A 256kb 6T Self-Tuning SRAM with Extended 0.38V-1.2V Operating Range using Multiple Read/Write Assists and V_{MIN} Tracking Canary Sensors

Arijit Banerjee, Ningxi Liu, Harsh N. Patel, and Benton H. Calhoun
University of Virginia
Charlottesville, VA 22904, USA
{ab9ca, nl6cg, hnpatel, bcalhoun}@virginia.edu

John Poulton and C. Thomas Gray
Nvidia Corporation
Durham, NC 27713, USA
{jpoulton, tgray}@nvidia.com

***Abstract*— A closed loop self-tuning 256kb 6T SRAM with 0.38V-1.2V extended operating range using combined read and write assists and canary-based V_{MIN} tracking is presented. 337X and 4.3X power reductions are achieved using multiple assists and V_{MIN} tracking, respectively; combining both saves 1444X in active power and 12.4X in leakage at the 0.38V.**

Keywords—self-tuning SRAM; combined assists; canary SRAM, V_{MIN} tracking;

I. Introduction

This paper presents an adaptive, closed loop memory system that leverages combinations of bias-based peripheral assists (CPA) for both read and write to expand the operating range of a 256kb 6T SRAM by over 67% to cover from 1.2V down to 0.38V. Assists are used in reverse to tune canary bitcells that allow a closed loop control of the V_{DD} to track the minimum operating voltage (V_{MIN}) at a desired operating frequency. The design uses CPA together with canary based V_{MIN} tracking to maximize the operating range that is compatible with the sub-threshold logic (6T SRAM usually has higher V_{MIN} than logic circuits across process, voltage, and temperature (PVT) variations [1][2][3][4]) and to minimize guard-banding. The design is thereby optimized for meeting the low power, and varying frequency needs of highly variable Internet of Everything (IoE) applications while retaining the density of 6T cells.

Since battery powered or energy harvested IoE devices mostly operate at lower frequencies (~10 kHz to 10 MHz) [5][6], there is a need to expand the 6T SRAM operating range to lower voltages to achieve low power operation. Bias-based assist techniques can lower SRAM V_{MIN} [1][2][4], but selecting the best CPA depends on the V_{DD} and can affect the power / performance tradeoff. Fig 1 (a) shows the measured cumulative distribution functions for the SRAM with three peripheral assists: (1) V_{DD} boosting (VDDB) for low-voltage readability and half-select [1][4] read-stability; (2) wordline (WL) boosting (WLB); and (3) negative bitline (NBL) for write-ability. Using all the three assists achieves 240mV of V_{MIN} improvement (at 90th percentile) and beats using other single or combinations (Fig 1 (a)), but using fewer assists can save power overhead when the target V_{DD} is higher for a given frequency.

Fig 1 (b) shows the measured Shmoo plot with the CPA-extended range highlighted for the 256kb SRAM. Using assists alone requires guard-banding to ensure that all chips function across PVT, reducing the potential power savings.

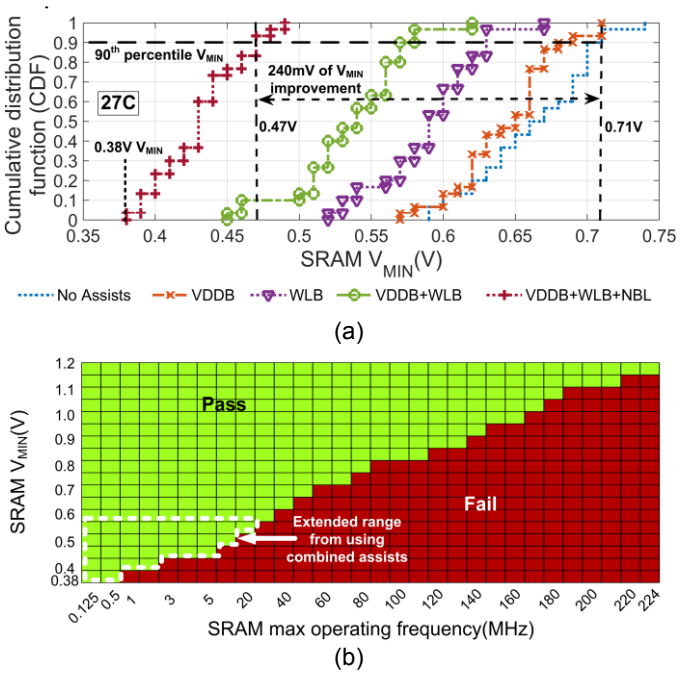

Fig. 1. a) Measured CDF of 256kb SRAM V_{MIN} showing 90th percentile V_{MIN} improvement of 240mV using combined assists [V_{DD} boosting (VDDB), WL boosting (WLB), negative bitline (NBL)], and b) measured V_{DD} Shmoo.

To maximize the benefits of CPA, runtime SRAM V_{MIN} determination [7] reduces the guard-banding of SRAM V_{MIN} at a given frequency. However, this technique has a huge penalty in the number of cycles for writing and reading the whole SRAM and in total energy for using a built-in-self-test (BIST). On the other hand, a smaller sized canary SRAM based V_{MIN} tracking [8] enables each chip to function at or near its V_{MIN} for much lesser clock cycles and energy.

A. Block Diagram of the System

Fig 2 (b) shows our full SRAM system comprising a 256kb SRAM in 4 sub-arrays (mats) each with 4 banks of 128x128 6T bitcells and 1 row of 128 canary bitcells per bank (2kb canary bitcells total), an assist controller (ASC), a frequency-to-digital converter (FDC), and a built-in self-test (BIST) block for the core SRAM and the canary bitcells (CBIST). The canary cells share the peripheral circuits such as write drivers, sense amplifiers, precharge circuits etc. with the SRAM array but have dedicated reverse assist (RA) controls [8] that tune write-ability and readability of the canaries by degrading the canary WL signal using eight programmable settings.

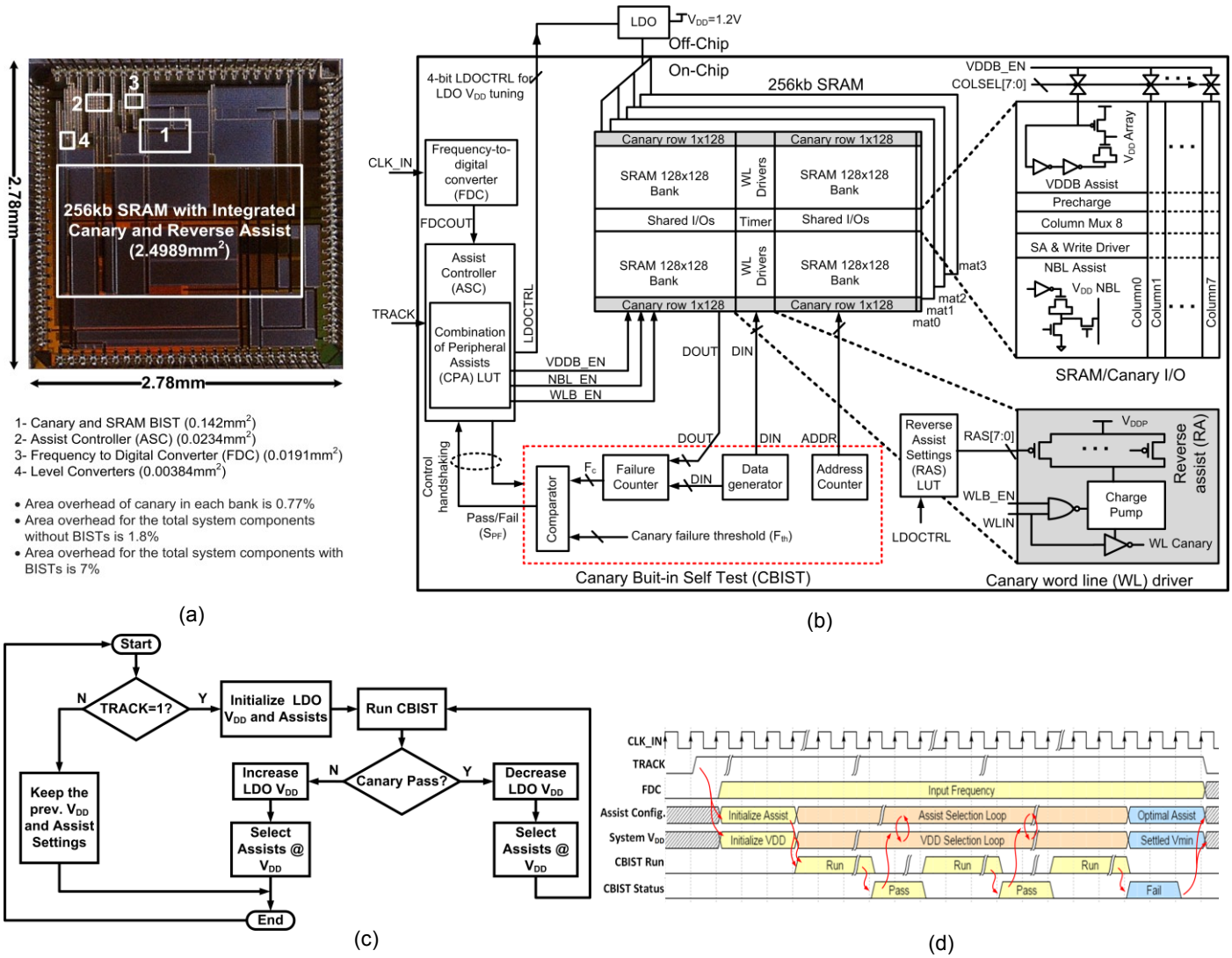

Fig. 2. a) Annotated micrograph of the SRAM chip, b) system level block diagram for the 256kb 6T self-tuning SRAM subsystem showing subcomponents, c) flowchart for canary V_{MIN} tracking, and d) system waveforms for V_{DD} self-tuning strategy.

The CBIST tests the canary to provide the number of failures to the ASC.

B. *Self-tuning Strategy of the System*

Fig 2 (c) and (d) present the self-tuning strategy for canary-based SRAM V_{MIN} tracking and dynamic control over assists, and V_{DD} selection. When tuning is enabled (TRACK=1), the FDC converts the input clock (CLK_IN) frequency to a 16-bit digitized output (FDCOUT) and initializes an (off-chip) Low-Dropout (LDO) regulator to an initial V_{DD} for the given frequency. Then, the ASC chooses an assist configuration for the current V_{DD} from a look-up table (LUT) for flexibly optimizing assist selection based on measured characterization across V_{DD}. The ASC then iterates to find the target V_{MIN} for the given frequency based on the canary outputs. The CBIST executes canary write and read operations across all canary addresses, calculates the number of canary failures (F_c), then compares F_c with a canary failure threshold value (F_{th}) to generate a pass/fail signal (SPF). If the CBIST passes, the ASC reduces V_{DD} by changing a 4-bit signal (LDOCTRL) controlling the off-chip LDO. The ASC repeats this process until the CBIST fails, then it raises V_{DD} to the last operational V_{DD}, which completes the closed-loop tracking for V_{MIN}. The SRAM retains its data through this process, and tuning can be re-run when the frequency changes or to periodically adjust for temperature changes.

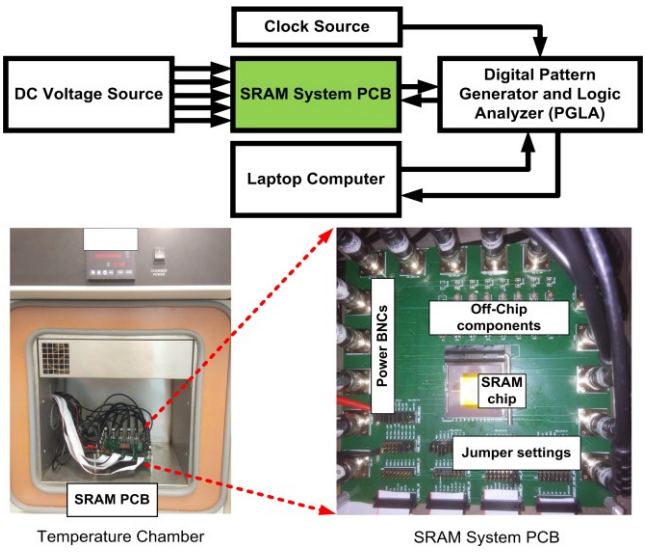

Fig. 3. Experimental setup for the chip measurements.

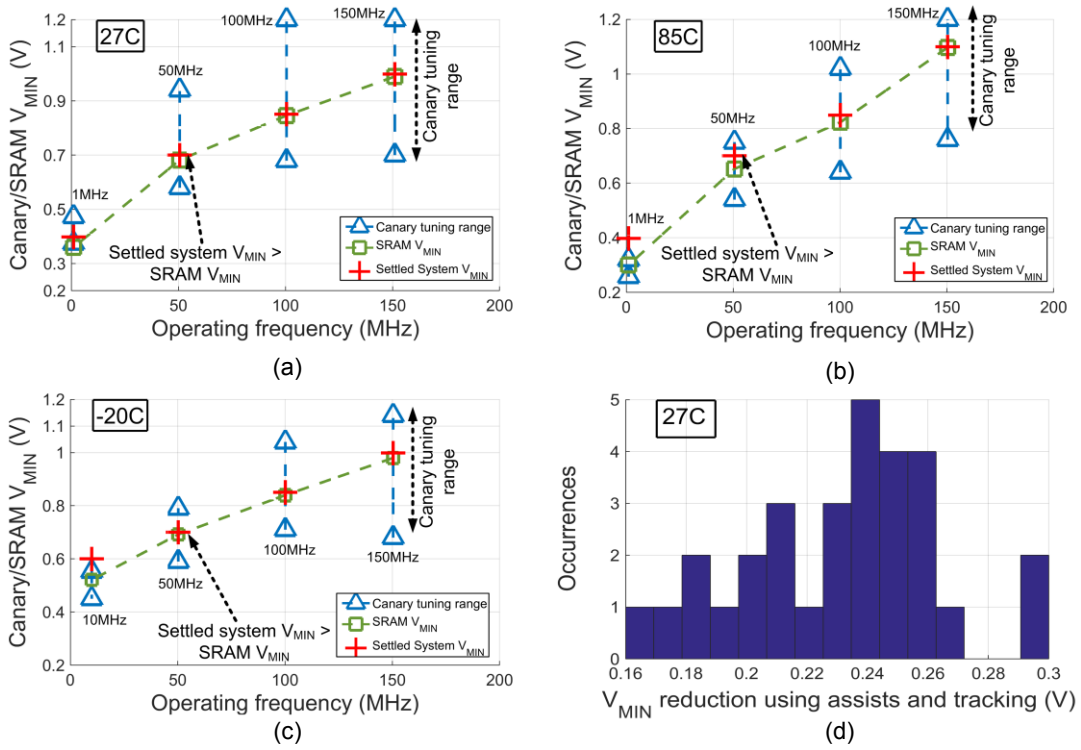

Fig. 4. Measured canary V_{MIN} tracking across clock frequencies [1 or 10, 50, 100, and 150] MHz and temperatures a) 27C, b) 85C, and c) -20C, showing V_{MIN} tuning range, and d) the distribution of overall V_{MIN} reduction using assist and tracking.

II. Canary Feedback Mechanism

While the assists expand the operating range, the canary feedback is critical to ensure that V_{DD} scaling stops before the core SRAM fails. The RA [8] forces failures in canary bits ahead of the core bits at eight programmable reverse assist settings (RAS). Since the canaries are core SRAM cells with RAS applied, their failure distribution is a shifted version of the core cells that tracks with frequency and temperature [8]. This allows us to set F_{th} based on CBIST results from a few dies to calibrate the canary failures relative to the core SRAM cells, thus all the chips are able to track their V_{MIN}.

III. Experimental Setup

Fig 3 shows the experimental setup for the measurement of data. A DC voltage source supplies power to the SRAM PCB. The digital pattern generator (PGLA) generates a waveform that controls the SRAM chip. An external clock source drives the PGLA to generate a clock signal to the PCB and the chip. Overall, the PGLA is controlled by a laptop computer for waveform generation and data collection.

IV. Measurement and Results

Fig 4 (a), (b), and (c) show the measured tunable range for canaries and the SRAM V_{MIN} across temperature and frequency. Fig 4 (d) shows the distribution of the V_{MIN} reduction using CPA and V_{MIN} tracking across 30 dies. The ASC sets the F_{th} and uses an LUT to select the RAS and sense amplifier delay based on the current V_{DD}, which allows the user to tradeoff guard-band margin with power savings. Fig 4 (a), (b), and (c) show settled V_{MIN} values based on settings that aggressively reduce the margin to maximize the power savings, but the flexible system allows including an arbitrary margin.

CPA and canary based V_{MIN} tracking work together to allow each chip self-tuning to its V_{MIN} for a given frequency, including expanded operating range and power savings for low V_{DD} IoE applications. Fig 2 (a) shows an annotated die photo

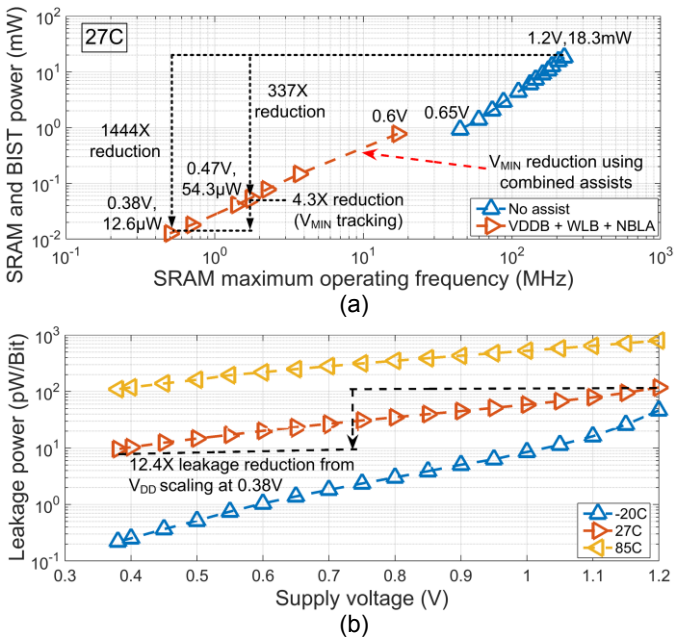

Fig. 5. a) Measured active power reduction of SRAM and BIST with combined peripheral assists and V_{MIN} tracking, and b) measured leakage reduction from V_{DD} scaling.

of the SRAM chip. The area overheads are 0.77% for the canary bits in each SRAM bank and 1.8% for the system components without BISTs in this design. The combined assist overhead in the SRAM is less than 2.8%. Fig 5 (a) shows the power savings from the combined approach, which extends the operating range down to 0.38V, and gives a 12.4X lower (Fig 5 (b)) leakage power (9.5pW/bit) than at 1.2V. If canary tracking were not available, process variation would require V_{DD} scaling to stop at 0.47V to ensure all chips work (achieves 337X active power reduction using CPA for SRAM and BIST), but V_{MIN} tracking allows an extra 4.3X power reduction by removing the

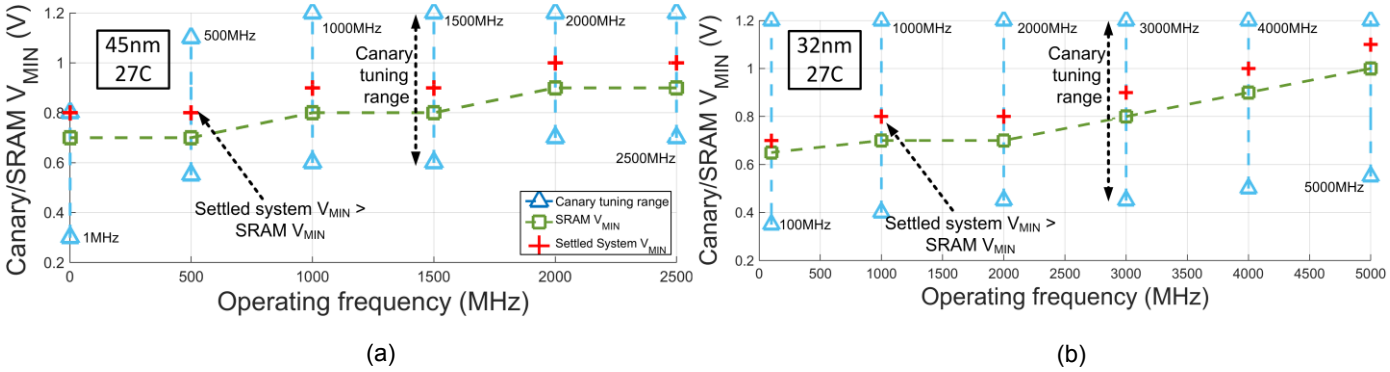

Fig. 6. Simulation results of canary tuning at a) 45nm and b) 32nm technology at TT_27C corner showing that canary based system V_{MIN} can be tuned above the SRAM V_{MIN}.

TABLE I. POWER BREAKUP FOR THE SRAM AND BIST.

Supply (V)	SRAM and BIST Power	SRAM Power	BIST Power
1.2	18.3mW	14.4mW	3.9mW
0.47	54.3µW	49.7µW	4.6µW
0.38	12.6µW	11.4µW	1.2µW

guard-band for those chips that can function at lower V_{DD} (up to 1444X active power savings (Fig 5 (a))). Table I includes the power breakup of the SRAM and BIST in the chip. It shows that these techniques reduce the SRAM power from 14.4mW to 11.4µW (1263X power reduction). Table II compares this work to recent wide voltage range SRAMs for low power applications. Fig 6 (a) and (b) show that the canary based V_{MIN} tracking is scalable to 45nm and 32nm technologies for a wide range of voltages and frequencies.

TABLE II. COMPARISON TABLE FOR SRAM SUB-SYSTEM WITH THE STATE-OF-THE-ART.

	This Work	ISSCC '10 [1]	ISSCC '12 [2]	VLSI '14 [3]	ISSCC '15 [4]
Technology	130nm	45nm	22nm	180nm	28nm
Cell Type	6T	8T	6T	8T	6T
Capacity	256kb	512kb	576KB	16KB	256kb
DVS range	1.2-0.38V (850mV)	1.2V-0.57V (630mV)	1V-0.625V (375mV)	1.8V-0.6V (1200mV)	0.9V-0.58V (320mV)
SRAM V_{MIN} tracking	Y	N	N	N	N
V_{MIN}	0.38V	0.57V	0.7V	0.6V	0.58V
Sub-V_T Operation	Y	-	-	-	-
Active Power	18mW @1.2Vand 12.6µW@ 0.38V	169mW @1.2V	-	250mW@ 1.8V, 15mW @0.82V	-
Power Reduction using Combined Assists	337X	-	-	16.4X	
Power Reduction using V_{MIN} Tracking	4.3X	N	N	N	N
Max Power Reduction	1444X	-	-	16.4X	-
Min Energy/op	25.3pJ/op @0.38V	-	-	690pJ/op @ 0.82V	-
Standby Leakage Power/bit	9.5pW/bit	644nW/bit	-	-	-
Standby Leakage Power Reduction to V_{MIN}	12.4X	9.4X	-	-	-
Sensor	Canary SRAM	DRV Sensor	-	Energy	-
Combined Assist	Any combination of NBL, WLB, VDDB	N	TVC-WA, WLUD	N	WLUD, NBL

V. CONCLUSIONS

This chip extends the 6T SRAM operating range by over 67% (from 1.2V-0.71V=0.49V to 1.2V-0.38V=0.82V, in sub-threshold) using three combined read/write assists and canary-based V_{MIN} tracking. The SRAM self-tunes to the V_{MIN} across process and temperature for a target frequency. This adaptive solution enables a range of IoE applications and achieves up to 1444X active power reduction. Our canary based V_{MIN} tracking technique is scalable to 45nm and 32nm technologies.

ACKNOWLEDGMENT

This work was funded in part by NVIDIA through the DARPA PERFECT program. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government.

REFERENCES

[1] M. Qazi, et al., "A 512kb 8T SRAM macro operating down to 0.57V with an AC-coupled sense amplifier and embedded data-retention-voltage sensor in 45nm SOI CMOS," ISSCC, Dig. Tech. Papers, pp. 350-351, 2010.

[2] E. Karl, et al., "A 4.6GHz 162Mb SRAM design in 22nm tri-gate CMOS technology with integrated active V_{MIN}-enhancing assist circuitry," ISSCC, Dig. Tech Papers, pp. 230-232. 2012.

[3] Y. Sinangil, et al., "A self-aware processor SoC using energy monitors integrated into power converters for self-adaptation," Symp. VLSI Circuits Dig. Tech. Papers, pp. 1-2, 2014.

[4] M. F. Chang, et al.,"17.3 A 28nm 256kb 6T-SRAM with 280mV improvement in V_{MIN} using a dual-split-control assist scheme," ISSCC Dig. Tech. Papers, pp. 1-3, 2015.

[5] Joyce Kwong, et al., "An Energy-Efficient Biomedical Signal Processing Platform," in IEEE Journal of Solid-State Circuits, vol. 46, no. 7, pp. 1742-1753, July 2011.

[6] Abhishek Roy, et al., " A 1.3uW, 5pJ/cycle Sub-threshold MSP430 Processor in 90nm xLP FDSOI for Energy-efficient IoT Applications," 17th International Symposium on Quality Electronic Design (ISQED), pp. 158-162, 2016.

[7] Y. C. Lai, S. Y. Huang and H. J. Hsu, "Resilient Self-V_{DD}-Tuning Scheme With Speed-Margining for Low-Power SRAM," in IEEE Journal of Solid-State Circuits, vol. 44, no. 10, pp. 2817-2823, Oct. 2009.

[8] A. Banerjee, et al., "A 130nm canary SRAM for SRAM dynamic write V_{MIN} tracking across voltage, frequency, and temperature variations," IEEE CICC, pp. 1-4, 2015.