

A 3T Gain Cell Embedded DRAM Utilizing Preferential Boosting for High Density and Low Power On-Die Caches

Ki Chul Chun, Pulkit Jain, Jung Hwa Lee, and Chris H. Kim, *Senior Member, IEEE*

Abstract—Circuit techniques for enabling a sub-0.9 V logic-compatible embedded DRAM (eDRAM) are presented. A boosted 3T gain cell utilizes Read Word-line (RWL) preferential boosting to increase read margin and improve data retention time. Read speed is enhanced with a hybrid current/voltage sense amplifier that allows the Read Bit-line (RBL) to remain close to VDD. A regulated bit-line write scheme for driving the Write Bit-line (WBL) is equipped with a steady-state storage node voltage monitor to overcome the data ‘1’ write disturbance problem of the PMOS gain cell without introducing another boosted supply for the Write Word-line (WWL) over-drive. An adaptive and die-to-die adjustable read reference bias generator is proposed to cope with PVT variations. Monte Carlo simulations compare the 6-sigma read and write performance of proposed eDRAM against conventional designs. Measurement results from a 64 kb eDRAM test chip implemented in a 65 nm low-leakage CMOS process show a 1.25 ms data retention time with a 2 ns random cycle time at 0.9 V, 85 °C, and a 91.3 μ W per Mb static power dissipation at 1.0 V, 85 °C.

Index Terms—Cache, logic-compatible eDRAM, low-power, low-voltage, 3T gain cell.

I. INTRODUCTION

POWER dissipation has become the chief performance limiter in modern microprocessors, triggering a flurry of research activities on low-power design techniques. One of the most effective ways to curb chip power is to integrate more memory: a larger cache memory improves micro-architectural performance with only a modest increase in CV^2f power. As a result, the past decade has seen a precipitous increase in the amount of on-die embedded memory. Approximately half the chip area is devoted to cache memory in state-of-the-art designs. For example, Intel’s 8-core Nehalem processor has 24 MB of shared L3 cache based on SRAM cells [1] while IBM’s POWER7 processor has a 32 MB L3 cache built in an embedded DRAM (eDRAM) technology [2]. The need for robust high-density embedded memories is projected to grow as designers continue to seek power-conscious ways to improve chip performance.

In order to maintain the historical growth in chip performance, designers must continue to deliver memory solutions

for achieving low static power and high operating speed. SRAMs have been the embedded memory of choice due to their logic compatibility and fast access time. Recently, embedded DRAMs (eDRAMs) have been gaining popularity in the research community due to features such as small cell size, low cell leakage, and non-ratioed circuit operation. There have been a number of successful eDRAM designs based on traditional 1T1C DRAM cells as well as logic-compatible gain cells [2]–[9]. 1T1C cells are denser than gain cells, but at the cost of a capacitor process, and the noise margin is reduced substantially at low voltages as the read operation is based on the charge sharing principle. Gain cells are made of logic devices allowing them to be built in a standard CMOS process with minimal alteration. The cell can be implemented using three transistors, or even two transistors when used with delicate read control circuits, achieving roughly 2x higher bit cell densities than SRAMs [7]–[9]. Furthermore, gain cells can have smaller cell leakage current than SRAMs in sleep mode due to the smaller number of devices and the super cut-off biasing condition. The write margin is better than SRAMs since there is no contention between the access device and the cross-coupled latch in an eDRAM cell. Despite these favorable features, conventional gain cells suffer from short data retention times due to the small storage capacitor and various leakage sources in the presence of process–voltage–temperature (PVT) variation [7], [8] requiring careful margin distribution, cell tracking, and reference voltage control. Short retention times result in larger refresh power and poor read performance. In this work, we address the aforementioned challenges by proposing various circuit techniques to improve the data retention time of gain cell based eDRAMs.

The remainder of this paper is organized as follows. Section II introduces the basic operation of a conventional 3T eDRAM gain cell. Section III presents the proposed circuit techniques to enhance the data retention time and improve the read speed of gain cell eDRAMs. Section IV compares access speeds and power dissipations of 6T SRAM and 3T eDRAM arrays by running Monte Carlo simulations. Section V describes hardware measurement results from a 65 nm test chip. Conclusions are given in Section VI.

II. BASIC OPERATION OF A CONVENTIONAL 3T eDRAM GAIN CELL

To aid the understanding of our proposed techniques, in this section, we first describe the basic operation of a conventional 3T eDRAM gain cell. Fig. 1(a) shows the cell schematic and Fig. 1(b) summarizes the signal conditions for each operating mode. PMOS devices are chosen over NMOS devices because

Manuscript received June 14, 2010; revised October 02, 2010; accepted January 23, 2011. Date of publication May 05, 2011; date of current version May 25, 2011. This paper was approved by Associate Editor Sreedhar Natarajan.

K. C. Chun, P. Jain, and C. H. Kim are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: chunx041@umn.edu).

J. H. Lee is with the DRAM Design Team, Memory Division, Samsung Electronics, Hwasung, Kyeonggi-Do 445-701, Korea.

Digital Object Identifier 10.1109/JSSC.2011.2128150

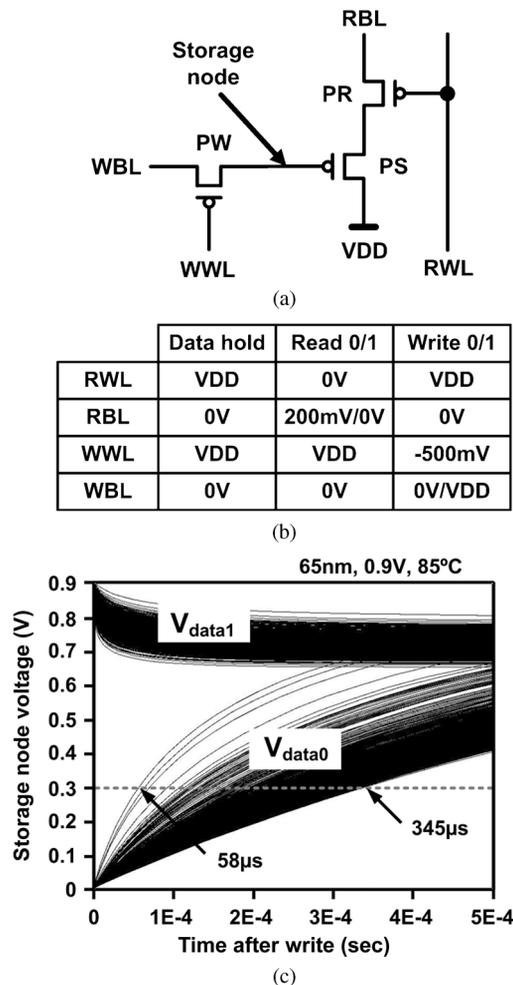


Fig. 1. (a) Conventional 3T PMOS eDRAM gain cell circuit diagram. (b) Signal voltages in each operating mode. (c) Monte Carlo simulation results of storage node voltage during data hold mode. Results are shown for 1024 Monte Carlo iterations which is equivalent to the cell-to-cell variation of a 1 kb array.

they have significantly less gate tunneling leakage current, which extends the data retention time [8], [9]. This preference may not hold in the future where high- k gate dielectrics become prevalent. The operating principle of an NMOS cell is identical to that of a PMOS cell with the only difference being the signal polarities. In the 3T PMOS cell, PW denotes the write access device, PS denotes the cell storage device, and PR denotes the read access device. In write (or write-back) mode, the write bit-line (WBL) data is written into the storage node through PW.

Similar to a 1T1C eDRAM cell, the write word-line (WWL) is negatively over-driven so that a 0 V can be written into the cell without the threshold voltage loss. In read mode, the pre-discharged read bit-line (RBL) voltage is pulled up only when the voltage stored in the gate of PS is low. In case the storage voltage is high, PS is off so RBL remains at the pre-discharged level. Cell data can be determined by comparing the RBL voltage with a reference RBL, whose level is between the data '1' and data '0' RBL levels, using a sense amplifier. During hold mode, PW and PR are turned off and the storage node is left floating. The sub-threshold, gate, and junction leakages in the surrounding

devices make the floating voltage change with time as shown in Fig. 1(c). Since the storage node is surrounded by high voltages in the PMOS cell, the retention time of data '0' is much shorter than data '1'. Similarly, the retention time of data '1' becomes critical in an NMOS cell where the surrounding signal voltages are 0 V during hold mode. The data retention time is directly related to the aggregated leakage currents flowing into the storage node.

In the presence of process variation, each cell in a memory array will have different retention characteristics so the cell with the shortest retention time (after applying any redundancy schemes to remove bad cells) will determine the refresh rate of the entire eDRAM array. Fig. 1(c) shows the simulation results of cell retention time variation. This plot was obtained by running Monte Carlo simulations in HSPICE with 1 k iterations, which gives a cell-to-cell variation equivalent to a 1 kb array. Results indicate that the time it takes for the data '0' voltage to rise to a specific voltage (0.3 V in this simulation to guarantee a 0.3 V gate over-drive voltage in the storage transistor which has a V_{TP} of 0.3 V) ranges from 58 μ s to 345 μ s at a 0.9 V supply voltage and 85 $^{\circ}$ C temperature. Poor retention characteristics of tail cells result in a large refresh current and decreased read performance. Therefore, increasing the cell retention time is the foremost challenge in low voltage gain cell eDRAMs.

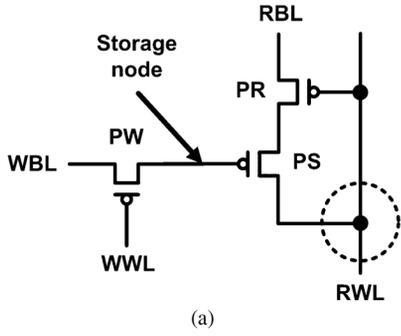
III. PROPOSED BOOSTED 3T eDRAM DESIGN

In this section, we present three circuit techniques to improve the eDRAM data retention time and ensure robust circuit operation under PVT variations.

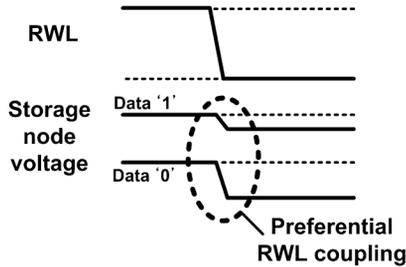
A. Boosted 3T eDRAM Gain Cell

The retention time and read speed of eDRAMs are highly dependent upon the storage node voltage at the time the cell is accessed. Even a small signal loss can cause severe speed degradation at low operating voltages. Fig. 2(a) shows the proposed 3T PMOS gain cell which can preferentially boost the storage voltage via capacitive coupling. Unlike the conventional design in Fig. 1(a), the drain of the storage device PS is connected to the RWL signal instead of the supply voltage. For read operation, RBL is first precharged to VDD and then the RWL switches from VDD to 0 V. The resultant bitline signal is detected by a sense amplifier.

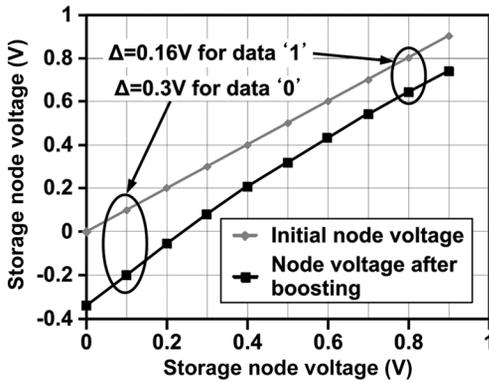
The central idea of the proposed cell is to preferentially boost the storage node voltage using the RWL signal for improving the cell's data retention capability. For example, consider the case when the storage node voltage is low (e.g., 0 V). This will make the gate-to-RWL coupling capacitance larger compared to when the storage node voltage is high (e.g., VDD). PS in inversion mode makes the entire oxide capacitance act as the coupling capacitance whereas PS in weak-inversion mode, the significantly smaller depletion capacitance acts as the coupling capacitance. Since a lower storage voltage has a larger coupling capacitance, it is coupled down more than a higher storage voltage when the RWL switches from high to low as shown in Fig. 2(b). This preferential boosting action amplifies the signal difference during read which allows the storage node voltage to decay further before it needs to be refreshed. This translates into a longer effective data retention time. A similar concept was



(a)



(b)



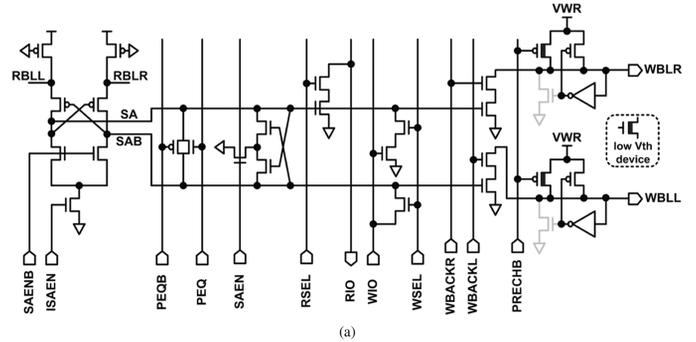
(c)

	RWL	RBL	WWL	WBL
Data hold	VDD	VDD	VDD	0V
Read 0/1	~0V	~VDD	VDD	0V
Write 0/1	VDD	VDD	-500mV	0V/VDD

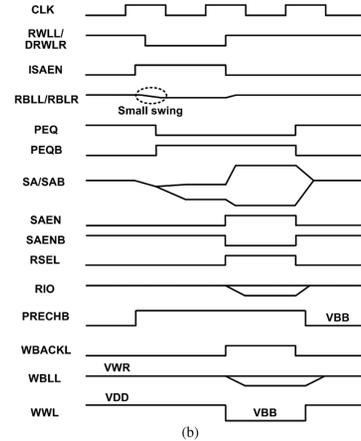
(d)

Fig. 2. (a) Proposed boosted 3T PMOS eDRAM gain cell. (b) Preferential RWL coupling effects of the proposed cell. (c) Simulation results of the storage node preferential boosting effects. (d) Signal voltages for each operating mode.

proposed by Luk *et al.*, where a 3T1D cell was used to boost the cell voltage [7]. However, this cell structure requires an additional diode device which increases the cell area as well as the gate tunneling leakage. It also has a limited signal amplification effect since the storage device acts as a parasitic capacitor limiting the amount of coupling that can be achieved. The proposed boosted 3T gain cell can provide a stronger coupling effect with only three transistors, increasing data retention time, enhancing the RBL margin and improving read performance. Simulation results in Fig. 2(c) verify that the data ‘0’ voltage is amplified by 0.3 V while the data ‘1’ voltage is coupled down by only 0.16 V. In addition to the amplification effect, the proposed cell can provide a ~2x larger current than conventional 3T gain cells since the boosted voltage provides a higher gate overdrive for PS. It



(a)



(b)

Fig. 3. (a) Hybrid bit-line current/voltage sense amplifier (S/A) with read port, write port, and write-back circuits. (b) Read and write-back timing diagram of the proposed S/A.

should be pointed out that the higher drive current is only observed when the RBL level is high, as the read current quickly diminishes as the RBL voltage drops due to the V_{TP} loss in the PMOS read device. To utilize the boosted read current of the proposed 3T cell, we employ a hybrid current/voltage sense amplification technique that keeps the RBL level close to VDD during the read operation [10], [11].

Fig. 3 shows the schematic and timing diagram of the bit-line sense amplifier (S/A) consisting of a hybrid current/voltage S/A, read port, write port and drivers for write-back. During read, the RBL signals to the current S/A are amplified and converted to voltage signals through a cross-coupled PMOS pair and a NMOS resistor pair while a load PMOS pair keeps the RBL swing small. After transferring the input differential current, the cross-coupled PMOS pair, in tandem with the cross-coupled NMOS pair, acts as a voltage S/A which generates a full CMOS swing signal. Dedicated timing control circuits are implemented for the equalizer to ensure stable current S/A operation as shown in Fig. 3(b). The write-back operation automatically follows the read cycle to refresh the cell data.

B. Regulated Bit-Line Write Scheme

When the WBL is driven to data ‘1’, the data ‘0’ levels in the unselected cells on the same WBL are pulled up by the sub-threshold leakage through the write access PMOS devices as shown in Fig. 4(a). Most DRAM designs use a boosted supply for the WWL to prevent the signal loss in the unselected cells by asserting a negative V_{gs} in the write access devices.

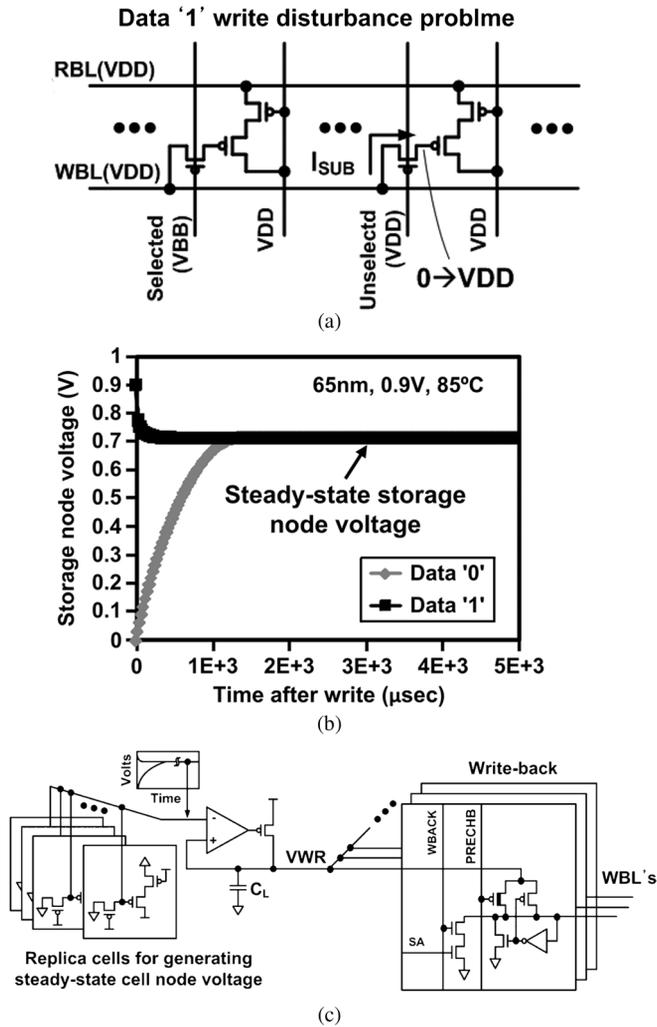


Fig. 4. (a) Storage node disturbance problem when writing data '1' to a cell sharing the same WBL. (b) Simulation results showing steady-state storage node voltage in case of no refresh. (c) Proposed regulated bit-line write bias generator based on replica cells.

However, this method incurs area and power penalty due to the large charge pump capacitors and poor pumping efficiency at low voltages. In this work, we propose a regulated bit-line write scheme which can eliminate the data '1' disturbance issue without having to generate an additional boosted supply.

Without a refresh, the storage node voltage eventually converges to a steady-state level close to VDD regardless of the initial cell voltage as shown in Fig. 4(b). In our design, we use this steady-state voltage level for writing data '1', as it will produce a negative V_{gs} in all the unselected cells without impacting the retention time of the selected cell. Note that the retention time is determined by the data '0' cell voltage rather than the data '1' voltage in a PMOS gain cell. A steady-state storage node voltage monitor shown in Fig. 4(c) is implemented with replica cells biased in hold mode, followed by a voltage down converter to drive the large WBL load. The speed loss due to the regulated bit-line write voltage (VWR) is prevented by pre-charging the WBL to VWR using the negative supply VBB as the gate signal, which is readily available on-chip for the WWL under-drive.

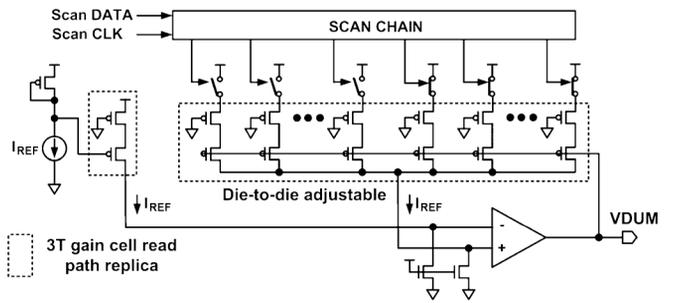


Fig. 5. PVT-tracking and die-to-die adjustable read reference bias (VDUM) generator.

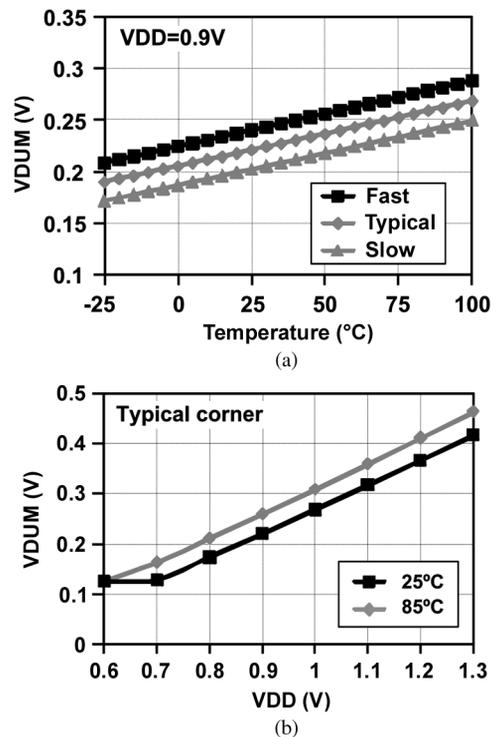


Fig. 6. (a) Simulation results of the proposed VDUM generator tracking temperature and process variations. (b) Simulation results showing the dependency of VDUM on VDD.

C. PVT-Tracking Read Reference Bias

An optimal bias voltage (VDUM) is applied to the reference dummy cells to maximize the read operating margin. VDUM must be carefully chosen as it affects both the data retention time and the read speed; a higher VDUM level improves the data retention time at a read speed penalty. Fig. 5 shows the proposed PVT-tracking and die-to-die adjustable read reference bias generator to cope with PVT variations. The negative feedback circuit tracks the desired cell read reference current (I_{REF} in the figure). Fig. 6 shows simulation results of the proposed VDUM level under PVT variations. Unlike previous designs which use a fixed VDUM level or a simple averaging scheme [8], our circuit can achieve the target retention time without sacrificing read speed by adaptively lowering the VDUM level at low leakage PVT conditions as shown in Fig. 6. For example, at lower temperatures or in slow corner dies, the excess retention time is traded off for faster read speed by lowering the VDUM level.

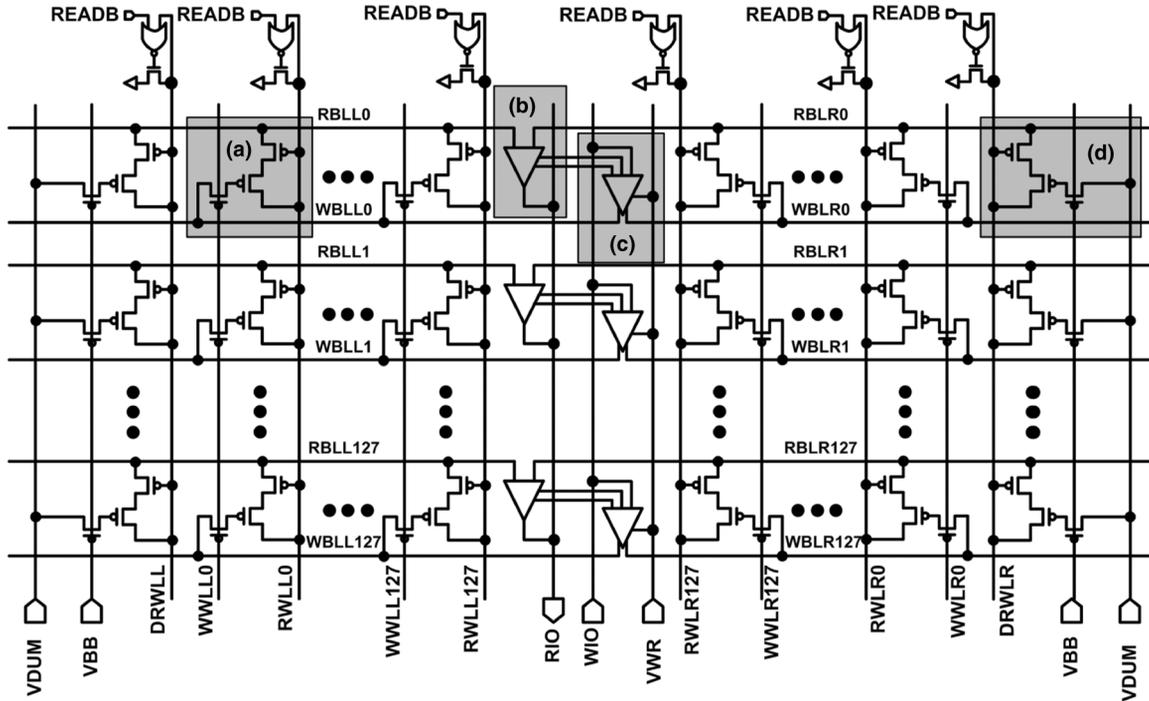


Fig. 7. A 32 kb array structure of the proposed eDRAM including (a) boosted 3T gain cell, (b) hybrid current/voltage S/A, (c) regulated bit-line write scheme, and (d) PVT-tracking read reference scheme.

Similarly, at low supply voltages, the VDUM level is shifted down since the reduced leakage make the storage node voltage lower compared to at high supply voltages for the same retention time. Binary weighted read path replica branches are implemented to precisely adjust the VDUM level according to the retention characteristics and read performance of each chip.

D. Architecture and Operation of a 32 kb Sub-Array

A detailed circuit diagram of the 32 kb boosted 3T array is shown in Fig. 7. The array has 128 cells per WL and 128 cells per split BL, which share a common BL S/A located at the center of the array. The proposed VDUM bias is connected to the dummy cells placed at both edges of the array, and the VWR bias is connected to the write-back circuitry of the BL S/A. The RWL pull-down keepers are located at the top row of the array to keep the ground noise of the activated RWL as small as possible. HSPICE simulations indicate a 66 mV RWL ground noise at 0.9 V, 85 °C when all cells connected to the same RWL contain data '0' which corresponds to the worst case scenario.

Fig. 8 shows simulation waveforms of read and write-back operations with a 2 ns random cycle time. A two-stage full pipeline structure was implemented to control read and write-back operations. At the first clock cycle, RWL is selected, and this amplifies the cell node by preferential coupling. When the current S/A control signal (ISAEN) is enabled, the current S/A amplifies its input signals to analog voltage signals with RBL held close to VDD. After achieving a recognizable voltage difference, the voltage S/A control signal (SAEN) is enabled. At the second clock cycle, read-out and write-back operations are

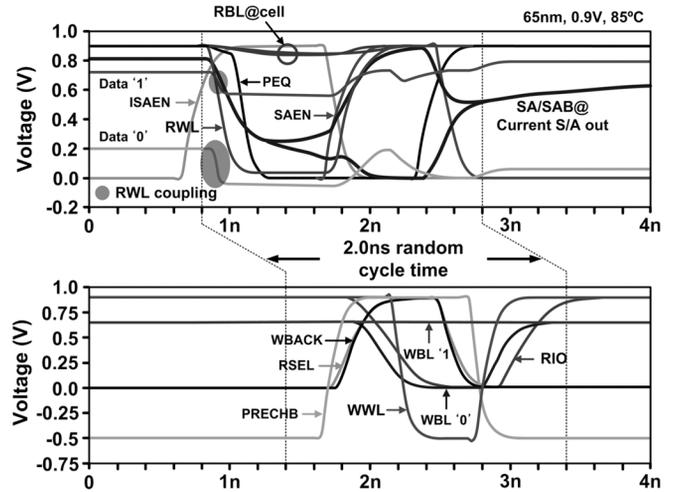


Fig. 8. Read and write-back simulation waveforms with a 2 ns random cycle time.

followed. After write-back, discharged WBLs are pre-charged using the negative supply VBB control signal (PRECHB).

IV. STATISTICAL SIMULATION RESULTS FOR 6T SRAM AND 3T EDRAM ARRAYS

This section presents Monte Carlo simulation results on megabit density SRAM and eDRAM arrays to estimate their speed and power in a practical scenario [12]. An operating voltage of 0.9 V was chosen (nominal operating voltage of the 65 nm process used is 1.2 V) so that cell failures exist in the small 32 kb unit test array. Table I summarizes the simulation

TABLE I
SIMULATION SETUP FOR 1 M MONTE CARLO ITERATIONS

		CONV 3T eDRAM	Proposed 3T eDRAM	6T SRAM
0.9V, 85°C, 1M Monte Carlo full array simulation using a 1.2V, 65nm, LP CMOS process				
Read operation	Cell node voltage	@100μs with voltage distribution under T_{OX} and V_{TH} variations		N/A
	Reference bias	Adaptive VDUM with 10% variations		N/A
	Cell	Device mismatches		
	Dummy cell	Dummy cell averaging scheme [8]	4X upsized device mismatches	N/A
	Current S/A	N/A	S/A pair mismatches	N/A
Write operation	Boosted supply	-0.5V with 10% variations		N/A
	Cell	Device mismatches		

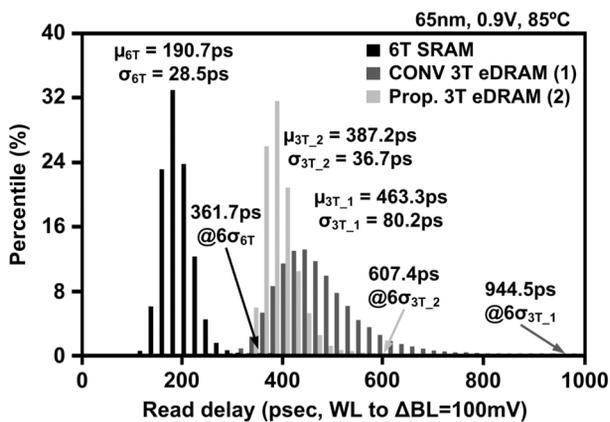


Fig. 9. Read performance comparisons between 6T SRAM and 3T eDRAM obtained from 2^{20} Monte Carlo iterations. Results are equivalent to the distribution of a 1 Mb macro array. 6T SRAM has the shortest bitline delay attributed to the differential swing nature and large drive current (361.7 ps @ 6σ) followed by the proposed 3T eDRAM (607.4 ps @ 6σ) and the conventional 3T eDRAM (944.5 ps @ 6σ).

setup for the Monte Carlo iterations including assumptions on the mismatch and voltage variations.

A. Read and Write Performance

Fig. 9 shows read bitline delay distributions with average and 6-sigma point delays annotated for the following three memory arrays; a 1 Mb SRAM, a 2 Mb conventional 3T, and a 2 Mb boosted 3T. Simulation results were obtained from 2^{20} Monte Carlo iterations. The peripheral circuit delay, which is a function of the unit sub-array size, and the global interconnect delay, which is a function of the total cache area, are identical for the three simulated arrays since we selected an SRAM with half the number of cells as the eDRAMs. Recall that an SRAM bitcell is about twice the area of an eDRAM bitcell. The single-ended sensing nature and the gradual loss in the storage node voltage of the conventional 3T eDRAM result in a 6-sigma read bit-line delay that is 2.6 times longer than a 6T SRAM as shown in

Fig. 9. The proposed 3T eDRAM with preferential amplification effect partially makes up for this performance shortfall, improving the bit-line sensing speed by 36% compared with the conventional 3T eDRAM. Although 6T SRAMs still have a 40% faster sensing delay than the proposed circuit, we will see later that their performance becomes worse than eDRAMs for large cache sizes due to the longer global interconnect delay. Fig. 10 shows detailed cell layouts of various logic-compatible embedded memory cells drawn using a standard 65 nm logic design rule. The dense bitcell design rules were not available to the authors but for area comparison purposes, using a logic design rule is generally sufficient. The four signal wire lines and the three transistors of the conventional and boosted 3T gain cells are marked in Fig. 10. The proposed boosted 3T gain cell is 47% smaller than a 6T SRAM cell. Fig. 11 shows latency comparison results between a 6T SRAM array and the boosted 3T eDRAM array for two different cache sizes. The latency of a cache shown in Fig. 11 consists of the bit-line sensing time (6-sigma value from Fig. 9), the peripheral circuit delay, and the global interconnect delay. The boosted 3T eDRAM achieves faster access times for cache sizes greater than 16 Mb (or 2 MB) owing the shorter interconnect delay made possible by the smaller bitcell.

Fig. 12 shows the 1 Mb write delay distributions of a 6T SRAM array and the proposed 3T eDRAM array. Here, the write delay is defined as the WL signal to the time when the cell node reaches 95% of the full voltage swing. The write speed of the gain cell is faster than the 6T SRAM since the latter is based on a ratioed operation. Note that the WWL of the gain cell must be sufficiently negative in order for the PMOS write devices to pass a good data '0' level. For a WWL under-drive voltage of -0.5 V, the 1 Mb Monte Carlo simulations show a write speedup of 17% (6-sigma point) for the boosted 3T eDRAM.

B. Static Power Consumption

Static power consumption of an eDRAM system consists of two main components: (i) the leakage current of the cell it-

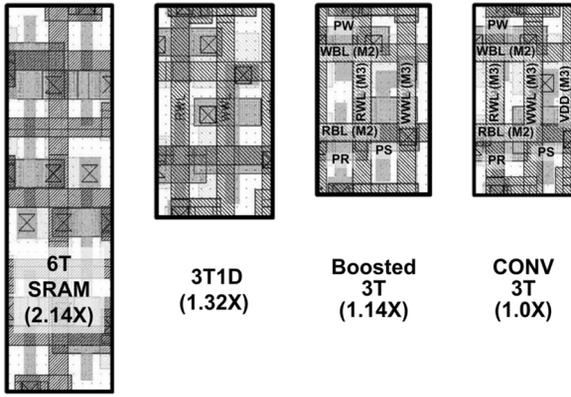


Fig. 10. Comparison of various logic-compatible embedded memory cell layouts using a 65 nm logic design rule (the authors did not have access to the dense bitcell design rule but for area comparison purposes, the logic design rule is generally acceptable). The outer box represents the cell boundary. Signal names, wire tracks, and device names are marked for the boosted 3T and conventional 3T cells.

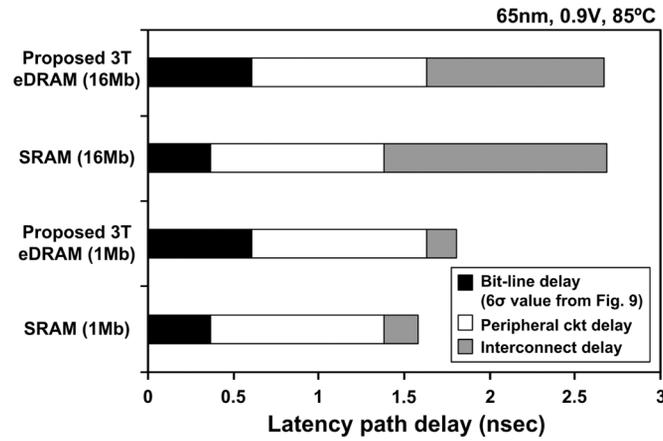


Fig. 11. Latency comparisons between SRAM and 3T eDRAM for 1 Mb and 16 Mb cache sizes. Gain cells have a shorter interconnect delay due to the smaller cell size making their performance favorable in larger arrays.

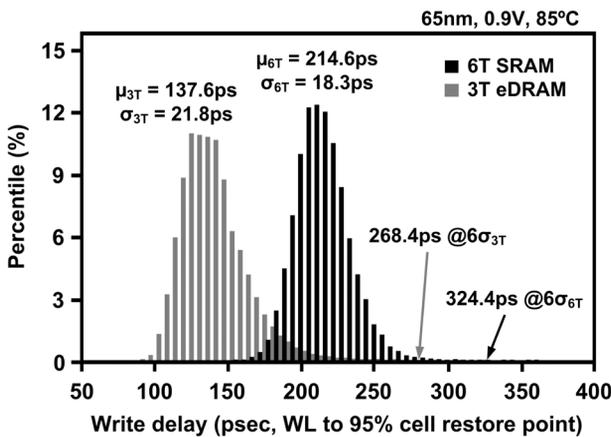
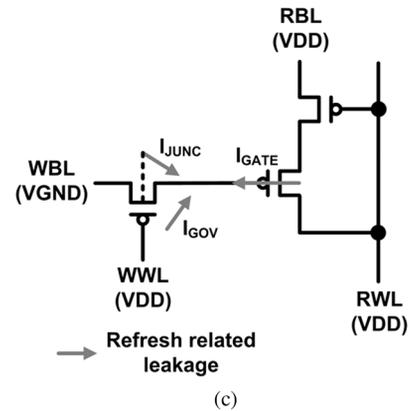
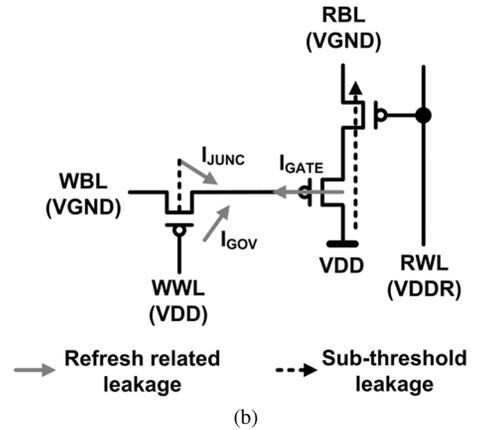
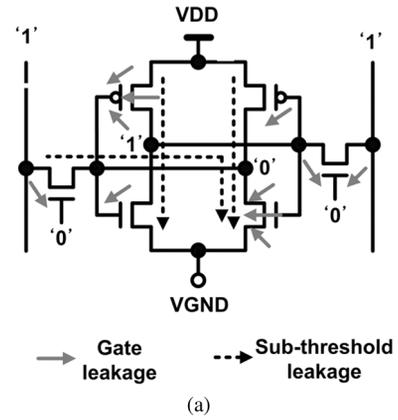


Fig. 12. Write delay distributions of 1 Mb arrays using 6T SRAM and 3T eDRAM.

self and (ii) the refresh power to keep the data “alive.” The refresh operation is a dummy read followed by a write-back cycle which simply reinforces the cell data. Hence, the refresh power is inversely proportional to refresh period. The data ‘0’ storage node voltage should be kept sufficiently low so that the PMOS read device can provide enough drive current that meets



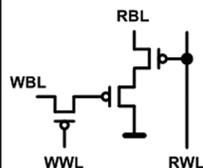
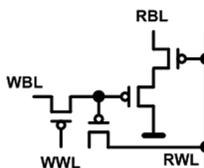
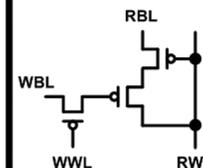
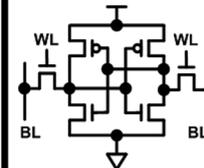
		6T SRAM	CONV 3T eDRAM	Proposed 3T eDRAM
Active mode	VDD	1.0V	1.0V	1.0V
	VGND	0V	0V	0V
	RWL	-	1.0V	1.0V
	Norm. cell leak.	1X	0.12X	0.00076X
Sleep mode (w/ data retention)	VDD	1.0V	1.0V	1.0V
	VGND	*0.4V (DRV)	0V	0V
	RWL	-	1.3V	1.0V
	Norm. cell leak.	0.22X	0.0019X	0.00076X

*Lowering the VDD is another option, albeit less effective compared to raising the VGND in the 65nm LP process used in this work

Fig. 13. Leakage components of a (a) 6T SRAM, a (b) conventional 3T eDRAM and the (c) proposed 3T eDRAM. (d) Bias conditions and normalized cell leakages of SRAM and 3T eDRAM in active and sleep modes.

the target read speed. This criterion determines the refresh period as pointed out in Section II. Fig. 13(a), (b), (c) illustrates

TABLE II
COMPARISON OF LOGIC-COMPATIBLE EMBEDDED MEMORIES

65nm, 0.9V, 85°C				
	CONV 3T (2T [8])	3T1D [7]	Proposed 3T [9]	6T SRAM [13]
*Cell schematic				
Features	Small size	Partial storage node amplification	Full storage node amplification	Fast
Issues	Short retention time	Additional device	RWL noise	Large size, low noise margin
**Cell size (ratio)	0.54x1.02=0.551μm ² (1.0X)	0.64x1.14=0.73μm ² (1.32X)	0.615x1.02=0.627μm ² (1.14X)	0.575x2.05=1.178μm ² (2.14X)
***RWL-BL delay (Δ=100mV)	945ps	794ps	607ps	362ps
***WWL-Cell 95% restore delay	-	-	268ps	324ps
Latency (simulated)	-	-	1.81ns @1Mb 2.67ns @16Mb	1.58ns @1Mb 2.69ns @16Mb
Retention time	110μs (measured)	200μs (simulated)	1.25ms (measured)	-
Static power	Large due to short retention time	Medium	Small	Large due to transistor leakage

* PMOS cells for low I_{gate} , ** 65nm logic design rule, *** Monte-carlo 6σ simulation results

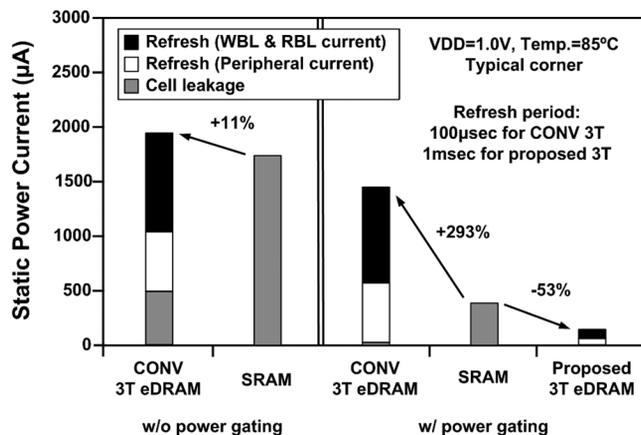


Fig. 14. Static power comparisons between a 1 Mb SRAM and a 2 Mb 3T eDRAM. Leakage power of the peripheral circuit is assumed to be negligible.

the leakage components in the three memory cells. Due to the higher number of devices per cell, there are more leakage paths from the supply to the ground in a 6T SRAM cell than in the 3T eDRAM cells. Since the leakage current through the storage node has to be extremely small in an eDRAM cell for it to be viable (e.g., $> 100 \mu s$ retention time), the main cell leakage component is through the read access device. In other words, the refresh related leakages shown in Fig. 13(b) and (c) are much smaller than the leakage current through the read access device.

Fig. 14 compares the static power consumption of a 1 Mb 6T SRAM array and a 2 Mb 3T eDRAM array with a $100 \mu s$ refresh period. HSPICE simulations were performed using a 65 nm

low-leakage CMOS process at 1.0 V, 85 °C (typical corner). Again, the number of cells of the 3T eDRAM array was chosen to be twice that of the SRAM array to account for the $\sim 50\%$ smaller cell size. Note that the eDRAM's higher density makes up for its longer latency improving the overall architectural performance [3], [4]. Simulation results show that the static power of a 2 Mb conventional 3T eDRAM array is similar to that of a 1 Mb SRAM during active mode. The refresh current consists of the RBL and WBL switching currents for the dummy read and write-back operations, as well as the refresh control power in the peripheral circuits. The refresh power constitutes 75% of the total eDRAM static power for a $100 \mu s$ refresh period.

Most embedded memories are now equipped with sleep mode capability, so it is important to compare the sleep mode power between SRAM and the proposed eDRAM. When power gating and wordline overdrive techniques shown in Fig. 13(d) are applied, the cell leakage component is reduced in both the SRAM and the eDRAM arrays [13], [14]. Since refresh power is not affected by these sleep techniques, the eDRAM's total static power becomes 3x larger compared to the SRAM's even with an additional boosted high supply for the RWL to suppress the read path sub-threshold leakage as shown in Fig. 13(b). Our proposed 3T eDRAM cell significantly reduces the refresh power component as it has a 10x longer retention time without any extra boosted supply. This makes the static power of the proposed eDRAM 53% less than that of a power gated SRAM, as shown in Fig. 14.

Table II summarizes simulation and layout results of various logic-compatible embedded memory cells.

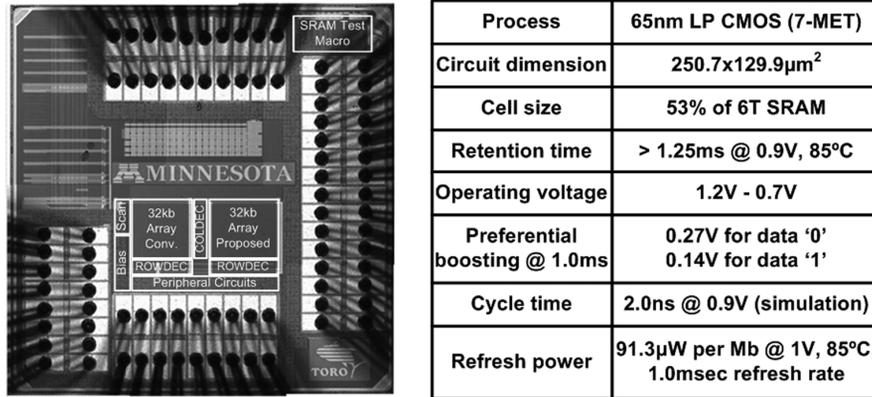


Fig. 15. Microphotograph of the 65 nm eDRAM test chip and feature summary.

V. TEST CHIP IMPLEMENTATION AND MEASUREMENTS

A proof-of-concept 64 kb eDRAM test chip was built in a 1.2 V, 65 nm low-leakage logic CMOS process to demonstrate the proposed circuit techniques. In order to fully verify the proposed techniques against the existing ones, each sub-array has a different combination of cell structure (boosted 3T vs. conventional 3T), reference scheme (proposed PVT-tracking vs. cell averaging [8]), and write scheme (conventional vs. regulated bit-line write). Fig. 15 shows the chip microphotograph and feature summary of the 64 kb eDRAM test chip fabricated in a 1.2 V, 65 nm low-leakage logic CMOS process. Fig. 16(a) shows the measured VWR levels at different supply voltages. The data '1' voltage (i.e., VWR) is high enough to keep the storage transistor off: the PMOS threshold voltage (V_{TP}) of this process is 0.315 V at 85 °C and the measured VWR level is slightly lower than $V_{DD} - V_{TP}$. The unselected cells undergoing the data '1' disturbance situation are not affected since a sufficient amount of negative V_{gs} is applied to the write access transistor. The VWR level is determined by the balance between the sub-threshold, gate, and junction leakage components. In most cases, sub-threshold leakage is the dominant factor in determining the VWR level. At high temperature and high VDD conditions however, the junction and gate leakage components have a stronger affect on the VWR level than the sub-threshold leakage component resulting in higher level over 1.1 V as shown in Fig. 16(a).

By externally adjusting the VDUM voltage, we can indirectly and noninvasively measure the storage node voltage at different data retention times. For example, read failure will happen for data '0' if the VDUM level is lower than the storage node voltage so the storage voltage can be measured by sweeping the VDUM voltage and measuring the failure point. It is worth mentioning that the storage node voltage measured using this method include effects such as process variation or transient noise (e.g., coupling noise or supply noise) providing us with an "effective" cell node voltage. Fig. 16(b) shows the measurement results of the storage node voltage of the proposed regulated write scheme compared with the conventional 3T gain cell under the data '1' disturbance condition. The data retention characteristics of the data '1' disturbance case and the data hold mode case are virtually identical when using the proposed regulated bit-line write scheme.

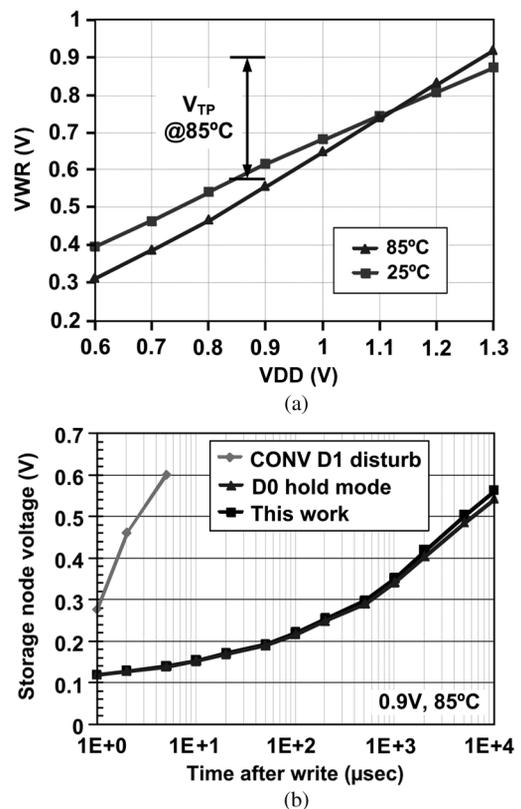


Fig. 16. (a) Measured regulated bit-line write bias (VWR) level. (b) Storage node voltage measurement results under data '1' disturbance conditions.

Fig. 17(a) shows the data retention characteristics of the conventional 3T and the proposed boosted 3T from the same test chip, including the cell-to-cell retention time variation. The retention time was for a read speed (i.e., RWL enable to voltage S/A enable interval) of 1.0 ns at 0.9 V and 85 °C. This translates into a 2.0 ns cycle time. The proposed boosted 3T design achieves a data retention time of 1.25 ms at 0.9 V, 85 °C, which is a 10 \times improvement over the conventional 3T cell measured from the same silicon die. Note that due to limitation in the test setup, only 32 cells were measured from each sub-array. As a point of reference, the target retention time of a 2T gain cell eDRAM was 10 μs in [8] and the measured retention time of a 1T1C eDRAM was 40 μs in [4].

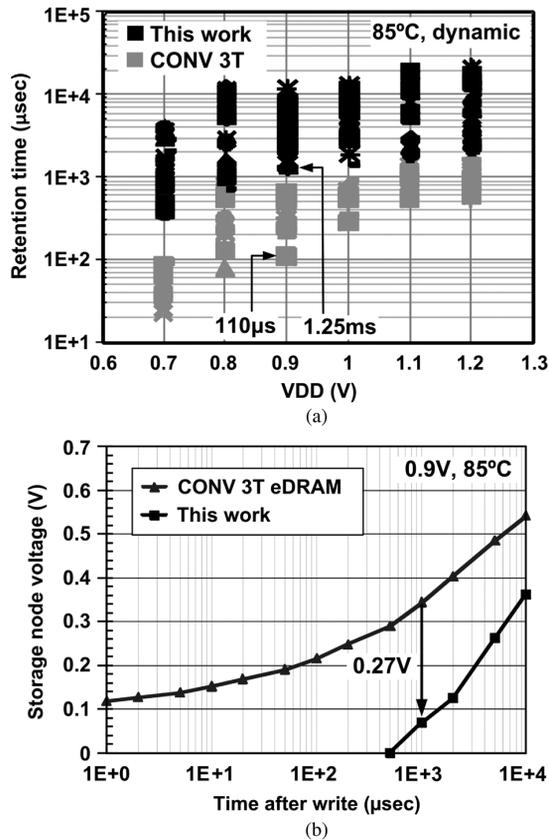


Fig. 17. (a) Measured retention time statistics. Due to limitations in the test setup, only 32 cells were measured from each sub-array. The measured cells were located evenly across the memory array. (b) Measured storage node voltage in the proposed boosted 3T cell and the conventional 3T cell. The cell voltage was indirectly and noninvasively measured by sweeping the reference cell node voltage.

Similar to Fig. 16(b), Fig. 17(b) shows the measured storage node voltage of the proposed boosted 3T and the conventional 3T gain cell. Due to threshold voltage variations between the read devices and the WWL coupling effect after the write-back, the data '0' voltage of the conventional 3T started at around 0.1 V. Read failures start to occur when the cell voltage is higher than around 0.2 V for the conventional 3T. The amount of cell node boosting of the proposed cell was 0.27 V after a 1.0 ms of hold time. The preferential boosting effect can be clearly observed in the measured data as the difference between the two curves diminishes at longer hold times. Note that the VDUM level could not be lowered below 0 V in the test chip, so although a large negative cell voltage is expected at short retention times, we were only able to measure the positive cell voltages as shown in Fig. 17(b). This is sufficient as we are more interested in measuring the positive storage node voltage region which is when the memory operation starts to fail.

Figs. 18(a) and (b) show the measured storage node voltage of data '1' and data '0' enabling a 2.0 ns random cycle time at 0.9 V, for high (85 °C) and room (25 °C) temperature corners, respectively. Optimal VDUM levels to achieve longer retention time with fixed read speed were 0.2 V for high temperature and 0.14 V for room temperature. Fig. 18(c) shows the measured VDUM level at high and room temperature corners

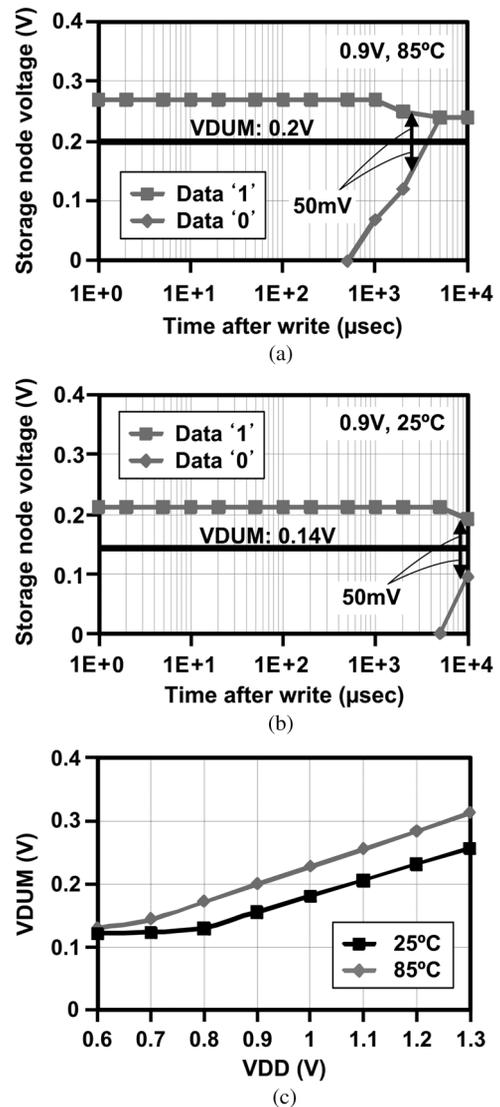


Fig. 18. Measured storage node voltages at (a) 85 °C and (b) 25 °C. (c) Measured PVT-tracking read reference (VDUM) level at different supply voltages.

for various supply voltages. VDUM level change across a temperature range of 25 °C to 85 °C and a supply voltage range of 0.8 V to 1.3 V was 50 mV. The 50 mV voltage difference is approximately the threshold voltage difference between the two temperature conditions.

VI. CONCLUSION

Circuit techniques have been presented for increasing the data retention time and enhancing the performance of gain cell eDRAMs. The proposed boosted 3T eDRAM cell preferentially boosts the cell voltage to obtain high performance and low static power dissipation, with a layout penalty of only 14% compared to a conventional 3T cell. The proposed regulated bit-line write scheme can eliminate the data '1' write disturbance problem without introducing another boosted supply for WWL. The measurement results show the 1.25 ms data retention time with 2 ns random cycle time at 0.9 V, 85 °C, which is a 10× improvement compared to a conventional 3T gain cell measured from the same silicon die. The measured static power

dissipation from a 64 kb test chip with the proposed schemes was $91.3 \mu\text{W}$ per Mb at 1.0 V, 85°C , and 1.0 ms refresh period, which is about 50% smaller compared with a power gated SRAM with half the number of cells.

REFERENCES

- [1] J. Chang, S. L. Chen, W. Chen, S. Chiu, and R. Faber *et al.*, "A 45 nm 24 MB on-die L3 cache for the 8-core multi-threaded Xeon Processor," in *Proc. VLSI Circuits Symp.*, 2009, pp. 152–153.
- [2] 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. 2010.
- [3] R. E. Matick and S. E. Schuster, "Logic-based eDRAM: Origins and rationale for use," *IBM J. Res. Devel.*, vol. 49, no. 1, pp. 145–165, Jan. 2005.
- [4] J. Barth, W. R. Reohr, P. Parries, G. Fredeman, and J. Golz *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.
- [5] S. Romanovsky, A. Katoch, A. Achyuthan, C. O'Connell, and S. Natarajan *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.
- [6] P. J. Klim, J. Barth, W. R. Reohr, D. Dick, and G. Fredeman *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, Apr. 2009.
- [7] 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.
- [8] D. Somasekhar, Y. Ye, P. Aseron, S. L. Lu, and M. Khellah *et al.*, "2 GHz 2 Mb 2T gain-cell memory macro with 128 GB/s bandwidth in a 65 nm logic process," in *IEEE ISSCC Dig. Tech. Papers*, 2008, pp. 274–275.
- [9] K. Chun, P. Jain, J. Lee, and C. H. Kim, "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.
- [10] E. Seevinck, P. J. van Beers, and H. Ontrop, "Current-mode techniques for high-speed VLSI circuits with application to current sense amplifier for CMOS SRAM's," *IEEE J. Solid-State Circuits*, vol. 26, no. 4, pp. 525–536, Apr. 1991.
- [11] J. Sim, H. Yoon, K. Chun, H. Lee, and S. Hong *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.
- [12] K. Agarwal and S. Nassif, "The impact of random device variation on SRAM cell stability in sub-90-nm CMOS technologies," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 16, no. 1, pp. 86–97, Jan. 2008.
- [13] K. Zhang, U. Bhattacharya, Z. Chen, F. Hamzaoglu, and D. Murray *et al.*, "A SRAM design on 65 nm CMOS technology with integrated leakage reduction scheme," in *Proc. VLSI Circuits Symp.*, 2004, pp. 294–295.
- [14] C. H. Kim, J. Kim, I. Chang, and K. Roy, "PVT-aware leakage reduction for on-die caches with improved read stability," *IEEE J. Solid-State Circuits*, vol. 41, no. 1, pp. 170–178, Jan. 2006.



Ki Chul Chun received the B.S. degree in electronics engineering from Yonsei University, Seoul, Korea, in 1998, and the M.S. degree in electrical engineering from KAIST, Daejeon, Korea, in 2000. He joined the Memory Division, Samsung Electronics, Hwasung, Korea, in 2000, where he has been involved in DRAM circuit design. Since 2007, he has been working towards the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis.

His research interests include digital, mixed-signal and memory circuit designs with special focus on embedded DRAM, PRAM, and STT-MRAM in scaled technologies.

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



Pulkit Jain received the B.Tech. degree in electrical engineering from the Indian Institute of Technology (IIT), Kanpur, India, 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. He has authored/coauthored several journal and conference papers.



Jung Hwa Lee received the B.S. degree in electronics engineering from Kyungbuk National University, Daegu, Korea, in 1991.

He joined the Memory Division, Samsung Electronics, Yongin, Korea, in 1991, where he has been engaged in the development of high-density DRAMs and PRAMs. From 2007 to 2008, he was with the University of Minnesota, Minneapolis, as a Visiting Researcher. His research interests include process technology for gigabit RAMs, circuit design and layout for DFM, and future memory development.



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, 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 of the University of Minnesota, Minneapolis, MN, in 2004, where he is currently an Associate Professor. His research interests include digital, mixed-signal, and memory circuit design for silicon and non-silicon technologies.

Prof. Kim is the recipient of a National Science Foundation 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 90+ journal and conference papers and has served as a technical program committee member for several circuit design conferences. He was the technical program committee chair for the 2010 International Symposium on Low Power Electronics and Design (ISLPED) and the guest editor for a special issue of the *IEEE Design and Test Magazine*.