

Citation for published version:

Daniele Rosi, Vasileios Tenentes, Sudhakar M. Reddy, Bashir M. Al-Hashimi, and Andrew Brown, 'Exploiting Aging Benefits for the Design of Reliable Drowsy Cache Memories', *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, July 2017.

DOI:

<https://doi.org/10.1109/TCAD.2017.2729399>

Document Version:

This is the Accepted Manuscript version.

The version in the University of Hertfordshire Research Archive may differ from the final published version.

Copyright and Reuse:

© 2017 IEEE.

Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Enquiries

If you believe this document infringes copyright, please contact the Research & Scholarly Communications Team at rsc@herts.ac.uk

Exploiting Aging Benefits for the Design of Reliable Drowsy Cache Memories

Daniele Rossi, *Senior Member, IEEE*, Vasileios Tenentes, *Member, IEEE*, Sudhakar M. Reddy, *Fellow, IEEE*, Bashir M. Al-Hashimi *Fellow, IEEE*, Andrew Brown

Abstract—In this paper, we show how beneficial effects of aging on static power consumption can be exploited to design reliable drowsy cache memories adopting dynamic voltage scaling (DVS) to reduce static power. First, we develop an analytical model allowing designers to evaluate the long-term threshold voltage degradation induced by bias temperature instability (BTI) in a drowsy cache memory. Through HSPICE simulations, we demonstrate that, as drowsy memories age, static power reduction techniques based on DVS become more effective because of reduction in sub-threshold current due to BTI aging. We develop a simulation framework to evaluate trade-offs between static power and reliability, and a methodology to properly select the “drowsy” data retention voltage. We then propose different architectures of a drowsy cache memory allowing designers to meet different power and reliability constraints. The performed HSPICE simulations show a soft error rate and static noise margin improvement up to 20.8% and 22.7%, respectively, compared to standard aging unaware drowsy technique. This is achieved with a limited static power increase during the very early lifetime, and with static energy saving of up to 37% in 10 years of operation, at no or very limited hardware overhead.

Index Terms—drowsy memory, dynamic voltage scaling, bias temperature instability, static power, soft error rate, noise margin

I. INTRODUCTION

As device sizes shrink, static power in VLSI circuits is increasing dramatically, to the point where it can be nearly as large as dynamic power [1]. Since cache memories in recent multi-core processors occupy a substantial portion of chip area, they are responsible for a large portion of static power consumption as caches may remain un-accessed for long periods [2], [3]. Power gating [4] and dynamic voltage scaling (DVS) [1], [5], [6] are two static power reduction techniques that can be successfully used for both memories and logic circuits [1], [7]. Among DVS solutions, the *drowsy* technique is proposed for on-chip caches [8], and is the focus of this paper. The low voltage of drowsy mode, denoted as *drowsy voltage* V_{dd}^D , degrades the reliability of the memory compared to active mode, and also memory cache lines could stay in drowsy mode for a large portion of cache’s lifetime [2]. Indeed, when the supply voltage is reduced, soft error

rate (SER) increases substantially due to critical charge Q_{crit} reduction [9], [10] where Q_{crit} is the minimum amount of charge collected by a node that is able to flip the state of an affected memory cell. Soft errors during reads can also be a concern, since the Q_{crit} can diminish due to the interference of the precharged bit lines [11]. In addition, memory robustness to noise decreases due to static noise margin (SNM) reduction in drowsy modes [12]. As a result, cache memory reliability can be severely impacted by its lower robustness against soft errors and noise during drowsy mode.

Low-power memory reliability is further undermined by device aging. Bias temperature instability (BTI), hot carrier injection and time dependent dielectric breakdown are the main aging mechanisms experienced by aggressively scaled devices [13]. Among these phenomena, negative BTI observed in pMOS transistors is the dominant one in the latest process technology [13], and is universally recognized as one of the primary parametric failure mechanisms for modern ICs [13]–[16]. Furthermore, with the use of high-k dielectric stacks, positive BTI exhibited by nMOS transistors is also significant and can no longer be neglected [17].

When transistors are biased in strong inversion [18], BTI causes an increase in their threshold voltage (V_{th}) over time. It has been shown that, both during stand-by and reads, the Q_{crit} and SNM degrade considerably due to BTI aging [11]. In [12], the negative effect of BTI on memory reliability has been considered for the selection of the minimum voltage that guarantees high reliable data retention in low-power memories. However, this technique ignores the beneficial effect of BTI aging on the sub-threshold current reduction, as shown in [19]–[21]. Moreover, in [22], a cross-layer SER analysis considering the effect of PVT variations have been performed.

In this paper, we show that BTI, besides causing detrimental effects on reliability, can also cause considerable static power saving in drowsy cache memories. Preliminary results on this beneficial effect have been presented in [23]. Although several other papers in the literature deal with the analysis of NBTI and power consumption in memories (for example, [24], [25]), to the best of authors’ knowledge this is the first time that the beneficial effect of aging on static power consumption is exploited to improve the reliability of a low power memory. We show that static power consumption in drowsy mode reduces considerably over time compared to the expected value determined at design time without considering the BTI effect. This extra power reduction can be beneficial for low power applications that do not have strict reliability requirements. Nevertheless, for applications in which power consumption

D. Rossi is with the Department of Engineering (DSP and VLSI Design group), University of Westminster, London, UK. E-mail: D.Rossi@westminster.ac.uk

V. Tenentes, B. M. Al-Hashimi and A. Brown are with the Department of Electronics and Computer Science (ECS), University of Southampton, Southampton, UK. E-mail: {V.Tenentes, bmah, adb}@ecs.soton.ac.uk

S. M. Reddy is with the Department of Electrical and Computer Engineering, University of Iowa, Iowa, USA. Email: sudhakar-reddy@uiowa.edu

Manuscript accepted May 31, 2017.

and reliability are both important constraints, the reliability deterioration in low power cache memories is an issue that needs to be addressed. Therefore, we propose to trade-off some of this extra power saving to improve memory reliability by increasing the drowsy voltage V_{dd}^D .

In order to assess the BTI effect on both reliability and static power consumption (P_{st}) of cache memories during drowsy mode, we first propose a DVS aware analytical model for aging accounting for the operating time spent in active mode or drowsy mode. Indeed, since the stress conditions in the two operating modes are different, the induced stress and consequently aging will also be different. The developed aging model is then embedded in to our HSPICE simulations. Considering a drowsy cache memory cell implemented with a 32nm high-k CMOS technology [26], we show that static power reduces by more than 35% during the first month, by more than 48% during the first year, and up to 61% in 10 years of operation.

We then develop a simulation framework for the design of reliable drowsy cache memories. The proposed framework allows us to evaluate possible trade-offs between power consumption and reliability, considering different drowsy voltages and access ratios, defined as the fraction of the operating time during which a cache line is being accessed. We propose a novel metric, referred to as *power efficiency for reliability*, which is defined as the ratio between a reliability figure (either Q_{crit} or SNM) and the static power for different operating conditions. This metric considers reliability and static power consumption jointly, thus enabling the identification of a drowsy voltage value that improves memory reliability, yet limiting static power consumption. Therefore, designers can evaluate the power efficiency in providing robustness against either soft errors (Q_{crit}/P_{st}) or noise sources (SNM/P_{st}). By using the proposed simulation framework, we derive three possible drowsy voltage selection policies, each characterized by a different static power and reliability trade-off. HSPICE simulations show a SER improvement up to 20.8% and SNM increase up to 22.7% during drowsy mode compared with a standard drowsy cache technique. A limited increase in static power consumption over the value expected by a standard, BTI-unaware drowsy technique, is exhibited during only the very early lifetime, while a static energy saving of up to 37% for 10 years of operation is achieved. We show that the reliability improvements are attained at no or very limited area overhead, and are valid also in case of process and temperature variations.

The rest of the paper is organized as follows. Section II introduces the basics of drowsy cache memories. In Section III, we develop an analytical model for the evaluation of the long-term threshold voltage degradation induced by BTI in a drowsy cache memory. In Section IV, through HSPICE simulations, we assess the beneficial effects of BTI aging on static power consumption of a drowsy memory cell, as well as BTI detrimental effects on SER and SNM. In Section V, we describe the proposed simulation framework allowing us to evaluate possible trade-offs between power consumption and reliability. Then, in Sec. VI, we describe and assess three derived drowsy voltage selection policies and

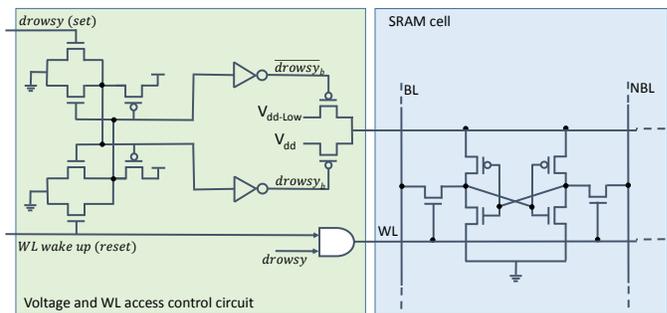


Fig. 1. Drowsy cache line architecture [8].

analyze the impact of process and temperature variations on a drowsy memory cell. Finally, concluding remarks are given in Section VII.

II. BACKGROUND

Cache memories are responsible for large part of static power in current ICs, due to their size. Approaches to reduce static power using power-gating techniques in [4], [5], and [27] selectively turn off cache lines storing data unlikely to be reused [5]. Power-gating is very effective in saving static power, but it does not support data retention. Therefore, data are retrieved from upper level memories in the memory hierarchy, causing performance penalties and potentially undermining energy savings of power gating.

The alternative approach of drowsy cache enables a lower reduction of static power compared to power-gating, but allows memories to retain their data. In this case, rather than turning off inactive lines, drowsy cache puts them into a low-voltage mode (drowsy mode). This way, the drowsy cache technique can allow up to 75% of energy reduction with no more than 1% of performance overhead [8], [28]. Drowsy cache line architecture is shown in Figure 1. It applies to the lines not being accessed a supply voltage V_{dd}^D lower than the that during active mode V_{dd}^A , but high enough to preserve stored data [8]. When a drowsy line is being referenced, that line is turned on again to a full power supply V_{dd}^A . More details on drowsy cache architecture can be found in [8].

In standard drowsy caches, the low V_{dd}^D value employed during the drowsy mode is determined without considering BTI degradation. It is approximately equal to 1.5 times the threshold voltage value of the memory cell transistors [8], a value that guarantees a considerable static power reduction, yet providing the design with adequate margins against noise and process variations [8]. Therefore, designers identify an expected static power consumption in drowsy mode as a target, which is kept constant during the entire memory lifetime. We refer to this value as *static power design constraint at t_0* , and denote it by $SPDC_0$.

III. DVS AWARE AGING MODEL FOR DROWSY CACHE MEMORIES

Bias temperature instability causes a threshold voltage increase in MOS transistors, denoted by ΔV_{th} , when they are ON (stress phase) [18]. BTI-induced degradation is partially

recovered when MOS transistors are in their OFF state (recovery phase). Negative BTI (NBTI) is observed in pMOS transistors, and it usually dominates the positive BTI (PBTI) observed in nMOS transistors [18].

In [17], [29], an analytical model based on the reaction-diffusion model in [18] has been proposed that allows designers to estimate long term BTI degradation. It is:

$$\Delta V_{th} = \chi K \sqrt{C_{ox}(V_{gs} - V_{th})} e^{-\frac{E_a}{kT}} (\alpha t)^{1/6} \quad (1)$$

The parameter C_{ox} is the oxide capacitance, t is the operating time, α is the fraction of the operating time during which a MOS transistor is under a stress condition (*stress ratio*), k is the Boltzmann constant, T the device temperature and E_a is the activation energy ($E_a \simeq 0.08\text{eV}$ [17]). The coefficient χ allows us to distinguish between PBTI ($\chi = 0.5$) and NBTI ($\chi = 1$). The parameter K lumps technology specific and environmental parameters. We have estimated it to be $K \simeq 2.7V^{1/2}F^{-1/2}s^{-1/6}$ by fitting the model with the experimental results reported in [30].

As can be observed from (1), BTI degradation depends on the voltage applied to the gate of a transistor and, consequently, to the supply voltage. Therefore, if a drowsy memory is considered, the expression in (1) can be profitably utilized when the supply voltage to a memory cell is constant, either equal to the full V_{dd}^A applied during active mode, or to the lower voltage V_{dd}^D supplied during drowsy mode. The first case $V_{dd} = V_{dd}^A, \forall t$ corresponds to a condition where a cache line is always being accessed, which is not a realistic assumption. The second case $V_{dd} = V_{dd}^D, \forall t$, instead, refers to a condition where a cache line is never accessed.

When a cache line switches from active mode to drowsy mode, and vice-versa, the model in (1) is not suitable to estimate BTI degradation. Therefore, to properly estimate the BTI degradation of a drowsy memory, we modified the model in (1) to account for the different degradations induced during active mode and drowsy mode. In this regard, it should be noted that transistors composing the memory cells and those utilized as power switches require the development of different models. Let us first define *access ratio* as the fraction of the total memory operating time and the time during which the considered cache line is operating in active mode, and denote it by γ . In turn, the proportion of the operating time during which the memory is operating in drowsy mode is $(1 - \gamma)$. As far as transistors composing memory cells are concerned, they are exposed to a degradation depending on V_{dd}^A during the fraction γ of their operating time. Conversely, their degradation is determined by V_{dd}^D during a fraction $(1 - \gamma)$ of their operating time. In addition, their degradation depends also on the data stored in the memory, which is accounted for by the stress ratio α . As a result, the new aging model for the low threshold voltage (V_{th}^L) memory cell transistors is:

$$\Delta V_{th}^L = \chi K \left\{ \gamma \sqrt{C_{ox}(V_{dd}^A - V_{th}^L)} + (1 - \gamma) \sqrt{C_{ox}(V_{dd}^D - V_{th}^L)} \right\} e^{-\frac{E_a}{kT}} (\alpha t)^{1/6}. \quad (2)$$

As for power switch transistors with high threshold voltage V_{th}^H that connect the V_{dd}^D to the memory cells, they are exposed to a stress time with a ratio $\alpha = (1 - \gamma)$. Therefore, the aging model for these transistors is:

$$\Delta V_{th}^H = K \sqrt{C_{ox}(V_{dd}^D - V_{th}^H)} e^{-\frac{E_a}{kT}} [(1 - \gamma)t]^{1/6}. \quad (3)$$

The proposed model does not consider the exact sequence between active and drowsy modes, as in [31], [32]. Instead, the history of the memory operation over a long period of time is embedded in the parameter γ , which represents an average access ratio. Moreover, the proposed model can be used at design time when the sequence of operating conditions is not known.

In Figure 2(a), we depict the trend over time of the threshold voltage degradation of the memory cell transistors, as given by (2), and of the power switch connected to the drowsy V_{dd}^D , as given by (3), for different V_{dd}^D values. The access ratio γ and stress ratio α have been set equal to 0.5. As expected, the degradation increases with V_{dd}^D , and this increase is more evident for the power switch than the memory cell transistors.

Figure 2(b) shows the threshold voltage degradation over time for different values of γ , for the standard drowsy voltage $V_{dd}^D = 0.65V$, which is approximately equal to $1.5 \times V_{th}^L$ [8], and $\alpha = 0.5$. Note that cell transistors (V_{th}^L) experience a larger degradation compared to power switch (V_{th}^H) connected to the drowsy V_{dd} , as well as a higher sensitivity to γ . In particular, the degradation of memory cell transistors increases with γ . Indeed, larger γ values represent longer time periods spent in active mode during which transistors are subjected to a larger stress. On the other hand, the degradation of the power switch connected to V_{dd}^D decreases with γ , since the stress ratio for this transistor is given by $(1 - \gamma)$.

In Figure 2(c), the ΔV_{th} trend over time for different stress ratio α is depicted, for $V_{dd}^D = 0.65V$ and $\gamma = 0.5$. It refers to memory cell transistors only, as the degradation of the power switches does not depend on α . We have assumed that α is the stress ratio of the pMOS transistor connected to the bit line BL and nMOS transistor connected to the complemented bit line NBL (Figure 1). The stress ratio for the other two memory cell transistors is $1 - \alpha$. The curve for $\alpha = 1$ and $\gamma = 0$ corresponds to a memory cell which is never accessed (always in drowsy mode), and so the stored bit never changes. This configuration exhibits the lowest degradation, since the electric stress is always at its minimum.

Finally, Figure 3 shows the ΔV_{th} for different aging temperatures. As expected, the degradation increases with temperature, and the difference between the three curves for the low- V_{th} transistors is considerably larger than for the high- V_{th} transistors.

IV. ANALYSIS OF BTI IMPACT ON LEAKAGE AND RELIABILITY OF A DROWSY MEMORY

A. Simulation setup

In order to assess the impact of BTI on a drowsy memory, we considered the memory cell scheme shown in Figure 4. It has been implemented with a 32nm high-k CMOS technology [26], with the supply voltage in active mode $V_{dd}^A = 1V$.

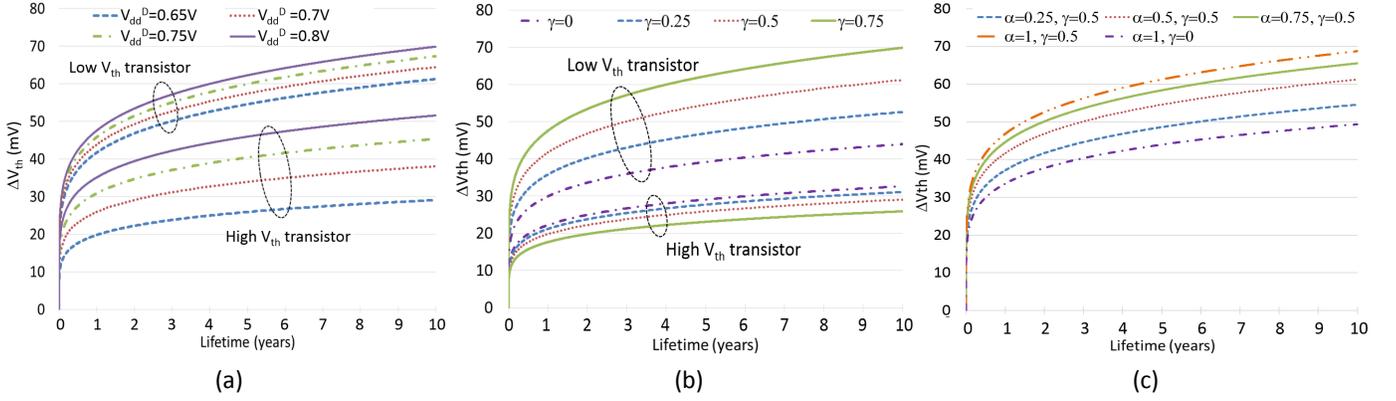


Fig. 2. Threshold voltage degradation profile over time for both low V_{th} and high V_{th} transistors ($T = 75^\circ C$), as a function of: (a) V_{dd}^D ($\gamma = 0.5$, $\alpha = 0.5$); (b) access ratio γ ($V_{dd}^D = 0.65V$, $\alpha = 0.5$); (c) stress ratio α ($V_{dd}^D = 0.65V$).

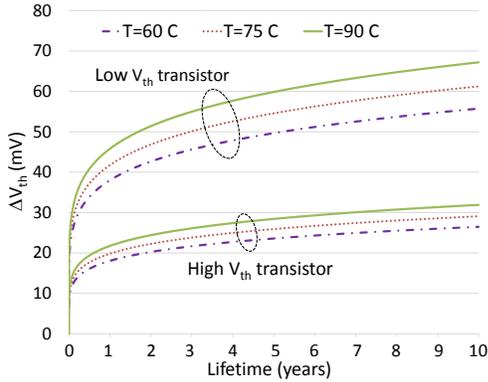


Fig. 3. Threshold voltage degradation profile over time for both low V_{th} and high V_{th} transistors, as a function of temperature ($V_{dd}^D = 0.65V$, $\alpha = 0.5$, $\gamma = 0.5$).

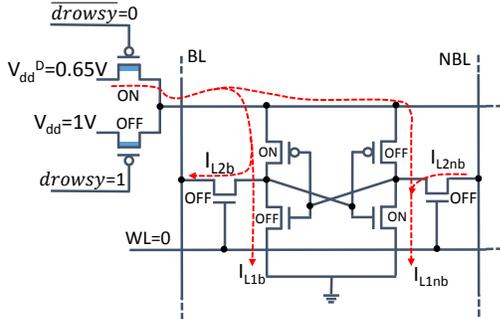


Fig. 4. Drowsy memory cell structure [8], and leakage current paths (dashed arrows) when the stored data is a logic 1.

Particularly, the $high-V_{th}$ (V_{th}^H) low power model has been adopted to implement the pMOS power switches, as suggested in [8], while all other transistors have been designed using the $low-V_{th}$ (V_{th}^L) high performance model. The value of the drowsy voltage is set to $V_{dd}^D = 0.65V$. In Figure 4 the leakage current paths are also highlighted (dashed arrows).

Because of the low voltage during drowsy mode, static power of inactive cells reduces considerably. On the other hand, memory reliability degrades noticeably when the power supply voltage is reduced. Indeed, the critical charge Q_{crit} and the static noise margin SNM decrease substantially in drowsy mode compared to active mode. As a consequence, the inactive memory cells are more prone to soft errors.

TABLE I
MEMORY CELL STATIC POWER CONSUMPTION (P_{st}), CRITICAL CHARGE (Q_{crit}) AND STATIC NOISE MARGIN (SNM) VALUES AT t_0 FOR ACTIVE (A) AND DROWSY (D) MODES, AND RELATIVE VARIATION ($\Delta\% = 100 \cdot (X^D - X^A)/X^A$, $X = P_{st}, Q_{crit}, SNM$).

	$V_{dd}^A = 1V$	$V_{dd}^D = 0.65V$	$\Delta\%$
P_{st} (pW)	4207	227.1	-94.6
Q_{crit} (fC)	10.44	1.305	-87.5
SNM (mV)	385	210	-45.5

These considerations are confirmed by the simulation results in Table I, where static power, Q_{crit} and SNM are reported for both active ($V_{dd}^A = 1V$) and drowsy ($V_{dd}^D = 0.65V$) modes. We can see a 94.6% static power reduction achieved by switching the memory in drowsy mode compared to active mode. On the other hand, Q_{crit} reduces (SER increases) and so does SNM, by 87.5% and 45.5%, respectively, as shown in Table I [9]. As a result, drowsy memories are much more susceptible to reliability threats when operated in drowsy mode than in active mode.

In addition, BTI aging has a considerable impact, either beneficial or detrimental, on static power consumption, SER and SNM. In Figure 5, we show the approach followed to embed BTI aging effects in our simulation flow. Given power supplies (V_{dd}^A and V_{dd}^D) and operating conditions (aging temperature T_A , access ratio γ and stress ratio α), the V_{th} degradation is estimated for both low- V_{th} and high- V_{th} transistors. The corresponding ΔV_{th} value obtained for each considered operating time interval is then utilized to customize the HSPICE device models and simulate the memory cell with the proper BTI degradation.

B. BTI Beneficial Effect on Static Power during Drowsy Mode

Static power P_{st} due to leakage current I_{leak} is becoming a major contributor to power consumption, especially in large cache memories. Leakage current I_{leak} has two main components [1]: sub-threshold current and gate current. Sub-threshold current dominates, since gate current can be well controlled by the use of high-k dielectrics. Therefore, I_{leak} can be approximated as follows [1]:

$$I_{leak} \simeq \mu C_{ox} \left(\frac{kT}{q} \right)^2 \frac{W}{L} e^{\frac{-q(V_{gs} - V_{th})}{m k T}} (1 - e^{\frac{-q V_{ds}}{k T}}). \quad (4)$$

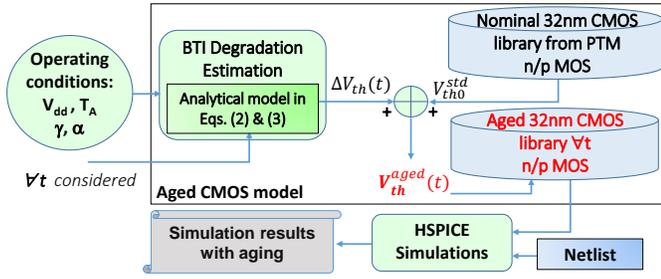


Fig. 5. HSPICE simulation flow embedding BTI degradation.

If the overdrive voltage ($V_{gs} - V_{th}$) reduces, I_{leak} decreases exponentially, and so does P_{st} . Therefore, when the drowsy mode is applied, a considerable saving in static power is experienced. As for the dependency on drain to source voltage V_{ds} , it is worth noting that $V_{ds} = V_{dd}^D$ in off transistors during drowsy mode. Since this value is considerably larger than the thermal voltage $kT/q \simeq 26mV$ at ambient temperature, the second exponential term in (4) can be neglected.

As previously discussed, a constraint $SPDC_0 = P_{st}(t = 0, V_{dd}^D = 0.65V)$ for static power consumption is identified at design stage, which remains constant over time. This value represents the static power per memory cell expected to be consumed by a standard drowsy technique not accounting for the beneficial effect of BTI. Instead, we expect that static power considerably decreases as memory ages due to BTI [20], [21]. This is confirmed by the simulation results shown in Figure 6(a), which shows also the relative reduction over time. After only 1 month of operations, P_{st} reduces by 31%; after 1 year, static power reduction exceeds 43% and, after 10 years, it reaches 56%. We observe that P_{st} decreases after 1 month of operation by more than 50% of the variation exhibited after 10 years. Therefore, as a result of BTI aging, P_{st} during stand-by decreases over time to a value considerably lower than $SPDC_0$ estimated by a standard BTI-unaware drowsy technique.

C. BTI-Induced Degradation of Soft Error Rate

In drowsy mode, Q_{crit} reduces by more than 87% (Table I) compared to active mode (from 10.4fC to 1.31fC at t_0). Moreover, Q_{crit} is further degraded by BTI. To evaluate Q_{crit} profile over time, we estimate ΔV_{th} by (2) for cell transistors and (3) for power switches. Similar to [16], [30], the estimated ΔV_{th} values for each considered lifetime have been utilized to customize the HSPICE device model, so that each transistor is simulated with the proper BTI degradation. In Figure 6(b), we depict the obtained simulation results for the Q_{crit} . The relative Q_{crit} reduction with respect to t_0 value is also shown. Note that the Q_{crit} decreases by more than 26% over 10 years, reaching 20% reduction after only 1 year. The Q_{crit} reduction exhibited by a drowsy memory cell is greater than in a standard SRAM cell operating at V_{dd}^A , for which we found a 11.4% Q_{crit} reduction over 10 years. This difference can be attributed to the presence of the power switch in a drowsy memory cell, whose BTI degradation exacerbates the Q_{crit} reduction.

We now evaluate the SER of a drowsy memory cell as a function of lifetime t , for $V_{dd}^D = 0.65V$ and $\gamma = \alpha = 0.5$. The SER of a memory cell is determined by the sum of several contributions, each referred to a node of the cell. In turn, the susceptibility of each node can be expressed as a function of the window-of-vulnerability (WOV), which is the time interval during which an energetic particle hitting the node can give rise to an SE, and the Q_{crit} of the considered node. The operation of the memory also impacts the SER calculation. Particularly, a memory cell is sensitive to SE when it is being read (active, denoted by A) and during stand-by (drowsy, denoted by D). As a result, the total SER is:

$$SER \simeq \sum_{mode} \sum_{i=1}^n \frac{WOV_i^{mode}}{t} \Phi_P A_{Si} e^{-\beta Q_{crit_i}^{mode}(t)}, \quad (5)$$

where $mode$ can be either *drowsy* (D) or *active* (A); $n = 2$ are the nodes of the memory cell that may produce an SE if hit by an energetic particle; A_{Si} is the susceptible area of the node i ; Φ_P is the flux of energetic particles ($\Phi_P \simeq 56.5s^{-1}m^{-2}$ at sea level [33]), and β is a parameter that depends on the considered technology and operating environment. For the considered 32nm CMOS technology, we have derived a value $\beta = 90 \times 10^{12} 1/C$ from [34].

In this paper, we assess the impact of BTI during drowsy mode only, and not during active mode. We have found that, in drowsy mode, Q_{crit}^D equals 1.31 fC, which is considerably lower than $Q_{crit}^A = 10.44$ fC during read operations. Moreover, WOV^D is usually much larger than WOV^A . As a result, $SER^D \gg SER^A$. It is worth noting that, by reducing the power supply from 1V to 0.65V, we have found that the SER increases by 63.4% at t_0 , thus confirming the higher reliability risks when the memory operates in drowsy mode. In Figure 6(c), we show the trend over time of the SER for a drowsy memory cell considering BTI aging, when the memory is in drowsy mode, together with its relative variation over the value exhibited at t_0 . We can see that, although the critical charge decrease over 10 years of operation exceeds 26%, SER increases by 3.6% only.

D. BTI-Induced Degradation of Static Noise Margin

Similar to Q_{crit} , static noise margin (SNM) degrades over time due to BTI aging. In a drowsy memory, this degradation adds on top of that induced by the use of a lower supply voltage. SNM profile during drowsy mode has been derived graphically from the butterfly curve, which is obtained by plotting the static characteristics of the two inverters in a memory cell. Then, the SNM is given by the size of the largest square that can be inscribed within the smallest butterfly lobe. The HSPICE simulation results are depicted in Figure 7. The SNM reduces by 9.5% over 10 years of operation, thus exhibiting a degradation over time which is considerably lower than Q_{crit} degradation. Indeed, the bistable structure of a memory cell helps reduce the SNM degradation.

V. PROPOSED SIMULATION FRAMEWORK FOR TRADING OFF POWER AND RELIABILITY IN DROWSY MEMORIES

The beneficial impact of BTI aging on static power has been ignored so far by DVS techniques. We propose to trade-

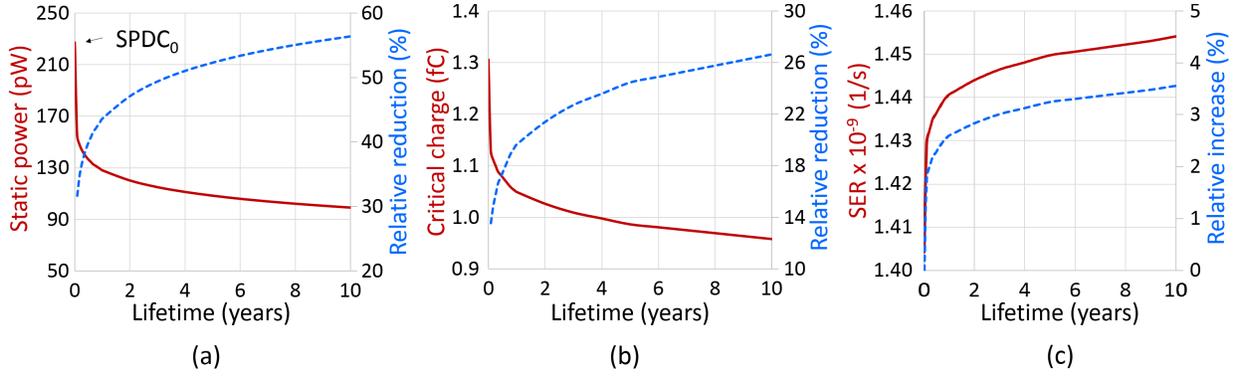


Fig. 6. Trend over time (solid line) and relative variation with respect to t_0 values (dashed line), with $V_{dd}(D_{st})=0.65V$ and $\gamma=\alpha=0.5$: (a) static power; (b) critical charge; (c) SER - Variation = $[X(t_0) - X(t)]/P_X(t_0)$, $X = P_{st}, Q_{crit}, SER$.

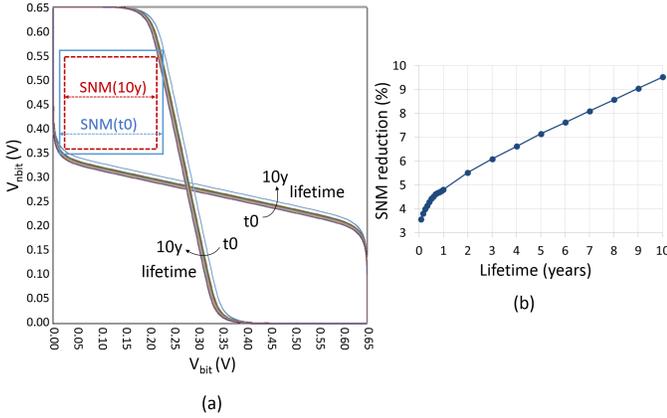


Fig. 7. SNM trend over time for $V_{dd}(D_{st})=0.65$ and access ratio $\gamma = \alpha = 0.5$: (a) butterfly plot; (b) SNM reduction over time with respect to SNM at t_0 : $[SNM(t_0) - SNM(t)]/SNM(t_0)$.

off some of this static power over-reduction to counteract the detrimental effect of BTI aging on SER and SNM, thus improving drowsy memory reliability. This can be achieved by selecting a higher drowsy voltage to be applied to cache lines not being accessed. Of course, different drowsy voltage values enable achieving different trade-offs between static power consumption and reliability. Therefore, we have developed a simulation framework allowing designers to evaluate static power and reliability trade-offs. In particular, we have analyzed the trend over time of P_{st} , Q_{crit} (SER) and SNM as a function of V_{dd}^D varying from 0.65V to 0.8V, and access ratio γ varying from 0 to 0.75. As for stress ratio α , we set it to 0.5. The obtained results are shown in Figure 8, whereas the flow of the simulation framework is depicted in Figure 9.

For the characterization of the memory cells accounting for BTI aging (“Aged library characterization” in Figure 9), the considered operating conditions feed the aging model, and the corresponding threshold voltage degradation is computed. In this regard, actual thermal profiles of the memory can be derived by means of the HotSpot tool [35], provided that information on the workload are available. If the actual workload is not known at design time, the statistically probable workload can be selected, as proposed in [36]. Afterwards, the considered metrics are evaluated by HSPICE simulations allowing designers to explore different trade-offs between

power consumption and reliability. Then, according to the design constraints, a drowsy V_{dd}^D value suitably higher than the value adopted in standard drowsy memory is selected.

In Figure 8(a), we show the P_{st} profile for a drowsy cache memory cell as a function of V_{dd}^D and lifetime, for $\gamma = 0.5$. As can be seen, the static power drops rapidly over time for all V_{dd}^D values. As expected, P_{st} values at t_0 are higher than $SPDC_0$ for $V_{dd}^D > 0.65V$. Nevertheless, while for $V_{dd}^D > 0.78V$ P_{st} is always greater than $SPDC_0$, this is not the case for lower values of V_{dd}^D . As an example, P_{st} drops below $SPDC_0$ after less than a month of operation for $V_{dd}^D = 0.7V$, and after approximately 3.5 years for $V_{dd}^D = 0.75V$.

Figure 8(b) depicts the contour curves for the considered values of access ratio γ . The four curves in each group have approximately the same value of P_{st} . We can see that static power variation depends noticeably on γ and, for a given V_{dd}^D value, the characteristic P_{st} value of each group is achieved in a shorter time for higher value of γ . Moreover, the sensitivity on γ increases with both V_{dd}^D and lifetime.

Figure 8(c) shows the Q_{crit} profile over time for the considered V_{dd}^D and lifetime. As expected, the Q_{crit} increases noticeably (and almost linearly) as V_{dd}^D raises. As an example, for $V_{dd}^D = 0.8V$, Q_{crit} is approximately 3.5X the value at $V_{dd}^D = 0.65$ (which is the standard value, as mentioned in Section III), and this ratio is maintained for all the considered lifetime range. In turn, Q_{crit} decreases over time due to BTI aging. The relative variation over time against the value exhibited at t_0 , computed as $[Q_{crit}(V_{ddi}^D, t) - Q_{crit}(V_{ddi}^D, t_0)]/Q_{crit}(V_{ddi}^D, t_0)$, is very similar for different $V_{ddi}^D \in [0.65V-0.8V]$. As an example, it ranges from 21.4% for $V_{ddi}^D = 0.65V$ to 21.7% for $V_{ddi}^D = 0.8V$ at 2 years, and from 24.5% for $V_{ddi}^D = 0.65V$ to 24.9% for $V_{ddi}^D = 0.8V$ at 5 years.

Different from P_{st} , Q_{crit} is almost constant with γ , and the four curves in each group in Figure 8(d) overlap. This is attributed to the higher sensitivity of P_{st} to V_{th} degradation compared to Q_{crit} . Indeed, the latter quantity is proportional to the driving strength (active current) of memory cell transistors, which depends almost linearly on the overdrive voltage ($V_{gs} - V_{th}$), whereas the sub-threshold leakage current, which is the dominant contributor to static power, varies exponentially with ($V_{gs} - V_{th}$), as given by (4).

SNM trend over time as a function of drowsy voltage V_{dd}^D is shown in Figure 8(e). SNM increases with V_{dd}^D

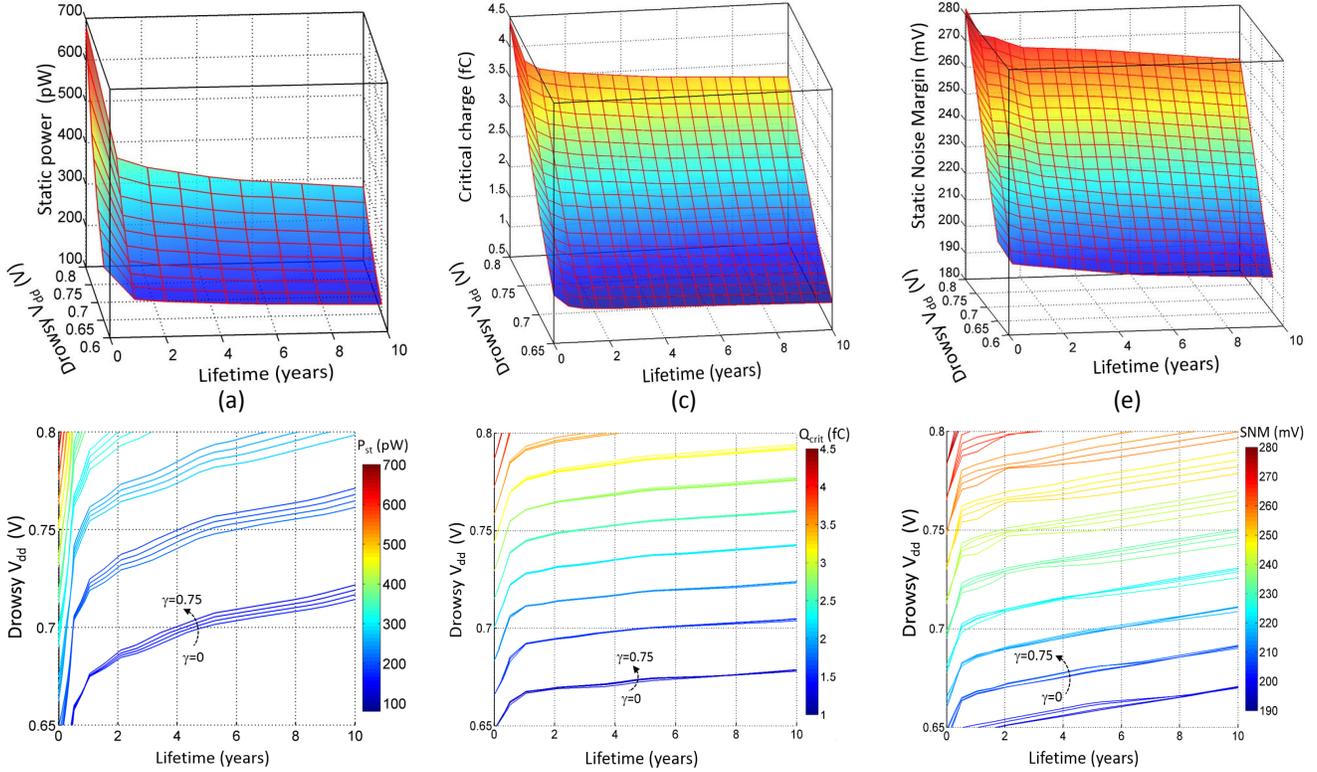


Fig. 8. Variation as a function of V_{dd}^D and lifetime ($\gamma = 0.5$) and contour curves for different values of access ratio γ (0, 0.25, 0.5 and 0.75): (a) static power variation; (b) static power contour curves; critical charge variation; (d) critical charge contour curves; (e) SNM variation; (f) SNM contour curves.

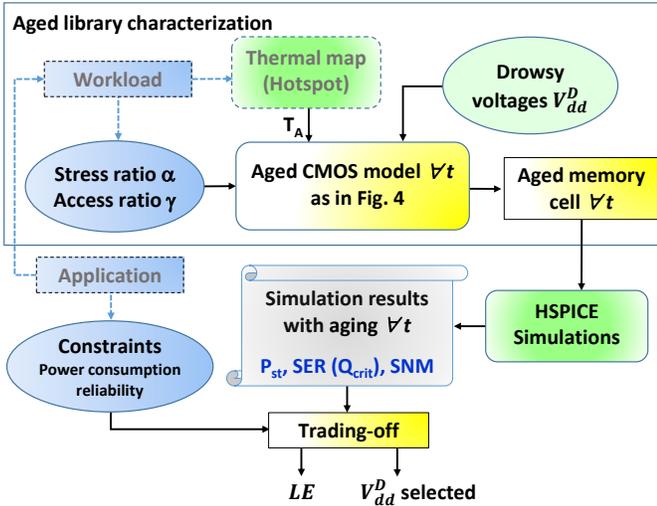


Fig. 9. Flow of the proposed simulation framework.

and decreases over time due to BTI aging. It is worth noting that the variation with V_{dd}^D is much less remarkable than for Q_{crit} : SNM improvement over the standard approach ($V_{dd}^D = 0.65V$) at a given time t_j , computed as $[SNM(V_{dd}^D, t_j) - SNM(V_{dd}^D, t_j)] / SNM(V_{dd}^D, t_j)$ ranges from 11.5% for $V_{dd}^D = 0.7V$ to 34% for $V_{dd}^D = 0.8V$. This SNM improvement is very similar for all t_j considered. In Figure 8(f), we show the contour curves, which evidence a greater dependence on γ for higher voltages.

Next, we evaluate the SER variation over the standard ap-

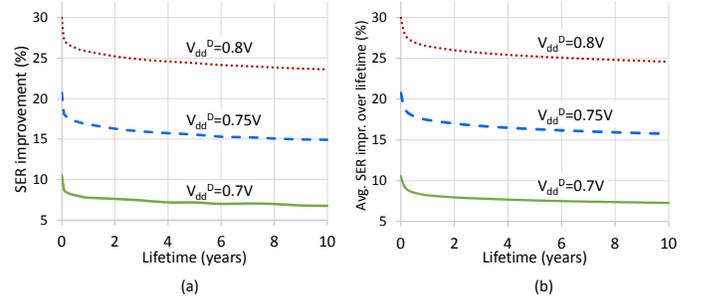


Fig. 10. SER improvement over the standard drowsy technique as a function of lifetime, for different V_{dd}^D value ($\Delta SER(V_{dd}^D, t) = SER(0.65V, t) - SER(V_{dd}^D, t) / SER(0.65V, t)$, $V_{dd}^D = 0.7V, 0.75V, 0.8V$).

proach ($V_{dd}^D = 0.65V$) when we increase the drowsy voltage. The SER is estimated using (5), and Figure 10(a) shows the trend over time of the SER improvement for different values of V_{dd}^D . Access ratio γ and stress ratio α are set to 0.5. In this regard, it is worth reminding that no appreciable Q_{crit} dependency on γ and α was found. As can be seen, SER improvement varies considerably with V_{dd}^D , since Q_{crit} is considerably affected by the drowsy voltage. Moreover, SER improvement is considerably higher during the early lifetime, reduces within the first months of operation and essentially stabilizes in two years. In Figure 10(b), we also show the trend of the average SER improvement evaluated over the lifetime, computed as reported in (6). Average SER improvement ranges from 10.4% for $V_{dd}^D = 0.7V$ to 30.0% for $V_{dd}^D = 0.8V$ at t_0 , from 8.1% to 26.7% after 1 year, and from 7.7% to 25.3%

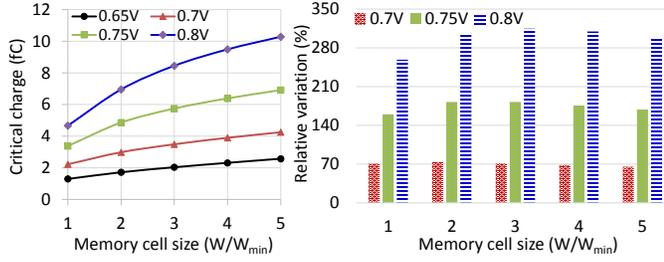


Fig. 11. (a) Trend over time of the critical charge as a function of cell size, for different drowsy voltage values; (b) relative variation over the standard drowsy voltage $V_{dd}^D = 0.65V$ ($\Delta = [Q_{crit}(V_{dd}^D) - Q_{crit}(0.65V)]/Q_{crit}(0.65V)$, $V_{dd}^D = 0.7V, 0.75V, 0.8V$).

after 5 years of operation.

$$\Delta SER_{avg}(V_{ddi}^D, t) = \frac{1}{t - t_0} \int_{t_0}^t \Delta SER(V_{ddi}^D, t) dt, \quad (6)$$

$$V_{ddi}^D = 0.7V, 0.75V, 0.8V; t \in [0, 10] \text{ years}$$

In our analysis so far, we have always considered minimum sized memory cells. As known, Q_{crit} is affected by cell parasitics, and particularly by the size of transistors. In this regard, as discussed in [37], Q_{crit} of a node is much more impacted by the conductance of the driving gate than the node capacitance. Therefore, we now evaluate the impact of memory cell size on Q_{crit} , and the obtained results are depicted in Figure 11. Q_{crit} increases noticeably with the memory cell size and with V_{dd}^D , as shown in Figure 11(a), whereas the relative variation over the standard $V_{dd}^D = 0.65V$ is depicted in Figure 11(b). It is worth noting that, for a fixed V_{dd}^D (0.7V, 0.75V or 0.8V), the relative Q_{crit} increase changes slightly with the lifetime.

So far, we have addressed the analysis of the impact of the considered drowsy modes on either static power or reliability features (Q_{crit} and SNM) separately. We now define a new metric allowing us to jointly evaluate static power consumption and either Q_{crit} or SNM. The new metric, which for Q_{crit} is defined as $PR_{eff}(Q_{crit}) = Q_{crit}/P_{st}$ (for SNM, it is $PR_{eff}(SNM) = SNM/P_{st}$), and referred to as *power reliability efficiency* for Q_{crit} (or SNM), represents the critical charge (SNM) offered by a solution per unit of static power consumed. It is therefore an evaluation of the power efficiency in providing resilience against soft errors (noise sources) during drowsy mode.

The variation of PR_{eff} as a function of lifetime and drowsy voltage, for different values of the access ratio γ is represented in Figure 12 and Figure 13, respectively. As we can see, the $PR_{eff}(Q_{crit})$ increases over time for all considered cases. Indeed, as discussed in Section IV, P_{st} decreases much faster with lifetime compared to Q_{crit} . Moreover, the depicted function exhibits a maximum for $V_{dd}^D = 0.75V$ for lifetime values up to 6.6 years. For longer lifetimes, the $PR_{eff}(Q_{crit})$ values for 0.75V and 0.8V almost coincide, with a slight prevalence of the latter. Moreover, we have found that the $PR_{eff}(Q_{crit})$ drops again monotonically for $V_{dd}^D > 0.83V$, and for $V_{dd}^D = V_{dd}^A$, $PR_{eff}(Q_{crit})$ drops well below the

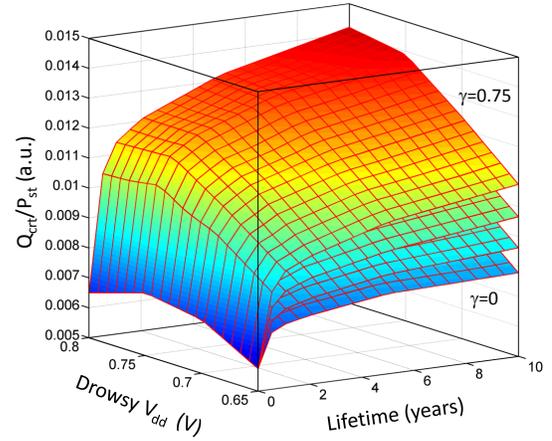


Fig. 12. Reliability power efficiency profile for Q_{crit} , as a function of time and drowsy voltage, with access ratio $\gamma = 0, 0.25, 0.5$ and 0.75 .

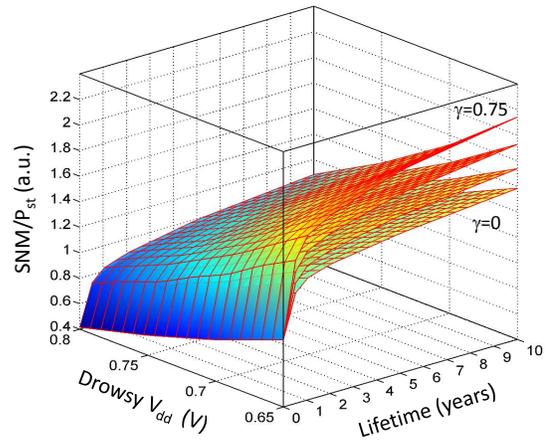


Fig. 13. Reliability power efficiency profile for SNM, as a function of time and drowsy voltage, with access ratio $\gamma = 0, 0.25, 0.5$ and 0.75 .

values obtained for $V_{dd}^D = 0.65V$ for all lifetimes. This can be explained by considering that P_{st} increases exponentially with V_{dd} , while Q_{crit} varies almost linearly. If for small value of the drowsy V_{dd} the Q_{crit}/P_{leak} metric is benefited by an increase of power supply, larger drowsy V_{dd} values turn-out to be a power inefficient approach for SER improvement.

The $PR_{eff}(SNM)$ profiles for different γ are depicted in Figure 13. Different from the analogous metric for Q_{crit} , it exhibits a monotonic behavior for all lifetime and voltage values. Particularly, it increases over time and decreases with the V_{dd}^D , showing that increasing the drowsy voltage is less an efficient approach to improve SNM compared with Q_{crit} improvement. On the other hand, we have shown that SNM is much less sensible to BTI and drowsy voltage variation compared to Q_{crit} (Figure 8). Therefore, in the next section, we consider only $PR_{eff}(Q_{crit})$ to jointly evaluate static power and reliability.

VI. PROPOSED DROWSY POLICIES AND ARCHITECTURE AND SIMULATION RESULTS

From the simulation results obtained with the proposed exploration framework, we derive and evaluate three drowsy V_{dd} selection policies and architectures, leading to three different power and reliability trade-offs, which are identified at

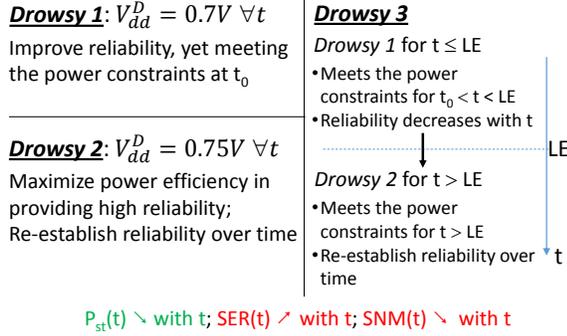


Fig. 14. Drowsy voltage selection policies rationale.

design time. As highlighted in Section III, BTI aging depends considerably on utilization and temperature, which therefore affect the trend over time of static power consumption, SER and SNM. Moreover, these metrics can also be affected by process variations, as we will show in Section VI-D. At design time, information on the workload can allow designers to build up a thermal map of the ICs with the temperature distribution [36]. Moreover, the impact of process variations can also be estimated at design time.

Figure 14 shows the rationale of the three proposed drowsy voltage V_{dd}^D selection policies that are discussed in the following subsections. In the first case, denoted by *Drowsy 1*, a static $V_{dd}^D = 0.7V$ is selected at t_0 . This is slightly higher than the standard drowsy voltage (0.65V), thus allowing us to improve reliability over the standard approach, yet limiting the static power consumption at t_0 , which represents the highest value over the memory lifetime. In the second case, denoted by *Drowsy 2*, a higher, static $V_{dd}^D = 0.75V$ is selected at t_0 . As highlighted in the previous section, this value allows us to increase considerably the reliability of the drowsy memory during stand-by, thus enabling to counteract the BTI detrimental effect on memory reliability. Moreover, it maximizes the power efficiency in increasing Q_{crit} . As a drawback, it increases the static power consumption compared with the policy *Drowsy 1*. Nevertheless, as shown in Section IV, static power substantially reduces over time. The third selection policy (*Drowsy 3*), exploits this reduction over time in the static power, and identifies a suitable time (elapsed time t_{el}), after which the drowsy voltages is increased from 0.7V to 0.75V, combining the advantages provided by *Drowsy 1* during the initial lifetime, and *Drowsy 2* for longer lifetime values.

The proposed V_{dd}^D selection policies have been validated through HSPICE simulations by evaluating the static energy saving with respect to the value expected for a standard, BTI-unaware drowsy technique, which is given by $E_{st}^{exp} = SPDC_0 \times lifetime$. Moreover, Q_{crit} , SNM and SER variation over the standard drowsy technique ($V_{dd}^D = 0.65V$) have been also considered as metrics for comparison, and evaluated as $[X(S_i, t) - X(S_{st}, t)]/X(S_{st}, t)$, with $X = (Q_{crit}, SNM, SER)$ and $i = 1, 2, 3$.

A. Drowsy 1: Improved Drowsy Memory

In the *Drowsy 1* solution, to increase memory reliability over the standard drowsy, yet meeting leakage power/energy

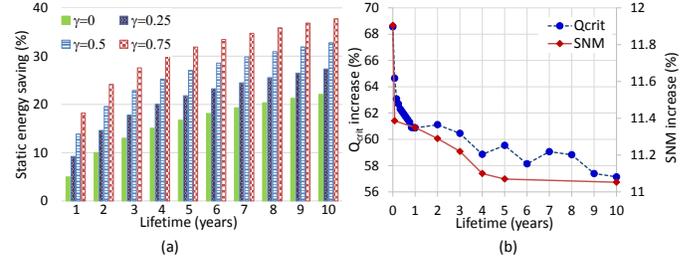
Fig. 15. Improved drowsy: (a) variation over time of static energy and (b) Q_{crit} and SNM increase over the standard drowsy technique.

TABLE II
 P_{st} INCREASE FOR 5 HIGH PERFORMANCE AND 2 LOW POWER MICROPROCESSORS WHEN THE DROWSY VOLTAGE INCREASES FROM 0.65V TO 0.7V.

μP	Cache size	ΔP_{st} (W)	$P_{\mu P}$ (W)	$\Delta\%$
i7-47XXS	8MB (L3)	2.95×10^{-3}	65	0.01229
i5-45XXS	6MB (L3)	2.21×10^{-3}	65	0.00921
i7-4790T	8MB (L3)	2.95×10^{-3}	45	0.01775
i5-4670T	6MB (L3)	2.21×10^{-3}	45	0.01331
i3-4130T	3MB (L3)	1.11×10^{-3}	35	0.00856
Cortex A9	32kB (L1)	1.15×10^{-5}	1.9	0.00164
Cortex A35	8kB (L1)	3.60×10^{-7}	1×10^{-2}	0.00975

constraints, power supply $V_{dd}^D = 0.7V$ is selected. Figure 15 depicts the obtained simulation results. As can be seen, in 10 years of operation the energy saving (Figure 15(a)) ranges from 22% for $\gamma = 0$ to 38% for $\gamma = 0.75$. As for the Q_{crit} increase (Figure 15(b)) over the standard drowsy memory ($V_{dd}^D = 0.65V$), it ranges from 68% at t_0 to 57% at 10 years, while SNM increase is in the interval 11%-12% for all lifetime values. The average SER improvement, computed as discussed in Section V, is 10.4% at t_0 , it decreases to 8.1% after a year, and reaches 7.7% after 10 years.

The *improved drowsy memory* does not introduce any hardware overhead over the standard drowsy technique. However, during the very early lifetime, the memory experiences higher static power consumption than the expected value $SPDC_0$ of the standard approach. Particularly, the average value per cell increases by 39%, from 227pW to 315pW. This leads to an excess static power consumption of 0.369mW for a 1MB cache memory. This extra static power consumption, which is completely recovered in approximately 15 days of operation, represents a very small fraction of the power consumption of a whole microprocessor. As an example, in Table II we report the power consumption increase when the drowsy voltage is raised from 0.65V to 0.7V for several microprocessor, and its variation over the total microprocessor power consumption $P_{\mu P}$. It is with noting that the power figures for the microprocessors come from different sources and refer to different technologies. As can be seen, the relative increase is negligible, ranging from 0.00061% for the low power Cortex A9 to 0.00656% for the high performance i7-4790T. This very small increase can be neglected when designing the power network and has a negligible impact on the thermal management system of the microprocessor.

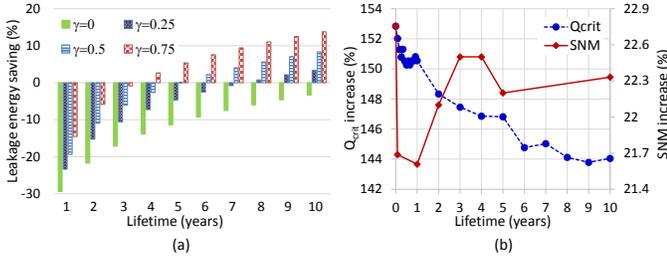


Fig. 16. Power efficient and reliable drowsy: variation over time of (a) static energy and (b) Q_{crit} and SNM, over the standard drowsy technique.

B. Drowsy 2: Power Efficient Reliable Drowsy Memory

This solution relies on the static selection of a drowsy power supply aiming to maximize the $PR_{eff}(Q_{crit})$ metric, thus the power efficiency in providing drowsy memory with soft error resilience. According to the results discussed in Section V, the drowsy voltage $V_{dd}^{D2} = 0.75V$ is selected. From the simulation results in Figure 16, we can see that the Q_{crit} (SNM) increase with respect to the standard drowsy technique is in the range 144%-153% (21.6%-22.7%) over the whole lifetime. The average SER improvement ranges from 20.8% at t_0 to 15.7% after 10 years of operation. This noticeable reliability improvement is achieved with no hardware overhead, but at the cost of an increase in static power consumption, whose amount varies with the access ratio γ . Particularly, the static power per cell is 315pW, which leads to an excess static power over the $P_{\mu P}$ ranging from 0.00164% for the low power Cortex A9 to 0.01775% for the high performance i7-4790T. This static power excess is recovered after 1.3 years of operation for $\gamma = 0.5$, whereas for longer lifetime the P_{ST} drops below $SPDC_0$.

C. Drowsy 3: Adaptive Drowsy Memory

To increase reliability, yet meeting static power/energy constraints, we proposed an *adaptive drowsy memory* architecture, as shown in Figure 17(a), where the memory switches from $V_{dd}^{D1} = 0.7V$ to $V_{dd}^{D2} = 0.75V$ during its lifetime. The *power control circuit* generates the three different supply voltages required, whereas the *supply voltage selection control* block is the same as in the original architecture (Figure 1). The additional control signal LE , which identifies the moment in time t_{el} when to switch from the two available drowsy supply voltages, is generated at system level so that $P_{st}(V_{dd}^D = 0.75V, t) \leq SPDC_0, \forall t \geq t_{el}$, thus always meeting the static power constraint $SPDC_0$. Through HSPICE simulations, we have found $t_{el} = 3.91$ years. The *supply voltage selection circuit* provides the cache line with the proper supply voltage. Its structure is expanded in Figure 17(b), where the logic equations of the control signals c_A , c_{D1} and c_{D2} for the three power switches are also reported. One additional power switch and a very small logic generating the control signals per cache line is required by the adaptive architecture.

Figure 18(a) shows the static energy saving over a standard drowsy technique. When the drowsy voltage switches from V_{dd}^{D1} to V_{dd}^{D2} , the energy saving over the expected value reduces, and then increases again up to 20% (for $\gamma = 0.75$) after 10 years of operation. Meanwhile, the Q_{crit} (SNM)

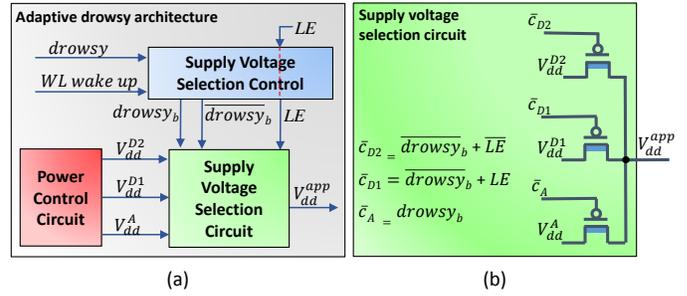


Fig. 17. Proposed adaptive drowsy architecture.

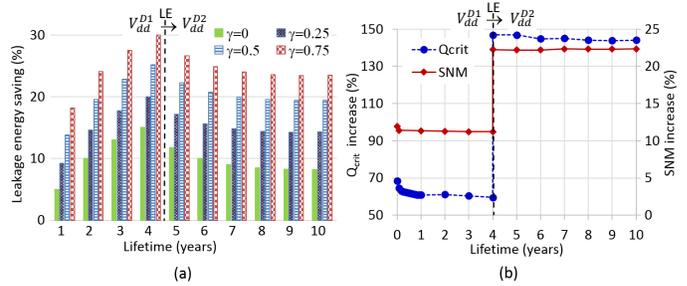


Fig. 18. Adaptive drowsy: variation over time of (a) static energy and (b) Q_{crit} and SNM, over the standard drowsy technique.

improvement over time (Figure 18(b)) increases from around 60% (11%) during the first 3 years, to slightly less than 150% (25%) for the rest of lifetime. The average SER improvement over the standard drowsy memory varies as for the *Drowsy 1* solution until t_{le} , upon which it reaches 7.9%. When the drowsy voltage switches from V_{dd}^{D1} to V_{dd}^{D2} , the average SER improvement jumps above 12%, and slightly reduces to 11.2% after 10 years of operation. Therefore, a higher soft error resilience over time is achieved at the cost of less energy saving over the standard drowsy technique. An excess power consumption over the expected value $SPDC_0$ is experienced only at the early lifetime, when the memory operates with $V_{dd}^D = 0.7V$, and the same considerations apply.

The described reliability improvement comes together with a small hardware cost, due to the on-chip generation of 2 different drowsy voltages, and one additional power switch and a very small *ad-hoc* control logic per cache line. Estimating the area of a cache memory as a summation of the gate area of all composing transistors expressed in terms of squares ($1 \text{ square} = W_{min}L_{min}$), we obtain the values reported in Table III for cache line sizes of 32B and 64B. For comparison, we show the area of the standard drowsy memory and the overhead of the proposed solution. The results for a standard cache (with no DVS technique adopted) are also reported as a reference. The area overhead of the proposed adaptive drowsy has been evaluated by estimating the extra area per cache line due to the *Supply Voltage Selection Control* (derived from the *Voltage and WL access control circuit* in Figure 1) and *Supply Voltage Selection Circuit* in Figure 17. The area of the voltage supply circuit has not been considered in this area overhead estimation, since it has been assumed already present in the system. Moreover, it should be noticed that it is shared by the whole memory array. As can be seen, the area overhead of the proposed adaptive drowsy architecture is very limited.

TABLE III
Area overhead of the proposed drowsy voltage selection policies for different cache line sizes ($1 \text{ square} = W_{\min} L_{\min}$).

Cache line size	Adap. Drowsy		Std Drowsy		Std memory	
	square	square	square	AO (%)	square	AO (%)
32B	4922	4812	2.29	4644	5.98	
64B	9530	9420	1.17	9252	3.01	

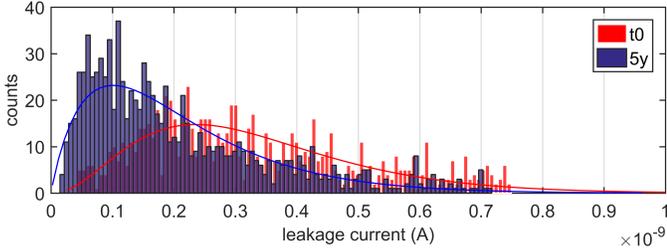


Fig. 19. Leakage current distribution with process variation at t_0 and after 5 years of operation.

D. Impact of Process and Temperature Variations on P_{st} , Q_{crit} and SNM

We evaluate the impact of process and temperature variations on P_{st} , Q_{crit} and SNM of a drowsy memory, considering $V_{dd}^D = 0.65V$. First, using Monte Carlo simulations, we assess the impact of process variations on the variability of the leakage current (thus static power) of a memory cell during drowsy mode. Threshold voltage, oxide thickness and mobility of each transistor follow a normal random distribution around the nominal values, with a standard deviation $3\sigma = 10\% \cdot Y_{nom}$, where Y_{nom} is the nominal value of the parameters varied during the simulations [38]. We perform 1000 permutations, collecting leakage current values during drowsy mode for $t = t_0$ and $t = 5y$. The obtained results are presented in Figure 19, where the x-axis is the possible leakage current values of a memory cell categorized in 100 bins and the y-axis is their counted occurrences. The obtained random data exhibit a mean and standard deviation values $m(t_0)=3.75E-10A$, $\sigma(t_0)=1.74E-10A$ and $m(5y)=1.84E-10A$, $\sigma(5y)=1.56E-10$ for $t = t_0$ and $t=5y$, respectively, thus validating that even in the presence of process variations, the average leakage current at drowsy mode is considerably reduced. Therefore, for the proper evaluation of the absolute static power consumption during the drowsy mode, the process variation effect should be considered. Nevertheless, the practicality of the proposed technique is not affected, since the relative power reduction at 5 years of operation over its value at t_0 incurs a negligible errors if process variation is not accounted for (50.9% power reduction with process variation and 52.1% using only nominal values). We expect to find similar results for different values of the V_{dd}^D .

We have evaluated the impact of process variation also on Q_{crit} and SNM, and it has been found to be considerably lower compared to that on the static power consumption. In particular, Q_{crit} varies in the range 0.93fC-1.91fC at t_0 , and in the range 0.55fC-1.44fC at 5 years. Since these values are very small, the presence of simulation inaccuracies should be considered. As for the SNM, it is in the range 196mV-229mV

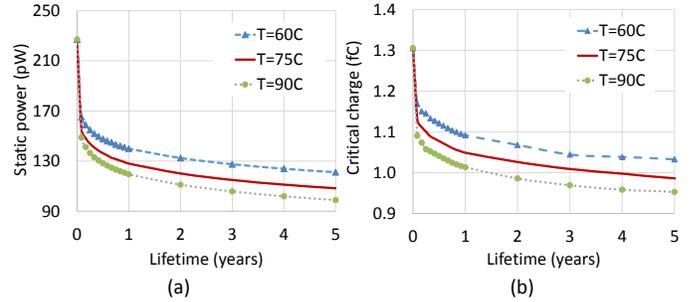


Fig. 20. Trend over time for different aging temperature: (a) static power; (b) critical charge.

TABLE IV
SNM FOR DIFFERENT AGING TEMPERATURES T_A .

lifetime	SNM (mV)		
	$T_A = 60^\circ C$	$T_A = 75^\circ C$	$T_A = 90^\circ C$
t_0	216	210	207
5y	197	195	192

at t_0 , and 179mV-208mV after 5 years of operation.

In Figure 20, we show the trend over time of static power and critical charge for three different temperatures, namely $60^\circ C$, $75^\circ C$ and $90^\circ C$. As expected, since the BTI aging increases with the temperature, to higher temperatures correspond lower static power (beneficial effect) and lower critical charge (detrimental effect). A larger impact of temperature on static power than critical charge is exhibited.

As for the SNM, similar to the results discussed in Section V, a lower impact of temperature has been found. The values obtained at t_0 and after 5 years of operation are reported in table Table IV.

We can conclude this analysis highlighting that, although the absolute values of the considered parameters are affected by process and temperature variations, the proposed approach is still valid, and the advantages provided by the proposed V_{dd}^D selection policies over the standard drowsy are approximately the same as in the nominal values, since they are all affected by the considered variations.

VII. CONCLUSIONS

In this paper, we showed that BTI-induced degradation can considerably benefit static power saving of drowsy cache memories. We developed a BTI analytical model for drowsy cache memories and a simulation framework allowing us to evaluate several trade-offs between power consumption and reliability. We proposed three drowsy architectures, allowing us to achieve different power/reliability trade-offs. Through SPICE simulations, we showed that, compared to standard aging unaware drowsy technique, a SER improvement up to 20.8% and a static noise margin increase up to 22.7% is enabled, with a limited increase in static power during only the very early lifetime, and with static energy saving up to 37% in 10 years of operation. These improvements are attained at no cost for the two static solution S_1 and S_2 , whereas the adaptive solution S_3 incurs a very limited area overhead. The practicality of the proposed approach is not affected by process and temperature variations.

ACKNOWLEDGMENTS

This work is partially supported by EPSRC (UK) under grant no. EP/K000810/1 and EP/K034448/1.

REFERENCES

- [1] D. Flynn, R. Aitken, A. Gibbons, and K. Shi, *Low Power Methodology Manual: For System-on-Chip Design*. NY, USA: Springer-Verlag, 2007.
- [2] N. S. Kim, K. Flautner, D. Blaauw, and T. Mudge, "Circuit and microarchitectural techniques for reducing cache leakage power," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 12, no. 2, pp. 167–184, 2004.
- [3] A. Nourivand, A. J. Al-Khalili, and Y. Savaria, "Postsilicon tuning of standby supply voltage in srams to reduce yield losses due to parametric data-retention failures," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 20, no. 1, pp. 29–41, 2012.
- [4] M. Powell, S.-H. Yang, B. Falsafi, K. Roy, and T. N. Vijaykumar, "Gated-vdd: A circuit technique to reduce leakage in deep-submicron cache memories," in *Proceedings of the 2000 International Symp. on Low Power Electronics and Design (ISLPED)*, July 2000, pp. 90–95.
- [5] M. J. Geiger, S. A. McKee, and G. S. Tyson, "Drowsy region-based caches: minimizing both dynamic and static power dissipation," in *Proc. of the 2nd conf. on Computing frontiers*. ACM, 2005, pp. 378–384.
- [6] A. Bardine, M. Comparesetti, P. Foglia, and C. A. Prete, "Evaluation of leakage reduction alternatives for deep submicron dynamic nonuniform cache architecture caches," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 1, pp. 185–190, 2014.
- [7] J. Wang and B. H. Calhoun, "Minimum supply voltage and yield estimation for large srams under parametric variations," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 19, no. 11, pp. 2120–2125, 2011.
- [8] K. Flautner, N. S. Kim, S. Martin, D. Blaauw, and T. Mudge, "Drowsy caches: simple techniques for reducing leakage power," in *Computer Architecture, 2002. Proceedings. 29th Annual International Symposium on*. IEEE, 2002, pp. 148–157.
- [9] V. Chandra and R. Aitken, "Impact of technology and voltage scaling on the soft error susceptibility in nanoscale cmos," in *Defect and Fault Tolerance of VLSI Systems, 2008. DFTVS'08. IEEE International Symposium on*. IEEE, 2008, pp. 114–122.
- [10] F. Firouzi, M. Salehi, F. Wang, and S. M. Fakhraie, "An accurate model for soft error rate estimation considering dynamic voltage and frequency scaling effects," *Microelectronics Reliability*, vol. 51, no. 2, pp. 460–467, Feb 2011.
- [11] Y. Shiyonovskii, A. Rajendran, and C. Papachristou, "A low power memory cell design for seu protection against radiation effects," in *Proc. of 2012 NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*, June 2012, pp. 288–295.
- [12] T. T.-H. Kim and Z. H. Kong, "Impact analysis of nbt/pbti on sram v min and design techniques for improved sram v min," *JSTS: Journal of Semiconductor Technology and Science*, vol. 13, no. 2, pp. 87–97, 2013.
- [13] H. Yi, T. Yoneda, M. Inoue, Y. Sato, S. Kajihara, and H. Fujiwara, "Impact of bias temperature instability on soft error susceptibility," *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 11, pp. 1951–1959, 2012.
- [14] D. Rossi, M. Omaña, C. Metra, and A. Paccagnella, "Impact of bias temperature instability on soft error susceptibility," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, no. 4, pp. 743–751, April 2015.
- [15] M. Agarwal, V. Balakrishnan, A. Bhuyan, K. Kim, B. C. Paul, W. Wang, B. Yang, Y. Cao, and S. Mitra, "Optimized circuit failure prediction for aging: Practicality and promise," in *Proc. of IEEE International Test Conf. (ITC)*, 2008, pp. 1–10.
- [16] D. Rossi, M. Omaña, C. Metra, and A. Paccagnella, "Impact of aging phenomena on soft error susceptibility," in *Proc. of IEEE International Symp. on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*, 2011, pp. 18–24.
- [17] K. Joshi, S. Mukhopadhyay, N. Goel, and S. Mahapatra, "A consistent physical framework for n and p bti in hkmg mosfets," in *Proc. of IEEE International Reliability Physics Symposium (IRPS)*, 2012, pp. 5A.3.1–5A.3.10.
- [18] M. A. Alam, H. Kufluoglu, D. Varghese, and S. Mahapatra, "A comprehensive model for pmos nbt degradation: Recent progress," *Microelectronics Reliability*, vol. 47, no. 6, pp. 853–862, 2007.
- [19] D. Rossi, V. Tenentes, S. Khursheed, and B. M. Al-Hashimi, "Nbt and leakage aware sleep transistor design for reliable and energy efficient power gating," in *2015 20th IEEE European Test Symposium (ETS)*, May 2015, pp. 1–6.
- [20] D. Rossi, V. Tenentes, S. Yang, S. Khursheed, and B. M. Al-Hashimi, "Reliable power gating with nbt aging benefits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 8, pp. 2735–2744, Aug 2016.
- [21] D. Rossi, V. Tenentes, S. Yang, S. Khursheed, and B. Al-Hashimi, "Aging benefits in nanometer cmos designs," *IEEE Transactions on Circuits and Systems II: Express Briefs (IEEE Early Access Article)*, vol. PP, no. 99, pp. 1–1, 2016.
- [22] B. Farahani, S. Habibi, and S. Safari, "A cross-layer ser analysis in the presence of pvta variations," *Microelectronics Reliability*, vol. 55, no. 7, pp. 1013–1027, Jun 2015.
- [23] D. Rossi, V. Tenentes, S. Khursheed, and B. M. Al-Hashimi, "Bti and leakage aware dynamic voltage scaling for reliable low power cache memories," in *Proc. of 2015 IEEE 21st International On-Line Testing Symposium (IOLTS)*, July 2015, pp. 194–199.
- [24] A. Ricketts, J. Singh, K. Ramakrishnan, N. Vijaykrishnan, and D. K. Pradhan, "Investigating the impact of nbt on different power saving cache strategies," in *2010 Design, Automation Test in Europe Conference Exhibition (DATE 2010)*, March 2010, pp. 592–597.
- [25] M. Basoglu, M. Orshansky, and M. Erez, "Nbt-aware dvfs: A new approach to saving energy and increasing processor lifetime," in *2010 ACM/IEEE International Symposium on Low-Power Electronics and Design (ISLPED)*, Aug 2010, pp. 253–258.
- [26] "Predictive Technology Model (PTM)," <http://www.ptm.asu.edu>.
- [27] Y. Wang, S. Roy, and N. Ranganathan, "Run-time power-gating in caches of gpus for leakage energy savings," in *Proc. of 2012 Design, Automation Test in Europe Conference Exhibition (DATE)*, March 2012, pp. 300–303.
- [28] M. Kulkarni, K. Sheth, and V. D. Agrawal, "Architectural power management for high leakage technologies," in *Proc. of 2011 IEEE 43rd Southeastern Symposium on System Theory (SSST)*,. IEEE, 2011, pp. 67–72.
- [29] M. Fukui, S. Nakai, H. Miki, and S. Tsukiyama, "A dependable power grid optimization algorithm considering nbt timing degradation," in *Proc. of IEEE Int'l New Circuits and Systems Conf.*, June 2011, pp. 370–373.
- [30] H.-I. Yang, W. Hwang, and C.-T. Chuang, "Impacts of nbt/pbti and contact resistance on power-gated sram with high-metal-gate devices," *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, vol. 19, no. 7, pp. 1192–1204, 2011.
- [31] L. Zhang and R. P. Dick, "Scheduled voltage scaling for increasing lifetime in the presence of nbt," in *2009 Asia and South Pacific Design Automation Conference*, Jan 2009, pp. 492–497.
- [32] N. Parihar, N. Goel, A. Chaudhary, and S. Mahapatra, "A modeling framework for nbt degradation under dynamic voltage and frequency scaling," *IEEE Transactions on Electron Devices*, vol. 63, no. 3, pp. 946–953, March 2016.
- [33] F. Wang and V. D. Agrawal, "Soft error rate determination for nanoscale sequential logic," in *Proc. of 2010 11th International Symposium on Quality Electronic Design (ISQED)*,. March 2010, pp. 225–230.
- [34] H. Singh and H. Mahmoodi, "Analysis of sram reliability under combined effect of nbt, process and temperature variations in nanoscale cmos," in *Proc. of 2010 5th International Conference on Future Information Technology*, May 2010, pp. 1–4.
- [35] W. Huang, K. Sankaranarayanan, K. Skadron, R. J. Ribando, and M. R. Stan, "Accurate, pre-rtl temperature-aware design using a parameterized, geometric thermal model," *IEEE Transactions on Computers*, vol. 57, no. 9, pp. 1277–1288, Sept 2008.
- [36] H. Chahal, V. Tenentes, D. Rossi, and B. M. Al-Hashimi, "Bti aware thermal management for reliable dvfs designs," in *2016 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*, Sept 2016, pp. 1–6.
- [37] D. Rossi, J. M. Cazeaux, M. Omana, C. Metra, and A. Chatterjee, "Accurate linear model for set critical charge estimation," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 17, no. 8, pp. 1161–1166, Aug 2009.
- [38] V. Tenentes, D. Rossi, S. Yang, S. Khursheed, B. M. Al-Hashimi, and S. R. Gunn, "Coarse-grained online monitoring of bti aging by reusing power-gating infrastructure," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, no. 4, pp. 1397–1407, April 2017.



Daniele Rossi received the Laurea degree in electronic engineering and the Ph.D. degree in electronics and computer engineering from the University of Bologna, Italy, in 2001 and 2005, respectively. He is currently a Senior Lecturer at the University of Westminster, UK, in the Applied DSP and VLSI research group. His research interests include fault modeling and design for reliability and test, focusing on low power and reliable digital design, robust design for soft error and aging resiliency, high quality test for low power systems and power electronics. He

has co-authored more than 90 papers published in international journals and conference proceedings, and holds one patent. Dr. Rossi is a Senior Member of IEEE.



Andrew Brown is Professor of Electronics at Southampton University, UK. He has held visiting posts at IBM Hursley Park (UK), Siemens NeuPerlach (Germany), Multiple Access Communications (UK), LME Design Automation (UK), Trondheim University (Norway), Cambridge University (UK) and EPFL (Switzerland). He is a Fellow of the IET and BCS, a Chartered Engineer, and a European Engineer.



Vasileios Tenentes received the B.Sc. degree in computer science from the University of Piraeus, Piraeus, Greece, in 2003. He also received the M.Sc. degree in computer science and the Ph.D. degree in computer science and engineering from the Department of Computer Science and Engineering, University of Ioannina, Ioannina, Greece, in 2007 and 2013, respectively. He has been a Research Fellow with the University of Southampton, U.K., since 2014 and an ARM Research Engineer, Cambridge, UK, since 2017. His research interests include electronic

design automation, design for testability, fault modeling, design for reliability and energy efficiency, wear-out effects analysis and modeling, and reliability assessment for IoT applications.



Sudhakar M. Reddy received the B.Sc. degree in Physics and the B.E. degree in Electronic Communications Engineering (ECE) from Osmania University, Hyderabad, the M.E. degree in ECE from the Indian Institute of Science, Bengaluru, India, and the Ph.D. degree from the University of Iowa, Iowa City, Iowa. He joined the faculty of the Department of Electrical and Computer Engineering at the University of Iowa in 1968 where he is currently a University of Iowa Foundation Distinguished Professor of ECE. He served as the chairperson of the ECE

Department from 1981 to 2000.

Professor Reddy has authored or co-authored over six hundred papers published in journals and proceedings of international conferences. Several papers co-authored by him received best paper nominations and awards. Professor Reddy has given tutorials and short courses to practicing engineers. He received a Von Humboldt Prize in 1995 and the first Life Time Achievement Award from the International Conference on VLSI Design. Professor Reddy is a Fellow of IEEE.



Bashir M. Al-Hashimi is an ARM Professor of Computer Engineering and Dean of the Faculty of Physical Sciences and Engineering, University of Southampton. In 2009, he was elected fellow of the IEEE for significant contributions to the design and test of low-power circuits and systems. He holds a Royal Society Wolfson Research Merit Award (2014-2019). He has published over 300 technical papers, authored or co-authored 5 books and has graduated 31 PhD students.