

Extension of DRAM Retention Time at 77 Kelvin by Replacing Weak Rows with Large GIDL Current

Ho-Jun Kim, Won-Cheol Lee, and Hong-June Park*

Abstract—At 77 Kelvin, it was found that the weak cells with large gate-induced drain leakage current are quite localized in six weak rows out of the entire 524,288 rows of a 8 Gb DRAM chip. By replacing the six weak rows, the DRAM retention time was increased from 8 to 4096 seconds at 77 Kelvin. Because this retention time is over 64,000 times longer than the JEDEC retention-time specification of 0.064 seconds, the DRAM refresh power can be reduced to a negligible level at the data center. To achieve this, the DRAM controller was modified to find weak rows by using three different data patterns at power-on reset and a reserved row of DRAM is used when access to a weak row address occurs during the normal operation.

Index Terms—DRAM controller, DRAM refresh, retention time, cryogenic memory, cryogenic temperature

I. INTRODUCTION

Dynamic random access memory (DRAM) cell has an access transistor and a small cell capacitor which stores a charge of $+0.5 \cdot C_s \cdot V_{DD}$ or $-0.5 \cdot C_s \cdot V_{DD}$ to represent a data of “1” or “0”, respectively. Because DRAM has a high density compared to other memory due to its simple design, DRAM is the main workhorse to store data at the data center, consuming over 25% of the total power of

the data center [1, 2]. Because a charge stored in the small cell capacitor is reduced significantly by even a small leakage current, all the cells in a DDR3 DRAM chip should go through the refresh operation every 64 ms. This refresh operation occupies around 20% of DRAM power in 16 Gb DRAM chip, and this percentage will reach around 50% as DRAM density increases [3]. Therefore, the extension of the refresh period beyond 64 ms would significantly reduce the DRAM power at the data center.

To extend the refresh period, reducing the leakage current of the DRAM cell is key. The leakage current of the DRAM cell consists of a subthreshold current, a junction leakage current, and a tunneling current such as gate-induced drain leakage (GIDL). It is known that the subthreshold current and the junction leakage current are dominant at room temperature (298 Kelvin) with the activation energy of 0.5 ~ 0.7 eV [4, 5] while the tunneling current is dominant at liquid nitrogen temperature (77 Kelvin) with the activation energy of 0.15 eV [6, 7]. In other words, by lowering the temperature, the leakage current of the DRAM cell can be reduced, and this can increase the retention time which is the longest time to keep the written data correctly without the refresh operation.

In this work, we measured the cumulative number of failure cells according to retention time at 298 Kelvin and 77 Kelvin to probe the temperature dependence of the DRAM retention time. As a result, the DRAM retention time was a few seconds at 77 Kelvin while it was below 0.128 seconds at 298 Kelvin. Besides, the measurements revealed that the weak cells with relatively large leakage current are quite localized in six rows out of the entire 524,288 rows of an 8 Gb DRAM chip at 77

Manuscript received Jan. 11, 2022; reviewed Apr. 20, 2022; accepted Jun. 3, 2022

Department of Electronic and Electrical Engineering, POSTECH (Pohang University of Science and Technology), Pohang 37673, Korea
E-mail : hjpark@postech.ac.kr
Corresponding Author : Hong-June Park

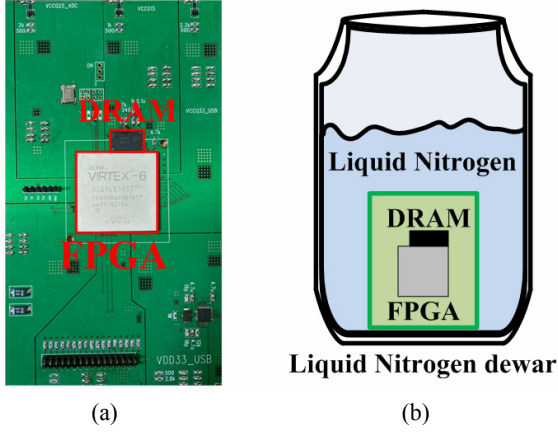


Fig. 1. Measurement setup: (a) DRAM controller board; (b) schematic depiction of liquid nitrogen measurement.

Kelvin; this ratio of 6 out of 524,288 is 11.4 ppm. The weak cells in the six weak rows had the retention time in the range from 8 to 4096 seconds while the retention time of the remaining rows was longer than 4096 seconds. Also, to probe the effect of capacitive coupling at 77 Kelvin, the DRAM retention time was measured using three different data patterns.

This work proposed a modified DRAM controller in Xilinx Virtex-6 FPGA that replaces access to the six weak rows with reserved rows, using the measurement results. The proposed DRAM controller increased the DRAM retention time and reduced the DRAM refresh power to a negligible level.

Section II shows the retention time measurement results according to the temperature of the DDR3 DRAM chip. Section III explains the architecture of the proposed DRAM controller and the weak row replacement method. Section IV and Section V show the measurement results and conclusion.

II. MEASUREMENT OF RETENTION TIME AT ROOM TEMPERATURE AND 77 KELVIN

To probe the effect of temperature to DRAM retention time, the cumulative number of failure cells of an 8 Gb DDR3 DRAM according to retention time was measured using a DRAM controller at 298 Kelvin and 77 Kelvin. The DRAM controller on Xilinx Virtex-6 FPGA and a DDR3 DRAM chip are placed on a printed circuit board (PCB) (Fig. 1(a)). The PCB is immersed in a dewar filled with liquid nitrogen for the measurement at 77 Kelvin

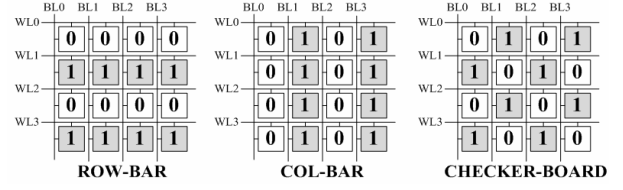


Fig. 2. Data patterns used for retention time measurement.

(Fig. 1(b)).

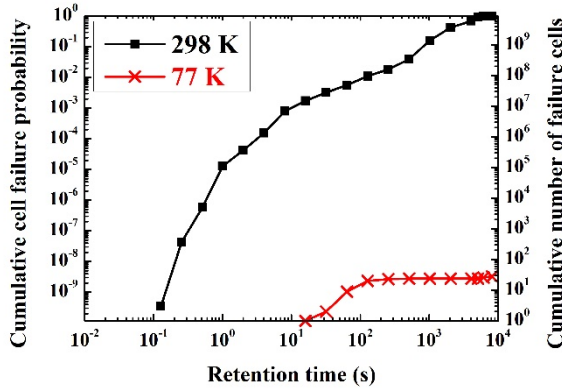
The DRAM controller in FPGA writes 8 Gb data to the DRAM chip, waits during the retention time without refresh, and reads back the 8 Gb data from the DRAM chip to find out failure cells with mismatches between the written and read data. The DRAM chip has parallel 16-bit DQ pins operating at 480 Mbps, and the DRAM controller sends out a burst of 8×16 bits data in two FPGA clocks of 120 MHz. So, it takes around 1.1 seconds to write the entire 8 Gb data ($524,288 \text{ rows} \times 16 \text{ kbit}$) to DRAM.

This measurement was performed for the retention times of 0.128, 0.256, 0.512, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 5120, 6144 and 8192 seconds, respectively. The measurement step, as mentioned above, was repeated for three data patterns to probe the worst-case retention time for different data patterns.; ROW-BAR(row-wise alternate), COL-BAR(column-wise alternate), CHECKER-BOARD(diagonal-wise alternate) (Fig. 2).

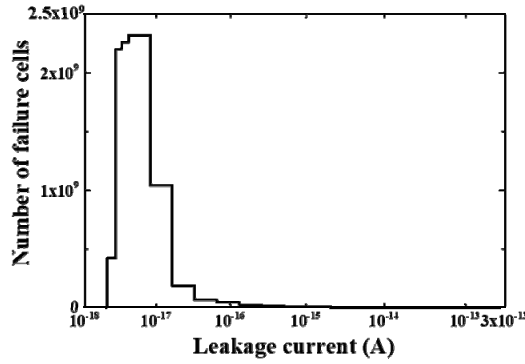
The cumulative number of failure cells according to retention time at 298 Kelvin and 77 Kelvin is shown in Fig. 3. In this measurement, the ROW-BAR data pattern was used. Since the failure cells are detected for 128 ms or longer retention times (Fig. 3(a)), the worst-case leakage current is estimated to be 0.26 pA as in (1); the cell capacitance (C_s) is assumed to be 25 fF, the supply voltage (V_{DD}) is 1.35 V. Also, the cumulative number of failure cells increases continuously with retention time at 298 Kelvin; this indicates the continuous magnitude distribution of leakage current (Fig. 3(b)).

$$\text{Retention time} = \frac{0.5 \times C_s \times V_{DD}}{\text{Leakage current}} \quad (1)$$

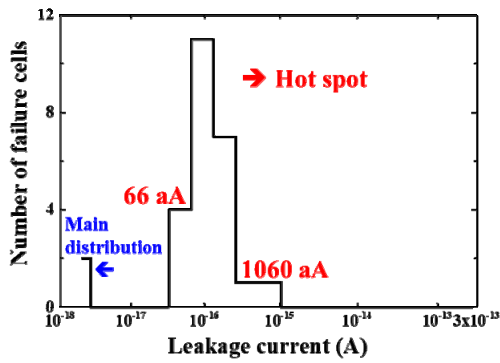
At 77 Kelvin, the cumulative number of failure cells is constant at 24 for retention times in the range from 256 to 5120 seconds; this indicates that only 24 cells have a relatively large leakage current in the range from 66 aA



(a)



(b)



(c)

Fig. 3. Measurement results at 298 Kelvin and 77 Kelvin: (a) Cumulative cell failure probability and cumulative number of failure cells vs. retention time in an 8 Gb DRAM; (b) Leakage current distribution at 298 Kelvin; (c) Leakage current distribution at 77 Kelvin.

to 1060 aA and the remaining 8,589,934,568 cells have a small leakage current less than 2.8 aA at 77 Kelvin (Fig. 3(c)). Therefore, one can observe a small set of hot spot cells with a relatively large leakage current.

The cumulative sum ('+' in Fig. 4) was measured at 77 Kelvin according to the retention time by applying three different data patterns (ROW-BAR, COL-BAR,

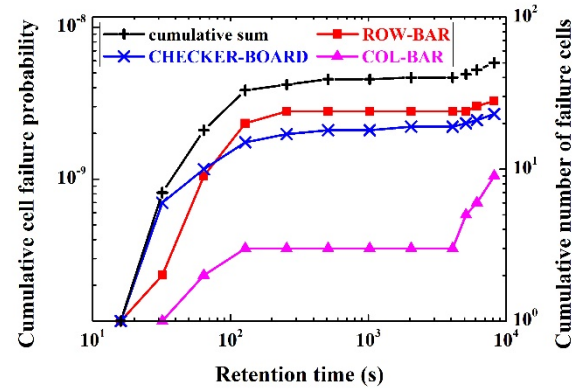


Fig. 4. Cumulative cell failure probability and cumulative number of failure cells with different data patterns at 77 Kelvin.

Table 1. The number of failure cells for each row at different retention times

Retention time (s)	Row address (bank,row)					
	0, 8500	0, 9960	0, 64297	0, 65535	5, 41283	7, 65534
64	1	0	0	0	0	17
256	1	1	1	1	1	31
1024	1	1	1	1	1	33
2048	1	1	1	1	1	34
4096	1	1	1	1	1	34

CHECKER-BOARD); the cells which failed at any of three data patterns were counted as the failure cells. The cumulative sum is almost constant at 39 ~ 40 for the retention time in the range from 512 to 4096 seconds; 40 cells failed at the retention time of 4096 seconds. This indicates that around 40 cells have a leakage current > 33 aA, and the remaining cells have a leakage current < 3.3 aA. The 40 failure cells are quite localized in six weak rows out of the entire 524,288 rows of the 8 Gb DRAM chip (11.4 ppm); 35 cells belong to a row, and the remaining five cells are distributed in five rows with a failure cell per row (Table 1). If these six weak rows are not used, the DRAM refresh time can be increased to 4096 seconds at 77 Kelvin.

III. WEAK ROW REPLACEMENT

A conventional DRAM controller circuit was modified to find out and replace the weak rows that include the cells with a short retention time at 77 Kelvin (Fig. 5). The 8 Gb DDR3 DRAM of this work has 16-bit parallel DQ pins, uses 3-bit bank address and 16-bit row address for row access, and 10-bit column address for column

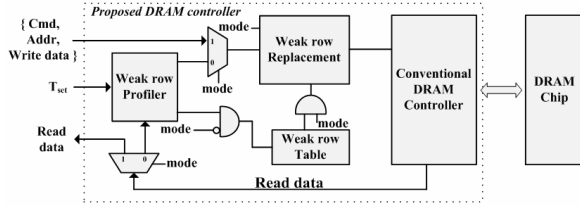


Fig. 5. Proposed DRAM controller architecture.

access. Out of 524,288 rows, 64 rows of the DRAM are reserved for processing any weak rows.

After the power-on reset, the mode signal is set to 0, and the weak row profiler generates one of the three data patterns (ROW-BAR, COL-BAR, CHECKER-BOARD). The data pattern is written into the entire DRAM chip. After waiting for the target minimum retention time (T_{set}), the weak row profiler reads back the DRAM data row by row. Even if the target minimum retention time (T_{set}) is longer than one second, the refresh power is reduced by at least 99.9% and can be a negligible level; T_{set} is set to 256 seconds in this work. If there is any mismatch between the written and read data of the row, the weak row profiler stores the corresponding bank/row address into the weak row table, including a reserved bank/row address for this weak row address. This procedure is repeated for the remaining two data patterns.

During the normal operation with the mode signal set to 1, the weak row replacement block compares the incoming bank/row addresses with those in the weak row table. It sends out a reserved bank/row address corresponding to the weak row address in the weak row table if the incoming bank/row address matches the one in the weak row table. Otherwise, the weak row replacement block sends out the incoming bank/row address to the conventional DRAM controller. It takes one system clock period for the weak row replacement operation. The additional one system clock for the weak row replacement reduces the DRAM bandwidth by around 0.3%, while the increase of the retention time enhances the DRAM bandwidth by around 17%. The DRAM timing parameters at room temperature are used for this DRAM bandwidth calculation. The FPGA implementation of the weak row replacement operation (Fig. 5) requires additional hardware of 2684 slice LUTs (about 16,104 NAND2 gates).

IV. MEASUREMENT RESULTS

By applying the proposed weak row replacement (Fig. 5) to an 8 Gb DDR3 DRAM chip at 77 Kelvin, six weak rows are detected at the target minimum retention time (T_{set}) of 256 seconds. The measured cumulative failure cells and rows with retention time (Fig. 6) demonstrate that the retention time was increased from 8 to 4096 seconds by the weak row replacement.

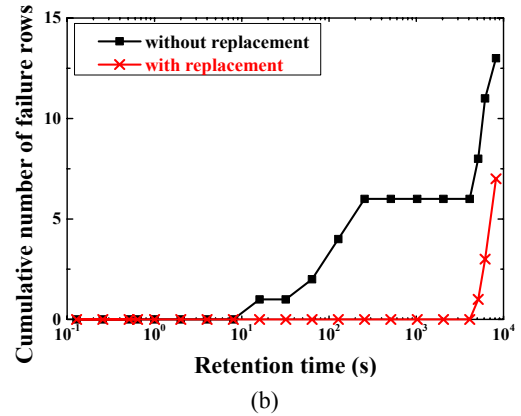
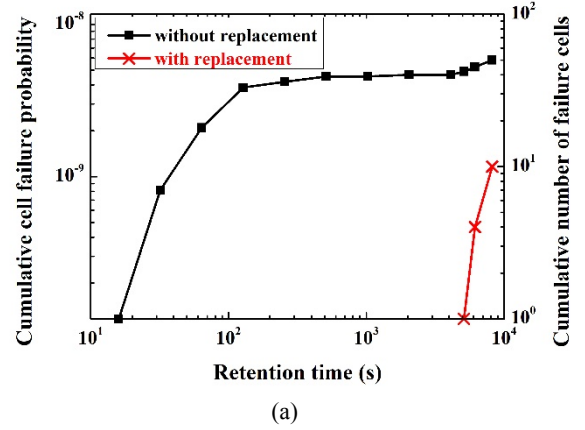


Fig. 6. Comparison of without replacement and with replacement when T_{set} is 256 seconds: (a) cumulative cell failure probability and cumulative number of failure cells; (b) cumulative number of failure rows vs. retention time.

V. CONCLUSIONS

According to the measurement of DRAM retention time at 77 Kelvin, only a few cells have a relatively large leakage current while all other cells have a much smaller leakage current. Based on this measurement, the weak rows are detected and their row addresses are stored in a weak row table. And then, the weak row replacement block replaces the weak row with the reserved row

during the normal operation. The weak row replacement operation was implemented on FPGA with 2684 slice LUTs (about 16,104 NAND2 gates) and had one system clock period. Applying the proposed weak row replacement to an 8 Gb DDR3 DRAM replaced six weak rows with reserved rows and increased the DRAM retention time from 8 to 4096 seconds at 77 Kelvin. This can reduce the DRAM refresh power to a negligible level while the DRAM refresh power accounts for almost 50% of total DRAM power, especially in large density DRAMs.

ACKNOWLEDGMENTS

This work was supported by National Research Foundation (NRF) grant funded by the Korea government (NRF-2019R1A5A1027055) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2022R1A2C2003451)

REFERENCES

- [1] K. T. Malladi, et al, "Towards energy-proportional datacenter memory with mobile DRAM," *2012 39th Annual International Symposium on Computer Architecture (ISCA)*, pp. 37-48, 2012.
- [2] K. Lim, et al, "System-level implications of disaggregated memory," *IEEE International Symposium on High-Performance Comp Architecture*, pp. 1-12, 2012.
- [3] J. Liu, B. Jaiyen, R. Veras and O. Mutlu, "RAIDR: Retention-aware intelligent DRAM refresh," *2012 39th Annual International Symposium on Computer Architecture (ISCA)*, pp. 1-12, 2012.
- [4] T. Hamamoto, S. Sugiura and S. Sawada, "On the retention time distribution of dynamic random access memory (DRAM)," *IEEE Transactions on Electron Devices*, vol. 45, no. 6, pp. 1300-1309, June., 1998.
- [5] F. Wang, T. Vogelsang, B. Haukness and S. C. Magee, "DRAM Retention at Cryogenic Temperatures," *2018 IEEE International Memory Workshop (IMW)*, pp. 1-4, 2018.
- [6] K. Rais, F. Balestra, G. Ghibaudo, "Temperature dependence of gate induced drain leakage current in silicon CMOS devices," *Electronics Letters*, 30, (1), p. 32-34, 1994.
- [7] A. Alnuaimi, A. Nayfeh, and V. Koldyaev, "Electric-field and temperature dependence of the activation energy associated with gate induced drain leakage," *J. Appl. Phys.*, vol. 113, no. 4, Jan., 2013



Ho-Jun Kim received the B.S. degree in Electronic and Electrical Engineering from Hongik University, Seoul, Korea, in 2019 and the M.S. degree in Electronic and Electrical Engineering from the Pohang University of Science and Technology (POSTECH), Pohang, Korea, in 2021. Currently, he is pursuing the Ph.D. degree at POSTECH. His research interests include DRAM controller.



Won-Cheol Lee received the B.S. degree in Electronic and Electrical Engineering from Pohang University of Science and Technology (POSTECH), Pohang, Korea, in 2015. Currently, he is pursuing the M.S and Ph.D. degree at POSTECH. His research interests include DRAM controller and hardware accelerators.



Hong-June Park (Senior Member, IEEE) received the B.S. degree from the Department of Electronic Engineering, Seoul National University, Seoul, Korea, the M.S. degree from the Korea Advanced Institute of Science and Technology (KAIST), Taejon, Korea, and the Ph.D. degree from the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, USA, in 1979, 1981, and 1989, respectively. He was a CAD Engineer with ETRI, Korea, from 1981 to 1984 and a Senior Engineer in the TCAD Department of Intel Corporation from 1989 to 1991. In 1991, he joined the

Faculty of Electronic and Electrical Engineering, Pohang University of Science and Technology (POSTECH), Pohang, Korea, where he is currently a Professor. His research interests include CMOS analog circuit design such as high-speed interface circuits, ROIC of touch sensors, and analog/digital beamformer circuits for ultrasound medical imaging. Dr. Park is a member of IEEK. He served as the Editor-in-Chief of the *Journal of Semiconductor Technology and Science*, an SCIE journal (<http://www.jsts.org>), from 2009 to 2012, as the Vice President of IEEK in 2012, and as a technical program committee member of ISSCC, SOVC, and A-SSCC for several years.