# A Write-Back-Free 2T1D Embedded DRAM With Local Voltage Sensing and a Dual-Row-Access Low Power Mode

Wei Zhang, Ki Chul Chun, and Chris H. Kim, Senior Member, IEEE

*Abstract*—A gain cell embedded DRAM (eDRAM) in a 65 nm LP process achieves a 1.0 GHz random read access frequency by eliminating the write-back operation. The read bitline swing of the 2T1D cell is improved by employing short local bitlines connected to local voltage sense amplifiers. A low-overhead dual-row access mode improves the worst-case cell retention time by 3X, minimizing standby power at times when only a fraction of the entire memory is utilized. Measurement results from a 64 kb eDRAM test chip in 65 nm CMOS demonstrate the effectiveness of the proposed circuit techniques

*Index Terms*—Dual row access, embedded DRAM, gain cell, local sense amplifier, low power, write-back-free read.

#### I. INTRODUCTION

**E** MBEDDED DRAM (eDRAM) technology has been drawing increasing attention in recent years as an alternative to the mainstream 6T SRAM, since it delivers higher bit-cell density and a practical random access time. 1T1C eDRAM has already been adopted for last level caches of high performance server chips [1]–[5]. Despite successful deployment of 1T1C eDRAM in recent server products, the complicated process steps involved in building the storage capacitor and the special access transistor, coupled with the limited signal swing at low supply voltages, make the scaling of this eDRAM technology unfavorable.

Gain cell eDRAM is considered as a promising embedded memory option with the potential of overcoming the scaling challenges encountered by SRAM and 1T1C eDRAM. It provides decoupled read and write paths which improve low voltage margin, while the cell size is approximately 2X denser than that of a 6T SRAM. Moreover, it is logic compatible and the separate read port enables non-destructive read and the capability of driving long bitline loads, making it competitive at low voltages. Table I compares the circuit parameters of interest for the three types of embedded memory.

Manuscript received December 17, 2012; revised January 19, 2013; accepted February 11, 2013. This paper was recommended by Associate Editor V. Chandra.

W. Zhang (corresponding author) was with the Department of ECE, University of Minnesota, Minneapolis, MN 55455 USA. He is now with Broadcom Corporation, Edina, MN 55435 USA (e-mail: zhangwei@broadcom.com; zhangweithu@gmail.com).

K. C. Chun was with the Department of ECE, University of Minnesota, Minneapolis, MN 55455 USA. He is now with Samsung Electronics, Hwasung 445-701, Korea (e-mail: kc.chun@samsung.com).

C. H. Kim is with the Department of ECE, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: chriskim@umn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSI.2013.2252652

TABLE I SRAM VERSUS EDRAM (1T1C AND GAIN CELL)

1

	6T SRAM [6]	1T1C eDRAM [1]	2T eDRAM [7]
Cell Schematic	S S S S S S S S S S S S S S S S S S S	<sup>B</sup> →	
<sup>(1)</sup> Reported array density	NA (70Mb)	3.01 Mbit/mm <sup>2</sup> (2Mb)	0.92 Mbit/mm <sup>2</sup> (2Mb)
<sup>(1)</sup> Reported cell size (ratio)	0.46x1.24= 0.5704 μm <sup>2</sup> (1X)	0.23x0.55= 0.1265 μm <sup>2</sup> (0.22X)	0.475x0.58= 0.2755 μm² (0.48X)
<sup>(2)</sup> Redrawn cell size (ratio)	0.575x2.05= 1.179 μm² (1X)	0.45x0.545= 0.245 μm² (0.21X)	0.48x0.925= 0.444 μm² (0.38X)
Low-VDD margin	Poor (ratioed)	Poor (destructive read)	Good (non-ratioed, gain function)
Storage cap.	irrelevant	<10fF	~1fF
Process	Logic compatible	Trench cap. + thick TOX access TR	Logic compatible
Retention time	NA	40µs @105°C, 99.99%	10µs @85°C
Random cycle	1ns <sup>(2)</sup>	2ns	2ns
Static power	1.0X	0.2X	NA

(1) All designs are in 65nm.

(2) Based on the same 65nm low power CMOS process.

Several recent gain cell eDRAM designs based on 2T or 3T cells have demonstrated practical retention times beyond 100  $\mu$ s [8], SRAM-like performance [9], [10], and true logic-compatibility by eliminating boosted voltages [10]. The most recent progress is a 2T1C (2 Transistors with 1 Capacitor) cell structure [10], which provides additional beneficial couplings through a capacitive device to enhance the read and write performance. Despite these innovations, the random access cycle in gain cell eDRAMs with practical retention time, which is no less than the 1.4 ns reported in 2011[10], is still relatively long compared to the GHz SRAMs. Nevertheless, one interesting and helpful feature of gain cells that has been largely overlooked in the past is the potential for write-back-free operations by taking advantage of the non-destructive read.

In this work, we have experimentally demonstrated for the first time, a gain cell eDRAM without write-back operation [11]. By removing the write-back from a read operation, the read access speed can be significantly improved. We also apply a local voltage sense amplifier (S/A) scheme to overcome the design complexities and variability issues prevalent in the existing current-sensing schemes used for 2T gain cells. Finally, a low-overhead low-power mode based on a dual-row-access scheme extends cell retention time by 3X to save refresh power in standby mode during periods when only a fraction of the cache memory is being used.



Fig. 1. (a) 1T1C eDRAM structure and read waveforms. Data is destroyed during a read due to charge sharing, and write-back is needed for restorage. (b) 3T gain cell structure and read waveforms. Although write-back is included for demonstration purpose, data is maintained after read operations and a write-back is theoretically not necessary.



Fig. 2. Random cycle time comparison between SRAM and eDRAM [11].

#### II. WRITE-BACK-FREE READ OPERATION

## A. Write-Back in Embedded DRAMs (EDRAMs)

Unlike SRAM operations, eDRAM requires a write-back during each read. In conventional 1T1C eDRAMs, the read operation relies on charge sharing between the storage capacitance and the bitline capacitance. Due to the destructive read nature demonstrated in Fig. 1(a), a write-back is needed to reinforce the cell data after the charge sharing operation.

Gain cells on the other hand, have a non-destructive read. A gain function is obtained by driving the gate without involving any charge sharing, as shown in Fig. 1(b). Therefore, this behavior theoretically eliminates the need for a write-back in gain cell read operations.

#### B. Write-Back-Free Read Benefits

A write-back-free read operation provides us a system speedup opportunity which has been neglected in previous gain cell designs [8]–[10], [12]. Fig. 2 compares the cycle time between a 6T SRAM and various 2T eDRAMs implemented



Fig. 3. 2T1D gain cell with preferential boosting [11].

in the same 65 nm low-power process. It shows that, by eliminating the write-back delay that accounts for approximately 30% of the read cycle, a significant improvement in operating frequency can be achieved. Although these benefits are only applicable to the read cycle, it is sufficient to improve the overall system performance significantly, because in general there are more reads than writes in a processor cache, and a read may stall the system and therefore degrade the system level performance, while a write will not.

# C. Active Power Discussion

It is worth noting that write-back will also consume additional dynamic power. For read operations, 1T1C eDRAM will have considerably more power consumption than SRAM, due to the additional WBL switching which is comparable to the read portion in terms of power. A gain cell design with write-back-free read operations, on the other hand, allows a similar read behavior as SRAM, and at the same time further lowers the power by reducing the interconnect load due to denser layout. Therefore, removing write-back from read operations makes gain cell eDRAMs more competitive in overall active power.

Nevertheless, since static power (equals to refresh power in 2T-based gain cell designs) is dominant for large caches, and the detailed active power comparison will involve building extensive simulation framework, in this work we only focus on the static power.

## D. Simulation Results Using A 2T1D Cell

A 2T1D (2 Transistor with 1 Diode) cell, shown in Fig. 3, was used in this work to demonstrate the write-back-free read operation. Fig. 4(a) shows its retention characteristics with the leakage currents in hold period, and the critical data "1" is maintained high due to the leakage profile dominating in the pull up direction. The gain cell design is a variant of the previous 2T1C cell [10], but is different in that the P-type coupling device shared between two adjacent cells is replaced by a separate N-type diode in each cell, due to the fact that an N-type diode is more layout efficient in our design. Without sharing the coupling device, it minimizes any coupling noise from the adjacent cells and is preferred for write-back-free read operations. This also provides the similar beneficial coupling up effect as described in [8], [12] during read.

The detailed operations, very similar to the 2T1C design in [10], are illustrated in Fig. 4(b). The regulated WBL scheme is adopted, which lowers the data "1" voltage on WBL instead of boosting up the WWL voltage to achieve similar subthreshold leakage suppression [8], [10] without a boosted supply. During a write or refresh operation, the PCOU signal first preferentially couples up the cell node voltage for increased read current, which compensates the coupling-down by the RWL signal.



Fig. 4. (a) Retention characteristics of 2T1D cell (left) and leakage currents during hold period (right). (b) basic operation illustrations of write/refresh (left) and write-back-free read (right) [10].

Note that the data "1" coupling is higher than data "0" because the N-type diode has higher capacitance when it is turned on [8], [12]. To write back, the PCOU beneficially couples down the data "0" voltage, which compensates the coupling-up by the WWL signal, and therefore results in a data "0" voltage close to 0 V without requiring a boosted negative supply [10]. Note that data "1" is not much affected by the PCOU coupling because the WWL remains low. Further details on the circuit operation of the 2T1C cell can be found in [10]. The major timing change is that during a write-back-free read operation, we pushed the PCOU falling edge so that the PCOU switching occurs within a single read cycle, allowing consecutive reads without a write-back phase. Although the write cycle also contains a read operation, which is required due to the row-wise write nature of eDRAMs to avoid overwriting the data of unaccessed cells on the same row, we kept the write cycle timing intact so that the beneficial write feature can be preserved.

Given the theoretical analysis, it is still necessary to evaluate the write-back-free read through simulations because potential couplings during cell access may have impact on the data, as depicted in Fig. 1(b). Simulation results in Fig. 5 examine the impact of write-back-free reads for access rates from 0.01% to 10%, where 10% means there is 1 read out of 10 cycles and the rest are idle. The voltage window between data "1" and data "0" remains unchanged for access rates up to 1%. For access rates greater than 1%, the cell voltage window slightly improves. Although the detailed reasons are still under investigation, this may be due to the minute difference between the couple up and couple down voltages that gets accumulated over a long retention period (e.g., 200  $\mu$ s), which compensates for the pull-up



Fig. 5. Cell retention characteristics without write-back for different read access rates. Coupling strength increases with cell voltage due to MOS capacitance change [11].



Fig. 6. Cell voltage w/o write-back vs. read access rate [11].

leakages in data "0." Another potential reason is that the PCOU coupling temporarily raises the data "0" voltage during a read cycle, which reduces the entire leakage profile and extends the retention time especially in cases with frequent reads. Data "1" has the similar behavior; however, the leakage profile change is negligible compared to that in data "0," because the leakage in data "1" cases is significantly lower. Fig. 6 plots the cell voltage after 100, 200, and 300  $\mu$ s, across different access rates up to 10%. Note that for an access rate more than a few percent, it is already an extreme case because it needs to keep accessing the same wordline for a relatively long time, given the typical retention time (200  $\mu$ s) and cache size (1Mb).

Both Figs. 5 and 6 show that for practical access rates, getting rid of the write-back does not have an adverse effect on the data "1" and data "0" levels. Our test chip design focuses on experimentally verifying the impact of write-back-free reads on overall eDRAM performance for practical access rates.



Fig. 7. Read disturbance in a typical 2T gain cell structure. The pull-up currents fight against the read current once RBL is pulled down. [9].



Fig. 8. (a) A typical voltage sense amplifier, and (b) a compact current sense amplifier, which is still much more complex and area consuming and cannot incorporate dummy averaging scheme.

## III. VOLTAGE SENSING W/ LOCAL SENSE-AMPLIFIER Architecture

#### A. Read Disturbance Issue

For 2T-based gain cell, despite the fast speed and compact cell size compared to its 3T counterpart, read disturbance has been a common issue [9] as shown in Fig. 7. When a data "1" is being read and RBL is being pulled down, the adjacent data "1" cells sharing the same bitline create disturbance currents in the pull up direction, and limit the the RBL swing. In worst case where all cells are storing data "1" in a 256-word-line configuration, the RBL swing is limited to around 0.2 V without considering variations. This is not sufficient for reliable voltage sensing, given the fact that eDRAM utilizes single-ended sensing and the sense-amplifier that cannot be shared among bitlines needs to be compact.

Common practice in 2T-based gain cells has been to implement a current-sensing scheme, which keeps the RBL level close to VDD during read [9], [10]. However, a current-sensing scheme is more complex and area consuming as shown in Fig. 8, and at the same time suffers from variation issues in the dummy cell due to the single-ended sensing while cannot utilize dummy averaging techniques because of the small input impedance requirement. This will introduce significant design



Fig. 9. Read disturbance mitigation of 2T1D cell using i) beneficial PCOU coupling, ii) regulated WBL, and iii) short local read bitlines [11].

overhead and scalability challenges in future technology nodes. Therefore, a voltage sensing solution, if applicable, is usually preferred.

## B. Read Disturbance Mitigation

As discussed above, the read disturbance issue originates in the contention between the read current and the multiple pull-up disturbance currents. Therefore, the final available voltage sensing window in the worst case is determined by three factors, namely the read current, the disturbance current, and the total number of disturbance.

To get around this problem, we apply three circuit techniques targeting at the three factors, respectively, as demonstrated in Fig. 9, eventually allowing a more robust voltage-sensing scheme to be used. First, the beneficial PCOU coupling in the accessed 2T1D cell provides stronger pull-down read current. Second, a regulated WBL scheme, mentioned in Section II, lowers the fresh data "1" voltage by 0.2 V (1.1 V to 0.9 V), significantly reducing the disturbance current. This has proven to have little impact on data retention since data "1" quickly stabilizes to around 0.85 V due to the cell leakage profile [10]. Finally, a local-sense-amplifier (L-S/A) scheme with short read bitlines [1] limits the maximum number of unselected cells to 63 which in turn reduces the worst case read disturbance current and provides a sufficient signal margin for reliable voltage sensing.

# C. Effectiveness Evaluation

For the three circuit techniques proposed to allow voltage sensing in 2T-based gain cells, namely the beneficial PCOU coupling, the regulated WBL and the local sense amplifier (L-S/A) architecture, it is meaningful to evaluate the effectiveness of each technique in mitigating the read disturbance issue.

Tables II and III summarize the simulated bitline voltage window improvement in the worst read disturbance scenario by incorporating individual schemes. Compared to the conventional scheme [9], the beneficial PCOU coupling contributes

TABLE II SIMULATED VOLTAGE WINDOW IMPROVEMENT

Scheme	Voltage Window Available	Improvement
Conventional	265mV	-
PCOU Coupling	285mV	PCOU Effectiveness +20mV
Regulated WBL	443mV	Reg. WBL Effectiveness +178mV

TABLE III Simulated Voltage Window Importent of L-S/A Architecture W/ Different # of Cells Per RBL

Scheme	Voltage Sensing Window	Improvement
PCOU + reg.WBL, 256/RBL	459mV	-
PCOU + reg.WBL, 128/RBL	490mV	2X Cell # Reduction +31mV
PCOU + reg.WBL, 64/RBL	523mV	4X Cell # Reduction +64mV
PCOU + reg.WBL, 32/RBL	560mV	8X Cell # Reduction +101mV

around 20 mV increase while the regulated WBL provides the most significant improvement of around 180 mV. The L-S/A architecture is examined with various numbers of cells per bitline. Around 30 mV improvement is observed for each 2X reduction in cell numbers, indicating a logarithm relationship between the voltage window and the cell number per bitline.

This logarithm relationship can be derived by assuming the read current being a constant, because the accessed data "1" cell is typically in its saturation region. At the steady state, the total subthreshold leakage from adjacent unaccessed cells is equal to the read current, as depicted in Fig. 9, which therefore determines the available bitline voltage window. Assume there are N cells per (local) read bitline, in the worst case we have:

$$I_{ACCESS} = (N-1)I_{UNACCESS}.$$
 (1)

Due to the exponential relationship between subthreshold leakage and gate voltage, the equation can be expressed as:

$$I_{ACCESS} = (N-1) \times C \times \exp\left(\frac{V_{GS}}{D}\right)$$
$$= (N-1) \times C \times \exp\left[\frac{(0.9V - V_{LRBL})}{D}\right]. (2)$$

where C and D are constants given by the device characteristics, and 0.9 V is the fresh data "1" voltages in unaccessed cells in the worst case. Based on our voltage setup at 1.1 V supply, (2) can be rewritten as:

$$I_{ACCESS} = (N-1) \times C$$

$$\times \exp\left\{\frac{\left[0.9V - (1.1V - V_{WINDOW})\right]}{D}\right\}$$

$$= (N-1) \times C$$

$$\times \exp\left[\frac{(V_{WINDOW} - 0.2 V)}{D}\right].$$
 (3)

where  $V_{\text{WINDOW}}$  is the available bitline voltage window. From (3) we can get:

$$V_{WINDOW} = -D\ln(N-1) + D\ln\left(\frac{I_{ACCESS}}{C}\right) + 0.2V = -D\ln(N-1) + Const. \approx -D\ln N + Const.$$
(4)

Equation (4) clearly reveals the logarithmic relationship between the voltage window ( $V_{WINDOW}$ ) and the cell number per bitline (N). A more detailed derivation based on the given device subthreshold swing can further provide the estimated slope value. For a subthreshold swing of 100mV/decade,

$$I_{UNACCESS} = C \times 10E\left(\frac{V_{GS}}{100 \ mV}\right)$$
$$= C \times \exp\left(\frac{\ln 10 \times V_{GS}}{100 \ mV}\right)$$
$$= C \times \exp\left(\frac{V_{GS}}{43.4 \ mV}\right). \tag{5}$$

Therefore, by comparing with definition of D in (2), we have

$$D = \frac{(Subthreshold\_swing)}{\ln 10}$$
  
= 43.4 mV. (6)

For a 2X reduction in the cell number per local bitline,

$$\Delta V_{WINDOW} = -D\ln(N) - (-D\ln(2N)) = D\ln 2 = 30.1 \ mV.$$
(7)

The result in (7) matches the simulated value of around 30mV as shown in Table III. It provides a better understanding of the effectiveness of local sense-amplifier architecture in terms of read disturbance mitigation from a theoretical view.

## D. Design Conclusion

According to the previous discussion, the regulated WBL is the most effective in mitigating the read disturbance issue while the PCOU coupling is the least. Although a reduced cell number per bitline is more beneficial, it has to be carefully selected due to the trade-off between performance and area overhead.

Fig. 10 shows the performance and area overhead with different number of cells per (local) bitline. Sixty-four serves as an optimized value with sufficient voltage window and sensing speed at a cost of 3.5% area overhead. Note that this overhead number is reduced to 1.6% considering the simple global voltage S/A compared to a current-sensing solution.

Fig. 11 shows the detailed architecture with local and global voltage S/As. The reduced load of the local RBLs ensures a fast voltage development while a simple global voltage S/A further speeds up the signal propagation. Dummy cell averaging which was not possible in previous current-sensing schemes, can now be implemented to enhance the robustness under PVT variations. Such voltage averaging is much easier than current averaging, and was implemented with negligible overhead by simply



Fig. 10. (a) Bitline voltage window and (b) sensing delay versus area overhead.



Fig. 11. Schematic and timing of local and global sense amplifier architecture [11].

connecting all the reference bitlines together, eliminating the variation across reference bitline voltages due to dummy transistor mismatch. The final available voltage sensing window is around 500mV in worst case as listed in Table III.

# IV. DUAL-ROW-ACCESS LOW POWER MODE

For applications that do not utilize the entire cache memory space, shutting down parts of the array is a practical way to save power. For gain cell eDRAMs, however, activating multiple rows at the same time during standby mode can potentially yield greater power savings, because the refresh power is determined by the worst case retention time of the tail cells, which



IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-I: REGULAR PAPERS

Fig. 12. Power comparison for different row numbers (a) without and (b) with additional timing adjustment.

2

1/Active-percentage (Simple powering down)

or Number of rows (Multi-row access)

4

8

can be repaired by enabling additional strong cells at the same time due to spatial randomness.

#### A. Multi-Row-Access Scheme Analysis

1

0.6

0.4

0.2

0

A tail cell typically occurs due to the weakness in either data storage or read path strength [13], resulting in a weak sensing current. By enabling multiple rows, sensing currents from two or more cells are combined, and the chance of having a weak combined sensing current is considerably lower due to the random spatial distribution of tail cells. Therefore, the tail cell situation can be significantly improved.

Fig. 12(a) compares the power consumption between simple powering down (single-row access) and multi-row-access approaches, with different numbers of rows accessed and no adjustment in timing or reference voltage. Note that the x-axis numbers for simple powering down approach (black) have a different meaning as the corresponding scaling factor, e.g., 4 means only 1/4 array remains enabled in order to keep the same amount of data as a 4-row access scheme. By not sharing the L-S/As, the multi-row-access mode can boost the sensing currents on the global bitlines without any changes in the timing or reference voltage, with a voltage swing limit of VDD. If we merely calculate the sensing currents regardless of the swing limit, the multi-row-access approach is estimated to be beneficial with up to 4 rows. However, this number is limited to 2 with a 1.1 V VDD swing. The reason is that the global read bitlines (both GRBL and GRBLBs) will be discharged much faster



Fig. 13. Dual-row access mode illustration (WL0 and WL64 selected) [11].

in the 4- or 8-row case. Without modifying the sensing timing, these bitlines are likely to reach 0 V and result in a 0 V sensing window before the global sense amplifier is enabled. Therefore, the data "0" in the 4- or 8-row case needs to be strong enough, so that the higher bitline will be maintained above 0 V by the time of launching the global sensing signal.

This swing issue can be solved by scaling voltage developing time accordingly and launching the global sensing signal earlier. Nevertheless, in this case it reduces time margin, and additional dedicated timing adjustment circuitry is required as design overhead. Fig. 12(b) indicates that such solution does not make 4 or 8 the favored number in our simulation vehicle, and for the 2-row case, the power saving is less than that in Fig. 12(a).

Although the actual power savings would vary in real chips, Fig. 12 provides a good estimate of the performance trend in our design. Activating more than two rows at a time (e.g., 4- or 8-row-access) has diminishing returns while incurring significant design overhead in forms of dedicated timing and reference circuitry. Meanwhile a dual-row access mode can be more widely used because the chance of requiring no more than a quarter of the cache is much lower. Therefore, we propose to use a simple dual-row-access mode.

#### B. Dual-Row-Access Mode

Fig. 13 shows our dual-row-access mode, where two wordlines with respective L-S/As are enabled at the same time in a refresh operation without any changes in the read reference circuit. Data are stored and refreshed on a dual-row basis. The weak cell is thus repaired by a stronger one according to spatial randomness, and in addition, mismatch between the L-S/As themselves gets averaged out. The cells in each pair are designed to have a significant distance from each other, so that local correlation is minimized to enhance special randomness. Moreover, the effective local sensing current to develop the global read bitline (GRBL) is doubled, improving the retention time even further under the same sensing window requirement and timing constraints. Thus, the worst retention time can be improved by more than 2X using a dual-row-access mode, while a simple powering-down approach may still suffer from the tail cell's retention time. Note here although a single refresh operation may incur up to double power consumption due to larger sensing current, the total number of refresh operations is cut to half due to the dual-row refresh basis.



Fig. 14. (a) Failure percentiles for a 1 kb sub-array and (b) detailed view of tail cells [11].

## V. EXPERIMENTAL RESULTS

A 64 kb eDRAM test chip was implemented in a 1.2 V, 65 nm logic CMOS process to demonstrate the proposed circuit techniques. An aggressive read cycle time of 1.0 ns was used to highlight the performance benefits of a write-back-free operation which is a 29% improvement compared to the previously reported cycle time of 1.4 ns [10]. The chip achieves a 99.9% retention time of 325  $\mu$ s and a refresh power of 234.1  $\mu$ W/Mb at 1.1 V, 85 °C. Figs. 14 and 15 show the failure percentile and retention map from a 1 kb sub-array, respectively, for a worst case read disturbance pattern and a 1.0 ns read cycle time. No noticeable changes in the retention time were observed across different access rates, which was as expected from simulations. Although our measurement setup supports only up to a 1% access rate, this is sufficient for real-world applications as discussed in Section II.

For a 99.99% bit yield, the extrapolated retention time was 90  $\mu$ s. Due to the dummy averaging feature of our proposed design, the measured retention map in Fig. 15 shows no significant signs of bitline dependency in the failure pattern which is in contrast to the results presented in our prior work [10]. It also confirms our assumption about the spatial randomness of tail cells, which is important for the dual-row-access mode.

Figs. 14 and 15 show that the worst case retention time improves from  $325 \ \mu s$  to  $900 \ \mu s$  using the dual-row-access mode.



Fig. 15. Measured retention maps for (a) single and (b) dual row access modes. No significant bitline dependency is observed in either case [11].



Fig. 16. Static power consumption comparison between a power gated SRAM with a 0.6 V retention voltage and various 2T1D eDRAM power down modes [11].

Fig. 16 compares the static power dissipation of SRAM and various eDRAM configurations. The proposed 2T1D design (single row access) achieves a 23.9% power saving compared to that of a power gated 6T SRAM with a 0.6 V retention voltage. Compared to a simple power down mode where half the eDRAM is unused, a 27.8% power reduction was achieved using the dual-row-access mode.

The measured VDD shmoo for retention time and cycle time is plotted in Fig. 17(a), and a static power comparison between 6T SRAM and the proposed eDRAM design for different supply voltages is shown in Fig. 17(b). The longer retention time at higher supply voltages makes the eDRAM refresh power to be



Fig. 17. (a) Measured VDD shmoo and (b) static power comparison [11].



Fig. 18. Retention time versus read random cycle time under different temperatures.

lower than the static power of an SRAM. Note that the optimal supply voltage of an eDRAM is usually higher than that of an SRAM due to the refresh power dominating the overall static power consumption.

Fig. 18 plots the retention time increase by relaxing the random cycle time at different temperatures. Fig. 19 shows the die microphotograph and summarizes the key features. With a 1.0 ns read and 1.5 ns write/refresh cycle time, the cell availability is calculated as 99.5% for a 1 Mb array.

## VI. CONCLUSION

We have presented several circuit techniques to enhance the performance and robustness of gain cell eDRAM. The proposed design was the first to experimentally verify writeback-free read operation in gain cells. No noticeable retention time difference was observed across a wide range of access rates, and a 1.0 ns read cycle time is achieved in embedded DRAM for the



Fig. 19. Test chip microphotograph and feature summary table [11].

first time. We also proposed various circuit techniques for mitigating read disturbance issues including a local-sense-amplifier scheme. Voltage sensing is therefore allowed in 2T-based gain cell designs and dummy averaging also becomes applicable. In addition, a dual-row-access low power mode was introduced to further reduce standby power in scenarios where no more than half the cache is being utilized. Test chip measurements were presented from a 64 kb eDRAM array implemented in a 1.2 V, 65 nm CMOS process, demonstrating 23.9% static power saving compared to a power gated SRAM, and an additional saving of 27.8% in dual-row-access mode.

#### REFERENCES

- J. Barth et al., "A 500 MHZ random cycle, 1.5 ns latency, SOI embedded DRAM macro featuring a three-transistor micro sense amplifier," *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 86–95, 2008.
- [2] J. Barth et al., "A 45 nm SOI embedded DRAM macro for POWER7TM processor 32 MB on-chip L3 cache," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 64–75, 2011.
- [3] P. Klim et al., "A 1 MB cache subsystem prototype with 1.8 ns embedded DRAMs in 45 nm SOI CMOS," *IEEE J. Solid-State Circuits*, vol. 44, no. 4, pp. 1216–1226, 2009.
- [4] M. Nakayama et al., "A 16 MB cache DRAM lsi with internal 35.8 GB/s memory bandwidth for simultaneous read and write operation," in Proc. Int. Solid-State Circuits Conf., 2000, pp. 398–399.
- [5] S. Romanovsky et al., "A 500 MHz random-access embedded 1 Mb DRAM macro in bulk CMOS," in Proc. Int. Solid-State Circuits Conf., 2008, pp. 270–612.
- [6] K. Zhang et al., "SRAM design on 65 nm CMOS technology with dynamic sleep transistor for leakage reduction," *IEEE J. Solid-State Cir*cuits, vol. 40, no. 4, pp. 895–901, 2005.
- [7] D. Somasekhar et al., "2 Ghz 2 MB 2T gain cell memory macro with 128 GBytes/sec bandwidth in a 65 nm logic process technology," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 174–185, 2009.

- [8] K. Chun et al., "A sub-0.9 V logic-compatible embedded DRAM with boosted 3T gain cell, regulated bit-line write scheme and PVT-tracking read reference bias," in Proc. VLSI Circuits Symp., 2009, pp. 134–135.
- [9] K. Chun, P. Jain, T. Kim, and C. H. Kim, "A 1.1 V, 667 MHz random cycle, asymmetric 2T gain cell embedded DRAM with a 99.9 percentile retention time of 110 μsec," in *Proc. VLSI Circuits Symp.*, Jun. 2010, pp. 192–192.
- [10] K. Chun, W. Zhang, P. Jain, and C. Kim, "A 700 MHz 2T1C embedded DRAM macro in a generic logic process with no boosted supplies," in *Proc. Int. Solid-State Circuits Conf.*, 2011, pp. 506–507.
- [11] W. Zhang, K. Chun, and C. Kim, "A write-back-free 2T1D embedded DRAM with local voltage sensing and a dual-row-access low power mode," in *Proc. Custom Integr. Circuits Conf.*, 2012, pp. 1–4.
- [12] W. Luk, J. Cai, R. Dennard, M. Immediato, and S. Kosonocky, "A 3-transistor DRAM cell with gated diode for enhanced speed and retention time," in *Proc. VLSI Circuits Symp.*, 2006, pp. 184–185.
- [13] W. Zhang, K. Chun, and C. Kim, "Variation aware performance analysis of gain cell embedded DRAMs," in *Proc. Int. Symp. Low Power Electron. Design*, 2010, pp. 19–24.



Wei Zhang received his B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2007, and his M.S. and Ph.D. degrees in electrical and computer engineering from the University of Minnesota, Minneapolis, MN, USA, in 2010 and 2012, respectively. He then joined Broadcom Corporation, Edina, MN, USA, in 2012, where he is currently a Staff Scientist.

He is now working on memory-related topics at Broadcom Corporation. When he was pursuing his Ph.D. degree, he was involved in 6 chip designs. He

mainly conducts research on memory-related topics, including SRAM, ROM, and embedded DRAM. He was also involved in a cooperative project on organic electronics. He is the author and coauthor of 10 journal and conference papers. His current research interests include memory circuit design and organic electronic technologies.



Ki Chul Chun received the B.S. degree in electronics engineering from Yonsei University, Seoul, Korea, in 1998, the M.S. degree in electrical engineering from KAIST, Daejeon, Korea, in 2000, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, MN, USA, in 2012. In 2000, he joined the Memory Division, Samsung Electronics, Gyeonggi—Do, Korea, where he has been involved in DRAM circuit design. After his Ph.D. study at the U of MN, he rejoined Samsung Electronics in 2012, where he has worked

for low-power DRAM development.

Dr. Chun is the recipient of ISLPED Low Power Design Contest Awards (2009 and 2012) and a Samsung Ph.D. Scholarship. His research interests include digital, mixed-signal and memory circuit designs with special focus on DRAM, PRAM, and STT-MRAM in scaled technologies.



**Chris H. Kim** (M'04–SM'10) received his B.S. and M.S. degrees from Seoul National University and a Ph.D. degree from Purdue University. He spent a year at Intel Corporation where he performed research on variation-tolerant circuits, on-die leakage sensor design and crosstalk noise analysis. He joined the electrical and computer engineering faculty at the University of Minnesota, Minneapolis, MN, in 2004 where he is currently an associate professor.

Prof. Kim is the recipient of an NSF CAREER Award, a Mcknight Foundation Land-Grant Pro-

fessorship, a 3M Non-Tenured Faculty Award, DAC/ISSCC Student Design Contest Awards, IBM Faculty Partnership Awards, an IEEE Circuits and Systems Society Outstanding Young Author Award, ISLPED Low Power Design Contest Awards, and an Intel Ph.D. Fellowship. He is an author/coauthor of 100+ journal and conference papers and has served as the technical program committee chair for the 2010 International Symposium on Low Power Electronics and Design (ISLPED). His research interests include digital, mixed-signal, and memory circuit design in silicon and non-silicon (organic TFT and spin) technologies.