Power-Gating: More Than Leakage Savings

Power-Gating: More Than Leakage Savings

Andrea Calimera
andrea.calimera@polito.it

Enrico Macii
enrico.macii@polito.it

Massimo Poncino
massimo.poncino@polito.it

Politecnico di Torino, 10129, Torino, Italy

***Abstract*—While CMOS technology keeps running towards the physical limit of "nanometer" lengths, many critical design issues have already appeared in today's technologies (65nm and 45nm). Among them, Power and Reliability are the most insidious, because they affect energy efficiency and life-time of digital systems. In this work we establish an important link between these two metrics. More specifically, we show how the most widely adopted leakage-reduction technique, that is power-gating, can go beyond its specific goal (i.e., saving leakage) by providing, per se, an effective solution to mitigate NBTI-induced aging. Based on this important property, we first present an automated methodology that allows a push-button estimation of the aging effects induced by NBTI on logic circuits (in terms of delay degradation) and SRAM memory cells (in terms of Static Noise Margin (SNM) reduction). Second, using an industrial 45nm technology, we quantify the actual capability of power-gating to further reduce the aging of CMOS devices and extend the lifetime of digital circuits.**

I. Introduction

As CMOS technology has shrunk to the nanometer regime, the design space of VLSI circuits has been populated by a multitude of new constraining variables. Among them, *power consumption* and *reliability* have been recognized as the most critical.

It is well known that the power consumed per unit area is converted into heat, which, in turn, induces sensible increase of the substrate temperature. This implies a double effect. On the one hand, static power (i.e., leakage), which has paired dynamic power [1] in modern technologies, grows exponentially with temperature exacerbating the problem of energy efficiency. On the other hand, high temperature accelerates various degradation effects, which induce *time-dependent* changes in the operating characteristics of devices, or, even worse, the occurrence of physical faults. Historically, Electromigration, Hot Carrier Injection (HCI), Time Dependent Dielectric Breakdown (TDDB), have been indicated as the main source of unreliability, but recently, Negative Bias Temperature Instability (NBTI) has emerged as the most critical issue in determining the lifetime of CMOS devices [2].

In CMOS logic, NBTI effects occur when a pMOS is negatively biased (i.e., a logic '0' is applied to the gate of the pMOS, resulting in $V_{gs} = -V_{DD}$), and manifests itself as an *increase of the threshold voltage V_{th} with time*. Increasing the V_{th} of pMOS devices reflects on logic circuits, in the form of delay degradation [3], but also on Static RAM (SRAM) memory cells, which lose on robustness and storage capability due to a reduced Static Noise Margin (SNM) [4]. In contrast, when a pMOS is positive biased (i.e., a '1' logic drives the pMOS, resulting in $V_{gs} = 0V$), NBTI stress is actually removed, resulting in a partial recovery (i.e., *decrease of the threshold voltage*).

The urge of *reliable low-power* circuits is driving the CAD community to develop new design techniques and circuit solutions able to address these issues. Unfortunately, this challenge is complicated by the fact that power and reliability are known to be intrinsically conflicting metrics: traditional solutions to improve reliability such as redundancy, increase of voltage levels, and up-sizing of critical devices do contrast with traditional low-power solutions, which rely on small devices and scaled supply voltages.

While in previous works [3]-[6] this conflicting relationship was completely ignored, our previous works [7], [8] tried to mitigate the incompatibility between power and reliability by establishing an important missing link between the two metrics. More specifically, we demonstrate how NBTI, that is the most insidious source of time-dependent variations (i.e., aging), can be effectively alleviated by *power-gating*, widely known to be the most effective technique to reduce static power consumption. We propose an automated flow for the assessment of NBTI induced aging effects in digital circuits (i.e., logic circuits and SRAM memory). The proposed methodology, which incorporates the characterization of the library gates, of the 6T SRAM cell, and the effects of power-gating for an industrial 45nm CMOS technology, allows the designers to quantify, at design time, the effects of NBTI as well as the capability of power-gating to further reduce the aging degradation. Experimental results show, for a sleep probability of 80%, life time extension of more than 2X for logic circuits and around 8.5x 6T-SRAM memory cells with respect to non power-gated solutions. Our methodology represents therefore an enabling technology of future aging-aware design flows, in which aging and power are included, together, in the optimization loop.

II. NBTI Effects

When negative biased, pMOS devices show a sensible increase of the threshold voltage V_{th} due to the generation of traps in the region close to the Si/SiO_2 interface. The Reactivation-Diffusion (R-D) model [2] has emerged as the more accredited model for pMOS NBTI. However, since a detailed treatment of NBTI effects and models is out of the scope of this paper, we limit our contribution to summarize the basic factors that impact NBTI effects.

Operating Conditions: for a given set of technological parameters and physical dimensions (e.g., nominal threshold voltage, effective channel sizing, oxide thickness) NBTI effects are mainly dependent on *(i) temperature T* (degradation increases with increasing T), and *(ii) supply Voltage V_{gs}* (delay degradation increases with increasing V_{gs}).

Signal Statistics: NBTI induced effects strongly depends on the actual amount of stress time (i.e., time in which the gate is negative biased). When stress is not applied (i.e., when $V_{gs} = 0V$), however, a partial *recovery* of of delay occurs. As shown in [9], thanks to some physical properties, it is possible to use *signal probabilities* for characterizing the actual stress/recovery waveform.

Idleness: in the most popular version [10], power-gating is based on the insertion of n-type MOS transistors, called sleep transistors, in the pull-down network of logic blocks. When the circuit is idle, the sleep transistors are turned off, thus sensible reducing static power consumption. During the stand-by periods, nodes with a logic value '1' will keep their value, whereas nodes with value '0' will be quickly charged to the logic '1' [11]. The '1' state corresponds to the NBTI recovery state, thus, from an aging perspective, this implies that logic blocks in a stand-by state are naturally immune to NBTI-induced aging [7]. Based on this observation, we can state that a larger idleness may help to reduce the NBTI-induced delay degradation.

III. Exploration Framework

A. Power-Gated Logic Circuits

Figure 1 shows the implemented NBTI-aware exploration framework for power-gated circuits. After obtaining a synthesized circuit, a post-synthesis simulation (i.e., VCD file) is needed to extract the statistical information of all the internal nodes. This also encompasses the analysis of the idleness periods of the circuits and the extraction of the sleep signal temporal distribution (i.e., P_{sleep}). Depending on

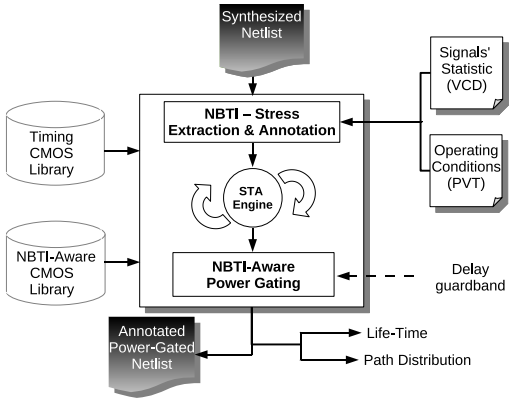

Fig. 1. Flow of the Proposed Exploration Framework for logic Circuits.

the operating PVT corner (i.e., Process, supply Voltage, and Temperature), and the static 0-probability of internal signals, the NBTI-induced delay degradation of each standard cell is extracted. This is done with the support of new NBTI-aware timing libraries that support time-dependent variations. Similarly to standard timing libraries, we filled look-up tables containing the NBTI-induced delay degradation of each cell under different PVT corners and elapsed time.

The netlist, annotated with the NBTI information, is then loaded into a standard Static Timing Analysis (STA) engine that provides timing information of the aged circuit. The collected aging curves are used to drive the design of the power-gating architecture, namely, to size the sleep transistor in order to provide the power-gated circuit with an extended lifetime. To be noticed that the lifetime is measured as the time at which the aging curve crosses a user defined *delay guard-band*.

The need of a power-gating design strategy that is aware of the NBTI effects relies on the fact that the application of power gating comes at a cost of some delay penalty, here indicated with γ. This penalty must be weighted against the potential benefit it offers in terms of lifetime extension. Even if power-gating is always beneficial in terms of aging (i.e., exploiting the idleness the aging curve always grows more slowly than the non power-gated case) when γ is too large, the aging curve of the power-gated circuit may intersect the one of the non power-gated circuit at a time that is well beyond the typical lifetime of a circuit.

B. Power-Gated SRAM Cells

Power-gating can be also applied to SRAM memories using different level of granularity: from cell level to block-level (single or multiple row/column). Since our analysis is focused on a single SRAM cell, we have considered a cell-based gating approach [12]. As depicted in Figure 2 (the SRAM cell in the left box), the ground terminals of the two cross-coupled inverters are tied together to the virtual ground signal, which is decoupled from the true ground by the sleep transistor. When turned-off (standby state), the cells is disconnected from the ground, and, both inverters' outputs will quickly reach the '1' value, corresponding to a NBTI-immune configuration.

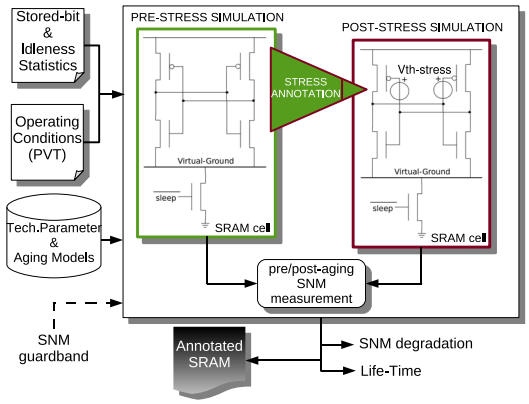

Fig. 2. Flow of the Proposed Exploration Framework for 6T-SRAM cells.

A good metric for the SRAMs aging is given by the Static Noise Margin (i.e., SNM). The SNM is defined as the minimum DC noise voltage necessary to change the state of an SRAM cell [13] and it measures the actual robustness of the cell: the larger the SNM, the larger the capability of the cell of storing a value. However, when the pull-up pMOS are negative biased, NBTI effects induce V_{th} shift over time, thus moving the static characteristics of the two inverters, and, in turn, reducing the SNM.

Our SPICE-based characterization framework, shown in Figure 2, allows to predict, under user-defined PVT conditions, the aging of a power-gated SRAM. Besides the technology

parameters, the automated flow receives, as inputs, the statistic profile of the stored bit (i.e., probability to store a '0' logic $0 - Probability$) and the idleness of the cell (i.e., P_{sleep}). It is important to note that, differently form logic circuits, a SRAM cell ages irrespective of the value it stores [4]. Due to the symmetric structure of a cell, in fact, one of the two inverters is always negatively biased. The best case (smallest degradation) happens when the value at the output of each inverter is '0' 50% of the time, i.e., both PMOS degrade of the same amount.

The analysis consists of a two-phase simulation. In the *pre-stress* simulation phase we compute the aging of the pMOS transistors based on HSPICE built-in aging models fitted to the technology parameters provided by silicon vendor. The aging information sampled during pre-stress simulation are then translated into device parameter degradation (i.e., threshold voltage degradation $V_{th} - stress$) and annotated into the SRAM cell netlist in the form of additional DC-controlled voltage sources on the gate terminal of each pMOS transistor. After the netlist has been annotated, we run the post-stress simulation and the collected results are used to measure the Static Noise Margin. More precisely, we refer to the *read SNM*, (i.e., when the cell is operating with lateral nMOS access transistors on), which represents the worst case condition for aging [4]. SNM is numerically extracted by using the graphical method described in [13]. By comparing the pre-stress and post-stress SNMs, the aging curves are profiled and the lifetime of the cell calculated. Similarly to logic circuits, the lifetime is measured as the time at which the SNM goes below a user defined *SNM guard-band*.

It is important to highlight that while in logic circuits the sleep transistor insertion induces delay penalty due to an increase of the Virtual-GND, in SRAM cells the value of virtual ground does not affect the SNM, which is by definition a DC quantity. In this sense, power-gating can always help to alleviate the aging of SRAM cell.

IV. Experimental Results

In order to demonstrate the natural property of power-gating of mitigating the effects of NBTI on digital circuits, we implemented the proposed exploration frameworks using standard EDA tools (*Synopsys Design-Compiler, Prime-Time, HSPICE* and *Mentor Graphics' ModelsSim*) and an industrial 45nm CMOS technology provided by STMicroelectronics.

A. Aging of Logic Circuits

The logic benchmarks have been taken from the public ISCAS85 suite, while the switching information of the internal nodes have been obtained running dedicated testbenches that emulate actual workloads. Concerning power-gating, in this work we consider an approach in which the entire circuit is connected to distributed sleep transistor cells driven by the same control sleep signal (i.e., *complete* power gating). The size of the sleep transistor has been calculated constraining the maximum virtual-ground voltage to 10% of V_{DD} [14].

Table I summarizes the results in terms of lifetimes extension, area overhead, and leakage saving for all the benchmarks. We used a delay guard-band value of 15% extra delay over the nominal delay, namely, we measure the lifetime as the time required by the circuit to degrade its performance of 15% beyond its nominal value. For the sake of space, we only report data for two extreme values of γ (1.5% and 7%), which are the extreme values we allowed in our characterization framework, and for a sleep probability P_{sleep} of 0.8 for all the benchmarks.

	Lifetime			*Leakage Saving*		*Area Penalty*	
Bench	*NO-PG*	*with PG*		*with PG*		*with PG*	
		$\gamma =$ 1.5%	$\gamma =$ 7%	$\gamma =$ 1.5%	$\gamma =$ 7%	$\gamma =$ 1.5%	$\gamma =$ 7%
c432	1.0	3.40	0.31	77.07	88.17	7.87	4.06
c499	1.0	2.20	0.36	81.66	90.54	11.49	5.93
c880	1.0	2.04	0.32	90.87	95.29	3.13	1.62
c1355	1.0	1.53	0.23	69.75	84.40	11.94	6.16
c1908	1.0	2.71	0.51	74.62	86.91	13.30	6.86
c2670	1.0	1.60	0.54	87.83	93.73	5.33	2.75
c3540	1.0	2.94	0.47	80.15	89.76	9.14	4.72
c5315	1.0	1.72	0.67	80.78	90.09	8.34	4.30
c6288	1.0	1.79	0.24	75.16	87.19	7.15	3.69
c7552	1.0	1.71	0.53	65.11	82.01	9.80	5.06
avg.	**1.0x**	**2.17x**	**0.42x**	**78.30%**	**88.81%**	**8.75%**	**4.51%**

TABLE I
RESULTS ON ISCAS85 BENCHMARKS.

The column labeled with *Lifetime – with PG* shows the lifetime values resulting from the application of power gating with different amount of "effort". The reported values have been normalized to the lifetime of the original, non power-gated circuit. As discussed in Section III-A a too large γ (7%), due to a too small sleep transistor, will imply a too high initial overhead, resulting in exceeding the guard-band delay earlier than the non power gated circuit and lifetime degradation (42% on average). Conversely, using a larger sleep transistor ($\gamma = 1.5\%$), we are able to extend the lifetime by more than 2X on average.

The reduction of γ comes of course at the cost of a larger sleep transistor, which, in turn, affects area. Reported values highlight an average area penalty (Column *Area Penalty*) of 8.75% and 4.51% for $\gamma = 1.5\%$ and $\gamma = 7\%$ respectively. Larger sleep transistors do not just affect layout dimension, but also power saving. In Table I, leakage figures (Column *Leakage Saving*) refer to the total stand-by leakage power saved thanks to the application of power-gating. We take as reference the ideal condition under which the gated logic consumes zero leakage when disconnected from ground (i.e., 100% of leakage saving). As a matter of fact, doubling the lifetime, namely, using $\gamma = 1.5\%$, reduces the total leakage saving down to 78.3%, just 10% less than the case of $\gamma = 7\%$ (for which the lifetime is halved).

It is also worth mentioning, that the idleness is a key variable in the aging-leakage space since it can sensible affect the figures of merit reported in Table I. For instance, a too small idle periods may induce negligible lifetime extension. Figure 3 shows the lifetime of the ISCAS benchmarks for different values of P_{sleep} (from 0.4 up to 0.9) and for a fixed $\gamma = 1.5\%$. The bars are normalized to the lifetime of the non power-gated circuit *NO-PG*. For all the circuits, power-gating is always beneficial and it increases the lifetime for any P_{sleep}. The only exception is the circuit `c1355`. In this case when $P_{sleep} = 0.4$, the delay penalty induced at time-zero exceeds the amount of delay saving during idle periods. A practical solution is to

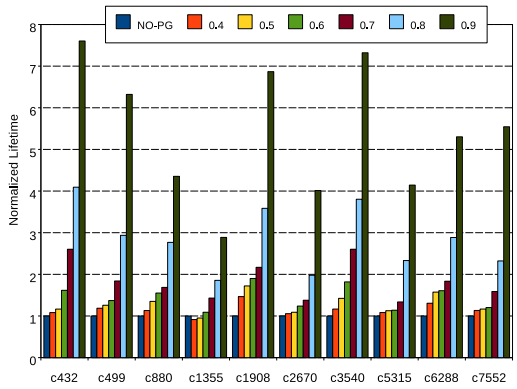

Fig. 3. Lifetime of Power-Gated ISCAS85 Benchmarks.

further reduce the value of γ, so that the crossing point of the aging curve with the guard-band threshold translates over time.

B. Aging of SRAM Cells

Our base SRAM cell consists of a conventional six-transistors cell. Transistor sizes have been determined as a reasonable tradeoff between cell density and robustness [4]. Concerning the sleep transistor size, the value has been chosen so as to guarantee a maximum virtual ground voltage V_{VGND} of 10% of the supply voltage (0.11V) during write operations. In

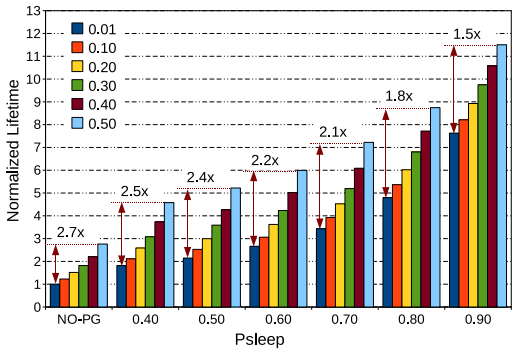

Fig. 4. Lifetime of a power-gated SRAM memory cell vs. 0-Probability.

Figure 4 we compare the lifetime of the non power-gated cell (NO-PG) with respect to the power-gated cell under different idleness conditions. Each set of bars corresponds to a different probability of the sleep signal, from 0.4 (40% of idleness) to 0.9 (90% of idleness)[1]. At each bar, instead, corresponds a different probability of storing a '0' logic (*0-Probability*), from 0.01 (1%) to (0.5) 50%. Values for higher 0-probabilities are symmetric with respect to the 0.5 value. The lifetimes, measured as the time required by the cell to reach a 15% of SNM degradation, have been normalized to the worst case, that is, non power-gated cells (NO-PG) with minimum 0-Probability (1%).

The graph shows that the application of power-gating provides, regardless of the 0-Probability, exceptional lifetime extensions. Let focus on a single case, namely, 0-Probability=50% (left-most bars of each bar set). Under 40% of idleness, the lifetime can be increased of around 80% (1.8x in the plot), while increasing the idleness up to 90%, the lifetime is 7.5 times larger than the NO-PG case. These results provide interesting insights on the possible applications in leakage-aware memory architectures. Playing with signal probabilities, however, can provide sizable benefits. Techniques that force 50% signal probabilities in the cells [5] and power-gating can be jointly applied resulting, in the best case ($0 - Probability = 0.5$ and 90% of idleness), on a lifetime extension of 11.5 times.

A less intuitive, thus more interesting, conclusion can be drawn by observing the variance of the bar sets in Figure 4 as a function of the idleness. Considering the non power-gated cell (NO-PG bar set) the ratio between the lifetime when 0-Probability=1% and the lifetime when 0-Probability=50% is 2.7, while for the power-gated cell with 90% of idleness (0.90 bar set), the ratio reduces to 1.5. This highlights that, *as idleness increases, the impact of controlling the bit probabilities is progressively less important*. A potential application of this property could be that bit-probability control techniques could be applied selectively, for instance only on those portions of a memory with a low "idleness" (e.g., with high access rate).

V. Conclusions

In this work we explored an new way to mitigate aging effects in circuits and in memory cells, that is, the possibility of exploiting low-power techniques for concurrent leakage and aging optimization. More specifically we showed how power-gating provides a natural way of reducing NBTI effects with minimal overheads. Experimental results conducted on an industrial 45nm technology show a life time extension of more than 2X for logic circuits, and around 8.5x for 6T-SRAM memory cells.

[1]The 100% case will exhibit no aging, since the cell is always power-gated.

References

[1] K. Roy, *et al.*, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," *Proceeding of the IEEE,* vol. 91, no. 2, pp. 305-327, 2003.
[2] M. Alam, "Reliability- and process-variation aware design of integrated circuits," *Microelectronics Reliability,* vol. 48, no. 8, pp. 1114-1122, 2008.
[3] R. Vattikonda, *et al.*, "Modeling and minimization of PMOS NBTI effect for robust nanometer design," *DAC-44: ACM Design Automation Conference,* pp. 1047-1052, 2006.
[4] K. Kang, *et al.*, "Impact of Negative-Bias Temperature Instability in Nanoscale SRAM Array: Modeling and Analysis," *IEEE Transactions on CAD*, vol. 26, no. 10, pp. 1770-1781, 2008.
[5] S. V. Kumar, *et al.*, "Impact of NBTI on SRAM read stability and design for reliability," *ISQED-06: International Symposium on Quality Electronic Design,* pp. 213-218, 2006.
[6] S. V. Kumar, *et al.*, "NBTI-Aware Synthesis of Digital Circuits," *DAC-45: ACM Desing Automation Conference*, pp. 370-375, 2007.
[7] A. Calimera, *et al.*, "NBTI-Aware power gating for concurrent leakage and aging optimization," *ISLPED-09: International Symposium on Low power Electronics and Design,* pp. 127-132, 2009.
[8] A. Calimera, *et al.*, "Analysis of NBTI-Induced SNM Degradation in Power-Gated SRAM Cells," *ISCAS-10: IEEE International Symposium on Circuits and Systems,* pp. 785-788, 2010.
[9] S. V. Kumar, *et al.*, "An analytical model for negative bias temperature instability," *ICCAD-06: 2006 IEEE/ACM international conference on CAD,* pp. 493-496, 2006.
[10] M. Anis, *et al.*, "Design and Optimization of Multi-Threshold CMOS (MTCMOS) Circuits," *IEEE Transactions on CAD,* vol. 22, no. 10, pp. 1324-1342, 2003.
[11] A. Calimera, *et al.*, "Optimal MTCMOS reactivation under power supply noise and performance constraints," *DATE-08: Design Automation and Test in Europe,* pp. 973-978, 2008.
[12] K. Nii, *et al.*, "A Low-Power SRAM using Auto-Backgate-Controlled MT-CMOS," *ISLPED-98: International Symposium on Low Power Electronics and Design,* pp. 293-298, 1998.
[13] E. Seevinck, *et al.*, "Static-noise margin analysis of MOS SRAM cells," *Journal of Solid-State Circuits,* vol. 22, no. 5, pp. 748-754, 1987.
[14] A. Sathanur, *et al.*, "Efficient Computation of Discharge Current Upper Bounds for Clustered Sleep Transistor Sizing," *DATE-07: IEEE Design Automation and Test in Europe*, pp. 1-6, 2007.