Microelectronics Journal 42 (2011) 101–112

Contents lists available at ScienceDirect

Microelectronics Journal

journal homepage: www.elsevier.com/locate/mejo

Impacts of gate-oxide breakdown on power-gated SRAM

Hao-I Yang, Wei Hwang *, Ching-Te Chuang

Department of Electronics Engineering and Institute of Electronics, National Chiao-Tung University, 1001 University Road, Hsinchu, Taiwan 300, ROC

ARTICLE INFO

Article history:
Received 3 March 2010
Received in revised form
18 August 2010
Accepted 20 August 2010
Available online 19 September 2010

Keywords:
Gate-oxide breakdown
Power-gating technology
SRAM

ABSTRACT

This paper presents a detailed analysis on the impacts of various gate-oxide breakdown (BD) paths in column-based header- and footer-gated SRAMs. It is shown that with gate-oxide BD, the read static noise margin (RSNM) and write margin (WM) degrade in general. Pass-transistor gate-oxide BD between WL and BL is shown to degrade read/write margin and performance, and to affect other healthy cells along the same column as well. The effects of gate-to-source BD of cell transistors are shown to confine to the individual cells, while multiple cells suffering cell transistor drain-to-drain BD in a column could cumulatively affect VVDD (header structure) or VVSS (footer structure), thus influencing other cells in the same column. In particular, we show that the gate-oxide BD of the power-switches has severe and even detrimental effects on the margin, stability, and performance of the SRAM array. Several techniques to mitigate the power-switch gate-oxide BD have been evaluated, including adding a gate series resistance to the power-switch, dual threshold voltage power-switch, thick gate-oxide power-switch, and dual gate-oxide thickness (dual-T_{OX}) power-switch. It is shown that the dual-T_{OX} power-switch improves the time-to-dielectric-breakdown (T_{BD}) of the power-switch while maintaining the performance without side effect.

1. Introduction

With technology scaling, the gate-oxide breakdown (BD) has become a major concern in CMOS circuit designs. It is particularly important to understand the impacts of gate-oxide BD on SRAMs as SRAMs occupy most area of a chip and the probability of gate-oxide BD is significantly higher than the logic circuits. Moreover, the margin and stability of SRAM cell deteriorate with technology scaling due to serious threshold voltage (V_{TH}) scatter caused by process variation, random dopant fluctuation (RDF), and microscopic effects such as line edge roughness (LER). When gate-oxide BD happens, a leakage path appears between the gate and source, gate and drain, or gate and channel. The leakage from gate-oxide BD exasperates the already poor margin and stability of scaled SRAM cells, leading to functional failure and yield loss.

In addition, most state-of-the-art SRAMs are designed with power-gating structures to reduce leakage in standby or sleep mode [1–4]. In active mode, the power-switches turn on to provide sufficient current to the cell array to maintain data access performance. The power-switches turn off to reduce standby leakage when SRAM is idle. Unfortunately, the power-switches can suffer gate-oxide BD as well and the function of power-gating structure would be affected.

Previous papers on gate-oxide BD [5,6] focused only on a single cell in a normal SRAM architecture. In this paper, we present a detailed analysis on the impacts of gate-oxide BD on power-gated SRAMs. We show that the gate-oxide breakdown of the power-switches have severe and even detrimental effects on the margin, stability, and performance of the SRAM array. As such, we evaluate several techniques to mitigate the gate-oxide BD, and propose a dual oxide thickness (dual-T_{OX}) power-gating structure to extend the lifetime of power-switches and to mitigate the negative impacts induced by power-switch gate-oxide BD. In the following sections, we first describe the details of our simulation model in PTM CMOS 32 nm technology. The read static noise margin (RSNM) and write margin (WM) of cells in a power-gated SRAM under various gate-oxide BD paths are examined. We then analyze the impacts of different gate-oxide BD paths on SRAM read/write delay, followed by the impacts of gate-oxide BD on virtual supply lines in standby or sleep mode, and the virtual supply bounce during wake-up transition. Furthermore, several techniques to mitigate the power-switch gate-oxide BD to extend the lifetime of power-gated SRAM are evaluated. Based on the evaluation results, a dual gate-oxide thickness power-gating structure is proposed and recommended.

2. Power-gated SRAM models

This section presents the details of the transistor gate-oxide BD model and SRAM power-gating structures used in our analyses.

* Corresponding author. Tel.: +886 3 572 9541; fax: +886 3 573 8390.
E-mail addresses: haoiyang@gmail.com (H.-I. Yang), hwang@mail.nctu.edu.tw (W. Hwang).

0026-2692/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.mejo.2010.08.020

2.1. Gate-oxide BD model

The post-BD characteristics of an NMOS can be explained by the location of a constant-size breakdown path and modeled as a narrow inclusion of highly-doped n-type silicon well with the equivalent circuit shown in Fig. 1 [7]. Under normal operating conditions, the model reduces to simple equivalent resistance

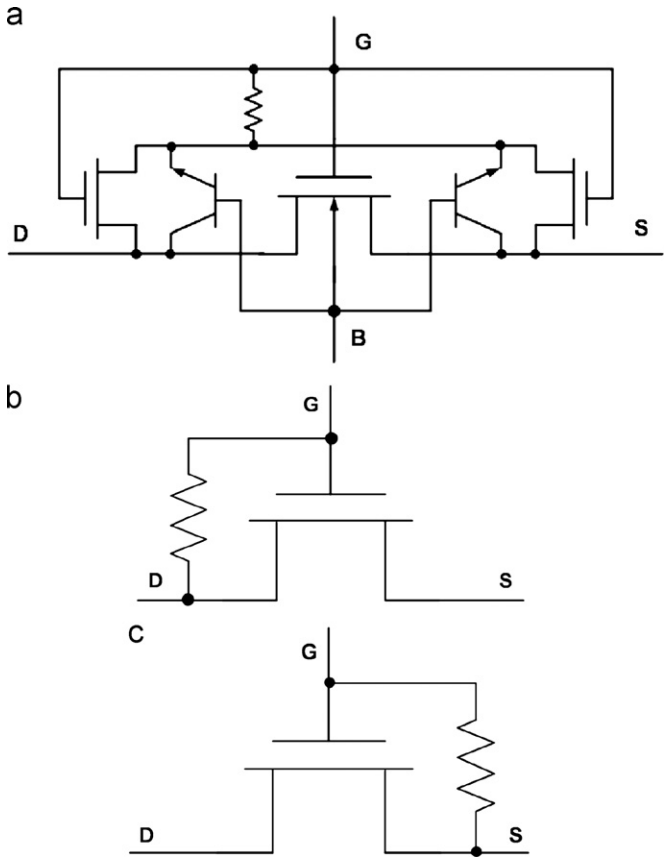

Fig. 1. (a) Equivalent gate-oxide HBD circuit model. This model can be simplified to (b) with gate-oxide HBD in drain and (c) with gate-oxide HBD in source [7].

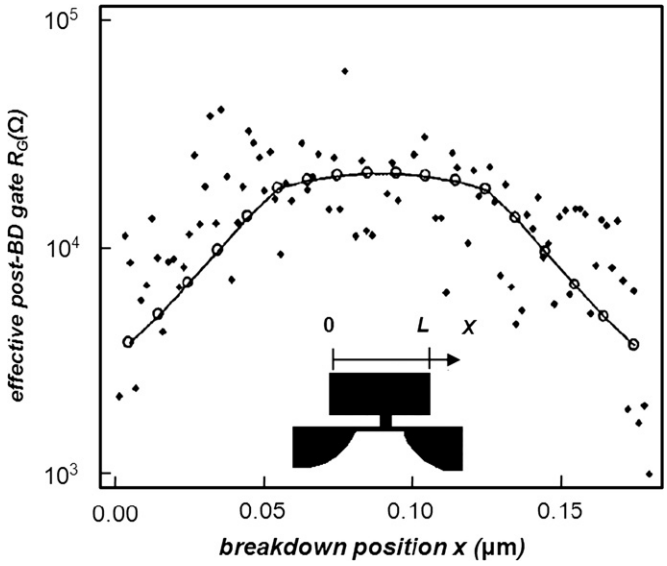

Fig. 2. Effective post-BD resistance vs. BD position [8].

Table 1
Possible gate-oxide breakdown path summary.

Acronym	Description
CPS BD	Cell PMOS gate-to-source breakdown
CDD BD	Cell drain-to-drain breakdown
CNS BD	Cell NMOS gate-to-source breakdown
CPG BD	Cell pass-gate breakdown
PHS BD	Header power switch gate-to-source breakdown
PHD BD	Header power switch gate-to-drain breakdown
PFS BD	Footer power switch gate-to-source breakdown
PFD BD	Footer power switch gate-to-drain breakdown

paths between the gate and source/drain. The gate leakage current after oxide hard BD can be modeled as $I=(1/R)V$; where V is the voltage across the gate-oxide and R is the equivalent post-BD resistance depending on the BD spot area and its position.

The equivalent post-BD resistance becomes smaller when the gate-oxide BD spot position is close to drain or source of a MOSFET (Fig. 2) [8]. Consequently, post-BD gate-to-drain and gate-to-source leakages affect SRAM most significantly. Thus, our analysis focuses on various gate-to-drain and gate-to-source post-BD current paths in the cell and the header/footer power-switches (Table 1).

2.2. SRAM power-gating structures

SRAM power-gating structures can be classified into header- and footer-gating structures. Fig. 3(a) shows a column-based header-gated SRAM structure, wherein PH1 is the header power-switch and MH1 is the clamping device to bias virtual supply (VVDD) for data retention in standby/sleep mode. Fig. 3(b) shows a column-based footer-gated SRAM structure. MF1 is the footer power-switch, and PF1 is the clamping device to bias virtual ground (VVSS) for data retention in standby/sleep mode. The power-switch (sleep transistor) can be implemented using a single lumped large or distributed small sleep transistors. The smaller area transistor has lower gate-oxide BD probability. However, with the total area of the distributed sleep transistors equal to the area of the lumped sleep transistor, TBD and fail rate of the lumped sleep transistor and the distributed sleep transistors are the same [9–11]. Moreover, package model [12] is also included in our analysis, and parasitic capacitance, inductance and resistance of the package are 5.32 pF, 8.18 nH, and 0.217 Ω, respectively. Fig. 4 shows the standard 6T SRAM cells with post-BD leakage paths in our analysis. The SRAM array block size is 128×128 cells. The SRAM supply voltage is 0.9 V, and the analysis temperature is 125 °C. The detailed simulation results based on BSIM Predictive Model PTM 32 nm [13]. In order to compare impacts of various gate-oxide BD paths, only one gate-oxide BD location/path is considered in each simulation. Table 1 summarizes these gate-oxide BD locations in our analysis. Moreover, the gate-oxide BD is assumed to happen only in one SRAM cell per column in each simulation.

3. Read/write operation

RSNM is defined as the minimum voltage difference between the SRAM inverter trip voltage and the read disturb voltage induced by voltage divider effect between the pass-transistor and the cell pull-down NMOS during read cycles. As shown in Fig. 5(a), RSNM of a header-gated SRAM degrades most when PHD (header power-switch gate-to-drain) BD happens. In contrast, the impact of PHS (header power-switch gate-to-source) BD on RSNM is negligible. The reason is as following: HPG (gate signal of header

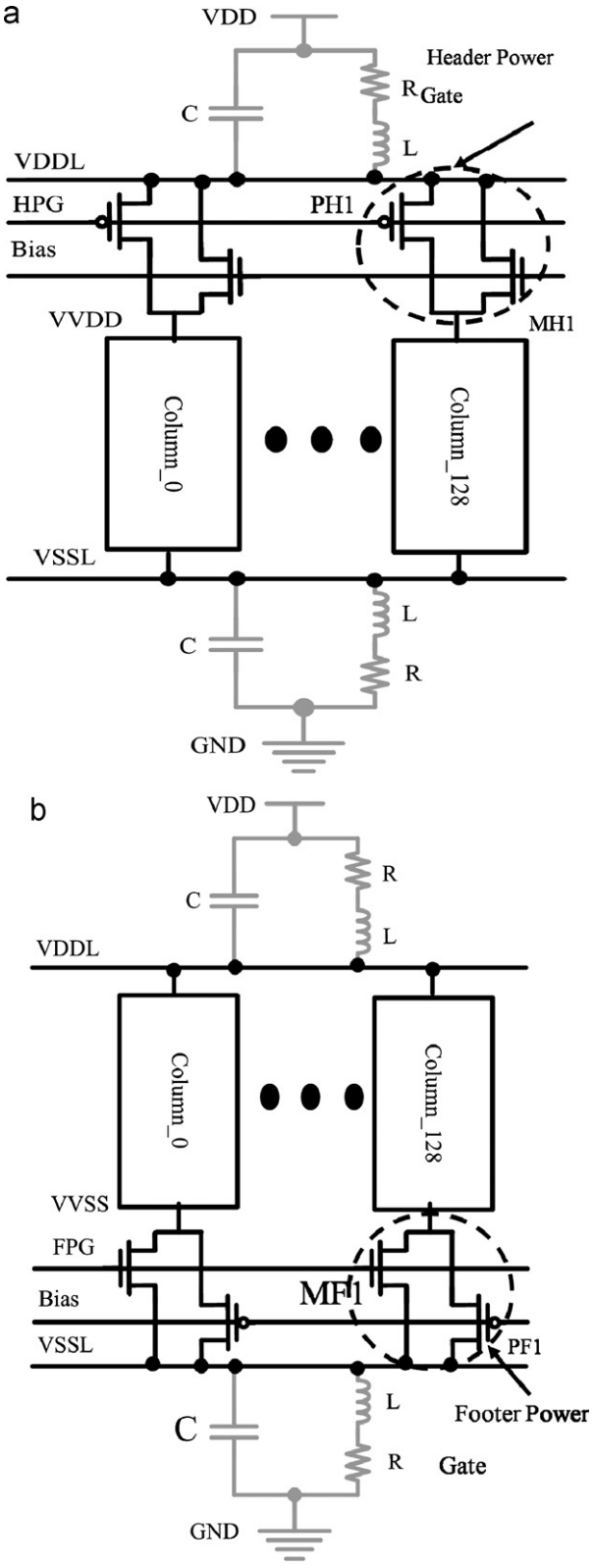

Fig. 3. Power-gated SRAM with (a) header, and (b) footer.

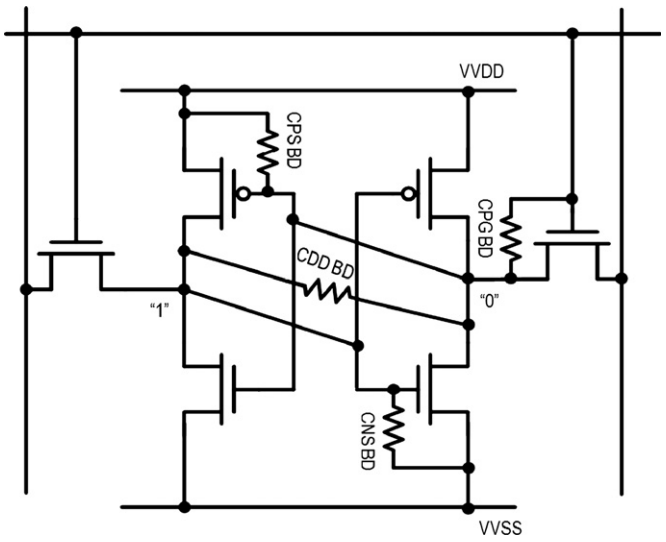

Fig. 4. 6T SRAM with possible BD paths.

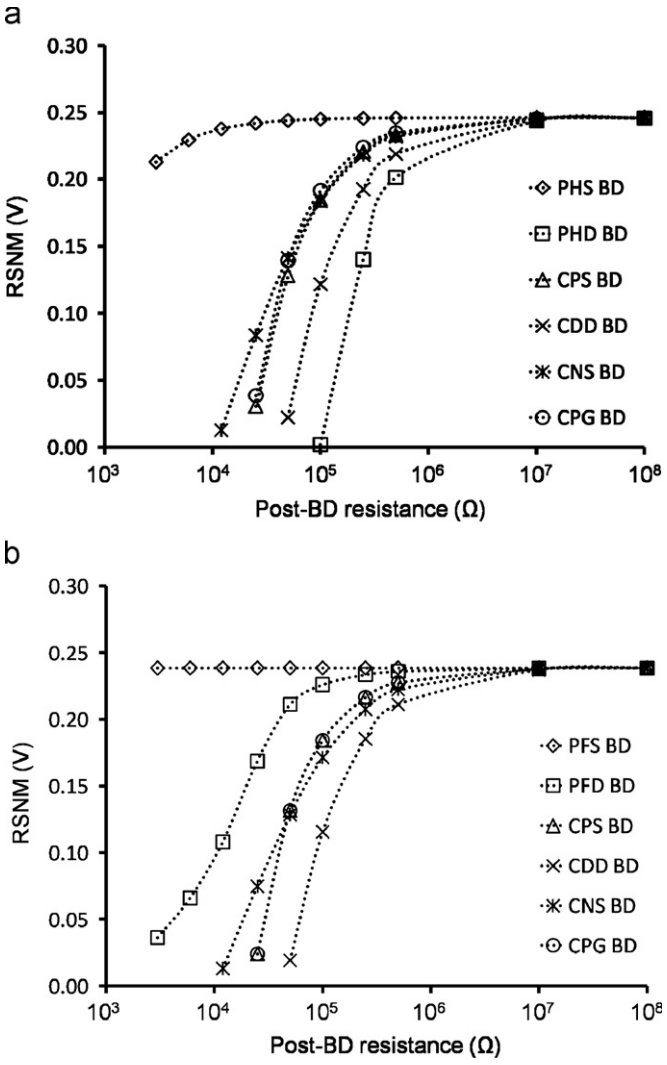

Fig. 5. RSNM vs. post-BD equivalent resistance (a) in a header-gated SRAM and (b) in a footer-gated SRAM.

power-switch) should be "low" to turn on power-switches to provide sufficient current for the SRAM array during access cycles. Thus, VVDD decreases seriously when PHD BD happens as it provides a resistive leakage path between the VVDD and the "low" HPG node, leading to RSNM degradation. On the other hand, when PHS BD happens, its post-BD resistive path shunts the gate-source of the power switch, and VVDD is not significantly affected until the post-BD resistance becomes very small. Additionally,

RSNM degradation induced by cell drain-to-drain (CDD) BD is worse than that induced by Cell PMOS gate-to-source (CPS) BD and Cell NMOS gate-to-source (CNS) BD. This is because the SRAM inverter trip voltage decreases and the read disturb voltage increases at the same time when CDD BD happens. On the other hand, only read disturb voltage increases when CPS BD happens, while only trip voltage decreases when CNS BD happens. Similar RSNM behavior can also be found in a footer-gated SRAM (Fig. 5(b)). The difference is that the post-BD leakage path of a cell plays more important roles than the corresponding cases for header-switches. Also, the NMOS footer-switch is stronger than the PMOS header-switch in our simulation model; hence, the "same" resistive leakage path in NMOS footer would have less effect compared with PMOS header.

In advanced process technology, the random dopant fluctuation (RDF) and line edge roughness (LER) induce V_{TH} mismatch among SRAM cell transistors, leading to SRAM RSNM and write margin degradation. Gate-oxide BD increases gate leakage currents, thus aggravating the already poor stability and margin and leading to SRAM reliability failure after long time of usage. Fig. 6(a) compares the impacts of V_{TH} mismatch and CDD BD on RSNM. The RSNM can be seen to degrade almost linearly with increased V_{TH} mismatch. On the other hand, the RSNM remains relatively constant as long as the post-BD resistance is larger than $10^8\,\Omega$. The RSNM starts to degrade as the post-BD resistance becomes smaller than $10^8\,\Omega$, and the degradation becomes very significant (the slope of deterioration becomes very steep) when the post-BD resistance is smaller than $10^6\,\Omega$. Fig. 6(b) shows the combined impact of gate-oxide BD and V_{TH} mismatch (of 30, 40, and 50 mV) on RSNM fail rate using Monte–Carlo analysis with 50,000 samples. As can be seen, when the post-BD resistance is larger than $10^6\,\Omega$, fail rates of these three cases approach zero. Once post-BD resistance becomes smaller than $10^6\,\Omega$, fail rates start to increase sharply. Furthermore, when post-BD resistance is larger than about $1.2\times10^5\,\Omega$, the case with higher V_{TH} mismatch has higher fail rate. However, when post-BD resistance is smaller than about $1.2\times10^5\,\Omega$, the differences in the fail rates for the three cases become less significant. This implies that the V_{TH} mismatch dominates the RSNM fail rate when the post-BD resistance is moderate to high, and gate-oxide BD dominates the RSNM fail rate when post-BD resistance is low ($<1.2\times10^5\,\Omega$ in our case).

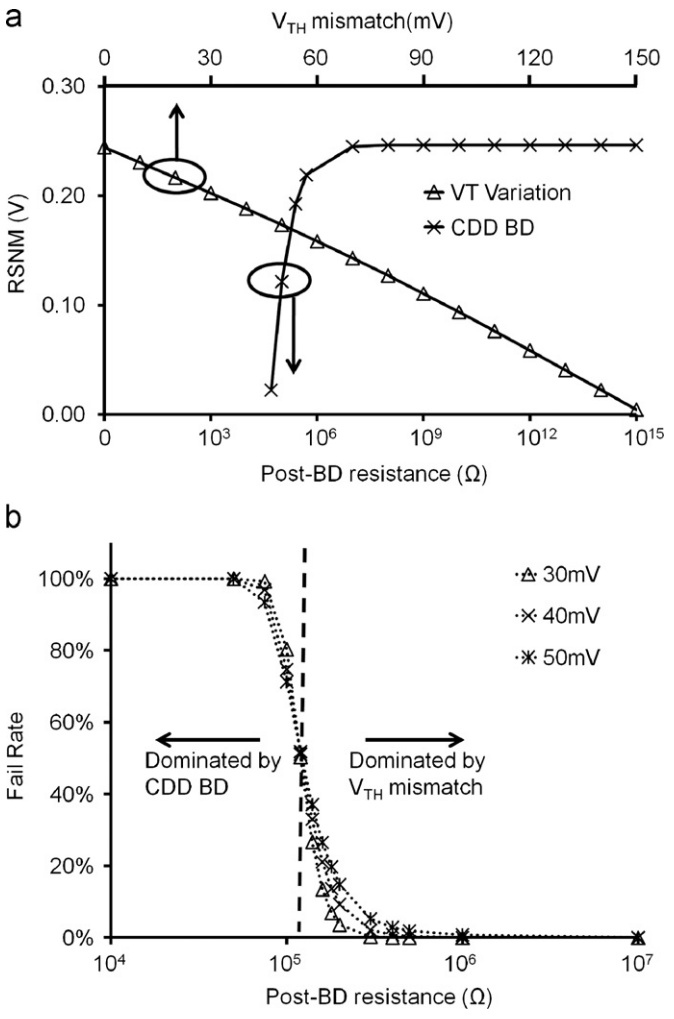

Fig. 6. (a) RSNM vs. post-BD equivalent resistance and V_{TH} mismatch, and (b) RSNM fail rate under the combined impact of CDD BD and V_{TH} mismatch using Monte–Carlo simulations (sample size=50000).

Impacts of gate-oxide BD on SRAM read performance are also analyzed. Recently, due to increased process variation in scaled technologies, large signal sensing structures have been widely used. Assuming an inverter is used to sense the bit-line (BL) signal, it is proper to define BL delay as the time from WL rises to half-VDD to the time BL discharges to half-VDD (for read "0"). Fig. 7(a) shows all possible gate-oxide BD paths that affect BL delay in SRAM cells. Especially, we consider the post-BD leakage path between WL and storage node (CPG BD) in the pass-gate, and the post-BD path between WL and BL (CPGD BD) in the pass-gate. The cell NMOS gate-to-source BD leakage path (CNS BD) is also divided into CNSQ BD and CNSQB BD.

Fig. 8(a) shows BL delays under different gate-oxide BD paths within a selected cell. The BL length in our analysis model is 128-bit. Most of the cases lead to increase in BL delay, except CNSQ BD. This is because for CPS BD and CPG BD leakage paths, the leakage currents flow into node Q, charging up node Q voltage level and thus degrading BL delay. When CDD BD happens, the leakage current not only reduces the voltage level of node QB to weaken the pull-down NMOS transistor for node Q, but also charges up the voltage level of node Q, with both effects degrading the BL delay When CNSQB BD happens, the leakage current causes a decrease of voltage level of QB, thus weakening the pull-down NMOS transistor for node Q and degrading the BL delay. Consequently, these gate-oxide BD paths make BL delay longer. In contrast, the CNSQ BD leakage path helps the cell to discharge BL, and hence reduces the BL delay. Fig. 8(b) shows the relation between BL delay and different gate-oxide BD paths in the header power switch. The figure shows that BL delay increases when PHD BD becomes more serious, but is relatively insensitive to PHS BD. This is because PHD BD lowers the voltage level of VVDD, thus reducing the voltage level of QB and the strength of pull-down NMOS transistor for node Q. Similar effect is observed in the footer structure. The PFD BD in the footer structure degrades BL read delay, while PFS BD has almost no effect on BL read delay. This is because PFD BD causes the voltage level of VVSS to increase, thus reducing the drive of the pull-down NMOS transistor. Fig. 8(c) compares BL delays between the case where the selected cell has CPGD BD and the case where a non-selected cell in the selected column has CPGD BD. Because the voltage of a non-selected WL is kept at zero, CPGD BD in a non-selected cell results in a leakage path from BL to the ground, thus reducing the BL voltage level. Notice that while the lower BL voltage level reduces the BL read delay, it also degrades the sensing margin. On the other hand, if a selected cell suffers CPGD BD, the post-BD leakage current flows from the selected WL to BL, thus charging up BL and leading to longer BL read delay.

WM is defined as the maximum bit-line voltage that can flip the state of a SRAM cell during write cycles. In the header-gated SRAM, PHD BD and CDD BD result in better WM (Fig. 9(a)). When

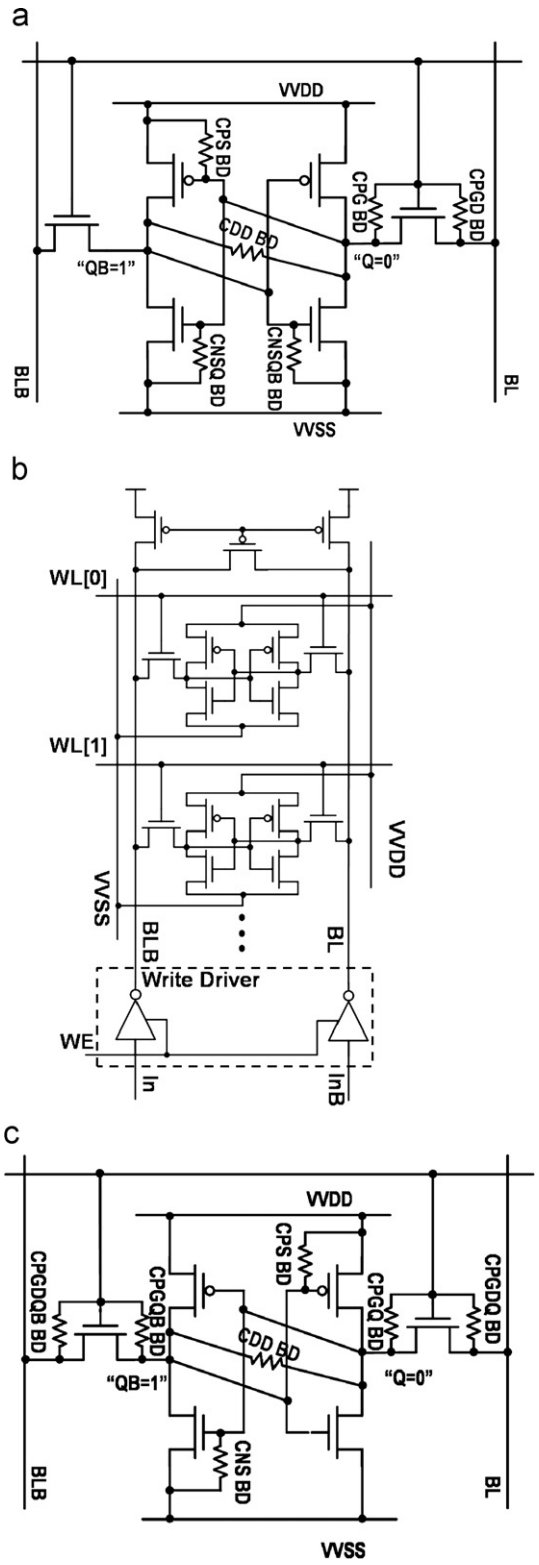

Fig. 7. (a) 6T SRAM with possible BD paths in Read analysis, (b) SRAM column with Write drivers, and (c) 6T SRAM with possible BD paths in write analysis.

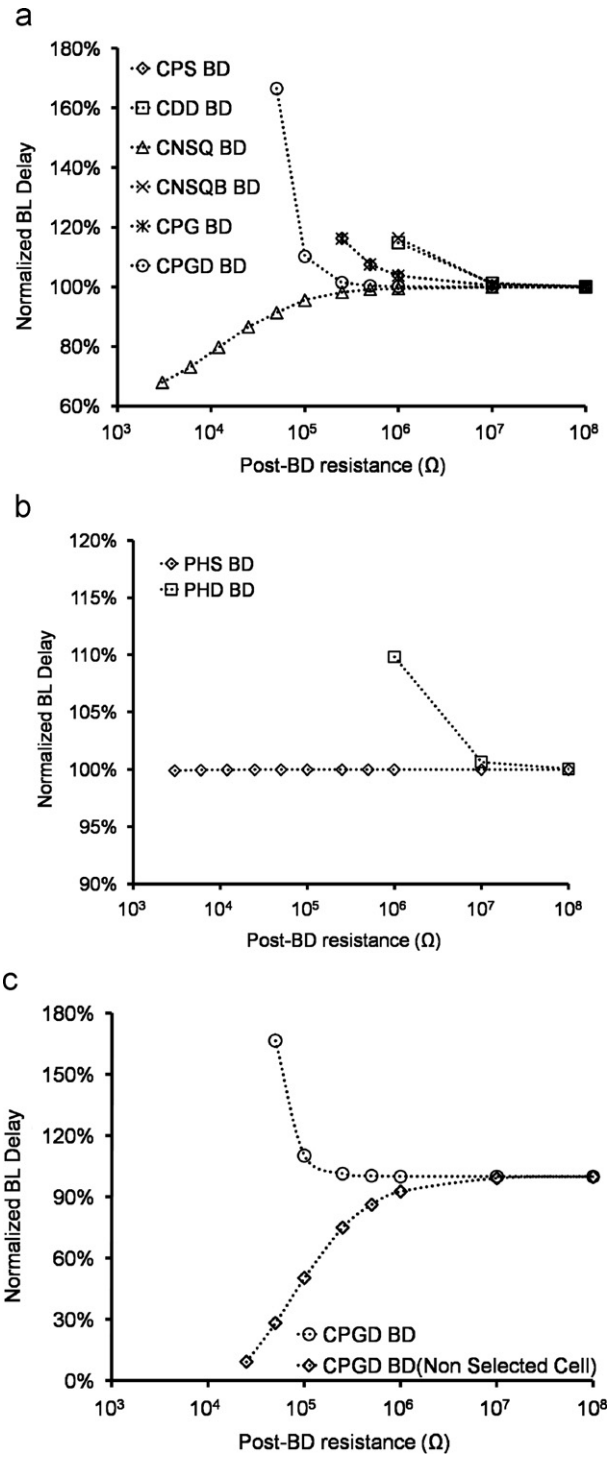

Fig. 8. Normalize BL delay vs. post-BD equivalent resistance in a header-gated SRAM (a) in SRAM cells, (b) in power-switches, and (c) in SRAM pass-gates.

CDD BD occurs, the node storing "high" would be dragged down by the node storing "low", rendering the "high" storage node easier to be written. When PHD BD happens, the VVDD is reduced. Thus, WM improves. However, CPS BD, CNS BD, and CPG BD make SRAM WM worse. When CPS BD and CPG BD happen, post-BD leakage paths charge up the node storing "high", leading to degradation of WM. When CNS BD happens, post-BD leakage path prevents the node storing "low" from charging to "high". Thus, WM also becomes worse. On the other hand, in the footer structure, CDD BD has the best WM, and impacts of the gate-oxide BD in footer switches on WM are less significant than that in header switches (Fig. 9(b)). This is because the NMOS footer-switch is stronger than PMOS header-switch in our analysis, and the impact of PFD BD in a footer structure is less significant

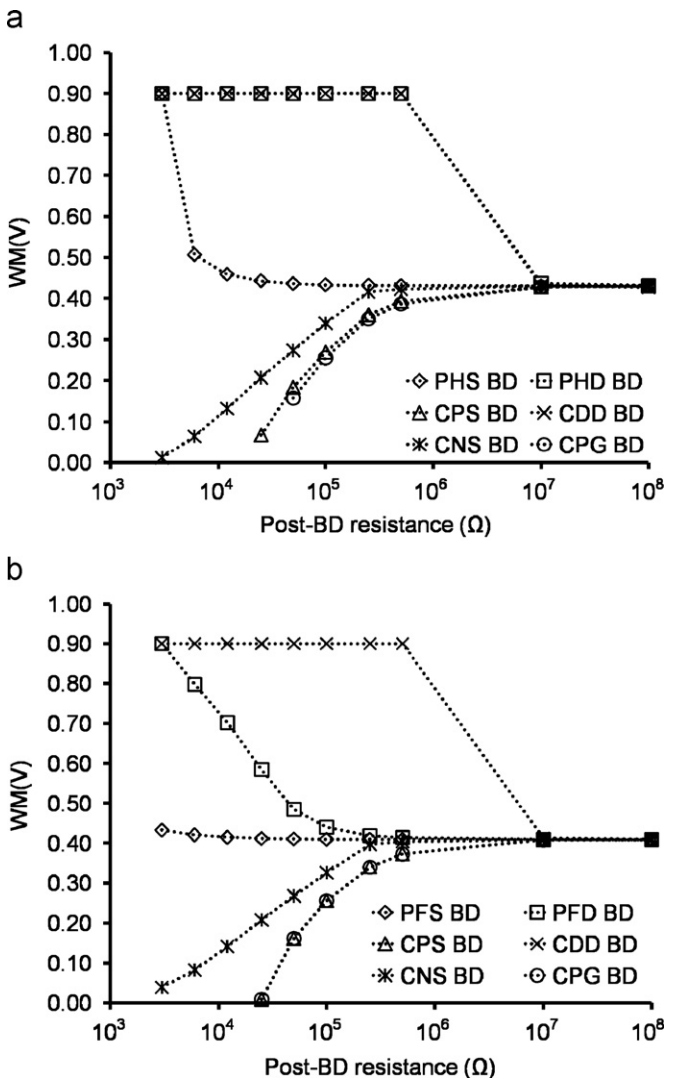

Fig. 9. WM vs. post-BD equivalent resistance (a) in a header-gated SRAM, and (b) in a footer-gated SRAM.

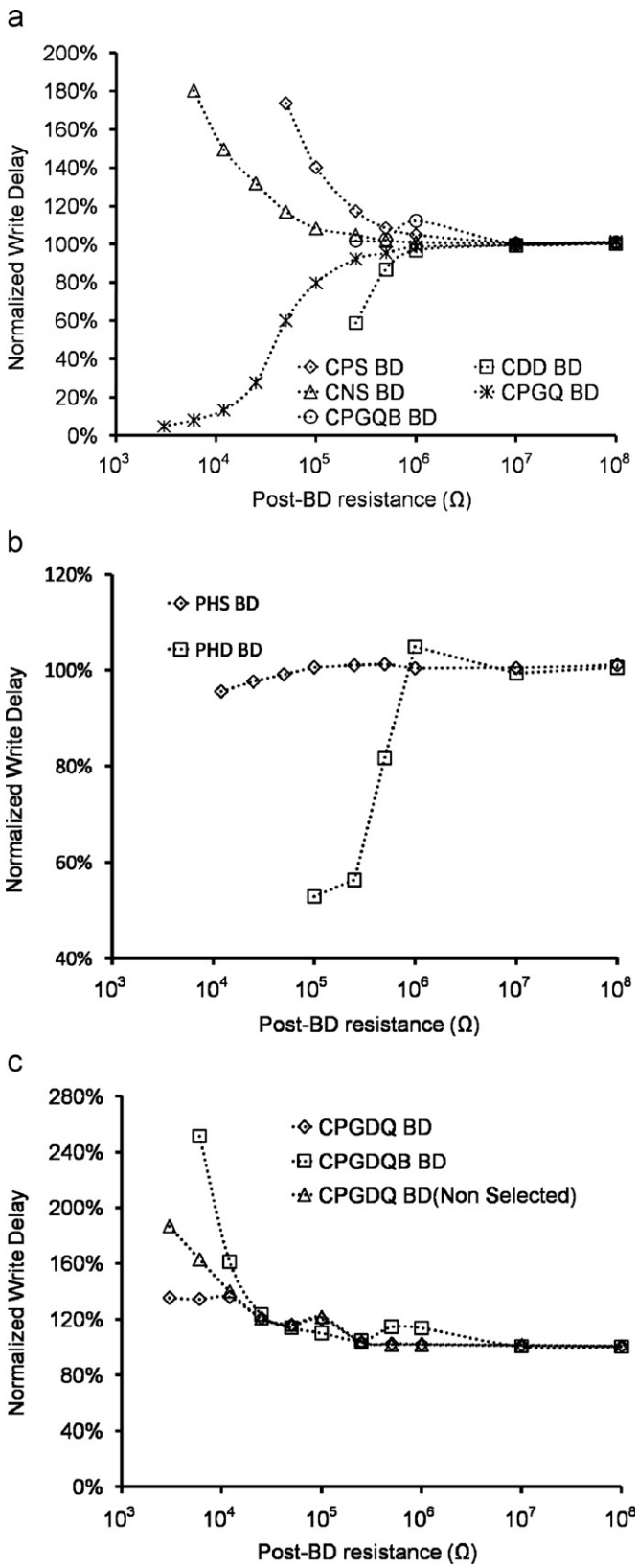

Fig. 10. Normalize write delay vs. post-BD equivalent resistance in a header-gated SRAM (a) in SRAM cells, (b) in power-switches, and (c) in SRAM pass-gates.

compared with PHD BD in a header structure with the same post-BD resistance.

Write delay can be defined as the latency between the time when the selected WL crosses 1/2 VDD to the time when voltage levels of Q and QB cross each other. Fig. 7(b) shows a column with its corresponding write drivers, and Fig. 7(c) shows a SRAM cell with various post-BD leakage paths. In Fig. 7(c), we separate the pass-gate BD leakage paths into CPGDQ BD, CPGQ BD, CPGQB BD, and CPGDQB BD to investigate their individual impact on write delay. The storage nodes, Q and QB, are assumed to store "low" and "high" respectively, and BL pairs are assumed to be ready (BL=1 and BLB=0) before the selected WL turns on.

Impacts of gate-oxide BD on SRAM Write delay in the header power-gating structure are shown in Fig. 10. Fig. 10(a) shows both CNS BD and CPS BD degrade write delay. When CNS BD happens, a leakage path appears between the storage node Q and ground, and it is more difficult to charge node Q to "high". For the post-BD leakage paths in the pass-gate transistor, while it seems that CPGQB BD would make write delay longer since it provides a path for the selected WL to charge QB, it actually causes decrease in the write delay because the voltage level of node QB (at "high") would be lower before the selected WL turns on. If CPGQB BD becomes more serious, node QB cannot maintain its stored "logic 1" anymore, and Write delay becomes negative. Fig. 10(b) shows both PHS BD and PHD BD improve write delay because these two gate-oxide BD paths cause VVDD to decrease. Fig. 10(c) shows

write delay vs. different gate-oxide BD paths in pass-gate transistors between the WL and BL. Before the selected WL turns on, CPGDQ BD discharges BL. Thus, BL voltage level (supposed to be kept at "high") is lower than VDD when the selected WL turns on, causing degradation in write delay. When the selected cell has CPGDQB BD, leakage current flows into BLB, and the voltage level of BLB (supposed to be at "low") increases, leading to write delay degradation. If an un-selected cell along the selected column suffers CPGDQ BD, write delay also becomes longer. This is because CPGDQ BD of the non-selected cell decreases BL voltage level since it provides a leakage path between BL and the un-selected WL of the unselected cell. Similar impacts of gate-oxide BD can also be observed in footer-gated structure.

The power-gating structure can be either fine-grain (row-based) or coarse-grain (sub-array based). When a fine-grain power-gated SRAM is accessed, only the power-switch of the selected row is active and others remain "off". The voltage across the cells in the un-selected rows remain low ($< V_{DD}$). Consequently, this architecture has lower gate-oxide BD probability. Furthermore, if gate-oxide BD happens on a row-based power-gated array, it only impacts the corresponding row. On the other hand, gate-oxide BD of a sub-array based power-gated array degrades the margin and stability of all cells in the sub-array. Therefore, the coarse-grain power-gating architecture is more sensitive to gate-oxide BD than the fine-grain architecture.

Another long term reliability issue in the advanced process technology is bias temperature instability (BTI). Bias temperature instability (BTI) can be classified into negative bias temperature instability (NBTI) for PMOS and positive bias temperature instability (PBTI) for NMOS. BTI increases V_{TH} of a transistor when the transistors turn on (Stress phase), but the V_{TH} drift partially recovers when the stress voltage is removed (recovery phase). On the contrary, the gate-oxide BD is not recoverable. Similar to gate-oxide BD, BTI degrades SRAM stability and WM after long time of usage. The degradation also becomes more severe with higher stress voltage [14]. Moreover, the degradation induced by BTI depends on the stored data during stress phase.

4. Virtual supply/ground line

In active mode, VVDD (header-gated structure) and VVSS (footer-gated structure) are impacted by gate-oxide BD. Fig. 11(a) shows the active mode VVDD of a header-gated SRAM. The impacts of PHD BD and PHS BD are larger than those cases where gate-oxide BD occurs in a cell. The VVDD decrease caused by PHD BD is the most serious because there is a direct post-BD current path between VVDD and the ground ("low" HPG node). Fig. 12 shows the active mode VVDD as functions of the number of SRAM cells having gate-oxide BD when the individual post-BD equivalent resistance is 6 KΩ. It shows that with CDD BD, VVDD decreases as more cells suffer gate-oxide BD. On the contrary, for CPS BD and CNS BD, the VVDD is quite independent of the number of cells suffering gate-oxide BD. When CDD BD happens, a current path occurs between two storage nodes of a cell, and the voltage difference between these two storage nodes becomes smaller. The pull-down NMOSs and pull-up PMOSs of the SRAM inverter pairs become weakly turned on, and currents flow through these inverter pairs to Ground. Hence, VVDD decreases with more SRAM cells suffering CDD BD. On the other hand, when CPS BD and CNS BD are serious, data stored in the cell would be flipped (causing error), and there would be no voltage across the gate-oxide leakage path, thus VVDD doesn't decrease. When a cell has one of these two BD locations, it does not affect other cells (although its own data may be destroyed). The dependence of active mode VVSS on gate-oxide BD location in a footer-gated SRAM is shown

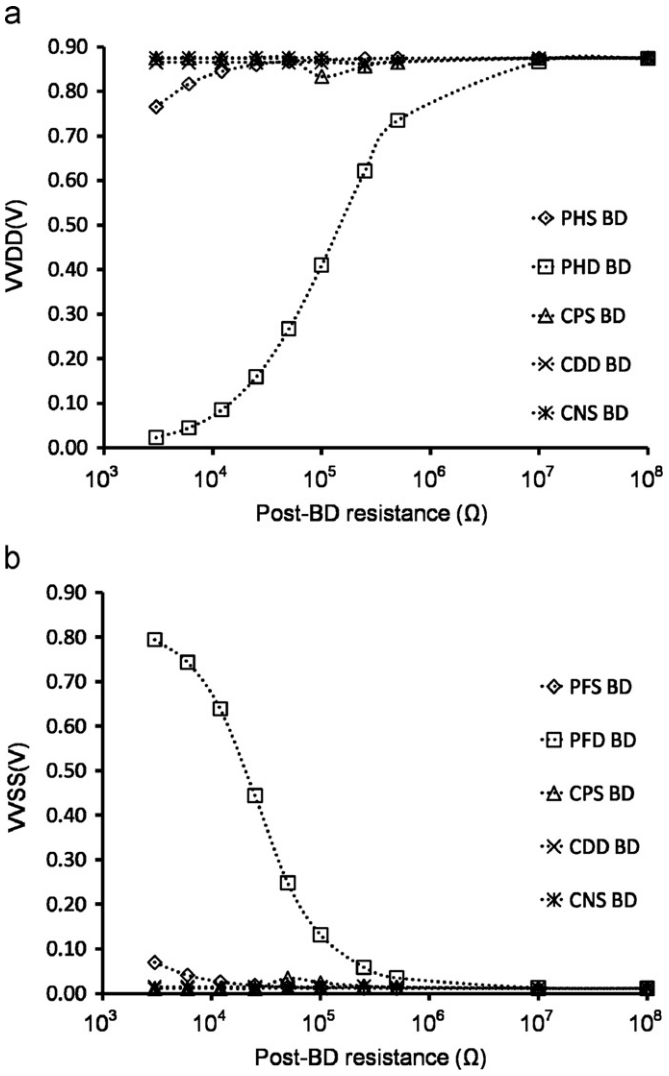

Fig. 11. (a) Active mode VVDD vs. post-BD equivalent resistance in a header-gated SRAM, and (b) active mode VVSS vs. post-BD equivalent resistance in a footer-gated SRAM.

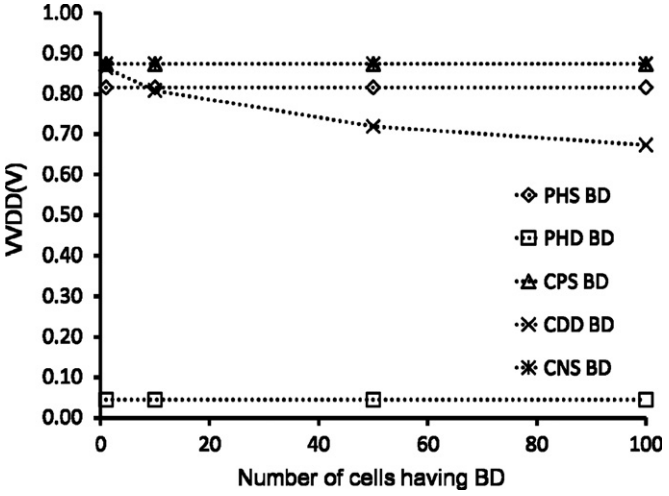

Fig. 12. Active mode VVDD vs. number of cells suffering BD with post-BD equivalent resistance of 6 KΩ.

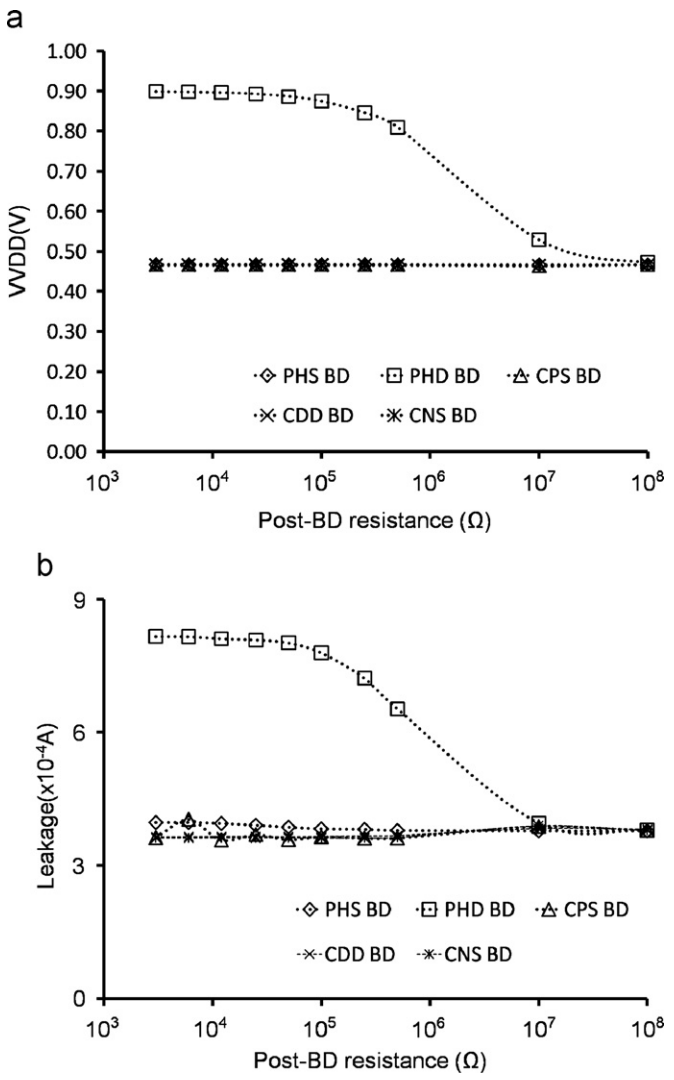

Fig. 13. (a) Sleep mode VVDD vs. post-BD equivalent resistance in a header-gated SRAM, and (b) sleep mode leakage vs. post-BD equivalent resistance in a header-gated SRAM.

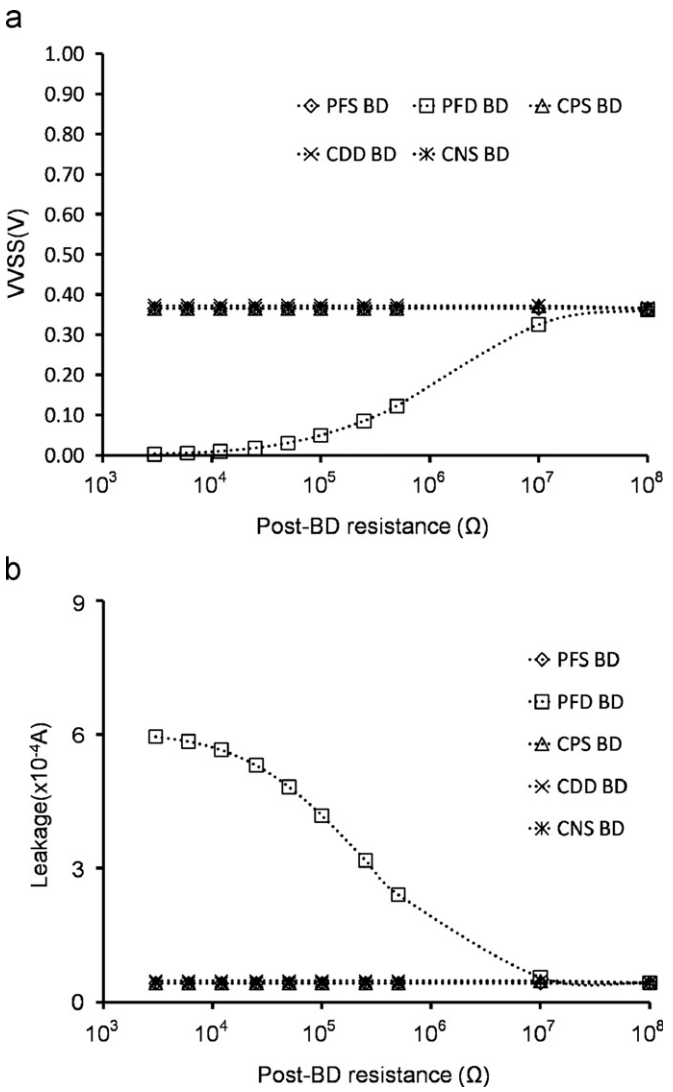

Fig. 14. (a) Sleep mode VVSS vs. post-BD equivalent resistance in a footer-gated SRAM, and (b) sleep mode leakage vs. post-BD equivalent resistance in a footer-gated SRAM.

in Fig. 11(b). PFD BD causes most increase in VVSS, while the effects of other gate-oxide BD locations are less significant.

When a header-gated SRAM is in sleep/standby mode, HPG is "high" to turn off the power-switch and VVDD is biased to proper level by the clamping device for data retention. As shown in Fig. 13(a), the sleep/standby mode VVDD increases significantly with more serious PHD BD, as there is a resistive path between VVDD and HPG (at "high"). Thus, the sleep/standby leakage increases significantly (Fig. 13(b)), and the power-gating structure loses its effectiveness. The effects of other gate-oxide BD locations are negligible. Similarly, when a footer-gated SRAM is in sleep/standby mode, FPG is "low" and VVSS is biased by the clamping device to proper level for data retention. PFD BD causes VVSS to decrease significantly (Fig. 14(a)), and footer-gating structure loses its ability to reduce sleep/standby leakage with serious PFD BD (Fig. 14(b)).

When the SRAM wakes up from sleep/standby mode, the power-switch turns on, and large wake-up current flows through parasitic capacitance, inductance, and resistance of the package, thus inducing supply line (node VDDL or VSSL in Fig. 3) and virtual supply line (VVDD or VVSS) bounce. In the header-gated structure, PHS BD and PHD BD induce significant VDDL bounce (Fig. 15(a)). This is because their post-BD current paths decrease the equivalent resistance of the power-switch. Similar behavior can be observed for VSSL bounce in the footer-gated structure (Fig. 15(b)).

5. Power-switch gate-oxide BD tolerance techniques

As shown in the previous analysis, the gate-oxide BD in the power-switches may lead to the collapse of virtual supply (ground) rails. In this section, we evaluate several techniques to mitigate the gate-oxide BD and extend the lifetime of the power-switches.

To mitigate the gate-oxide BD and limit the leakage current through the gate-oxide of the power-switch, a series resistance R_{Series} could be added between the gate of the power switch and its control signal, as shown in Fig. 16 for a header-gated structure. According to analysis in previous sections, PHD BD may cause VVDD collapse and degrade SRAM stability most significantly. Thus, we evaluate the VVDD and RSNM improvement in a header structure with R_{Series} when PHD BD happens. Fig. 17(a) shows the relation between VVDD and post-PHD BD resistance with various values of R_{Series}. In region I (post-PHD BD resistance smaller than $2.5 \times 10^5\ \Omega$) where the BD is serious, VVDD voltage level

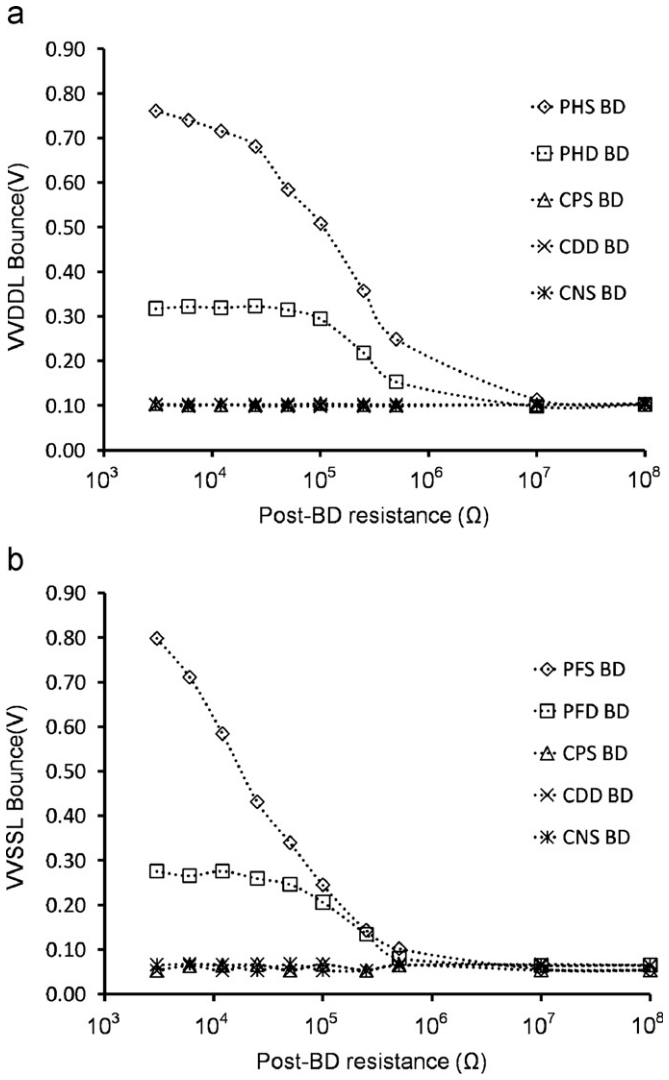

Fig. 15. (a) VVDDL bounce during wake-up vs. post-BD equivalent resistance in a header-gated SRAM, and (b) VVSSL bounce during wake-up vs. post-BD equivalent resistance in a footer-gated SRAM.

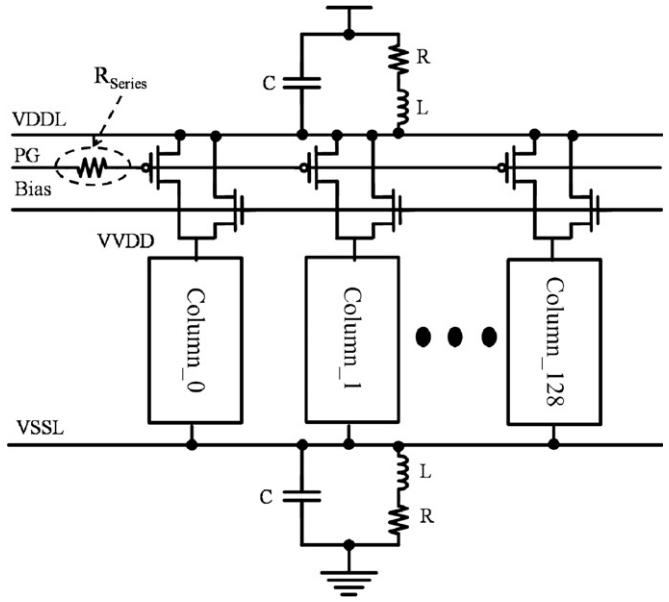

Fig. 16. A header-gated SRAM with R_{Series}.

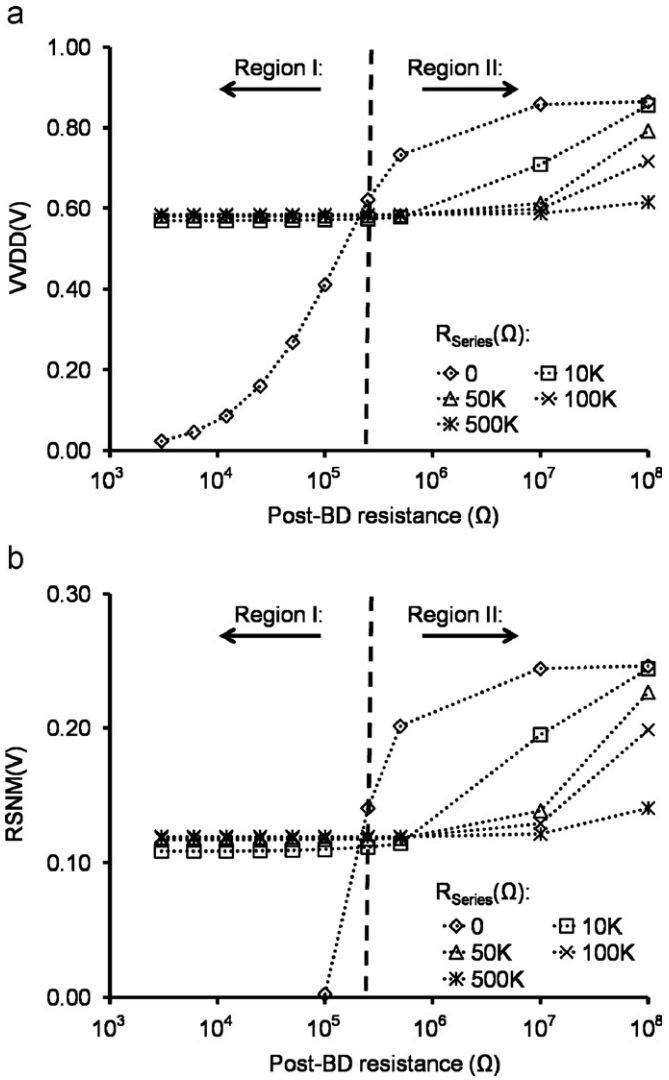

Fig. 17. (a) Active mode VVDD vs. post-BD equivalent resistance, (b) RSNM vs. post-BD equivalent resistance in a header-gated SRAM with R_{Series} when PHD BD happens.

degradation is mitigated using R_{Series} since it limits the post-BD leakage current. However, in region II, where the BD is mild to moderate, the VVDD voltage level decreases with R_{Series} inserted due to smaller V_{GS} of the power switch caused by the voltage drop across R_{Series} (the gate voltage of power switch increases as the post-BD leakage current flows through R_{Series}). Thus, the power-switch has less V_{GS} drive, and the voltage level of VVDD degrades. Fig. 17(b) shows the relations between RSNM and post-PHD BD resistance with various R_{Series}. Clearly, RSNM improves in region I but degrades in region II since RSNM depends on VVDD voltage level. To limit the gate-oxide BD leakage current of a footer power switch a series resistance can also be used in a footer-gated structure. The impact on a footer-gated SRAM would be similar to a header-gated SRAM.

Another way to mitigate the impacts of power-switch gate-oxide BD is to prevent gate-oxide BD from happening. Time to breakdown (T_{BD}) of gate-oxide can be formulated by anode hole injection model [15] or thermochemical model [16]. Conventionally, anode hole injection model is used to describe gate-oxide T_{BD} at high gate-oxide field (E_{OX}), and thermochemical model is used to describe gate-oxide T_{BD} at low E_{OX}. Because E_{OX} of our analysis model is in the low field region [17], we use thermochemical model to calculate T_{BD} in our analysis. In thermochemical model,

T_{BD} can be formulated as:

$$T_{BD} \propto \exp(\Delta H_0/k_B T - \gamma E_{OX}) \quad (1)$$

where ΔH_0 is the observed activation energy of the bond breakage, k_B is Botzmann constant, T is temperature, and γ is the field acceleration factor. According to Eq. (1), the T_{BD} ratio with different E_{OX} can be formulated as

$$T_{BD2}/T_{BD1} = \exp[r(E_{OX1} - E_{OX2})] \quad (2)$$

Moreover, the difference of E_{OX} with different gate-oxide thickness (T_{OX}) and threshold voltage (V_{TH}) can respectively be formulated as

$$E_{OX1} - E_{OX2} = \frac{V_{GS} - V_{TH}}{T_{OX1}}\left(1 - \frac{T_{OX1}}{T_{OX2}}\right) \quad (3)$$

$$E_{OX1} - E_{OX2} = \frac{V_{TH1}}{T_{OX}}\left(\frac{V_{TH2}}{V_{TH1}} - 1\right) \quad (4)$$

Therefore, according to Eq. (2), Eq. (3), and Eq. (4), the T_{BD} ratio with different T_{OX} and V_{TH} of NMOS and PMOS can respectively be formulated as

$$T_{BD2}/T_{BD1} = \exp\left[\frac{V_{GS} - V_{TH}}{T_{OX1}}\left(1 - \frac{T_{OX1}}{T_{OX2}}\right)\right] \quad (5)$$

$$T_{BD2}/T_{BD1} = \exp\left[\frac{V_{TH1}}{T_{OX}}\left(\frac{V_{TH2}}{V_{TH1}} - 1\right)\right] \quad (6)$$

Fig. 18(a) and (b) show the T_{BD} ratio with different T_{OX} and V_{TH} of NMOS and PMOS, respectively, according to Eq. (5) and Eq. (6). Clearly, devices with higher V_{TH} or thicker T_{OX} have longer gate-oxide T_{BD}. Furthermore, thicker T_{OX} devices offer better T_{BD} improvement than higher V_{TH} devices. Therefore, we can use higher V_{TH} devices or thicker T_{OX} devices to extend the lifetime of power-switches and prevent gate-oxide BD from happening.

Although thick T_{OX} power-switch and/or high-V_{TH} power-switches can prevent gate-oxide BD from happening, they degrade the wake-up transition performance. To resolve this problem, we propose a dual-T_{OX} power-gating structure and a dual-V_{TH} power-gating structure. In these structures, power-switches with thick T_{OX} (or high-V_{TH}) provide sufficient currents to SRAM array during active mode, while power-switches with regular T_{OX} (or regular-V_{TH}) are used to maintain SRAM wake-up performance. Moreover, according to Fig. 18, T_{BD} improvement using high-V_{TH} is limited, and it may not be practical to increase V_{TH} to two times of regular V_{TH}. Hence, we focus on analyzing the dual-T_{OX} power-gating structure. An example of our proposed dual-T_{OX} header-gating structure is shown in Fig. 19(a). In this structure, M1 is a thin gate-oxide device, and M2 is a thick gate-oxide device. Both of them serve as power-switches During sleep/standby mode, both M1 and M2 turn off. During wake-up transition, M1 turns on first to charge up VVDD quickly, while M2 remains off. After VVDD voltage level approaches VDD, M1 turns off, and M2 turns on as SRAM enters active mode. In active mode, M1 keeps off but M2 turns on to provide sufficient current to the SRAM cell array. Fig. 19(b) shows the timing control circuit for signals PPG and PG of M1 and M2. By using inverter chain delay, PG becomes "logic 0" right after PPG becomes "logic 1". Fig. 20 compares pertinent waveforms and wake-up currents during wake-up transition for three different header-gating structures: namely, single regular gate-oxide power-switch, single thick gate-oxide power-switch, and dual gate-oxide power-switch. In our analysis, the thick T_{OX} (T_{OX2}) is twice the thickness of the regular T_{OX} (T_{OX1}). We also maintain the same W/L ratio for both power-switch devices in the dual-T_{OX} power-gating structure as that of the power-switch device in the single T_{OX} power-gating structure. As can be seen in Fig. 20, the wake-up transition performance of dual-T_{OX} power-gating structure is comparable to the case with single regular T_{OX} power switch. In contrast, the use of single thick T_{OX} header significantly degrades the wake-up transition performance. Finally, according to Fig. 18, T_{BD} of dual-T_{OX} header can be extended to about 51 times the T_{BD} of the single regular T_{OX} case. Notice that thick T_{OX} device typically requires longer channel length in order to contain the short channel effect. The area overhead of the proposed dual-T_{OX} power-gating structure depends on the organization and partition of the overall power-gating structure, and should not be significant compared with the SRAM cell array area.

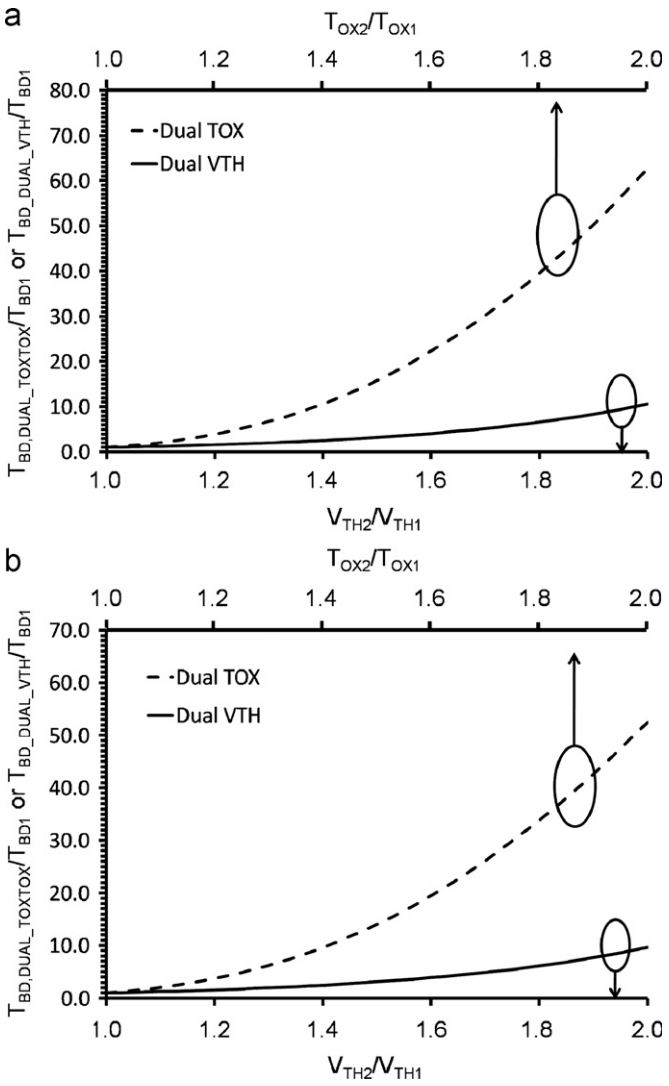

Fig. 18. (a) T_{BD} ratio of NMOS devices with different V_{TH} or T_{OX} ratio, and (b) T_{BD} ratio of PMOS devices with different V_{TH} or T_{OX} ratio.

Notice that in a power-gated SRAM, during standby/sleep mode, the power switch turns off and the voltage level of virtual supply (ground) rail is maintained by the clamping devices. The voltage across SRAM cell decreases, so is the E_{OX} of each and every cell transistors. Consequently, by properly controlling power-gating structures, the lifetime of SRAM cells can be extended. Moreover, circuit techniques such as boosted WL should be avoided to mitigate the gate-oxide BD of pass-gate transistors. Instead, suppressed WL with transient negative bit-line (NBL) [18,19] can be used to maintain RSNM, write margin, and write performance while minimizing the pass-transistor gate-oxide stress. To recover the Read performance,

transient negative cell virtual ground voltage can be used. The durations of transient negative bit-line and virtual ground are shorter than the WL pulse width. Thus, the stress time of pass-gate transistors and cell transistors also decreases and the lifetime increase. The WL pulse width can be optimized by using read/write tracing circuit to avoid excessively long WL pulse width. Increasing T_{OX} of cell transistors is another way to extend the lifetime of SRAM cells. Based on Fig. 18(a) and (b), T_{BD} can be extended significantly even if T_{OX} only slightly increases. Although thick T_{OX} devices would require longer gate length, increasing T_{OX} only slightly may contain the increase in gate length, and hence increase in cell area. Finally, hierarchical BL structure with short local BL can mitigate the impacts of the breakdown cells on other healthy cells. If a WL-to-BL BD (in pass-gate transistor) happens, only those cells sharing the same local BL pair are affected.

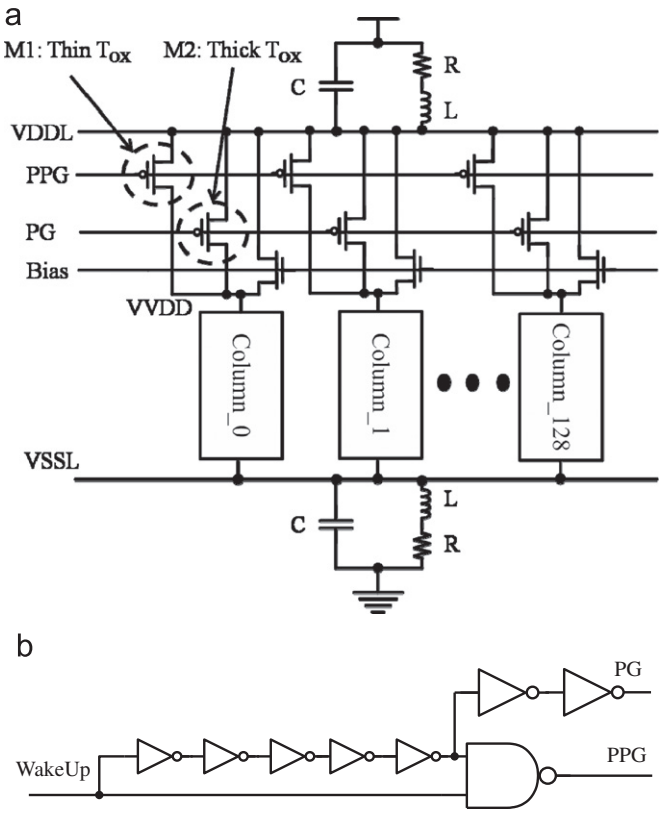

Fig. 19. (a) A dual-T_{OX} header power-gating structure, and (b) timing control circuit for signals PPG and PG of M1 and M2.

6. Conclusions

We have investigated impacts of various gate-oxide BD paths on column-based header- and footer-gated SRAMs based on BSIM PTM 32 nm node model. It was shown that with gate-oxide BD, the RSNM and WM degraded in general. Pass-transistor gate-oxide BD between WL and BL was shown to degrade read/write margin and performance, and to affect other healthy cells along the same column as well. Although stored data could be flipped, CPS BD and CNS BD in individual cell will not affect other cells. In contrast, if there were multiple cells suffering CDD BD in a column, their impacts on VVDD (header structure) or VVSS (footer structure) would be cumulative and would affect other cells in the same column. Most important, we showed that gate-oxide breakdown of the power-switch caused significant degradation of the cell stability and margin, and could lead to the detrimental collapse of the active mode supply across the array (thus affecting the functionality of the entire array), large increase of sleep/standby voltage across the array (thus increasing the leakage and negating the effectiveness of power-gating), and unacceptable virtual supply bounce during wake-up. We evaluated several techniques to mitigate the power-switch gate-oxide BD, and proposed dual-T_{OX} power-gating structure to extend the lifetime of power-switches. We also discussed some design techniques to alleviate gate-oxide BD in SRAM cells and their impacts.

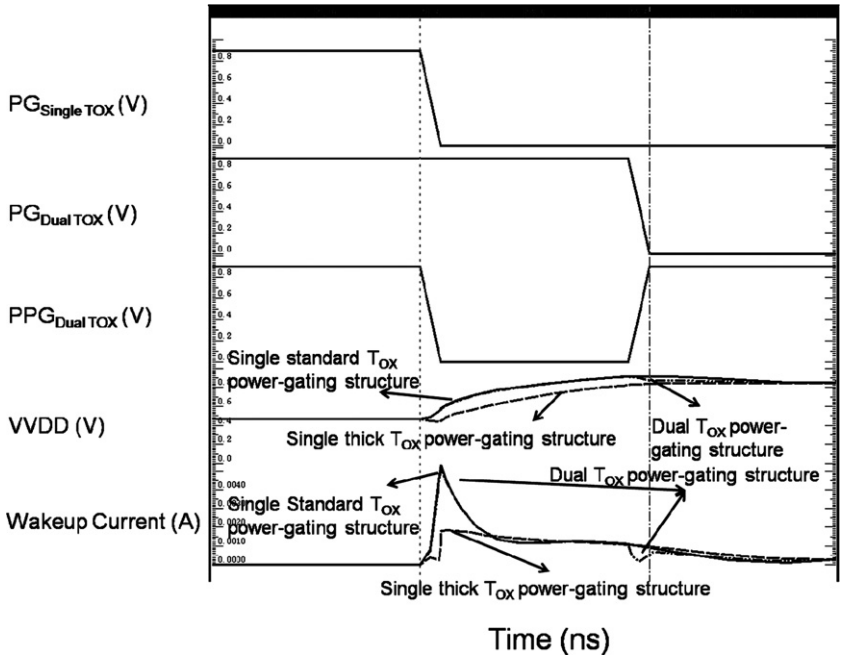

Fig. 20. Pertinent waveforms and wake-up currents during wake-up transition for three different header-gating structures: single regular gate-oxide power-switch, single thick gate-oxide power-switch, and dual gate-oxide power-switch.

Acknowledgement

This work was supported by the National Science Council of Taiwan, under Contract NSC 98-2221-E-009 -112, and Ministry of Economic Affairs, R.O.C., under the Project MOEA 98-EC-17-A-01-S1-124.

References

[1] F. Hamzaoglu, et al., A 3.8 GHz 153 Mb SRAM design with dynamic stability enhancement and leakage reduction in 45 nm High-k metal gate CMOS technology, IEEE Journal of Solid-State Circuits 44 (1) (2009) 148–154.
[2] Y. Wang, et al., A 1.1 GHz 12 μA/Mb-leakage SRAM design in 65 nm ultra-low-power CMOS technology with integrated leakage reduction for mobile applications, IEEE Journal of Solid-State Circuits 43 (1) (2008) 172–179.
[3] M. Sharifkhani, et al., Segmented virtual ground architecture for low-power embedded SRAM, IEEE Transactions on Very Large Scale Integration Systems 15 (2) (2007) 196–205.
[4] Y. Wang, U. Bhattacharya, F. Hamzaoglu, P. Kolar, Y. Ng, L. Wei, et al., A 4.0 GHz 291 Mb voltage-scalable SRAM design in 32 nm high-κ metal-gate CMOS with integrated power management, IEEE International Solid-State Circuits Conference (2009) 456–458.
[5] R. Rodríguez, et al., The impact of gate-oxide breakdown on SRAM stability, IEEE Electron Device Letter 23 (9) (2002) 559–561.
[6] R. Rodriguez, et al., Oxide breakdown model and its impact on SRAM cell functionality, Simulation of Semiconductor Processes and Devices (2003) 283–286.
[7] Ben Kaczer, et al., Consistent model for short-channel nMOSFET after hard gate oxide breakdown, IEEE Transactions on Electron Devices 49 (3) (2002) 507–513.
[8] Ben Kaczer, et al., Impact of MOSFET gate oxide breakdown on digital circuit operation and reliability, IEEE Transactions on Electron Devices 49 (3) (2002) 500–506.
[9] G. Cellere, et al., Micro breakdown in Small-Area Ultrathin Gate Oxides, IEEE Transactions on Electron Devices 49 (8) (2002) 1367–1374.
[10] K. Cheung, et al., Plasma charging damage in deep-submicron CMOS technology and beyond, IEEE International Solid-State and Integrated Circuit Technology Conference (2001) 315–320.
[11] Y.-H. Lee, et al., Prediction of Logic Product Failure due to Thin-Gate Oxide Breakdown, in: Proceedings of the IEEE International Reliability Physics Symposium, 2006, pp. 18–28.
[12] S. Kim, et al., Understanding and minimizing ground bounce during mode transition of power gating structures, Low Power Electronics and Design (2003) 22–25.
[13] 〈http://www.eas.asu.edu/∼ptm/〉.
[14] H.-I. Yang, et al., Impacts of NBTI/PBTI and contact resistance on power-gated SRAM with high-k metal-gate devices, IEEE Transactions on Very Large Scale Integration Systems, in press, doi:10.1109/TVLSI.2010.2049038.
[15] Y.-C. Yeo, et al., MOSFET gate oxide reliability: anode hole injection model and its applications, International Journal of High Speed Electronics and Systems 11 (3) (2001) 849–886.
[16] J.W. McPherson, H.C. Mogul, Underlying physics of the thermochemical E model in describing low-field time-depandent breakdown in SiO_2 thin films, Journal of Applied Physics 84 (3) (1998) 1513–1523.
[17] M.A. Alam, et al., Field acceleration for oxide breakdown—can an accurate anode hole injection model resolve the E vs. $1/E$ controversy? in: Proceedings of the IEEE International Reliability Physics Symposium, 2000, pp. 21–26.
[18] M. Yabuuchi, et al., A 45 nm 0.6 V cross-point 8T SRAM with negative biased read/write assist, in: Proceedings of the IEEE Symposium on VLSI Circuits, 2009, pp. 158–159.
[19] S. Mukhopadhyay, et al., Capacitive coupling based transient negative bit-line voltage (Tran-NBL) scheme for improving write-ability of SRAM design in nanometer technologies, in: Proceedings of the IEEE International Symposium on Circuits and Systems, 2008, pp. 384–387.