Built-in self-test for bias temperature instability, hot-carrier injection, and gate oxide breakdown in embedded DRAMs

Dae-Hyun Kim*, Soonyoung Cha, Linda S. Milor

School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

Abstract

As CMOS technology scales, frontend wearout mechanisms, such as bias temperature instability, hot-carrier injection, and gate oxide breakdown significantly degrade transistor performance. We investigate the impact of these frontend wearout mechanisms on embedded DRAM cells, which are used for last-level caches owing to their high-density characteristics. Our results show that a cell transistor of eDRAM is more susceptible to gate oxide breakdown than bias temperature instability and hot-carrier injection. The impact of all such wearout mechanisms on a cell capacitor is negligible. Based on observations from our simulations, which estimate the performance degradation of eDRAMs resulting from frontend wearout mechanisms, we propose monitoring methods to proactively detect the failures of eDRAMs caused by frontend wearout mechanisms.

1. Introduction

For better performance with less cache misses, the last-level cache (LLC) requires high capacity. As an LLC, embedded dynamic random access memory (eDRAM) is attractive because it has much higher density than static random access memory (SRAM), widely used as cache memories. The small cell area causes eDRAM to be more vulnerable to reliability issues. Scaling the feature sizes of eDRAM, the path to higher density also raises further reliability concerns.

Among the wearout mechanisms, frontend wearout resulting from bias temperature instability (BTI), hot-carrier injection (HCI), and gate oxide breakdown (GOBD) degrades the performances of transistors [1–3]. BTI and HCI increase the threshold voltages ($V_{th}$) of MOSFET devices. Such increases in $V_{th}$ reduce the operating currents of devices and cause performance degradation of circuits. Meanwhile, GOBD creates paths from the gate of a transistor to either its drain or source, also causing malfunctions and degradation of circuit performances.

In this paper, we analyze the impact of frontend wearout mechanisms on eDRAM. To investigate the impact of BTI and HCI on eDRAM cells, we simulate eDRAM operations with the increased $V_{th}$ of cell transistors using SPICE. To model GOBD in circuit simulation, we introduce a gate-level model involving resistors that connect a wordline (WL) to either a bitline (BL) or a storage node (SN), whose resistance decreases as the device ages. Based on observations from the analysis, we propose a proactive test algorithm for eDRAM that can be implemented as a built-in self-test (BIST) circuit that monitors the degree of degradation resulting from frontend wearout in eDRAM. Using our test algorithm, we not only monitor the degradation but also distinguish GOBD between a gate and a storage node from one between a gate and a bitline. The contributions of the paper are the following: (1) we investigate the frontend reliability issues in eDRAM; (2) we develop a method for detecting aging errors proactively by monitoring eDRAM in the field; and (3) we propose a method of identifying GOBD errors in various sites for feedback to circuit designers or device fabrication engineers.

2. The impact of frontend wearouts on eDRAM

We utilize the IBM 90 nm process design kit (PDK) in our simulations and exploit the structures and operations of eDRAM based on prior publications from IBM [4,5]. As depicted in Fig. 1(a), eDRAM has a basic cell structure with one transistor and one capacitor (1T-1C). Since BTI and HCI cause shifts in the threshold voltage ($V_{th}$) of a transistor, we can model the effects of BTI and HCI in eDRAM by changing the $V_{th}$ of a cell transistor in the netlist for SPICE simulations. To model GOBD, we place a 15 Ω resistor ($R_{res}$) between a gate and a wordline to model the gate poly-resistance and a 10$^7$ Ω–10$^8$ Ω resistor between a gate and either a bitline ($R_{bgd}$) or a storage node ($R_{res}$) to model paths resulting from GOBD. Illustrated in Fig. 1(b), Fig. 1(c) shows the architecture of an eDRAM sub-array including the three-transistor micro-sense amplifier.

2.1. Impact of BTI, HCI, and GOBD on a cell transistor

Fig. 2 depicts the effect of the gate voltage on the threshold voltage variation caused by BTI. Since we exploit the charge trapping and
cycle is 0.2. Based on prior work[9], we assume that the ratio of stress time, time gets stressed from BTI during the active mode of eDRAM operations.

Because a cell transistor in this research is an NMOS transistor, as shown in Fig. 1, we assume that positive bias temperature instability (PBTI) affects the cell transistor. Since the positive bias on a gate of an NMOS cell transistor as a single-ended sense amplifier (sSA) [4].

Before an NMOS transistor is the stress condition under PBTI, a cell transistor [7,8].

Because a cell transistor in this research is an NMOS transistor, as shown in Fig. 1, we assume that the stress condition under PBTI, a cell transistor gets stressed from BTI during the active mode of eDRAM operations. Based on prior work [9], we assume that the ratio of stress time, time spent in active mode, to the total operating time is 20%, or the duty cycle is 0.2.

To obtain a sense of how much $V_{th}$ varies in a microprocessor with our assumptions, we exploit simulations from prior work [10] with various workloads, whose experimental setup and simulation methodology are also based on the charge trapping and detrapping model for BTI. Fig. 3 shows variations of $V_{th}$ resulting from BTI in a last-level cache as a function of stress time. Using the $V_{th}$ distribution, we conduct a Monte Carlo SPICE simulation of eDRAM circuits. Fig. 4 illustrates basic write and read operations in an eDRAM cell. During write operations, RBL and WBL turn on the appropriate transistors in the read head, so that LBL is driven to either ‘1’ or ‘0’. When WL is turned on, the state on the LBL is transferred to the storage node. During read operations, RBL and WBL disconnect LBL from the read head. The state of the storage node is transferred to LBL. If the result is ‘1’, after LBL is charged up partially, the read head turns on and helps to pull up LBL to ‘1’.

Fig. 5 shows the impact of BTI on a cell transistor with Monte Carlo simulations. The results demonstrate that since the increased $V_{th}$ of a cell transistor degrades its on-current by 9.1%,1 BTI leads to variation in the storage node cell voltage of eDRAM cells during write ‘1’ operations with a mean shift of 48.1 mV (6.1%) and a standard variation of 15.4 mV. The lower written voltage level of the storage node also lowers the local bitline voltage level during the read ‘1’ operation. However, no more degradation is found after bitline sense amplification. After all, an NMOS read-head transistor as a single-ended sense amplifier has enough sensing margin to tolerate such degradation caused by BTI because of the high voltage swing on the local bitline owing to the high transfer ratio of 84%.2

The $V_{th}$ degradation of a transistor resulting from HCI depends on the stress time ($t$) as follows [11]:

$$\Delta V_{th} = A \cdot t^\alpha. \quad (5)$$

$$A = \exp(-\alpha/V_{DS}). \quad (6)$$

where $A$ and $\alpha$ are the fitting constants and $V_{DS}$ is the drain voltage. From the equations, we know that the shift of the $V_{th}$ resulting from HCI becomes more severe as the drain voltage of a transistor increases.

1 If the voltage between a gate and a source is 1.7 V and $V_{th}$ is 450 mV, the shift in $V_{th}$ from 0 V to approximately 5.8 mV (at 8 years stress) makes only a 9.1% decrease in drive current based on the saturation current equation of a MOSFET:

$$\Delta I_{th}/I_{th,0} = 1 - (V_{DS} - V_{th})/(V_{DS} - V_{th,0}) \overset{2}{=} 0.91.$$  

2 The transfer ratio is

$$C_s/(C_s + C_{th})$$  

where $C_s$ (18 fF) and $C_{th}$ (3.5 fF) are the capacitance of a storage node and a local bitline, respectively [4].

---

Fig. 2. The dependence of the threshold voltage shift on duty cycle, temperature, and gate voltage [5].

---

Fig. 3. Simulated $V_{th}$ variation caused by BTI, which is generated based on the simulation methodology described in [10].
The probability of the number of defects generated in the percolation model, which number of conduction paths. The number of conduction paths is a function of time under stress, hard breakdown occurs. We determine the stress, temperature, and time. We plot the probability of different values of GOBD resistance as a function of time in

To understand the effect of GOBD on a cell transistor of eDRAM, we model paths formed by GOBD in the gate oxide of a transistor as a resistor whose value decreases as soft-breakdown events occur. For our simulations, we exploit the percolation model (PM) illustrated in Fig. 7 [12]. Using the PM model, the gate is partitioned into a grid and we statistically generate defects in the dielectric based on the stress conditions, especially the gate voltage. A vertical sequence of defects, a conduction path, forms a leakage path from a gate to either a source or a drain. Once a conduction path is formed, which is referred to as soft breakdown, the resistance between a gate and either a source or a drain reduces. As the number of soft-breakdown events increases as a function of time under stress, hard breakdown occurs. We define hard breakdown when a device no longer functions as a device or when the circuit violates performance specifications [13]. We exploit the number of conduction paths to calculate the resistance of both soft-breakdown and hard-breakdown using the quantum point contact (QPC) model [14].

The resistance through the gate oxide is estimated by counting the number of conduction paths. The number of conduction paths is a function of the number of defects generated in the percolation model, which is determined by the stress, temperature, and time. We plot the probability of different values of GOBD resistance as a function of time in

Fig. 6 shows the drain voltage of a cell transistor during basic write and read operations. For eDRAM, a write operation of the same data as written in the cell and the write/read operations of data ‘0’ have a zero drain voltage, resulting in no HCI concerns. Hence, only write operations with a data transition from low to high are of interest. However, even during a write operation with a low-to-high data transition, less than 0.6 V is applied to the drain voltage for 250 ps. Since this is too small to generate hot carriers, HCI in a cell transistor is negligible.

To understand the effect of GOBD on a cell transistor of eDRAM, we model paths formed by GOBD in the gate oxide of a transistor as a resistor whose value decreases as soft-breakdown events occur. For our simulations, we exploit the percolation model (PM) illustrated in Fig. 7 [12]. Using the PM model, the gate is partitioned into a grid and we statistically generate defects in the dielectric based on the stress conditions, especially the gate voltage. A vertical sequence of defects, a conduction path, forms a leakage path from a gate to either a source or a drain. Once a conduction path is formed, which is referred to as soft breakdown, the resistance between a gate and either a source or a drain reduces. As the number of soft-breakdown events increases as a function of time under stress, hard breakdown occurs. We define hard breakdown when a device no longer functions as a device or when the circuit violates performance specifications [13]. We exploit the number of conduction paths to calculate the resistance of both soft-breakdown and hard-breakdown using the quantum point contact (QPC) model [14].

The resistance through the gate oxide is estimated by counting the number of conduction paths. The number of conduction paths is a function of the number of defects generated in the percolation model, which is determined by the stress, temperature, and time. We plot the probability of different values of GOBD resistance as a function of time in

Fig. 6. Simulated waveform of the drain voltage (V_DS) during eDRAM basic operations.
less than that of a cell transistor, which experiences 0.9 V with data and the boosted wordline voltage of 1.7 V with data hot carriers. Therefore, we ignore HCI in the cell capacitor. Since degrading by BTI depends on bias, the impact of BTI on a cell capacitor is much less severe than on a cell transistor. Hence, we neglect the impact of BTI on a cell capacitor.

HCl occurs when a hot carrier is formed by the bias between a source and a drain, causing it to penetrate the oxide and fill traps. However, because no drain voltage is applied, the cell capacitor does not suffer from hot carriers. Therefore, we ignore HCI in the cell capacitor.

in the gate voltage results in more variation in the threshold voltage. Note that the voltage between a cell storage node and a cell plate node is about 0.8 V with data ‘1’ and zero with data ‘0’, which is much less than that of a cell transistor, which experiences 0.9 V with data ‘1’ and the boosted wordline voltage of 1.7 V with data ‘0’. Since degradation by BTI depends on bias, the impact of BTI on a cell capacitor is much less severe than on a cell transistor. Hence, we neglect the impact of BTI on a cell capacitor.

For the calculation of lifetime degradation using Eqs. (7) and (8), we employ the dimensions of a cell transistor and a cell capacitor for the 65 nm node are summarized in Table 1 as a case study. The area (WL) of the cell transistor is 0.0144 μm² [16] and that of the cell capacitor is 0.205 μm² [17]. Because of GOBD, the lifetime of a cell capacitor is expected to degrade 5.05 times as fast as that of a cell transistor. In addition, since a cell capacitor is stressed during almost all of the execution time and a cell transistor is stressed during write, read, and refresh operations, a cell capacitor degrades four times as fast as a cell transistor, assuming that the portion of the operation time out of total runtime is 20%. However, note that the gate voltages of 1.7 V for data ‘0’ and 0.9 V for data ‘1’ are applied to a cell transistor, and the bias of 0 V for data ‘0’ and 0.8 V for data ‘1’ are applied to a cell capacitor. Assuming that the signal probability of having data ‘1’ is 35% in an LLC [9], the reduction in the lifetime of a cell capacitor resulting from GOBD is 0.0035 times less than that of a cell transistor. All in all, the lifetime degradation resulting from GOBD in a cell capacitor is only 7.15% of that in a cell transistor. Since a cell capacitor may have 14 times as long a lifetime as a cell transistor, we neglect the impact of GOBD on a cell capacitor in this work.

To estimate lifetime degradation resulting from GOBD, we exploit equations from [15] and [16]:

\[ t_f = \left( \frac{1}{W} \right)^{\frac{1}{\beta}} \tag{7} \]

\[ t_f = \exp\left( -\gamma \times V_{gs} \right) \tag{8} \]

where \( t_f \) is the time-to-failure at the 63% cumulative percentile of failure, \( W \) and \( L \) are the width and the length of a device, respectively, \( \beta \) is the Weibull shape parameter (1.64), \( \gamma \) is the voltage acceleration factor (e.g., 5.6 at 125 °C), and \( V_{gs} \) is the gate voltage.

For the calculation of lifetime degradation using Eqs. (7) and (8), we employ the dimensions of a cell transistor and a cell capacitor for the 65 nm node summarized in Table 1 as a case study. The area (WL) of the cell transistor is 0.0144 μm² [16] and that of the cell capacitor is 0.205 μm² [17]. Because of GOBD, the lifetime of a cell capacitor is expected to degrade 5.05 times as fast as that of a cell transistor. In addition, since a cell capacitor is stressed during almost all of the execution time and a cell transistor is stressed during write, read, and refresh operations, a cell capacitor degrades four times as fast as a cell transistor, assuming that the portion of the operation time out of total runtime is 20%. However, note that the gate voltages of 1.7 V for data ‘0’ and 0.9 V for data ‘1’ are applied to a cell transistor, and the bias of 0 V for data ‘0’ and 0.8 V for data ‘1’ are applied to a cell capacitor. Assuming that the signal probability of having data ‘1’ is 35% in an LLC [9], the reduction in the lifetime of a cell capacitor resulting from GOBD is 0.0035 times less than that of a cell transistor. All in all, the lifetime degradation resulting from GOBD in a cell capacitor is only 7.15% of that in a cell transistor. Since a cell capacitor may have 14 times as long a lifetime as a cell transistor, we neglect the impact of GOBD on a cell capacitor in this work.

3. Detection of frontend wearout in eDRAM

To detect GOBD in eDRAM, our analysis demonstrates that the resistance of paths from GOBD should be smaller than 10⁵ Ω. However, note that at least one failure in a 32 MB LLC after two and a half years of stress time is expected with the number of GOBD paths corresponding to 10⁶ Ω resistance, as shown in Fig. 8. To enhance the sensitivity of our GOBD detection scheme, we utilize read operations with a long duration. A read operation as long as 13 ns can detect read ‘0’ failures resulting from a 10⁶ Ω resistance. The signals of the GOBDG-to-BL fault, illustrated in Fig. 10(a), show similar results to those of the GOBDG-to-SN fault, as shown in Fig. 10(b).

Because of leakage from the storage node with a high voltage level to a wordline precharged to a negative bias during the standby mode, a GOBDG-to-SN fault, unlike a GOBDG-to-BL fault, causes read ‘1’ to fail with a long interval between write ‘1’ and read ‘1’ operations, as shown in Fig. 11. Therefore, we can distinguish GOBDG-to-BL from GOBDG-to-SN using intervals between write ‘1’ and read ‘1’ operations.

Because current on a wordline (Iwp) during the write ‘0’ operation increases as the resistance of GOBD decreases (see Fig. 12), we can monitor GOBD with Iwp. Current flows from a gate (connected to an activated wordline) to either a bitline or a storage node written as ‘0’ through the GOBD resistor. This current increases as the resistance decreases. In the two GOBD cases, when the resistance is under 10⁵ Ω, the difference in

Please cite this article as: D.-H. Kim, et al., Built-in self-test for bias temperature instability, hot-carrier injection, and gate oxide breakdown in embedded DRAMs, Microelectronics Reliability (2015), http://dx.doi.org/10.1016/j.microrel.2015.06.077
the current $I_{pp}$ as a function of resistance results from the limited charge capacity of the cell capacitor for the GOBDG-to-BL case.

4. Reliability test algorithm for eDRAM

To design BIST for detecting failures resulting from frontend wearout, note that the impact of BTI and HCI is negligible and GOBD is dominant, which is detected as summarized in Table 2. Since boosting the voltage level of a wordline enhances the driving current of a cell transistor in eDRAM with a high $V_{gs}$, a bias generator is used for the wordline bias ($V_{pp}$) [18]. By measuring the current, $I_{pp}$, we monitor GOBD in the wordline direction using a current detector added to the generator, whose general implementation is shown in Fig. 13. To detect GOBD in the bitline direction, we can exploit the read ‘0’ operation with different read durations. By increasing the duration of the read ‘0,’ the detectable resistance resulting from GOBD increases. To differentiate GOBDG-to-BL from GOBDG-to-SN, we use write ‘1’ and read ‘1’ operations at various intervals between write and read operations. A larger interval also increases the detectable resistance of GOBDG-to-SN.

Fig. 14 presents the algorithm for testing eDRAMs under the impact of frontend wearout. While writing ‘0’s to eDRAMs, we detect GOBD in a wordline by monitoring $I_{pp}$. If we detect a difference larger than several $\mu A$ in $I_{pp}$, we proceed to read the ‘0’ data to detect the read ‘0’ fails in the wordline to identify the position of the failure in the array. Since only GOBDG-to-SN results in a read ‘1’ failure, testing of the write ‘1’ and read ‘1’ operations with an interval distinguishes GOBDG-to-BL from GOBDG-to-SN in the failed bit. To detect wearout in eDRAMs, the interval for monitoring GOBD can initially be coarse (e.g., once every six months), but fine (e.g., once every month) after a certain amount of time (e.g., one year), based on the expected probability of failure, as suggested in Fig. 8, for a reduction of the total test time.

5. Conclusion

This paper has analyzed the impact of frontend wearout mechanisms, namely as BTI, HCI, and GOBD, on eDRAMs. Since an eDRAM cell consists of one transistor and one capacitor, we have investigated the impact of these mechanisms on both the cell transistor and the cell capacitor.

Our simulation results have demonstrated that degradation of the transistor from BTI is negligible because of the robust micro-sense amplifier with a high transfer ratio. The impact of BTI on the cell capacitor is also negligible because the gate bias of the cell capacitor is smaller than that of the cell transistor.

For HCI, since the cell transistor has a low drain voltage during write operations, degradation of the cell transistor resulting from HCI can be ignored. Since no drain voltage is applied to the cell capacitor, the impact of HCI on the capacitor is also negligible.

Table 2

| Test modes and test patterns for GOBD monitoring. |
|-------------------|-----------------|-----------------|
| Mode              | Test point       | Test patterns   |
| WL-direction      | Vpp current      | (w0)            |
| BL-direction      | Dataline P/F     | (w0, r0 with increasing read time) |
| Differentiation   | Dataline P/F     | (w1, increasing interval, r1) |
With GOBD, a current path is formed from a wordline to either a drain or a source. This study has found that the current sensing of \(I_{PP}\) helps monitor and detect GOBD in the wordline direction. With detection schemes for both the bitline and the wordline direction, one can determine the fault location. We also found that a test pattern of write ‘1’ and read ‘1’ with increasing intervals between operations helps to distinguish if the fault comes from GOBD \(G\) to \(S\) or GOBD \(G\) to \(B\), since only GOBD \(G\) to \(S\) results in a read ‘1’ fail.

After investigating the impact of frontend wearout mechanisms on an eDRAM cell, we have proposed a built-in self-test algorithm for detection of frontend wearout mechanisms, to detect GOBD in both the bitline and the wordline directions and to determine if GOBD occurs between the gate and either its drain or source.

Acknowledgement

The authors would like to thank the National Science Foundation under Award #116786 for financial support.

References