implementation uses the silicon area of 550 um^2, which is around 10% of a 1kB 45nm SRAM memory silicon area.

TABLE I. ADAPTIVE PROACTIVE MONITORING DESIGN AREA

Unit	Area μm^2
1kB 45nm 6T SRAM	6000 [14]
Our register files and counters	300
Our monitoring circuit	250
Overall monitoring units	550~ 10%

Our implemented methodology in the SRAM memory has slight effect on the memory cache performance. In each column reconfiguration step, the CPU copies the working column's data that goes into recovery mode in the spare column and this copied data is written back in the column before the next column reconfiguration step. One complete proactive reconfiguration among all the memory columns can take up to couple of the days and the frequency of reconfiguration process among the columns is very low, which allows the copying process to have enough time in order to let the two columns contain the same data, therefore the small performance loss would be only at the switching time of a column to another, once per couple of the hours or days.

IV. CONCLUSIONS

This work presents the design and implementation of an adaptive proactive reconfiguration approach to be utilized in SRAM memories. The proposed technique tracks the time zero process variation and BTI aging of SRAM cells in memory columns using an on chip-monitoring scheme. The information from the monitoring technique is used by the CPU in order to balance the work distribution and recovery periods between the memory columns and to adapt the operation of columns in such a way that the memory columns arrive to the failure point all together. Finally, our presented technique enlarges the total memory lifetime between 2X to 4.5X times with around 10% area overhead and a negligible impact on the memory performance.

ACKNOWLEDGMENT

This work was supported by the European TRAMS project (FP7 248789), and the Spanish MINECO (JCI-2010-07083 and TEC2008-01856).

REFERENCES

[1] Kunhyuk Kang ; et al. " NBTI induced performance degradation in logic and memory circuits" , ASPDAC ,pp 726-731,2008.

[2] K. Kang; et al. "Impact of negative-bias temperature instability in nanoscale SRAM array," *IEEE TCAD*, pp. 1770-1781, 2007.

[3] A. Bansal; et al. "Impacts of NBTI and PBTI on SRAM static/dynamic noise margins," *Microelectronics Reliability*, pp. 642–649, 2009.

[4] T. Grasser; et al. "Simultaneous Extraction of Recoverable and Permanent Components Contributing to Bias-Temperature Instability," *IEEE IEDM*, pp 801-804, 2007.

[5] S. Ghosh; et al. "Parameter Variation Tolerance and Error Resiliency: New Design Paradigm for the Nanoscale Era," *IEEE Proceedings*, pp 1718-1751, 2010.

[6] P.Pouyan; et al."Process-Variability Aware Proactive Reconfiguration Technique for Mitigating Aging Effects in Nano-scale SRAM Lifetime"*Proc VTS*, pp 240-245 ,2012.

[7] J. Shin; et al. "A Proactive Wearout Recovery Approach for Exploiting Microarchitectural Redundancy to Extend Cache SRAM Lifetime," *Proc of 35th ISCA*, pp. 353-362, 2008.

[8] I. Kim; et al. "Built in Self Repair for Embedded High Density SRAM,"*Proc Test*, pp. 1112-1119, 1998.

[9] K. Kang; et al. "Characterization of NBTI induced temporal performance degradation in nano-scale SRAM array using IDDQ," *IEEE Int. Test*, pp. 1-10, 2007.

[10] Z. Qi; et al. "SRAM-Based NBTI/PBTI Sensor System Design,"*IEEE DAC*, pp 849-852, 2010.

[11] F. Ahmed; et al. "Reliable Cache Design with On-Chip Monitoring of NBTI Degradation in SRAM Cells using BIST,"*Proc VTS*,pp 63-68,2010.

[12] Ceratti, A.; et al. "On-chip aging sensor to monitor NBTI effect in nano-scale SRAM,"*Proc DDECS*,pp 354-359,2012.

[13] "Predictive Technology Models (PTM) [On line]," Http://ptm.asu.edu.

[14] S. Barasinski; et al. "A 45nm single power supply SRAM supporting low voltage operation down to 0.6V,"*ESSCIRC*,pp 502-505,2008.