Aging and Performance Sensor for SRAM

H. Santos, J. Semião
University of Algarve / INESC-ID
Faro, Portugal
{a35539, jsemiao}@ualg.pt

Ruben Cabral, André Romão
Silicongate
Lisbon, Portugal
{rcabral, aromao}@silicongate.pt

M. Santos, I. Teixeira, P. Teixeira
IST / INESC-ID
Lisbon, Portugal
{marcelino.santos, isabel.teixeira, paulo.teixeira}@ist.utl.pt

Abstract—CMOS memories occupy a significant percentage of the Integrated Circuits footprint. With the development of new manufacturing technologies to a smaller scale, issues about performance and reliability exist, namely caused by parametric variations such as process variations (P), voltage (V) and temperature (T), or considering all these variations, and in a general perspective, PVTA (Process, Voltage, Temperature and Aging). This work aims to present an aging and performance sensor, for CMOS SRAM memory cells, sensing and signaling PVTA variations on SRAM memory cells. The detection strategy consists on the active monitoring of the read and write operations performed by the memory cell on the bit lines. In the presence of PVTA degradations, the memories read and write operations have slower transitions, which indicate performance degradations and also increases the error occurrence probability. Thus, when transitions doesn’t occur during the expected time frame, an error signal is signaled to the output due to a slow transition. The sensors’ operation is shown using SPICE simulations for 65nm and 22nm technologies, allowing to show their effectiveness on monitoring performance and aging on SRAM memory circuits.

Keywords—Performance Sensor, BTI, CMOS Memories, SRAM, Slow Transitions.

I. INTRODUCTION

As CMOS technologies continue to scale down to deep sub-micrometer levels, devices are becoming more sensitive to noise sources and other external influences. Systems-on-a-Chip (SoCs) and other integrated circuits, today are composed of nanoscale devices that are crammed in small areas, presenting reliability issues and new challenges. In critical system applications (for example: medical industry, automotive electronics, or aerospace applications), the performance degradation and an eventual failure can’t occur. A system error on this critical applications can lead to the loss of human lives. Thus, time is a key factor in critical-safety systems and, under disturbances, the unexpected increasing of propagation delays may lead to delay faults.

CMOS circuits’ performance is affected by parametric variations, such as process, power-supply Voltage and Temperature (PVT) [1], as well as aging effects (PVT and Aging – PVTA). The circuit’s aging degradation is pointed to the follow effects: BTI (Bias Temperature Instability), HCI (Hot-Carrier Injection), Electromigration (EM) and TDDB (Time Dependent Dielectric Breakdown) [2]. The most relevant aging effect is the BTI, namely the Negative Bias Temperature Instability (NBTI), which affects mainly the PMOS transistors, resulting in a gradual increase of absolute threshold voltage over time ($|\Delta V_{th}|$). As the high-k dielectrics started to be employed from the sub-32nm technologies [3], the BTI also affects significantly the NMOS transistors – Positive Bias Temperature Instability (PBTI), resulting in a rise of the threshold voltage $V_{th}$. These effects degrade the circuit’s performance over time, increasing the variability in CMOS circuits, mainly in nanometer technologies. The decrease of performance results in a decrease of switching speed, leading to potential fault delays and consequent chip failures.

Therefore, variability, regardless of their origin, may lead to chip failures [4], especially when several effects occur simultaneously, or when cumulative degradations pile up. Variability also decreases circuit dependability, i.e., its ability to deliver the correct functionality within the specified time frame. Hence, smaller technologies tend to be more susceptible to parametric variations, which lower circuit’s reliability and performance [5][6]. As a result, the new node SoC chips have: (i) higher performance, but with increased reliability issues; (ii) higher integration, but with increased power densities. These issues place difficult challenges on testing and reliability modelling.

Moreover, today’s Systems-on-Chip (SoC) face the rapidly increasing need to store more information. The increasing need to store more and more information has resulted in the fact that Static Random Access Memories (SRAMs) occupy the greatest part of the System-on-Chip (SoC) silicon area, being currently around 90% of SoC density [7]. Therefore, SRAM’s robustness is considered crucial in order to guarantee the reliability of such SoCs over lifetime [7]. And the trends indicate that this number is still growing in the next years. Consequently, memory has become the main responsible of the overall SoC area, and also for the active and leakage power in embedded systems.
The main purpose of this work is to develop an Aging and Performance Sensor for CMOS Memory Cells. The proposed aging and performance sensor allows to detect degradation on SRAM memory cells. The new aging sensor will be connected to the memories’ bit lines, to monitor transitions occurred in these signals during read/write operations. The purpose is to show that, by monitoring the bit lines’ operation, it is possible to monitor memory aging and memory’s performance with a very low overhead. The aging and/or performance monitoring is achieved by detecting slow transitions due to a reduction of performance caused by PVTA variations (or any other effect) in the memory cells or in the circuitry (like the sense amplifier, also connected to the bit lines). Besides, by monitoring bit line transitions, the same sensor architecture can be implemented in almost any cell structure or type.

II. BACKGROUND

A. Previous Work on Aging Sensors

One of the major issues in the design of an SRAM cell is stability. The cell stability determines the sensitivity of the memory to process tolerances and operating conditions. It must maintain correct operation in the presence of noise signals, to ensure the correct read, write and hold operations. Due to NBTI and PBTI effects, the memory cell aging is accelerated, resulting in degradation of its stability and performance.

Previous works dealing with aging sensors for SRAM cells, especially focused on BTI (Bias Temperature Instability) effect, are attempts to increase reliability in SRAM operation. Authors in [14] proposed an approach as a means of alleviating the NBTI-induced aging effects. In another approach, [15], authors propose an analog on-chip NBTI monitoring scheme to be embedded in the memory. However, as the block is analog, it may be very sensitive to process variations and aging effects. Moreover, it requires modifications of the row decoder block and write circuitry, and it doesn’t work on-line in normal operation. Another approach, [16], presents a compact on-chip sensor design that tracks NBTI for SRAMs. However, this approach is complex and has serious reliability problems. The On Chip Aging Sensor (OCAS) proposed in [7], detects the aging state of an SRAM array caused by the NBTI effect.

All these works deal with aging in SRAMs, but none is a general sensor to deal with performance and, simultaneously, PVTA variations in memories. In fact, performance sensors for memory applications are still a long way to go, and existing solutions are in an initial stage, when compared to existing ASIC performance sensor solutions. An example is the Scout Flip-Flop sensor [8][9], which acts as a performance sensor for tolerance and predictive detection of delay faults in synchronous circuits. This local sensor creates two distinct guard-band windows: (1) tolerance window, to increase tolerance to late transitions, (2) a detection window, which starts before the clock edge trigger and persists during the tolerance window, to inform that performance and circuit functionality is at risk.

Consequently, the next years will bring additional challenges that will need to be addressed with new approaches for memory applications dealing with memories’ reliability and power reduction. Therefore, there is a need for R&D work on performance sensors for memories, to deal with Process, power-supply Voltage, Temperature and Aging variations.

B. SRAM Memory Structure

The most common CMOS SRAM cell uses six transistors (6T) as seen in Fig. 1. Transistors M1, M2, M3 and M4 form a pair of cross-coupled inverters, M5 and M6 transistors are the access ones and they are connected when the word line goes high, with a bidirectional stream of current between the cell and the bit lines.

Fig. 1. CMOS 6T SRAM Memory Cell [12].

Considering a Read operation, let us assume a value of ‘1’ stored. This means that Q will be high (VDD) and /Q will be low with 0 V. Before the operation starts, the bit lines (BL and /BL) are pre-charged with VDD due to the precharge circuit. When the word line is selected (WL = VDD), M5 and M6 transistors are connected, and the current flows from VDD thru M4 and M6, charging the BL associated capacitance. On the other side of the circuit, the current from pre-charged /BL thru M5 and M1 transistors, discharging its associated capacitance. During this operation the voltage in BL rises and the voltage in /BL lowers. This creates a differential voltage between BL and /BL and the sense amplifier will detect the presence of logic ‘1’ stored into the cell.

Considering a write operation, to write a value on the SRAM cell, the column decoder selects the bit line and injects the data value (logic ‘0’ or ‘1’) intended to store on the memory cell. Both bit line are pre-charged with VDD and supposing that the cell is storing the logic value ‘1’ and it will be written the logic ‘0’, the BL is set to 0V and /BL is set to VDD. Then WL is activated and set to VDD, selecting the cell by connecting the access transistors.

In the writing operation of a logic ‘0’, a current from Q node to BL will flow, decreasing the voltage on Q from VDD to 0. On the other part of the circuit, from /BL will flow a current to the node /Q rising the node voltage to VDD. When the voltage on Q and /Q equals to VDD, the positive feedback starts and then the circuit from the previous analysis will not be applied. The new logic value is then stored.
III. PERFORMANCE SENSOR FOR CMOS MEMORY CELLS

An important module on memories is the Sense Amplifier, responsible for the identification of small differences on the bit lines and the reestablishment of digital signals (full swing), allowing the correct read of the stored values. With the aging of the transistors that compose the memory cells or even the Sense Amplifier, the physical properties and consequently the conduction of some transistors is affected, affecting the response time from the Sense Amplifier.

Fig. 2. Transitions: a) Fast transition b) Slow transition.

In Figure 2, a fast transition and a slow transition are presented. When the circuits are new and their switching times are fast, the transitions are also fast. But, when aging occurs and the transistors physical properties suffer a degradation, the result is an increase of the transistors switching time, which can result in slower transitions or degradations in nodes’ logic levels. Thus, monitoring the response time of a cell and measuring the switching times of the bit line signals, allows to measure memory cells performance and, consequently allows aging monitoring. Therefore, the proposed approach intends to detect when a slower transition occurs in a cell read/write operations, regardless of its origin (e.g., PVTA effects), allowing to monitor and detect aging degradations.

Fig. 3. Aging and performance sensor block diagram.

In Figure 3 it is shown the block diagram of the proposed aging and performance sensor. The sensor is composed by a transition detector, described in more detail in the following section, which generates pulses in the presence of a signal transition, on the memory cell bit lines, and the pulse detector block indicates if the generated pulse (which has a duration proportional to the transition time) exceeds a defined value in the pulse duration, indicating a slow transition and, consequently, a critical performance of the memory cell that could lead to a fault. In this case, an error output is generated.

A. Transition Detector

The implementation of the transition detector is presented in Fig. 4, and consists in two paths with 4 inverters each, 2 inverters with a more conductive NMOS MOSFET and 2 with a more conductive PMOS MOSFET, converging to a XOR gate with pass-transistor logic (and using transmission gates). This configuration with 8 inverters with different switching voltages, allows to detect fast and slow transitions, generating pulses with similar pulse widths when the rising and the falling edge occur at the input. Moreover, the XOR gate used is not a classic CMOS XOR gate, but rather a pass-transistor logic XOR gate, which includes an inverter at its output, and ensures good performance without logic levels degradation when logic ‘0’ or logic ‘1’ are passed to the output (Fig. 5).

Fig. 4. Transition detector’s implementation.

Fig. 5. Pass-transistor XOR gate implementation

TABLE I. TRANSITION DETECTOR TRANSISTORS’ SIZES.

<table>
<thead>
<tr>
<th>Path</th>
<th>Inverter</th>
<th>Nmos</th>
<th>Pmos</th>
<th>L</th>
<th>V_{th,n}</th>
<th>V_{t,p}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Xinv1</td>
<td>5*WNmin</td>
<td>WPmin</td>
<td>65n</td>
<td>0.423V</td>
<td>-0.365V</td>
</tr>
<tr>
<td></td>
<td>Xinv2</td>
<td>WNmin</td>
<td>5*WPmin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xinv3</td>
<td>5*WNmin</td>
<td>WPmin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xinv4</td>
<td>WNmin</td>
<td>5*WPmin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Xinv1</td>
<td>WNmin</td>
<td>5*WPmin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xinv2</td>
<td>5*WNmin</td>
<td>WPmin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xinv3</td>
<td>5*WNmin</td>
<td>5*WPmin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xinv4</td>
<td>5*WNmin</td>
<td>WPmin</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B. Pulse Detector

The proposed solution for the pulse detector is to use a stability checker (Fig. 10) to detect transitions in the pulses created by the transition detector, after a specific time instant defined by the clock. The idea is to reuse the main concepts of the Scout Flip-flop, previously presented in [8][9], which detects all transitions in the data input that reaches the stability checker during the active pulse of the clock. The delays introduced by the sensor circuitry should allow to define when the sensor is activated or not. Moreover, reusing Scout Flip-flop’s solution here should bring robustness to the solution and improve pulse detector’s sensitivity in the presence of PVTA degradations, because delays are also sensitive to PVTA degradations, and because a reliable operation of a digital circuit is directly related to the clock frequency used (the heartbeat of the entire system), i.e. if the clock frequency is reduced (increased), the performance is relaxed (excited) and the error probability is alleviated (aggravated).

The new pulse detector solution is presented in Fig. 6, and it comprises a delay element, an inverter, and a stability
checker. The delay element is basically a buffer, to provide a

time delay to the input signal, and its architecture was already

presented in [8][9][10]. According with the time delay needed

and the clock frequency, one or more elements can be used

from the three solutions presented in [9]. The stability checker
(Fi. 6) is used here to detect transitions in the delayed pulses

obtained from the delay element. The difference from Scout

Flip-flop’s solution is that now the clock feeds the stability

checker through an inverter, which means that it detects

transitions during the low state of the clock.

Fig. 6. Stability-checker implementation.

![Stability-checker implementation](image)

The basic idea is to use the main clock as a fixed reference
to detect abnormal delays in the pulses generated by the
transition detector. In the memory (and in a common digital

circuit), all the control signals and all the instructions are
generated synchronously with the main clock. Therefore,

considering that pulses in the transition detector are generated
during the active state of the clock, if pulse duration and the

propagation delay of the delay element makes the delayed
pulse to reach the stability checker during the low state of the

clock, an error signal will be generated. This means that by
design we have two parameters where we can control the
delays in the sensor and the error/non-error decision: one is the

sensibility of the transition detector and the width of the pulses

generated; the other one is the time delay introduced in the

signal in the delay element. Fig. 7 resumes the operation of the

pulse detector presented in Fig. 6. Note that the grey areas
represent the operation when a slow-transition occurs in the

bit-line signal (or a degraded logic level in the bit-line),
indicating a performance reduction. In these cases, the pulses

generated by the transition detector are wider, the delay added
in the delay element is larger, which results in an error signal at
the output of the sensor (when the delayed pulses reaches the

stability checker during the low level of the clock).

IV. SENSOR ROBUSTNESS RESULTS

The simulation environment for the sensor robustness tests
will submit the circuitry through aging effects, by shifting the

Vth on the PMOS and NMOS MOSFETs, using Berkeley
Predictive Technology Models (BPTM) for 65nm transistor
models [17].

A. Transition Detector’s Robustness

To test the transition detector’s robustness, the HSPICE
netlist was developed and a VDD sweep was performed at the

input, from 0.8V to nominal VDD of 1.1V, with a step of 0.05,

and using a square pulse. The simulation results are presented
in Fig. 8(a), and as it can be seen, the generated pulses increase

its pulse width when the voltage value is reduced to 0.8V. It’s

also noted that the pulses are stable and symmetrical, with

almost the same pulse width in both pulses generated in each

simulation.

Fig. 8. Transition detector implementation response: (a) to voltage variation;
(b) to temperature variation.

![Transition detector implementation response](image)

A similar test was done for temperature, using a sweep in
the between 25 °C and 80 °C, with a step of 10 °C. As it can be
observed in Fig. 8(b), as temperature rises, the pulse width of
the generated pulses increase, due to a transistors’ conduction
degradation caused by the temperature raise.

The last robustness test was related with the aging variable.
In the following simulations, results are presented for the
impact of NBTI and PBTI effects, separately, on the transition
detector performance, changing Vthp and Vthn in the

transistors. It is important to note that aging insertion is very
complex if we want to simulate real situation, as aging is
affected by many parameters (and differently along time). For
instance, it is impossible to predict circuit workload, because it
depends on unknown variables such as human interaction.
However, as done in other related works ([10], [11]), it can be
statistically modeled and, assuming some workload, aging
degradation can be computed (in the form of Vth modulation,
if BTI is considered) for each circuit’s transistor. Also, in a

simpler approach and as done also in precious works ([7]), a
single Vth modulation can be assumed for all transistors,
allowing an easier methodology for aging insertion and
producing a first approximation of circuit aging. In this work,
this simpler aging insertion criteria was adopted, and equal Vth
modifications are done for NMOS and PMOS transistors when
considering, respectively, PBTI and NBTI effects. Fig. 9(a)
presents the results for NBTI aging effect, with a sweep in

Vthp between -0.365 to -0.7, with steps of 0.2. Similarly, Fig.
9(b) presents the results for PBTI aging effect, for a sweep in

Vthn between 0.365 and 0.6, with steps of 0.2. It’s noted in
both pictures that the pulses increase when [Vth] is raised. In
other words, transistors’ aging increases pulse widths in the transition detector.

Fig. 9. Transition detector implementation response: (a) to Vthp variation; (b) to Vthn variation.

Let us give a brief note on process variations. These are random variations in transistors parameters, which usually are modeled by random Vth, Leff (transistors effective channel length), and other, and can produce a degradation or an improvement in performance. Commonly, performance degradation caused by process variations are considered when worse case design corners are met, resulting in specific process variations that lead to a worst case degradation. Therefore, similar results to aging insertion are obtained when considering a worse case design corner for process variations.

In conclusion, and from the performed test results to transition detector implementation in the presence of PVTA variations, we can conclude that this implementation is a robust solution. It was demonstrated that, when transistors suffer a performance degradation due to PVTA variations, the transition detector generates larger pulses, and this result can be used to build a robust aging and performance sensor.

B. Pulse Detector’s Robustness

To test and implement the stability-checker based pulse detector, a SPICE netlist was used and an HSPICE simulation was performed and the results are presented in Fig. 10. A VDD of 1.1V and a temperature of 25˚C was used in the simulation for Fig. 10(a), with an input with two different pulses. The upper graph shows the input pulse, the bottom graph shows the output signal. As it can be seen, the output is not activated for the first pulse (the shorter pulse), but it is activated for the larger pulse, indicating the correct functionality of the implementation. However, when considering VT variations with the temperature raised to 80˚C and the power-supply voltage decreased to 0.8V, the simulation results of Fig. 10(b) show that in the presence of VT degradations both pulses are now signalized as errors. This result confirms the stable and robust behavior of the stability checker in the presence of VT variations.

In a second robustness test, aging degradations were introduced, in the form of 30% increase in |Vth| values for the transistors. The results are presented in Fig. 11, and as we can see, similar results are obtained for VT and Aging degradations (which again confirms the robustness of the solution for VTA degradations).

C. Complete Sensor Robustness

Fig. 12. Results for complete sensor simulation in nominal conditions.

Considering the complete solution for the sensor, i.e., using the transition detector and the pulse detector, simulations were performed to validate the entire sensor solution. Now, two different situations were tested: (1) an input signal with fast up and down transitions (small rise and fall times), and (2) an input signal with slow up and down transitions (large rise and fall times). Fig. 12 shows the results for nominal conditions, Fig. 13(a) for sensor’s operation with VT degradations (T=80˚C, VDD = 0.8V) and Fig. 13(b) for operation with aging degradations (30% increase in |Vth|). In these simulations, the first graph in each figure shows the input signal, representing the memory bit lines; the second graph shows the transition detector output signal, showing the pulses generated for each input signal’s transition; and the third graph in each figure shows the output of the pulse detector (sensor’s output), indicating with an high state when a performance error has occurred. As it can be seen, for nominal conditions the error is
only signalized for the slower transition (Fig 12). However, when VTA variations occur, both slower and faster transitions are signalized at the output as errors, due to the performance reduction achieved with VTA variation (Fig. 13).

V. SRAM AND SENSOR RESULTS

For the SRAM and sensor results, the test environment will include the sensor, an SRAM memory cell and all of its peripheral circuitry, namely the sense amplifier, the pre-charge and the equalizer circuit. The test SRAM cell is a 6T cell, and the transistor sizes (namely: (i) the ratio between pull-down and access transistors, (ii) the ratio between pull-up and access transistors (iii) access transistors) were determined to ensure its robustness. In this section, a 22nm technology was used to implement all the layouts and simulations.

Fig. 14 shows a 64 (8x8) bit SRAM implementation, 8 rows and 8 columns. Each bit line column uses an aging and performance sensor, to monitor each column of the 8 bit SRAM cell. For larger memories, the same principle should be used, i.e, 1 sensor connected to each bit line column.

The netlist was extracted from the layout design and then simulated on HSPICE. Fig. 15 shows a situation of no detection. The inverted pulses generated (blue color) are contained inside of the clock pulse (which is not seen in this figure), thus not entering on the detection window, meaning the presence of a fresh memory cell.

The SRAM cell was aged, by changing the Vth nominal values in 10%, for PTM 22nm transistor model [17]: Vthp = -0.63745V and Vthn = 0.68858V. In Fig. 16, it can be observed that the aging degradations of 10% increase in |Vth| cause the increase of pulse width and also shifts the position of the inverted pulses, due to the decrease of performance caused by the transistors aging. This way the pulses enter the detection window, and the sensor responds ‘1’ to the output, which means an error caused by circuits’ aging.

As a final analysis, sensor showed similar results for all PVT degradations, which shows its robustness, being more sensible to when these degradations occur. This result is very important, because it brings additional improvements to memory sensors, namely its adaptive sensibility to PVT variations, and changes the paradigm of memory sensors. In the presence of enhanced variability in PVT degradations, the sensor increases its sensitivity to detect slower performances in memory circuits’ operation, by increasing locally the pulse duration in the pulse detector, and by increasing the delays in the transition detector, according with the local PVT variations. This feature simplifies also sensor design, as the sensor itself does not need to be more robust to PVT variations than the circuit where it is installed. On the contrary, sensor’s performance degradation due to PVTA improve its sensibility to detect slower transitions in the bit-lines. Thus, sensor aging works in favor of its sensibility. Moreover, it can be always active, monitoring every transition that occurs in the memory. However, considering that transistors age more when they are statically in stress mode, it is recommended that, sparsely in time, at least a memory read operation should be executed, to stimulate sensor operation.

It is important to note that, no test was performed to measure memory’s performance degradation due to the additional capacitance connected to the memory resulted by installing the sensor. However, we believe that the impact of the sensor in the cell performance is completely negligible, because bit-lines have high parasitic capacitances, due to the high number of cells connected, and the sensor connection uses a small delay element (which is a simple buffer) to connect to the bit-lines. Therefore, the degradation is similar to an additional memory bit, which is negligible if we consider a high capacity memory.

VI. CONCLUSION

This work focused on the development of an aging and performance sensor for SRAM memories, to detect and signalize aging and performance degradations. The main goal was to develop a new sensor, aiming the identification and signalization of slow transitions occurrences on memory cells caused by PVT degradations. Results showed that the aging
and performance sensor, connected to the bit-lines and actively monitoring the transition delays of the read and write transitions, can effectively detect PVT degradation.

For future work, improvements are expected in the pulse detector block, to reduce its complexity and improve operation. Sensor applicability to DRAM cells should be studied in the future. Also, silicon validation is important and should be pursued.

REFERENCES


