Design of Nonvolatile SRAM with Ferroelectric FETs for Energy-Efficient Backup and Restore

Xueqing Li, *Member, IEEE*, Kaisheng Ma, Sumitha George, Win-San Khwa, John Sampson, *Member, IEEE*, Sumeet Gupta, *Member, IEEE*, Yongpan Liu, *Senior Member, IEEE*, Meng-Fan Chang, *Senior Member, IEEE*, Suman Datta, *Fellow, IEEE*, and Vijaykrishnan Narayanan, *Fellow, IEEE*

***Abstract*— Nonvolatile SRAM (nvSRAM) has emerged as a promising approach to reducing the standby energy consumption by storing the state into an *in situ* nonvolatile memory element and shutting down the power supply. Existing nvSRAM solutions based on a nonvolatile backup in magnetic tunnel junction and ReRAM, however, are costly in backup and restore energy due to static current. This cost results in a long break-even time (BET) when compared with a lowered voltage standby volatile SRAM. This brief proposes an nvSRAM based on ferroelectric FETs (FeFETs) that are capable of fully avoiding such static current. A simple differential backup and restore circuitry is proposed, achieving sub-*fJ*/*cell* total energy per backup and restore operation at the 10-nm node. This leads to hundreds of times BET improvement over existing ReRAM nvSRAM solutions. This nvSRAM also indicates the future FeFET design trends for such memory-logic synergy.**

***Index Terms*— Ferroelectric FET (FeFET), nonvolatile computing, nonvolatile SRAM (nvSRAM), power gating, SRAM.**

I. Introduction

LEAKAGE current in SRAM contributes to significant energy cost, from the high-performance server centers to low-power energy-harvesting sensor nodes [1]–[3]. To reduce the leakage, various mitigating methods have been proposed, such as using low-standby-power transistors, proper transistor sizing, and supply voltage modulation especially in the idle phase. Using intrinsically nonvolatile memory (NVM) such as ReRAM, spin-transfer torque magnetic random-access memory, and phase-change random-access memory, leads to zero-standby cell leakage, but is constrained in use as an SRAM due to limited speed, endurance, or dynamic energy or power [4].

In contrast, nonvolatile SRAM (nvSRAM) embeds NVM to store an *in situ* nonvolatile copy of a CMOS SRAM bit and saves energy by completely shutting OFF the leaky supply [18]–[23]. However, the energy consumed in an existing nvSRAM to *backup* and *restore from* the NVM, $E_{B\&R}$, is high. For example, the state-of-the-art $E_{B\&R}$ in experiments is ∼200 fJ [18]. The fundamental causes include: 1) the high static current during a backup or restore operation; b) long backup time to survive variations; and c) long restore time to cover the supply recovery ramp. As a result, the minimum power-OFF time to save static leakage energy to compensate $E_{B\&R}$, i.e., break-even time (BET), is high for existing nvSRAM designs. This prohibits energy savings with frequent data access.

This brief proposes an nvSRAM, shown in Fig. 1(a), with sub-fJ $E_{B\&R}$ and thus significant BET improvement over existing nvSRAMs. It is achieved with a novel backup and restore circuitry using Ferroelectric FETs (FeFETs) that completely get rid of the static current in prior nvSRAMs. FeFET is an emerging beyond-CMOS transistor for Boolean logic and memory [5]–[11], [16]. It fundamentally differs from other NVM devices in its capability of concurrent logic and memory operation with the drain–source conducting route separated

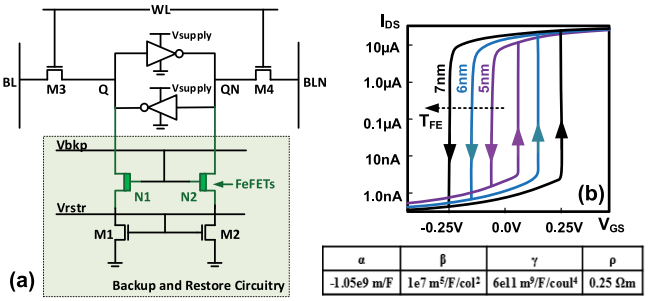

Fig. 1. Proposed FeFET nvSRAM. (a) Circuitry. (b) FeFET *I*–*V* characteristics.

Manuscript received March 14, 2017; revised April 28, 2017; accepted May 17, 2017. Date of publication June 5, 2017; date of current version June 19, 2017. This work was supported in part by GRC under Grant 2657.001, in part by NSF under grant CCF-1409798 subaward KK1522, in part by ASSIST, in part by NSFC under Grant 61674094, and in part by the Center for Low Energy Systems Technology, one of the six SRC STARnet Centers, sponsored by MARCO and DARPA. The review of this brief was arranged by Editor J. Kang. *(Corresponding authors: X. Li; V. Narayanan.)*

X. Li, K. Ma, S. George, J. Sampson, S. Gupta, and V. Narayanan are with the School of Electrical Engineering and Computer Science, The Pennsylvania State University, University Park, PA 16802 USA (e-mail: lixueq@cse.psu.edu; kxm505@cse.psu.edu; sug241@csepsu.edu; sampson@cse.psu.edu; skg157@engr.psu.edu; vijay@cse.psu.edu).

W. S. Khwa and M. F. Chang are with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan (e-mail: vince.khwa@gmail.com; mfchang@mx.nthu.edu.tw).

Y. Liu is with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: ypliu@tsinghua.edu.cn).

S. Datta is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA (e-mail: sdatta@nd.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/TED.2017.2707664

from the gate control signal. It also works under a low voltage with very high distinguishability in the drain–source resistance between different FeFET memory states. Such an FeFET nvSRAM also indicates the significance of designing future FeFETs with embedded logic-memory synergy in a single device for more energy-efficient computing.

II. Proposed FeFET nvSRAM Operations

This section first describes FeFET device for nvSRAM and the operation theory of FeFET nvSRAM in different modes. Then device-circuit codesign and optimizations are discussed.

A. FeFET Device Characteristics

An FeFET is implemented with a ferroelectric (FE) material layer electrically and preferably physically sandwiched in the gate-stack [6]. Recent FE development promises good FeFET scaling capability with low-voltage operation [8]. Various FeFETs have been reported. While some of them exhibit hysteresis-free switching with a steep slope, others are highlighted with tunable I–V hysteresis curves [6]–[9]. The proposed nvSRAM in this brief makes use of the FeFET I–V hysteresis covering zero gate bias, as shown in Fig. 1(b). This could be achieved by tuning the FE layer thickness T_{FE} to obtain proper capacitance matching and coercive voltage [6].

In the FeFET I–V hysteresis, the positive and negative polarization could be set by applying a positive and negative gate–source voltage V_{GS}, respectively. When V_{GS} stays away from the rising and falling edge in the two directions, the polarization state does not change. Even if the gate voltage is removed, the stable remnant positive or negative polarization of the FE material maintains a nonvolatile state of low or high FeFET drain–source resistance, respectively [5].

In this brief, simulations are based on the 10-nm LSTP PTM model in [14] and the calibrated FeFET model based on L-K equation in [15] that has been adopted by [10], [11], and [16]. The FeFET device model parameters for the L-K equation are included in Fig. 1(b).

B. SRAM-Mode Operation

During normal SRAM-mode operations, the supply voltage V_{supply} and the gate voltage of the two access transistors (M1 and M2) V_{rstr} are kept stably at *VDD* and *GND*, respectively. As M1 and M2 are turned OFF, the backup and restore circuitry now acts as a high-resistance OFF-state route from Q and QN to ground, regardless of the polarization states of N1 or N2.

The operation of the rest of the nvSRAM is similar to an existing volatile 6-T SRAM in read and write. For read, the bit lines *BL* and *BLN* are both precharged to be high in voltage and then the word line signal *WL* turns on the access transistors M3 and M4. The internal inverter that outputs zero discharges the corresponding bitline and causes a lower sensed voltage at the bitline. The sense amplifier reads the bitline voltage difference and obtains the Boolean value. To write the nvSRAM, the external bit lines are kept stable at high or low voltage, according to the values to be written. Then M3 and M4 are turned on and the internal inverters are regulated accordingly.

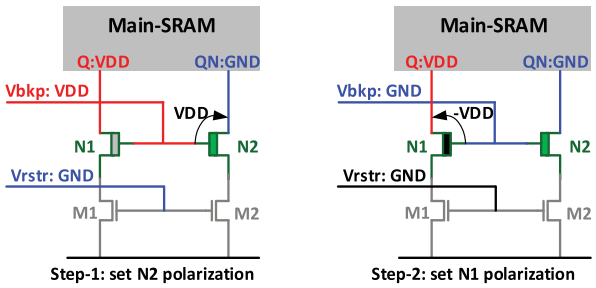

Fig. 2. Steps to back up SRAM states to FeFETs in proposed nvSRAM.

It is noted that the additional backup and restore circuitry adds some minor parasitics to the internal nodes Q and QN. This reduces the read access time as these parasitics help discharging the bit lines, and also increases the write access latency. The impact is very minor because the extra parasitics are orders lower in magnitude than bitline capacitance. Read and write noise margins are dc performance metrics, and will be nearly the same as a conventional 6-T SRAM as the extra circuitry is floating (OFF from *GND*) in the SRAM mode.

As the two FeFETs (N1 and N2) are floating in the SRAM-mode, their gate voltage V_{bkp} does not affect the SRAM-mode functionality. If the SRAM access is not frequent and the FeFET endurance is sufficiently high, V_{bkp} can be set to *VDD* or *GND* or somewhere in between. In contrast, when endurance is a concern, V_{bkp} should be kept at an appropriate voltage so that the FeFETs do not wear out soon due to too many polarization switching activities with frequent SRAM state updates in some applications. A possible V_{bkp} for this purpose could be *VDD*/2, to make sure that the largest FeFET voltage stress $|V_{GS}|$ is minimized to be *VDD*/2 to avoid unnecessary polarization switching.

For the same reason, given an FeFET nvSRAM design, *VDD* should be within a certain range. For the backup purpose, *VDD* should be higher than the coercive voltage to update the polarization. For FeFET endurance considerations, *VDD* should be less than twice the coercive voltage so that updates of the SRAM data do not flip the FeFET polarization. If a dynamic *VDD* is available, *VDD* could also be further lowered in the SRAM-mode to save more power and then raised temporarily to enable backup operations when necessary.

C. Backup-Mode Operation

Fig. 2 illustrates the 2-step backup operation. In this mode, V_{supply} is stable at *VDD*, and V_{rstr} is stable at *GND*. The SRAM external access transistors (M3 and M4) are also turned OFF by a low *WL* voltage. The state of the SRAM is assumed to be *VDD* at Q and *GND* at QN. At step 1, V_{bkp} is raised to *VDD*, above the FeFET coercive voltage, to ensure that N2 is positively polarized with low drain–source resistance. At step 2, V_{bkp} is lowered to *GND* to ensure that FeFET N1 is negatively polarized to have high drain–source resistance. Note that N2 polarization stays stable at step 2.

If the nvSRAM backup operation is followed by a power supply cutoff, V_{bkp} could stay as *GND*. If the nvSRAM backup is a run-time backup followed by subsequent SRAM-mode operations, V_{bkp} could be set back to its SRAM-mode value.

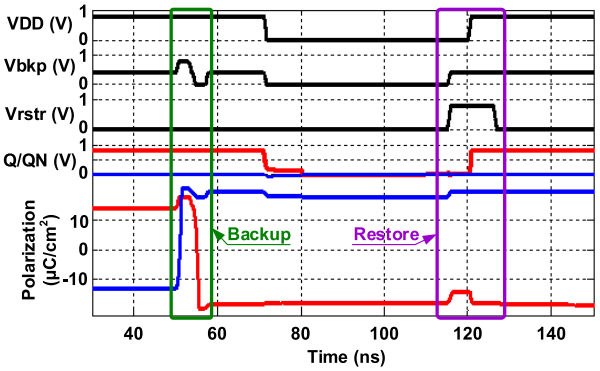

Fig. 3. Backup and restore transient waveforms of proposed nvSRAM.

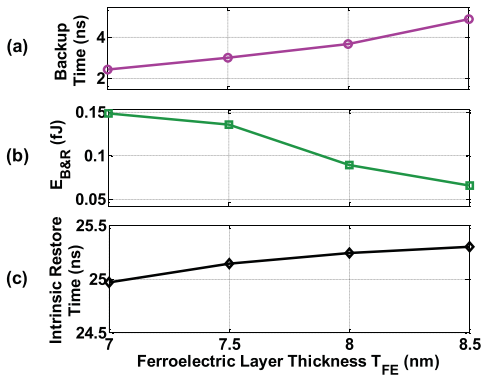

Fig. 4. Impact of FeFET T_{FE} at 0.6 V *VDD*: (a) Backup time. (b) EB&R. (c) Intrinsic restore time.

D. Restore-Mode Operation

Fig. 3 shows the transient waveforms for backup and restore operations before and after a power failure, respectively. Before a restore, all nodes are assumed to be zero in voltage. The polarization of N1 and N2 is assumed to be negative and positive, respectively. At first, V_{rstr} is raised to *VDD* (V_{bkp} remains zero). Accordingly, the positive N2 polarization leads to a low-resistance path from *QN* to ground, and the negative N1 polarization results in a high-resistance path from *Q* to ground. After that, V_{supply} is raised to *VDD*. The different pull-down-to-ground strength at nodes *Q* and *QN* by the restore circuitry regulates the main SRAM inverters, so that *Q* follows V_{supply} and *QN* stays zero. After *Q* and *QN* settle down, V_{rstr} is grounded. After a restore, if *VDD* is higher than twice the coercive voltage, V_{bkp} could be adjusted to its SRAM-mode value for endurance considerations. Note that restore operations do not change the FeFET polarization.

E. Device-Circuit Co-Design and Optimization

The main nvSRAM performance features include backup and restore energy and delay, retention time, yield, and reliability. The FE layer thickness T_{FE} could be tuned accordingly for optimizations. Fig. 1(b) has shown how T_{FE} affects I–V. Related key FeFET features include the hysteresis window width and ON- and OFF-state currents.

Increasing T_{FE} increases the FeFET retention time with a higher coercive voltage, which also leads to longer time till the polarization is flipped, as shown in Fig. 4(a) for the backup

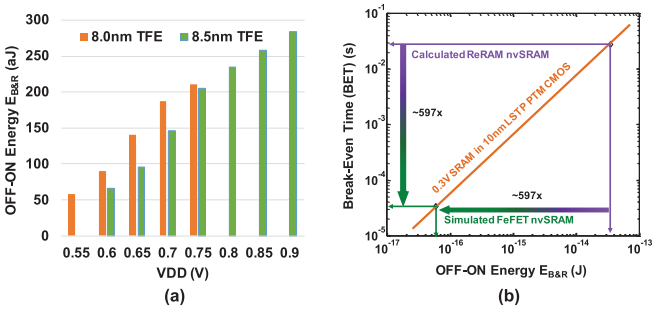

Fig. 5. Proposed FeFET nvSRAM. (a) Power OFF–ON energy $E_{B\&R}$. (b) BET comparison. Simulated with 1-μs supply voltage ramp-up time.

latency. T_{FE} also affects the backup and restore energy $E_{B\&R}$, showing more energy consumed with a smaller T_{FE}, as shown in Fig. 4(b). For Fig. 4(a) and (b), the V_{supply} wakeup ramp is set to be 1 μs to mimic typical scenarios. At the array or processor level, restore speed is usually limited by *VDD* recovery. For fine-grained power-gating scenarios, a fast recovery of a local supply is possible, and the FeFET nvSRAM intrinsic restore speed becomes meaningful. Fig. 4(c) shows the restore latency in such a scenario that is slightly affected by T_{FE} under a 0.1-ns-recovery *VDD*. This is significantly faster than ReRAM nvSRAM in [18] because of the single-step restore operation with a positive feedback loop in the nvSRAM.

III. Performance Evaluations and Discussions

A. Performance Evaluations

When operating at different supply voltages, Fig. 5(a) shows FeFET nvSRAM $E_{B\&R}$. A different T_{FE} values lead to a different operating range of *VDD*. The $E_{B\&R}$ is well below 1.0 fJ. For example, $E_{B\&R}$ is only 58.7 aJ at 0.55-V *VDD* with 8-nm T_{FE}. This corresponds to BET as low as 33.8 μs, against 0.3-V standby CMOS SRAM built with 1-fin LSTP PTM MOSFETs.

For comparison, the state-of-the-art 7T1R ReRAM nvSRAM in [18] is also evaluated with the PTM technology. The low-resistance state of the ReRAM is set to be 5.5 kΩ, to provide sufficient pull-down strength against the preset pull-up PMOS in the SRAM (pull-up resistance with 1 fin is ~22 kΩ at 0.7-V *VDD*). The adopted switching time t_{SW} is 0.3 ns at 0.8-V SET/RESET from an ultrafast energy-efficient HfOx ReRAM in [22]. Even if only write energy is considered here, the ReRAM nvSRAM consumes ~597 × higher energy (more than 35 fJ), assuming no ReRAM device variations. The huge $E_{B\&R}$ difference is due to the required static current for at least t_{SW} to flip the ReRAM states [18], [22]. When ReRAM device variations are considered, $E_{B\&R}$ will be much higher as the write pulse duration will be determined by the slowest ReRAM [25].

The $E_{B\&R}$ advantage also leads to BET benefits, as shown in Fig. 5(b). This indicates that the minimum power-OFF window width during power-gating to win energy savings is only ~0.2% of that in ReRAM nvSRAM, which benefits frequent-access scenarios. Scaled magnetic tunnel junction (MTJ)-based nvSRAM is far outside the plot due to similar backup energy to ReRAM nvSRAM and large wakeup energy at a typically > μs V_{supply} wakup ramp [22], [23].

B. Area Overhead Analysis

The proposed FeFET nvSRAM needs four additional transistors in a cell for nonvolatility, indicating more area overheads than the prior 7T1R design in [18] in which ReRAM could be placed on top of the SRAM cell. Note that this ReRAM nvSRAM has extra area overhead due to the need for various voltage power supplies, which is significant when the SRAM array is small.

Meanwhile, density may not be the key priority for SRAM in some scenarios, such as in 8T and 10T SRAM designs, for which the portion of area overhead reduces. Therefore, the proposed design shows the most advantage when the backup and restore energy overhead is significant, as the proposed FeFET nvSRAM is far lower than the existing nvSRAMs.

More importantly, the area overhead of FeFET nvSRAM can be significantly reduced by moving the accessory circuitry to another layer using new fabrication methods such as the monolithic 3-D process [24], [26].

C. Retention, Reliability, and Endurance

If the number of FeFET polarization switching activities exceeds beyond the device endurance, the FeFET may be worn out and will exhibit a different ON–OFF resistance and affect the restore behavior. Thanks to the huge ON–OFF state resistance ratio (up to seven orders in magnitude as reported in [6]), its degradation of up to even 90% is still acceptable for functionality as long as ON-state resistance is sufficiently low. Future work on retention time and reliability analysis will be meaningful for FeFET device design and comprehensive evaluation when a proper model or experiment becomes available.

IV. Conclusion

This brief has proposed FeFET nvSRAM design by exploiting the embedded logic-memory integration in the FeFETs. With ultralow energy consumption in backup and restore operations, it enables a new SRAM paradigm for more energy savings.

Acknowledgment

The authors would like to thank S. Srinivasa, Prof. S. Salahuddin, Prof. A. Khan, Prof. F. Catthoor, and Prof. P. Asbeck.

References

[1] K. Ma *et al.*, "Architecture exploration for ambient energy harvesting nonvolatile processors," in *Proc. IEEE 21st Int. Symp. High Perform. Comput. Archit. (HPCA)*, Burlingame, CA, USA, Feb. 2015, pp. 526–537.

[2] Y. Liu *et al.*, "Ambient energy harvesting nonvolatile processors: From circuit to system," in *Proc. 52nd ACM/EDAC/IEEE Design Autom. Conf. (DAC)*, San Francisco, CA, USA, Jul. 2015, pp. 1–4.

[3] X. Li, U. D. Heo, K. Ma, V. Narayanan, H. Liu, and S. Datta, "RF-powered systems using steep-slope devices," in *Proc. IEEE 12th Int. New Circuits Syst. Conf. (NEWCAS)*, Trois-Rivieres, QC, Canada, Jun. 2014, pp. 73–76.

[4] Y. Xie, *Emerging Memory Technologies: Design Architecture and Applications*. New York, NY, USA: Springer, 2014.

[5] A. I. Khan, C. W. Yeung, C. Hu, and S. Salahuddin, "Ferroelectric negative capacitance MOSFET: Capacitance tuning & antiferroelectric operation," in *IEDM Tech. Dig.*, Washington, DC, USA, Dec. 2011, pp. 11.3.1–11.3.4.

[6] A. I. Khan *et al.*, "Negative capacitance in short-channel FinFETs externally connected to an epitaxial ferroelectric capacitor," *IEEE Electron Device Lett.*, vol. 37, no. 1, pp. 111–114, Jan. 2016.

[7] J. Jo and C. Shin, "Negative capacitance field effect transistor with hysteresis-free sub-60-mV/decade switching," *IEEE Electron Device Lett.*, vol. 37, no. 3, pp. 245–248, Mar. 2016.

[8] K.-S. Li *et al.*, "Sub-60 mV-swing negative-capacitance FinFET without hysteresis," in *IEDM Tech. Dig.*, Washington, DC, USA, Dec. 2015, pp. 22.6.1–22.6.4.

[9] M. H. Lee *et al.*, "Steep slope and near non-hysteresis of FETs with antiferroelectric-like HfZrO for low-power electronics," *IEEE Electron Device Lett.*, vol. 36, no. 4, pp. 294–296, Apr. 2015.

[10] S. George *et al.*, "Nonvolatile memory design based on ferroelectric FETs," in *Proc. 53rd ACM/EDAC/IEEE Design Autom. Conf. (DAC)*, Austin, TX, USA, Jun. 2016, pp. 1–6.

[11] S. George *et al.*, "Device circuit co design of FEFET based logic for low voltage processors," in *Proc. IEEE Comput. Soc. Annu. Symp. VLSI (ISVLSI)*, Pittsburgh, PA, USA, Jul. 2016, pp. 649–654.

[12] A. I. Khan *et al.*, "Negative capacitance in a ferroelectric capacitor," *Nature Mater.*, vol. 14, no. 2, pp. 182–186, 2015.

[13] D. E. Nikonov and I. A. Young, "Overview of beyond-CMOS devices and a uniform methodology for their benchmarking," *Proc. IEEE*, vol. 101, no. 12, pp. 2498–2533, Dec. 2013.

[14] *Predictive Technology Model*, accessed on Mar. 14, 2017. [Online]. Available: http://ptm.asu.edu

[15] A. Aziz, S. Ghosh, S. Datta, and S. K. Gupta, "Physics-based circuit-compatible SPICE model for ferroelectric transistors," *IEEE Electron Device Lett.*, vol. 37, no. 6, pp. 805–808, Jun. 2016.

[16] D. Wang, S. George, A. Aziz, S. Datta, V. Narayanan, and S. K. Gupta, "Ferroelectric transistor based non-volatile flip-flop," in *Proc. Int. Symp. Low Power Electron. Design*, San Francisco, CA, USA, Aug. 2016, pp. 10–15.

[17] J. Li, B. Nagaraj, H. Liang, W. Cao, C. H. Lee, and R. Ramesh, "Ultrafast polarization switching in thin-film ferroelectrics," *Appl. Phys. Lett.*, vol. 84, no. 7, pp. 1174–1176, Feb. 2004.

[18] A. Lee *et al.*, "RRAM-based 7T1R nonvolatile SRAM with 2x reduction in store energy and 94x reduction in restore energy for frequent-off instant-on applications," in *Proc. Symp. VLSI Circuits (VLSI Circuits)*, Kyoto, Japan, Jun. 2015, pp. C76–C77.

[19] P.-F. Chiu *et al.*, "A low store energy, low VDDmin, nonvolatile 8T2R SRAM with 3D stacked RRAM devices for low power mobile applications," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2010, pp. 229–230.

[20] S. Yamamoto, Y. Shuto, and S. Sugahara, "Nonvolatile SRAM (NV-SRAM) using functional MOSFET merged with resistive switching devices," in *Proc. IEEE Custom Integr. Circuits Conf.*, Sep. 2009, pp. 531–534.

[21] W. Wei, K. Namba, J. Han, and F. Lombardi, "Design of a nonvolatile 7T1R SRAM cell for instant-on operation," *IEEE Trans. Nanotechnol.*, vol. 13, no. 5, pp. 905–916, Sep. 2014.

[22] H. Y. Lee *et al.*, "Evidence and solution of over-RESET problem for HfO_X based resistive memory with sub-ns switching speed and high endurance," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2010, pp. 19.7.1–19.7.4.

[23] T. Ohsawa *et al.*, "A 1 Mb nonvolatile embedded memory using 4T2MTJ cell with 32 b fine-grained power gating scheme," *IEEE J. Solid-State Circuits*, vol. 48, no. 6, pp. 1511–1520, Jun. 2013.

[24] C.-H. Shen *et al.*, "Heterogeneously integrated sub-40 nm low-power epi-like Ge/Si monolithic 3D-IC with stacked SiGeC ambient light harvester," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2014, pp. 3.6.1–3.6.4.

[25] M.-F. Chang *et al.*, "Embedded 1 Mb ReRAM in 28 nm CMOS with 0.27-to-1 V read using swing-sample-and-couple sense amplifier and self-boost-write-termination scheme," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2014, pp. 332–333.

[26] M. M. Shulaker *et al.*, "Monolithic 3D integration of logic and memory: Carbon nanotube FETs, resistive RAM, and silicon FETs," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2014, pp. 27.4.1–27.4.4.

Authors' photograph and biography not available at the time of publication.