A 2T1C Embedded DRAM Macro With No Boosted Supplies Featuring a 7T SRAM Based Repair and a Cell Storage Monitor

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

Abstract—A truly logic-compatible gain cell eDRAM macro with no boosted supplies is presented. A 2T1C gain cell implemented only with regular thin oxide devices consists of an asymmetric 2T cell and a coupling PMOS capacitor. The PMOS capacitor ensures proper operation even without a boosted supply by utilizing a beneficial coupling for read and a preferential boosting for write. A repair scheme based on a single-ended 7T SRAM has features such as a local differential write and shared control with the main 2T1C array. A storage voltage monitor is proposed to track the retention characteristics of a gain cell eDRAM under PVT variations and to adjust its refresh rate adaptively. A 128 kb eDRAM test chip implemented in a 65 nm Low-Power (LP) process operates at a random access frequency of 714 MHz with a static power dissipation of 161.8 μ W per Mb for a 500 μ s refresh rate at 1.1 V and 85°C.

Index Terms—2T, 2T1C gain cell, 7T SRAM, cache, embedded memory, logic-compatible eDRAM, repair scheme, retention time, storage monitor, temperature sensor.

I. INTRODUCTION

N-DIE cache memory is a key component in advanced processors since it can boost micro-architectural level performance at a moderate static power penalty. Demand for denser memories only going to increase as the number of cores in a microprocessor goes up with technology scaling. A commensurate increase in the amount of cache memory is needed to fully utilize the larger and more powerful processing units. 6T SRAMs have been the embedded memory of choice for modern microprocessors due to their logic compatibility, high speed, and refresh-free operation [1]–[3]. However, the relatively large cell size and conflicting requirements for read and write at low operating voltages make aggressive scaling of 6T SRAMs challenging in sub-22 nm. Recently, 1T1C embedded DRAMs (eDRAMs) have replaced SRAMs in several server applications reducing the footprint and improving performance [4]-[7]. Difficulties in scaling the trench capacitor and the additional process steps involved in manufacturing the thick oxide (T_{OX}) access devices are currently limiting the wide spread adoption of 1T1C technology.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: kichul.chun@gmail.com; chunx041@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/JSSC.2012.2206685

Gain cell eDRAMs are considered as an alternative for high density cache memories with the potential for overcoming the scaling challenges associate with 6T SRAMs and 1T1C eDRAMs. Gain cells can be implemented using three transistors, or even two transistors when used with delicate read control circuits, achieving roughly a $2 \times$ higher bit-cell density than SRAMs [8]. Furthermore, gain cells can have smaller cell leakage current than power-gated SRAMs due to the fewer number of devices and the negative-Vgs biasing condition. The cell write margin is better than SRAMs since there is no contention between the access device and the cross-coupled latch in a gain cell. The key challenge for gain cells is the limited retention time due to the small storage capacitor and various leakage currents that depend exponentially on the Process-Voltage-Temperature (PVT) condition. A shorter retention time leads to higher refresh power and poor read current. The former is a result of frequent refresh operation while the latter is due to the rapid loss of cell storage voltage. Several circuit techniques have been proposed and demonstrated to enhance the retention time of gain cell eDRAMs to be comparable to that of 1T1C eDRAMs, making them a strong contender for future embedded memories [9]–[11]. However, the boosted high and low supplies, often referred to as VPP and VBB in traditional DRAM literature, needed to ensure basic DRAM cell operation necessitate thick oxide devices to prevent oxide reliability issues. This will result in a larger bit-cell size and a reduced macro level performance, not to mention a modification in the logic process, however this fundamental limitation have been overlooked in past research. In this paper, we present the following circuit techniques for realizing a truly logic compatible (i.e., thin oxide only implementation) gain cell eDRAM with no boosted supplies; (i) a 2T1C gain cell featuring a beneficial read and a preferential write, (ii) a single-ended 7T SRAM to repair weak gain cells, and (iii) a storage voltage monitor capable of tracking cell retention time under PVT variations for adaptive refresh control [12]. A 64 kb test macro implemented in a 65 nm Low-Power (LP) process achieves a random cycle of 714 MHz and a retention time of 500 μ s after applying the repair scheme at 1.1 V and 85°C. To the best of our knowledge, the proposed design outperforms all previous gain cell eDRAMs implemented using boosted supplies. The remainder of this paper is organized as follows. Section II describes the proposed 2T1C gain cell with no boosted supplies. The single-ended 7T SRAM for weak gain cell repair and the storage voltage monitor for adaptive refresh control are presented in Sections III and IV, respectively.

Manuscript received February 15, 2012; revised May 12, 2012; accepted May 14, 2012. Date of current version October 03, 2012. This paper was approved by Associate Editor Peter Gillingham.

Section V describes hardware measurement results from a 65 nm test chip, and conclusions are given in Section VI.

II. 2T1C GAIN CELL WITH NO BOOSTED SUPPLIES

DRAMs typically require two boosted supplies: a boosted high voltage (VPP) to suppress the subthreshold leakage (assuming a PMOS write device) and a boosted low voltage (VBB) to prevent V_{TH} drop during write. Fig. 1 illustrates how the boosted supply level affects the performance of a 2T gain cell eDRAM. Here we consider an asymmetric 2T cell [11] with a PMOS write device and an NMOS read device, although a similar analysis can be made for other types of gain cells. Write Word-Line (WWL) is biased at VPP during data retention mode in order to suppress the subthreshold leakages flowing into unselected cells. The subthreshold leakage is worst when writing '1' to another cell on the same the Write Bit-Line (WBL). The VPP level modulates the gate-overlap and gate-induced drain leakages and therefore the optimal retention time can be achieved by considering the retention times of both data '1' and data '0' as shown in Fig. 1(a). The VBB level on the other hand affects the data '0' restore time during write. The simulated data '0' restore time dependency on VBB level in Fig. 1(b) indicates that VBB should be -0.4 V or below to ensure a practical write time. The above analysis shows that a WWL voltage swing from -0.4 V to 1.6 V is required for optimal memory cell operation which is 67% higher than the nominal supply level of 1.2 V in this 65 nm LP CMOS process. Boosted high and low supplies can only be used with special devices with a thicker T_{OX} to avoid voltage overstress. Alternatively, I/O devices can be considered; however, this will increase the bit-cell area considerably and in turn degrade system level performance. Layout comparison between several embedded memory bit cells are presented in Section III.

In order to realize a truly logic-compatible eDRAM with a competitive bit-cell size and higher macro level performance, we propose a 2T1C gain cell that can be implemented with regular thin oxide devices. The new cell structure consists of an asymmetric 2T cell [11] and a separate coupling MOS capacitor controlled by a control signal. Fig. 2 shows the proposed bit-cell schematic along with the signal conditions for each operating modes. It's important to note that none of the voltage levels exceed the nominal VDD. The bit-cell may look similar to the previous 3T1D cell that has an additional gated-diode controlled by Read Word-Line (RWL) in order to enhance speed and retention time by signal amplification [9]. However, the structure and operating principle of the 2T1C cell are considerably different from prior work. The timing diagram shown in Fig. 3 illustrates the operation principle of the proposed cell. The capacitor control signal (PCOU) is pre-discharged to 0 V during hold mode introducing only a small amount of gate-overlap leakage through the coupling device (PC). At the beginning of the read access when the RWL is activated, PCOU is also switched to VDD. This couples up both data '1' and '0' storage voltages. The higher voltage levels increase the drive current for the NMOS read access device (PS) enhancing the read performance. After the Sense Amplifier (S/A) samples the Read Bit-Line (RBL) data, a write-back operation follows which drives the WWL to 0 V instead of the usual negative boosted supply. Using a data '1' WBL voltage that is slightly lower than VDD



Fig. 1. Impact of boosted supply level on 2T eDRAM performance [11]. (a) Boosted high supply (VPP) level vs. retention time (measured). (b) Boosted low supply (VBB) level vs. write performance (simulated).



Fig. 2. Proposed 2T1C gain cell based on thin oxide devices with no boosted supplies. (a) Schematic. (b) Signal conditions for each operating mode.

(i.e., VDD- α in Fig. 3), the subthreshold leakage in the unselected cell can be effectively cut off without using a boosted high supply for WWL [10]. Data '1' can be easily written back to the cell with a PMOS write device (PW). However, without a boosted negative supply, data '0' will not be fully restored due



Fig. 3. Timing diagram of the proposed 2T1C cell for read and write-back operations.

to the V_{TH} drop in PW. To resolve this issue, PCOU is switched to 0 V immediately after write back. This couples down the data '0' voltage while the data '1' voltage is not affected since PW remains on when WBL is high. Finally, WWL is switched back to its precharge level of VDD and this slightly couples up both data '1' and '0' voltages through the gate-overlap capacitance, fully restoring the cell storage levels. Fig. 4 shows the simulated waveforms of read and write-back operations. The proposed 2T1C with no boosted supplies achieves a similar data '1' voltage level during read and a similar data '0' level after write-back operations compared to the asymmetric 2T with boosted supplies.

The initial voltage levels and data windows between '1' and '0' in Fig. 4 are based on retention simulations for a 1 Mb macro (Fig. 5). The data window at 200 μ s for the 2T1C eDRAM is 150 mV smaller than that of a 2T eDRAM with boosted high and low supplies. A narrower data window reduces the margin between data '1' and '0' resulting in worse retention time and increased static power due to the frequent refresh operation. To cope with this issue, we propose two circuit techniques: (i) a single-ended 7T SRAM for weak gain cell repair and (ii) a storage voltage monitor for adaptive refresh by tracking the retention characteristics under PVT variations. Fig. 6 shows the schematic diagram of a 64 kb 2T1C eDRAM macro including the 7T SRAM repair cells (details in Section III) and the storage voltage monitor (details in Section IV) that are seamlessly integrated into the array. Two adjacent WL's share a single PCOU signal in order to minimize the cell size overhead. Since gain cells have a non-destructive read, the shared PCOU has no effects on the retention time of the unselected cells when the WL is activated. The shared PCOU reduces the bit-cell size by 21% compared to the separated layout shown in Fig. 7(a). Simulated waveforms in Fig. 7(b) confirm that the signal loss due to the redundant PCOU activation is negligible. Note that the storage capacitance of a gain cell is very small (< 1 fF), so the additional power consumption due to the shared PCOU is also insignificant.



Fig. 4. Simulated waveforms of read and write-back operations for (a) a conventional 2T eDRAM and (b) the proposed 2T1C eDRAM.



Fig. 5. Comparison of retention characteristics between (a) a conventional 2T eDRAM with boosted supplies and (b) the proposed 2T1C eDRAM with no boosted supplies.

III. DECOUPLED 7T SRAM REPAIR CELL WITH SHARED CONTROL

Outlier cells having poor retention times are usually repaired using the same type of cell as the main array. However, gain cells have a very small storage capacitance, so the probability



Fig. 6. Schematic diagram of a 64 kb 2T1C gain cell eDRAM macro with no boosted supplies.

of having a failure cell in a redundant row or column is also high compared to a 1T1C DRAM. The proposed 2T1C eDRAM has a narrower data window due to the reduced WWL voltage swing aggravating this situation. In order to improve the retention time of the 2T1C eDRAM, we devise a single-ended decoupled 7T SRAM based repair scheme. The proposed 7T SRAM consists of a decoupled read by replicating the 2T1C gain cell and a differential write using a locally generated complementary WBL signal (WBLB) as shown in Fig. 8. The pitch matched 7T SRAM cell shares control signals (i.e., RBL, WBL, WWL, RWL) with the main 2T1C array minimizing the area overhead. Note that WBLB is generated by an inverter inside the 7T SRAM cell while WBL is connected to every cell in the bitline direction as shown in Fig. 6. Therefore, the local differential write minimizes power dissipation incurred by the additional signal switching during memory access. Fig. 9 shows the comparison of signal-to-noise margin (SNM) between the proposed 7T SRAM and a conventional 6T SRAM. The decoupled read structure of the 7T SRAM improves the read SNM by 113% than a 6T SRAM, and the write SNM of the 7T SRAM having a lower WBL voltage can be made comparable to that of a 6T SRAM by sizing optimization. As explained in Section II, the data '1' WBL voltage is lower than VDD by 0.2 V in order to suppress the subthreshold leakages flowing into unselected cells during write in the absence of a boosted WWL voltage [10].

Fig. 10 shows the transistor dimensions and layouts of the following memory cells: 6T SRAM, 2T gain cell, 3T gain cells using thin and thick $T_{\rm OX}$ devices, 7T SRAM, and 2T1C gain cell. All bit-cells were designed and drawn in a generic 65 nm

LP process. Dense bit-cell design rules were not available to the authors but for area comparison purposes, using a logic design rule is a generally accepted practice. The 2T and 3T gain cells are $2.4 \times$ and $2.2 \times$ denser than a 6T SRAM, respectively. Similar cell area ratios have been reported in industry designs based on dense design rules; for example, the 2T gain cell in [8] is $2.1 \times$ denser than the 6T SRAM in [13], both implemented in Intel's 65 nm process. However, the density advantage of gain cell over 6T SRAM claimed in prior literature is misleading since the boosted supply voltages will cause oxide reliability concerns. One way to get around this problem is to use 1.8 V I/O devices in which case the bit cell area density improvement compared to SRAM is reduced to around $1.2 \times$. Since the array efficiency of gain cell eDRAMs is typically lower than SRAM due to charge pumps and the complex peripheral circuitry (e.g., RWL and WWL decoders for the decoupled bit-cell access, S/A with write-back circuits in each RBL and WBL [11]), gain cells no longer have an area advantage at the macro level when implemented using I/O devices. Conversely, the proposed 2T1C gain cell implemented using regular thin T_{OX} devices is 1.7× denser than a 6T SRAM without having an oxide reliability concerns. For a 1 Mb macro including all peripheral circuitry, a 2T1C eDRAM is still $1.6 \times$ denser than a 6T SRAM array making it a viable alternative to conventional SRAM for last level caches.

IV. CELL STORAGE MONITOR

Retention time of commodity DRAMs varies exponentially with temperature since it is highly sensitive to the junction and subthreshold leakages. Therefore, DRAM products have



Fig. 7. Shared coupling signal (PCOU). (a) Bit-cell schematic and layouts. (b) Simulated waveforms show negligible disturbance in an unselected cell.



Fig. 8. Proposed decoupled 7T SRAM repair cell shares BL and WL signals with the 2T1C cell.

on-chip temperature sensors to control the refresh period adaptively according to the chip operating temperature [14]. Similarly, retention time of gain cells is also dependent on operating temperature since the storage node voltage changes according to the junction, subthreshold and gate leakages. "However, gate leakage has a weaker dependency on temperature, and the various coupling effects originated from RWL, WWL, and PCOU in Fig. 4 make a simple temperature sensor based refresh control ineffective for gain cell designs."

To overcome this problem, we propose a gain cell based temperature sensor that directly measures the storage node voltage using a cell access pattern generator and 2T1C replica cells. Fig. 11 shows the proposed storage voltage monitor and its timing diagram. The SCAN signal triggers the cell



Fig. 9. Signal-to-noise margin (SNM) of a 6T SRAM and the proposed 7T SRAM. (a) Read SNM. (b) Write SNM.

access pattern generator (PG) that provides control signals (WBL, WWL, PCOU, and RWL) to the 2T1C replica cells. The repetitive access patterns have the same timing as the main array in order to track storage node voltages under a realistic memory access condition. The operating clock frequency of the PG generated by the VCO-1 indicates the current retention time setting. The merged storage node voltage of the 256 replica cells is captured by the sample-and-hold circuit. The buffered storage voltage using a unity gain amplifier is temporarily stored in MOS capacitors implemented with standard thick T_{OX} I/O devices whose gate leakage is negligible. The final storage node voltage is utilized to adaptively control the refresh rate of the 2T1C gain cell eDRAM. In this design, the measured storage voltage is translated in the form of frequency for convenient off-chip measurement, and the corresponding storage voltage can be found out using a calibration procedure. To remove any systematic error that may have been introduced while merging the 256 cells, the calibration step is needed to obtain the relationship between the measured storage voltage and the actual retention characteristic.

In real systems, operating temperature of cache memories is strongly related with the activity of the nearby cores. Fig. 12(a) shows the thermal map of an 8-core processor with a 24 MB L3 cache [1]. On-chip thermal sensors readily available across the microprocessor can be utilized to control the storage voltage monitor. For example, the thermal sensor can trigger the monitor when there is a predetermined temperature change such as 10°C in the core area. The measured storage voltage is then sampled and the retention information is sent to the refresh rate control and event scheduler as shown in Fig. 12(b). When the refresh information is updated by 2T1C storage monitor, the event

Write access device operates under nominal voltages so no significant reliability concern



Fig. 10. Bit-cell layout comparison (6T SRAM, 3T, 2T, 2T1C cells). All bit-cells were drawn in a generic 65 nm LP process.



Fig. 11. Proposed storage voltage monitor for adaptive refresh control.

scheduler of the processor updates the refresh rate in the next refresh cycle. Repetitively sampling the storage node voltage removes any residual voltages in the sample-hold capacitors built using thick T_{OX} devices. During normal chip operation, two consecutive samples are enough to for a stable captured storage voltage. For a retention time of 500 μ s, the average power dissipation of the monitor circuit can be made less than 1% of the total operating power dissipation by reducing the sampling rate. This is possible because thermal conduction has a very long time constant in the order of hundreds of milliseconds [15]. In our design, the current consumption of the monitor circuit is 849 μ A.

V. TEST CHIP MEASUREMENTS

A 128 kb test macro implemented in a 1.2 V, 65 nm low-power logic CMOS process comprises a conventional 3T array and the proposed 2T1C array for performance comparison. Fig. 13 shows the chip microphotograph and key features of the 65 nm eDRAM test chip. Our design achieves a 1.4 ns (= 714 MHz) random cycle and a 500 μ s retention time (after a single-BL repair scheme) at 1.1 V and 85°C without using a boosted supply. Fig. 14 shows the measured retention time distribution of a three eDRAM implementations: conventional 3T, the previous 2T [11], and the proposed 2T1C. The amount of boosting (Δ) above VDD and below GND is 0.5 V for the 2T and 3T eDRAMs whereas the 2T1C operates under a nominal power supply level. The single-ended sensing nature and the small storage capacitance of conventional 3T eDRAMs result in the poor retention characteristics. The asymmetric 2T eDRAM achieves a 400 μ s retention time for a 99.9% bit yield condition at 1.1 V and 85°C. The retention time for a 99.99%



Fig. 12. (a) Thermal map of an 8-core processor with a 24 MB L3 cache [1]. (b) Block diagram of the proposed adaptive refresh control.

bit yield is estimated to be 80 μ s at 105°C which is 2× longer than that reported for a commercial 1T1C eDRAM under the same yield and temperature condition [5]. Therefore, it is fair to say that an asymmetric 2T eDRAM has a retention time that is comparable to real product eDRAMs. However, the need for a special thick T_{OX} device would limit the wide spread adoption of asymmetric 2T eDRAMs, especially for fabless companies. The proposed 2T1C eDRAM achieves similar performance as previous designs but without any boosted supplies.

Single-ended sensing methods usually exhibit more BL failures than WL failures since variation in the dummy reference cells and the BL-S/A offset impacts the read margin of the entire BL. A decoupled 7T SRAM array was implemented to evaluate the effectiveness of a repair scheme under variation effects in the dummy cell and BL-S/A. The measured retention bit-map of a 1 kb 2T1C sub-array shows weak bit-lines as well as randomly located weak cells (Fig. 15). The proposed 7T SRAM sharing the same BL-S/A shows better stability compared to a 2T1C cell under the same operating condition. Based on the measured retention time distribution of a 2T1C array in Fig. 16(a), we can estimate the effectiveness of various repair schemes. For a target retention time of 500 μ s, a single BL repair scheme using a redundant 2T1C bitline will not work for 6.25% of the time. On the other hand, a single BL repair scheme based on a 7T SRAM for an array with 128 BLs ensures a 500 μ s retention time (150%) improvement from 200 μ s to 500 μ s) with an area overhead of just 1.23% as summarized in Fig. 16(b). Note that the cell retention time is determined by aggregated leakage current (i.e., the sum of gate-overlap, GIDL, reverse junction, and subthreshold



Process	65nm LP CMOS			
Ckt dimension	556x345µm²			
Array size	2x64kbits (Conv. 3T & Prop. 2T1C)			
Cell size	58% of 6T SRAM			
Retention time	500µs @ 1.1V, 85°C			
Random cycle time	1.40ns (714MHz) @ 1.1V			
VMIN	0.7V @ 10µs retention			
Refresh power	*161.8µW per Mb (0.28X of **6T SRAM)			
*@ 1 11/ 95% 500ucoo refresh rate				

*@ 1.1V, 85°C, 500µsec refresh rate **@ Retention voltage of 0.6V

(b)

Fig. 13. (a) Microphotograph of the 65 nm 2T1C eDRAM test chip. (b) Chip feature summary.



Fig. 14. Measured retention time distribution.

leakage) that has an exponential dependence on the PVT condition. This results in a significant variation in the measured retention time as shown in Fig. 16(a).

Fig. 17(a) shows the measured retention characteristics of the 2T1C eDRAM at 25°C and 85°C, respectively indicating a $5 \times$ retention time difference in the tail cells between the two temperatures. This implies that a significant reduction in refresh power dissipation can be achieved at lower temperatures by adjusting the refresh rate accordingly. Storage voltages were measured using the proposed monitor scheme at various temperatures and retention times as shown in Fig. 17(b). The measured



Fig. 15. Measured retention bit-map of 2T1C and decoupled 7T SRAM arrays.



Fig. 16. (a) Measured retention time distribution of the 2T1C array. (b) Effectiveness of various repair schemes.

storage voltage includes all coupling effects during memory access as well as the change in leakage currents at different temperatures. Fig. 18 shows the static current comparison between a 1 Mb SRAM in power down mode and proposed 1 Mb 2T1C eDRAM. The data retention voltages of the 6T SRAM and the 2T1C eDRAM are 0.6 V and 1.1 V, respectively. The static current of the 6T SRAM decreases exponentially at lower operating temperatures. Similarly, the static current of 2T1C eDRAM can be reduced exponentially by adjusting the refresh rate using the proposed storage voltage monitor. The static current of the proposed 2T1C eDRAM is 72% and 83% smaller than that of 6T SRAM at 85°C and 105°C, respectively. Without an adaptive refresh control, the refresh rate of the eDRAM would have to be fixed based on the worst case condition (i.e., 105°C in Fig. 18). In this case, the 2T1C eDRAM would have a larger static power



Fig. 17. (a) Measured retention time distribution of data '1' and '0' at 25° C and 85° C. (b) Measurement storage voltage for different temperature and retention time.



Fig. 18. Comparison of static current between SRAM (with power-gating) and 2T1C eDRAM (with adaptive refresh control).

dissipations than the 6T SRAM at low operating temperatures such as below 65°C as verified in our tests.

The measured VDD shmoo of random cycle time, retention time and the corresponding static power dissipations in Fig. 19 shows a wide operating voltage range from 1.4 V down to 0.8 V. One unique aspect of eDRAMs (both 1T1C and gain cell) that may not be very obvious to SRAM designers is that a lower

65nm CMOS	6T SRAM [13]	1T1C eDRAM [5]	3T eDRAM [12]	2T eDRAM [11]	This work
Cell Schematic					WBL PCOU RWL
Process	Logic compatible	Logic compatible +2 (FEOL)+3 (Cap)	Logic compatible +2 (FEOL)	Logic compatible +2 (FEOL)	Logic compatible
Boosted supplies	Not required	Required (High & Low)	Required (High & Low)	Required (High & Low)	Not required
⁽¹⁾ Reported cell size (ratio)	135F ² (1X)	30F ² (0.22X)	NA	65F ² (0.48X) [8]	NA
⁽²⁾ Redrawn cell	0.575x2.05=	0.45x0.545=	0.52x1.015=	0.48x0.995=	0.50x1.37=
size (ratio)	1.179µm ² (1X)	0.245µm² (0.21X)	0.528µm² (0.45X)	0.478µm² (0.41X)	0.685µm² (0.58X)
(2) Redrawn 1Mb	1.377x1.124=	0.632x0.739=	1.250x0.649=	1.168x0.638=	1.191x0.807=
macro (ratio)	1.548mm ² (1X)	1.548mm ² (1X) 0.467mm ² (0.30X)		0.746mm ² (0.48X)	0.961mm ² (0.62X)
Data storage	Latch (Static)	Capacitor (20fF)	MOS gate (<1fF)	MOS gate (<1fF)	MOS gate (<1fF)
Cell access	(+) Differential read	(-) Destructive read	(+) Decoupled read and	(+) Decoupled read and	(+) Decoupled read and
	(-) Ratioed operation	(-) Refresh	write, (-) Refresh	write, (-) Refresh	write, (-) Refresh
Random cycle	⁽²⁾ 1GHz	500MHz	⁽²⁾ <500MHz	⁽²⁾ 667MHz	⁽²⁾ 700MHz
Retention time @99.9% yield	NA	NA	<100µs @85ºC (<i>Meas</i> .)	400µs @85°С (<i>Meas</i> .)	500µs @85°C (<i>Meas</i> .)
Retention time @99.99% yield		40µs @105ºC (<i>Meas</i> .)	NA	80µs @105ºC (<i>Est</i> .)	50µs @105ºC (<i>Est</i> .)
Static power	1X	0.2X @500MHz	>0.95X @500MHz	0.19X @500MHz	0.28X @700MHz

 TABLE I

 Comparison Between Proposed Design and Several Other Embedded Memory Options.

⁽¹⁾All designs are in 65nm, ⁽²⁾Based on the same 65nm low power CMOS process



SRAMs where the static power goes down at lower supply voltages making $V_{\rm MIN}$ the chief design parameter. Static power in eDRAM is dominated by refresh power which can be lower at higher supply voltages due to the robust cell retention characteristics. So the operating voltage for eDRAMs should be chosen not based on the functional $V_{\rm MIN}$ of the memory, but based on the lowest power consumption point which tends to be higher than an SRAM $V_{\rm MIN}$.

Table I compares the proposed 2T1C eDRAM with several other embedded memory options in the same 65 nm LP process. The measured random cycle time of the 2T1C eDRAM is 40% faster than that of a 1T1C eDRAM while achieving a similar retention time at 105°C under a 99.99% bit yield condition. The 1T1C eDRAM has replaced 6T SRAMs in IBM's POWER7™ microprocessor [6]. Although the cycle time of the 1T1C eDRAM is $2 \times$ longer than that of 6T SRAMs, the smaller memory footprint and shorter global interconnect delay leads to a high overall cache performance. Bit-cell size and random cycle time of the proposed 2T1C eDRAM stands between those of 6T SRAM and 1T1C eDRAM, and the read and write paths can be optimized separately allowing gain cells to scale favorably in future technology nodes. Our experimental results show that gain cell based eDRAMs can be a strong contender for future embedded memories.

VI. CONCLUSION

Fig. 19. Measured VDD shmoo. (a) Random cycle time and retention time of the 2T1C eDRAM. (b) Static power dissipations of a 6T SRAM and the proposed 2T1C eDRAM.

operating voltage does not necessarily result in a lower static power consumption as shown in Fig. 19(b). This is contrary to Several circuit techniques have been presented for enabling a truly logic-compatible gain cell eDRAM with a competitive bit-cell size and improved memory performance. The proposed 2T1C gain cell utilizes a beneficial coupling that enhances read margin and a preferential boosting that improves write margin. This unique feature allows us to achieve robust DRAM operation without any boosted supplies. A decoupled 7T SRAM was seamlessly integrated as part of the array by sharing control signals with the main 2T1C array. The retention time of the 2T1C eDRAM was improved by 2.5× using the 7T SRAM based repair scheme while the repair failure rate was 6.25% when using redundant 2T1C cells. The array overhead of the 7T SRAM repair is 1.23% for a single redundant BL for every 128 BL's. The storage voltage monitor tracks the retention characteristics of the 2T1C gain cell under PVT variations while capturing realistic coupling effects during memory access. Measurement results show a 714 MHz random cycle using a 500 μ s refresh period for a 1 BL repair scheme at 1.1 V, 85°C. The static power dissipation including refresh currents and cell leakages was 161.8 μ A/Mb at 1.1 V and 85°C which is 72% lower than that of a power gated SRAM with a data retention voltage of 0.6 V.

REFERENCES

- S. Rusu et al., "A 45 nm 8-core enterprise Xeon® processor," *IEEE J. Solid-State Circuits*, vol. 45, no. 1, pp. 7–14, Jan. 2010.
- [2] S. Rusu et al., "A 65-nm dual-core multithreaded Xeon® processor with 16-MB L3 cache," *IEEE J. Solid-State Circuits*, vol. 42, no. 1, pp. 17–25, Jan. 2007.
- [3] R. J. Riedlinger et al., "A 32 nm 3.1 billion transistor 12-wide-issue Itanium® processor for mission-critical servers," in *IEEE ISSCC Dig. Tech. Papers*, 2011, pp. 84–85.
- [4] R. Kalla, B. Sinharoy, W. J. Starke, and M. Floyd, "POWER7: IBM's next-generation server processor," *IEEE Micro*, vol. 30, no. 2, pp. 7–15, Mar.-Apr. 2010.
- [5] 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, Jan. 2008.
- [6] J. Barth et al., "A 45 nm SOI embedded DRAM macro for the POWER™ processor 32 MByte on-chip L3 cache," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 64–75, Jan. 2011.
- [7] S. Romanovsky et al., "A 500 MHz random-access embedded 1 Mb DRAM macro in bulk CMOS," in *IEEE ISSCC Dig. Tech. Papers*, 2008, pp. 270–271.
- [8] 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, Jan. 2009.
- [9] W. K. Luk, J. Cai, R. H. Dennard, M. J. Immediato, and S. V. Kosonocky, "A 3-transistor DRAM cell with gated diode for enhanced speed and retention time," in *Proc. VLSI Circuits Symp.*, 2006, pp. 184–185.
- [10] K. Chun, P. Jain, J. Lee, and C. H. Kim, "A 3T gain cell embedded DRAM utilizing preferential boosting for high density and low power on-die caches," *IEEE J. Solid-State Circuits*, vol. 46, no. 6, pp. 1495–1505, Jun. 2011.
- [11] K. Chun, P. Jain, T. Kim, and C. H. Kim, "A 667 MHz logic-compatible embedded DRAM featuring an asymmetric 2T gain cell for high speed on-die caches," *IEEE J. Solid-State Circuits*, vol. 47, no. 2, pp. 547–559, Feb. 2012.
- [12] K. Chun, W. Zhang, P. Jain, and C. H. Kim, "A 700 MHz 2T1C embedded DRAM macro in a generic logic process with no boosted supplies," in *IEEE ISSCC Dig. Tech. Papers*, 2011, pp. 506–507.
- [13] K. Zhang et al., "A 3-GHz 70-Mb SRAM in 65-nm CMOS technology with integrated column-based dynamic power supply," *IEEE J. Solid-State Circuits*, vol. 41, no. 1, pp. 146–151, Jan. 2006.
- [14] J. Sim et al., "A 1.8-V 128-Mb mobile DRAM with double boosting pump, hybrid current sense amplifier, and dual-referenced adjustment scheme for temperature sensor," *IEEE J. Solid-State Circuits*, vol. 38, no. 4, pp. 631–640, Apr. 2003.
- [15] F. J. M. -Martinez, E. K. Ardestani, and J. Renau, "Characterizing processor thermal behavior," in *Proc. ACM ASPLOS*, 2010, pp. 193–204.



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 the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2000, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, 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 University of Minnesota, he rejoined

Samsung Electronics in 2012, where he has worked on low-power DRAM development. His research interests include digital, mixed-signal and memory circuit designs with special focus on DRAM, PRAM, and STT-MRAM in scaled technologies.

Dr. Chun is received a Samsung Ph.D. Scholarship for outstanding employees and the ISLPED 2009 Low Power Design Contest Award.



Wei Zhang (S'11) received the B.S. degree in electrical engineering from Tsinghua University, China. He is currently working towards the Ph.D. degree at the University of Minnesota, Minneapolis, where he has been involved in six chip designs since joining Prof. Chris Kim's research group in 2007.

He spent eight months at Broadcom Corporation as an intern engineer where he conducted research on memory-related designs in the summers of 2009 through 2011. His work is focused on low-power circuit design techniques as well as SRAM and

embedded DRAM design. He is also involved in a cooperative project on organic electronics. He is an author or co-author of nine journal and conference papers.



Pulkit Jain received the B.Tech. degree in electrical engineering from the Indian Institute of Technology (IIT) Kanpur in 2007. For the M.S. degree, his research involved power delivery issues in three-dimensional integrated circuits. He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering, University of Minnesota, Minneapolis, where he is working on circuit techniques to monitor aging and variation in circuit design.

Mr. Jain is the recipient of an IBM scholarship award and has authored/coauthored more than ten journal and conference papers.



Chris H. Kim (M'04–SM'10) received the B.S. and M.S. degrees from Seoul National University, Seoul, Korea, and the Ph.D. degree from Purdue University, West Lafayette, IN.

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, in 2004, where he is currently an Associate Professor. His research interests include

digital, mixed-signal, and memory circuit design in silicon and non-silicon (organic and magnetic) technologies.

Prof. Kim is the recipient of an NSF CAREER Award, a Mcknight Foundation Land-Grant Professorship, 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, an Intel Ph.D. Fellowship, and the Magoon's Award for Excellence in Teaching. He is an author/coauthor of over 100 journal and conference papers and has served as a technical program committee member for numerous circuit design conferences. He was the technical program committee chair for the 2010 International Symposium on Low Power Electronics and Design (ISLPED) and guest editor for a special issue of the *IEEE Design and Test Magazine*.