AnaRefresh: A 6-bit 64-by-64 Capacitor-based Analog Memory with a Digital-Sample-Analog-Refresh Technique Achieving 26-ns/row and 24.6-pJ/row Refresh

Linfang Wang*, Mosom Jana*, Rentao Wan, Chuan-Tung Lin and Mingoo Seok
Department of Electrical Engineering, Columbia University, New York, NY, USA
Email: {lw3206, mj3204, rw2899, cl4030, ms4415}@columbia.edu

***Abstract*—Capacitor-based analog memory is essential for end-to-end analog computing hardware for deep neural networks. Due to various leakage currents, however, the existing analog memory test chips exhibit only limited retention time, and to retain data longer, they often require making a backup and reloading data from it, which severely increases area, latency, and energy overhead. In light of this, we present AnaRefresh, a 6-bit 64-by-64 capacitor-based analog memory macro featuring a first-of-a-kind refresh technique, which can extend retention time indefinitely without backup memory. The careful circuit design makes the refresh operation fast and efficient: taking 26 ns/row and consuming 24.6 pJ/row.**

***Index Terms*—Analog computing, analog memory, refresh, digital-sample-analog-refresh (DSAR)**

I. Introduction

Capacitor-based analog memory has emerged as a promising solution for end-to-end analog computing hardware for deep neural networks (DNN) [1]–[6]. It can store intermediate data in analog form, thereby avoiding the use of digital memory and the associated data-conversion overhead [1], [4], [5]. However, due to the various leakage currents of a capacitor-based cell, analog memory exhibits only a limited retention time of tens to hundreds of microseconds, imposing significant constraints on the design of an analog processor. Especially if a memory macro intends to retain data beyond its native retention time, it must make a backup and reload the data from it, which severely increases area, latency, and energy overhead, as well as dataflow complexity (Fig. 1).

In light of this limitation, we aim to develop new analog memory circuits capable of retaining data indefinitely. To achieve this goal, we develop AnaRefresh, a capacitor-based analog memory macro that employs a novel digital-sample-analog-refresh (DSAR) technique. As shown in Fig. 2, the proposed DSAR technique periodically samples a cell voltage and selects the closest analog reference voltage, thereby recovering leakage-induced voltage drift. Note that simply sensing the cell voltage and writing it back cannot recover the drift, since it cannot replenish the lost charge.

This work is supported in part by the DARPA ScAN program, Catalyst Foundation, and National Science Foundation (2425331).

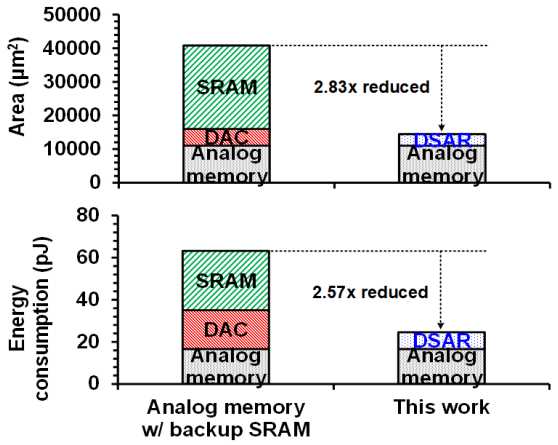


Fig. 1. Comparison on area and energy between analog memory with backup and this work.

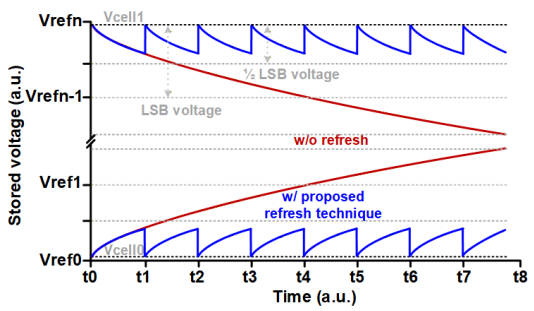


Fig. 2. Leakage-induced information loss in analog memory. We propose a DSAR technique to periodically refresh cell voltage.

We prototype the AnaRefresh test chip in 28-nm. AnaRefresh can refresh cells every 43.5 μs, just before any of its cell voltage can drift by half of the least-significant-bit (LSB) voltage (∼3 mV), thereby retaining the voltages indefinitely. It takes 26 ns and 24.6 pJ to refresh one row (64 cells) of the macro. AnaRefresh is the first capacitor-based analog memory having refresh capability in a modern nanometer technology.

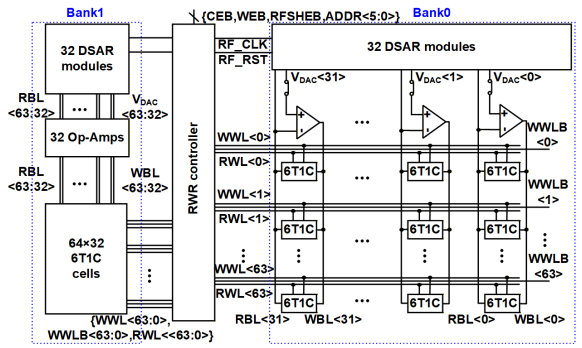


Fig. 3. Block diagram of the proposed 64X64 AnaRefresh memory macro.

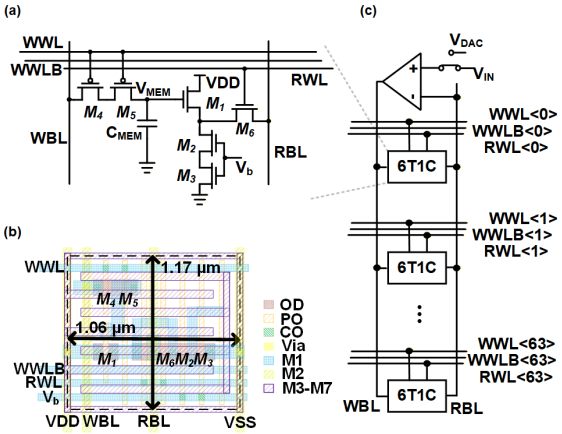


Fig. 4. (a) Schematic and (b) layout of the proposed 6T1C cell. (c) The write-with-feedback scheme for analog memory.

II. Proposed AnaRefresh Macro Design

A. Macro Circuits

Fig. 3 depicts AnaRefresh's block diagram. It consists of two banks and a shared read-write-refresh (RWR) controller in the middle. Each bank contains i) 64 rows and 32 columns of 6-transistor 1-capacitor (6T1C) memory cells, ii) 32 operational amplifiers (Op-Amps), and iii) 32 DSAR modules. Each row shares a write word-line (WWL) and a read word-line (RWL), while each column shares a write bit-line (WBL) and a read bit-line (RBL).

Fig. 4(a) shows the proposed 6T1C cell. It comprises a storage capacitor (C_{MEM}), a source follower (M_1, M_2, M_3), a write access switch (M_4, M_5), and a read access switch (M_6). To reduce cell leakage, we stack the devices (M_2, M_3, M_4, M_5) and employ the floating-body and floating-bitline techniques. To minimize the coupling noise, we add the dummy WWLB. All the transistors are minimum-sized. C_{MEM} is ~4.8 fF, implemented in an M3-M7 metal-oxide-metal (MOM) structure and positioned above the transistors to save area. The cell occupies 1.24 µm² (Fig. 4(b)).

We adopt the write-with-feedback (WFB) scheme to write a precise voltage in a cell. During the write operation, as shown in Fig. 4(c), we form a closed loop using the Op-Amp to ensure that the write-in voltage and the RBL voltage are equal. The closed loop is designed to support 6-bit precision. The cell voltage ranges from 120 to 508 mV with an LSB voltage of about 6 mV.

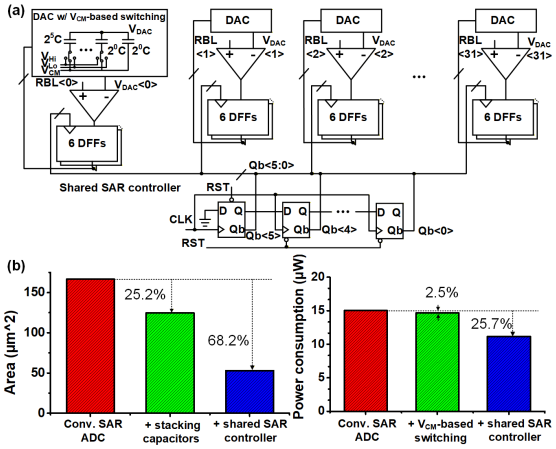


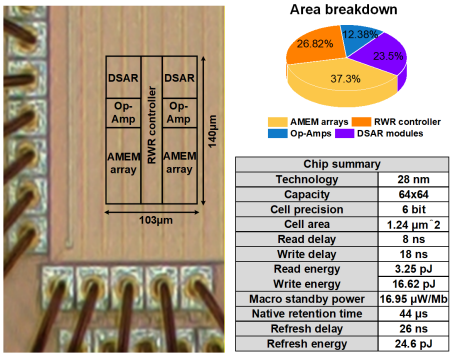
Fig. 5. (a) Schematic of the proposed DSAR modules. 32 of them are shown. (b) Circuit-level optimizations improve the area and power consumption of the DSAR module.

B. Digital-Sample-Analog-Refresh (DSAR) Module

We propose the DSAR technique to refresh the cell voltage. The worst-case native retention time (a time until a cell loses half of the LSB voltage, ~3 mV, without refresh) at a typical process, voltage, and temperature (PVT) condition is measured to be 44 µs. However, if we refresh cells periodically before they lose half of their LSB voltage (i.e., every 44 µs or slightly less), we can effectively extend the retention time indefinitely.

Fig. 5(a) shows the schematics of 32 DSAR modules. Each module consists of a voltage comparator, a 6-bit digital-to-analog (DAC) converter, and six D-flip-flops (DFFs). The 32 modules share a single successive approximation register (SAR) controller. The DSAR modules can refresh all cells in a row simultaneously. Each DSAR module successively samples the RBL voltage and finds the closest reference level. Then, it writes this approximated reference level to the cell using the WFB scheme.

We optimize the DSAR modules for area and power. We use a single controller for 32 DSARs to eliminate redundant logic. We also employ a common-voltage (V_{CM})-based capacitor-switching scheme [7]. The basic scheme switches two capacitors per cycle, requiring seven D flip-flops (DFFs). However, the V_{CM}-based switching scheme can switch only the bottom plate of one capacitor, requiring six DFFs. We also implement the DAC capacitors using the MOM in M4-M7. We place them above the transistors to reduce the area overhead. Those techniques reduce the area by 68.2% and the power consumption by 25.7% (Fig. 5(b)).



Chip summary	
Technology	28 nm
Capacity	64x64
Cell precision	6 bit
Cell area	1.24 μm^2
Read delay	8 ns
Write delay	18 ns
Read energy	3.25 pJ
Write energy	16.62 pJ
Macro standby power	16.95 μW/Mb
Native retention time	44 μs
Refresh delay	26 ns
Refresh energy	24.6 pJ

Fig. 6. Die micrograph, area breakdown, and chip summary.

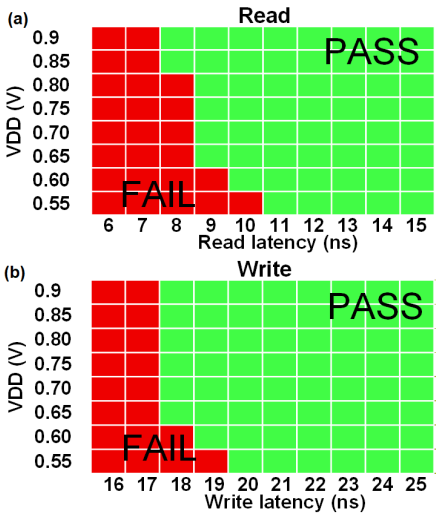


Fig. 7. Shmoo plots of (a) read operation and (b) write operation. (Measured).

III. Measurement Results

We prototyped the AnaRefresh test chip in a 28-nm CMOS process. The active area is 0.0144 mm^2 (Fig. 6). The cell array takes the largest portion of the area, followed by an RWR controller. The macro can operate across 0.55V to 0.9V. At 0.9V, AnaRefresh can operate at 110 MHz. It can perform a read operation in 8 ns (1 cycle) and a write in 18 ns (2 cycles) (Fig. 7). It consumes 3.25 pJ to read one row (64 cells) and 16.62 pJ to write one row.

We extensively test the refresh capability. At 0.9V and 25°C, the native retention time ranges from 1,488 to 44 μs across stored cell voltages (Fig. 8). Therefore, we set the refresh period to 43.5 μs to achieve near-zero error (Fig. 9(a)). The macro successfully refreshes over multiple refresh periods (Fig. 9(b)). At 0.9V, to refresh one row, AnaRefresh takes 26 ns (3 cycles) and 24.6 pJ. It takes a slightly higher delay and energy to refresh one row of cells than to write, since the refresh utilizes both the DSAR modules and the Op-Amps. Fig. 10 presents the transfer curve of write-in voltages to

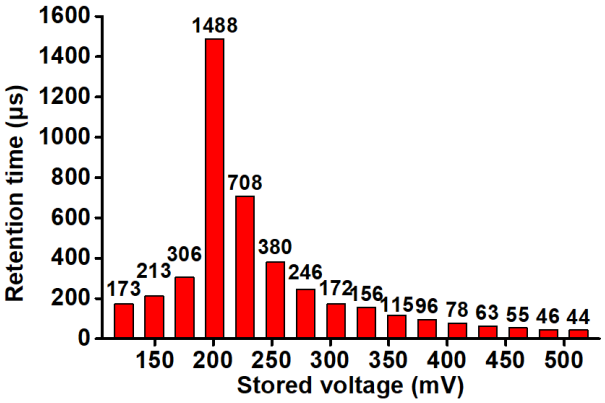


Fig. 8. Retention time measurement across different cell voltages.

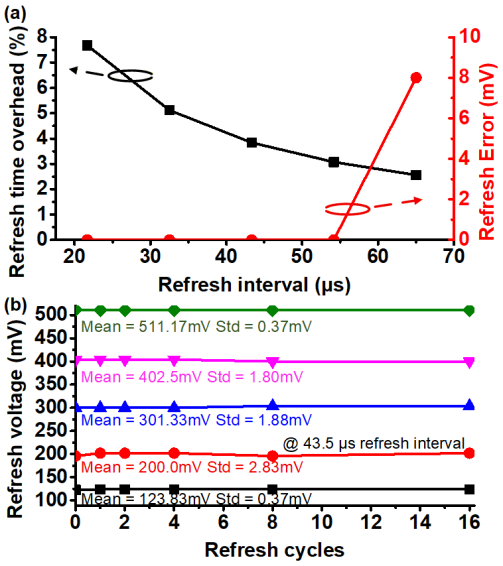


Fig. 9. (a) Refresh time overhead (total time to refresh all rows divided by refresh interval) measurement. (b) The cell voltages remain steady over multiple refresh cycles. (Measured).

post-refresh cell voltages. It achieves a small root-mean-square error (RMSE) of 1.85%.

We further characterize the native retention time. We measure 10 samples. The average native retention time is 43.32 μs, and the standard deviation is 7.25 μs. We pick the worst sample (Chip #6) and sweep the temperature from -30°C to 85°C. The native retention time of Chip #6 at 85°C is 17.6 μs (Fig. 11).

TABLE I presents a comparison with the prior works [1]–[5]. AnaRefresh targets the highest (6-bit) precision while achieving one of the smallest cell areas. It also achieves state-of-the-art delay and energy consumption for read and write operations. Most notably, it is the only capacitor-based

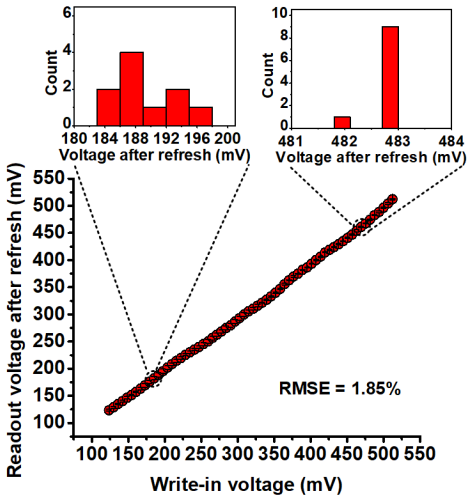


Fig. 10. Write-in voltage and post-refresh voltage measurement. We performed write-and-refresh operations for 10 different cells at each write-in voltage. The measurement exhibits a tight distribution.

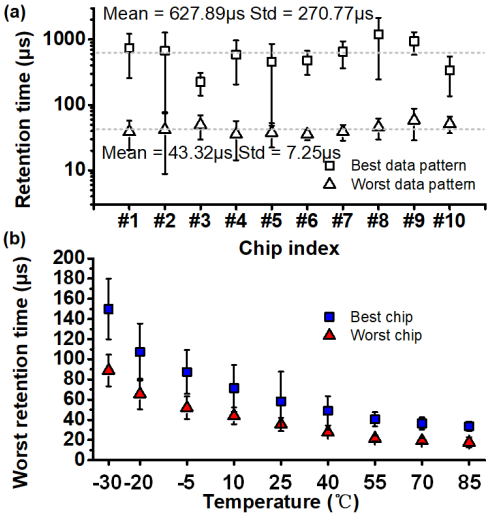


Fig. 11. (a) Retention time across 10 chips for the best- and worst-case data patterns. (b) Retention time across a temperature range of -30°C to 85°C for the best- and worst-case chips.

analog memory in nanometer CMOS that can refresh its cells' voltages. Other works assume that data is retained only for a short time or ignore the area, latency, and energy overhead associated with backup memory. Note that analog memory differs from multi-level DRAMs [8]–[11], as multi-level DRAMs in sub-micron technologies target lower bit precision (2-3 bits), have a digital-in, digital-out interface, and cannot be used for end-to-end analog computing hardware.

TABLE I
COMPARISON WITH OTHER WORKS

	This work	ISSCC 2021 [2]	JSSC 2024 [1]	ESSERC 2024 [3]	CICC 2025 [4]	ISSCC 2026 [5]
Technology (nm)	28	65	28	28	180	55
Macro capacity capacity	64x64, 4k	64x128, 8k	28x112, 3k	64x33, 2k	8x128, 1k	32x32, 1k
Cell type	6T1C	3T1C	6T1C	3T1C	5T1C	6T1C
Cell precision (bits)	**6**	4	5	4	8	N/R
LSB voltage (mV)	6.16	N/R	15.6	13.33	4.71	N/R
Cell area (µm2)	1.24	1.51[3]	1.3	1.05	130	2.4
Operating frequency (MHz)	110	105	200	47	10	100
Read delay[1] (ns)	8	9.52	N/R	5.3	100	10
Write delay[1] (ns)	18	9.52	N/R	21.1	100	10
Read energy[1] (pJ)	3.25	N/R	N/R	N/R	21.96[3]	N/R
Write energy[1] (pJ)	16.62	N/R	N/R	11.38[3]	329.5[3]	N/R
Standby power (µW/Mb)	**16.95**	N/R	N/R	N/R	N/R	N/R
Refresh capable?	**Yes**	No	No	No	No	No
Native retention time[2] (µs)	44	360[4]	11.2[3]	208[4]	180	N/R
Refresh period[2] (µs)	**43.5**	N/A	N/A	N/A	N/A	N/A
Refresh delay[1] (ns)	**26**	N/A	N/A	N/A	N/A	N/A
Refresh energy[1] (pJ)	**24.6**	N/A	N/A	N/A	N/A	N/A

[1]per row (64 cells); [2]at a typical corner;
[3]estimated; [4]at typical data pattern.

REFERENCES

[1] J. -O. Seo et al., "A 44.2-TOPS/W CNN Processor With Variation-Tolerant Analog Datapath and Variation Compensating Circuit," in IEEE Journal of Solid-State Circuits, vol. 59, no. 5, pp. 1603-1611, May 2024.

[2] Z. Chen et al., "15.3 A 65nm 3T Dynamic Analog RAM-Based Computing-in-Memory Macro and CNN Accelerator with Retention Enhancement, Adaptive Analog Sparsity and 44TOPS/W System Energy Efficiency," 2021 IEEE International Solid-State Circuits Conference (ISSCC), 2021, pp. 240-242.

[3] R. Wan et al., "AACIM: A 2785-TOPS/W, 161-TOP/mm2, ¡1.17%-RMSE, Analog-In Analog-Out Computing-In-Memory Macro in 28nm," 2024 IEEE European Solid-State Electronics Research Conference (ESSERC), 2024, pp. 581-584.

[4] R. Wan et al., "AJPEG: A 26.4-pJ/pixel, 252-fps, 128×128 Image Sensor with an In-Sensor Analog DCT Processor for Data Compression," 2025 IEEE Custom Integrated Circuits Conference (CICC), 2025, pp. 1-3.

[5] Z. Yang et al., "7.8 A 55nm Intelligent Vision SoC Achieving 346TOPS/W System Efficiency via Fully Analog Sensing-to-Inference Pipeline," 2026 IEEE International Solid-State Circuits Conference (ISSCC), 2026, pp. 126-128.

[6] J. Song et al., "A 4-bit Calibration-Free Computing-In-Memory Macro With 3T1C Current-Programmed Dynamic-Cascode Multi-Level-Cell eDRAM," in IEEE Journal of Solid-State Circuits, vol. 59, no. 3, pp. 842-854, March 2024.

[7] Y. Wu, X. Cheng and X. Zeng, "A split-capacitor Vcm-based capacitor-switching scheme for low-power SAR ADCs," 2013 IEEE International Symposium on Circuits and Systems (ISCAS), 2013, pp. 2014-2017.

[8] M. Horiguchi, M. Aoki, Y. Nakagome, S. Ikenaga and K. Shimohigashi, "An experimental large-capacity semiconductor file memory using 16-level/cell storage," in IEEE Journal of Solid-State Circuits, vol. 23, no. 1, pp. 27-33, Feb. 1988, doi: 10.1109/4.252.

[9] J. C. Koob, S. A. Ung, B. F. Cockburn and D. G. Elliott, "Design and Characterization of a Multilevel DRAM," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 19, no. 9, pp. 1583-1596, Sept. 2011, doi: 10.1109/TVLSI.2010.2051569.

[10] J. Lynch and P. P. Irazoqui, "A Low Power Logic-Compatible Multi-Bit Memory Bit Cell Architecture With Differential Pair and Current Stop Constructs," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 61, no. 12, pp. 3367-3375, Dec. 2014.

[11] G. Yan et al., "First Demonstration of True 4-bit Memory with Record High Multibit Retention >103s and Read Window >105 by Hydrogen Self-Adaptive-Doping for IGZO DRAM Arrays," 2023 International Electron Devices Meeting (IEDM), 2023, pp. 1-4.