4.7 A 65nm ReRAM-Enabled Nonvolatile Processor with 6× Reduction in Restore Time and 4× Higher Clock Frequency Using Adaptive Data Retention and Self-Write-Termination Nonvolatile Logic

Yongpan Liu[1], Zhibo Wang[1], Albert Lee[2,3], Fang Su[1], Chieh-Pu Lo[2], Zhe Yuan[1], Chien-Chen Lin[2], Qi Wei[1], Yu Wang[1], Ya-Chin King[2], Chrong-Jung Lin[2], Pedram Khalili[3], Kang-Lung Wang[3], Meng-Fan Chang[2], Huazhong Yang[1]

[1]Tsinghua University, Beijing, China,
[2]National Tsing Hua University, Hsinchu, Taiwan,
[3]University of California, Los Angeles, CA

With the rising importance of energy efficiency, zero leakage power and instant-on capability are highly desired features in energy harvesting sensors, as well as "normally off" high performance processors. However, intermittent power in such systems requires nonvolatile memory (NVM) to hold intermediate data and avoid rollbacks. Previous work has adopted FeRAM and STT-MRAM to achieve zero-standby power and fast-restore nonvolatile processors (NVPs) [1-3]. Previous NVPs, however, suffer from several drawbacks: 1) Various power interrupt periods are not considered; 2) the 2-macro memory architecture slows access speed; 3) worst-case store/restore operations are always performed. We present a 65nm fully-CMOS-logic-compatible ReRAM-based NVP achieving time/space-adaptive data retention. A 1-macro nvSRAM with self-write-termination (SWT) is integrated to boost clock frequency and reduce store energy. The adaptive retention and SWT strategy relieve the ReRAM write endurance challenge (10^6-10^{12}), making it sufficient for most applications. The NVP operates at 100MHz with 20ns/0.45nJ restore time ($T_{RESTORE}$)/energy ($E_{RESTORE}$), realizing 6× reduction in $T_{RESTORE}$, >6000× reduction in $E_{RESTORE}$ and 4× higher clock frequency compared with existing designs.

Figure 4.7.1 shows the challenges of a conventional NVP. When power failures happen, the data in volatile registers and memory is stored into nonvolatile flip-flops (nvFFs) and a NVM macro. However, challenges exist: 1) Previous nonvolatile controllers (NVCs) lack time-domain-adaptive retention for variable power-interrupt periods. For power disruptions longer than a "breakeven time", NVM store should be used for zero leakage, while fast but leaky data retention by lowering V_{DD} is preferred for short power interrupts. Furthermore, faster restore speed is always desired to catch critical events, while store speed can be relaxed by an energy buffer capacitor C_{bulk}; 2) The 2-macro architecture attaches slow NVM on the bus, leading to lower performance in normal mode. In restore operations, data has to be read from NVM to SRAM sequentially, degrading restore speed/energy by 2-to-4 orders of magnitude; 3) Conventional NVPs adopt worst-case data retention, assuming all contents in registers and memory should be updated. However, in real applications, data values in up to 90% NVM bits match previous ones and many bits are in fact unused.

Figure 4.7.2 shows the architecture of the fabricated NVP, consisting of an adaptive NVC, a code ReRAM macro, adaptive nvFFs and a configurable nvSRAM. The NVC workflow is described as follows. The time-domain controller detects the sleep signal, and uses a 2b predictor to determine whether the retention or store operation should be performed. If the store operation is adopted, the NVP backs up volatile data and enters the OFF state. Otherwise, it goes to RETENTION mode. If the retention time exceeds a specific threshold, a timeout signal is generated and the NVP enters STORE mode. When the wakeup signal is active, the NVP goes back to NORMAL mode. During the store/restore operation, the space-domain controller manages the memory size by generating the variable ADDR for nvSRAM.

Figure 4.7.3 presents the adaptive nvFF with SWT. The ReRAM device is switched to a low/high resistance state (LRS/HRS) by applying a SET/RESET voltage (V_{SET}/V_{RESET}) for period T_{SET}/ T_{RESET}. The nvFF has 4 modes: NORMAL, STORE, RESTORE and RETENTION. In NORMAL mode (RSWL=0), it acts as a typical flip-flop. In STORE operation, the data is moved into two ReRAM devices (RL and RR). As ReRAM devices suffer from a wide distribution in T_{SET}/T_{RESET} [5], SWT senses the NX/Q voltage and terminates the STORE operation, suppressing wasted store energy and degraded reliability due to over-RESET/SET for fast-switch cells. In the RESTORE operation, Q becomes logic-0/1 when RL=LRS/HRS. In RETENTION mode, the clock is gated and V_{DD} is lowered to 0.4V to reduce leakage, while most of other domains are power gated. This mode reduces store energy and extends the lifetime of ReRAM devices by avoiding frequently writing to ReRAM in short power interruptions. A scan chain is used to access nvFFs for testability.

Figure 4.7.4 shows the adaptive nvSRAM, which has three modes: SRAM, STORE and RESTORE. In SRAM mode (RSWL=0), the 7T1R-nvSRAM [4] has the same high-speed read/write behavior as a nominal 6T-SRAM. In RESTORE mode, by using a dual-supply-initialization pulse-overwrite (DSI-POW) scheme, Q will become logic-0 (logic-1) if ReRAM is LRS (HRS). The adaptive restore controller trades off restore speed and peak current by configuring restore parallelism (1/4/16WL, i.e. 1×16B, 4×16B, or 16×16B). In STORE mode, nvSRAM cells in the same page store data from SRAM to ReRAM in parallel. However, there would be a large store-DC-current causing nontrivial wasted energy if the store conditions are continuously applied for the worst-case store, determined by the slowest cell. This work incorporates the SWT circuit into each column of nvSRAM array for energy savings. At the beginning of storing, CVDDQ=CVDDQB=V_{SET}-V_{RESET} and BL=BLB=1, leading to SWT being initialized. When WL=1, the BL of a 0-cell (Q=0) drops, causing RSL=V_{SET} for a SET operation. After ReRAM devices finish the SET operation and switch to LRS, the large $I_{STORE\text{-}DC}$ raises the voltage at node Q, and QB flips from 1 to 0. BLB drops and Driver_EN=0, which terminates the SET operation. Similarly, the RESET operation is performed for a 1-cell.

Figure 4.7.5 shows store/restore time/energy of the adaptive NVP. The store energy is reduced via three adaptive strategies: nvSRAM sizing, SWT, and time-domain adaptive retention, which can be applied in a stacking way. The nvSRAM sizing configures the capacity (0/16B/256B/1KB/4KB) according to application requirements. It achieves up to 28× energy savings by eliminating store operations in unused memory. The SWT circuits in the nvFF and nvSRAM avoid store operations in matched ReRAM cells and reduce store energy by up to 172×. A matched cell is defined as a ReRAM device, whose previous value equals the data to be written. The typical ratio of matched cells in NVPs is larger than 80% for embedded benchmarks. Finally, the time-domain adaptive retention saves as much as 2× store energy if most power gating periods are shorter than the breakeven time.

Figure 4.7.6 shows the measurement results. An 8b NVP was implemented using a 65nm CMOS process and logic-process-compatible contact ReRAM [6]. It runs a counter program at 100MHz with a data waveform displayed by GPIOs. The restore time is 170ns under a 4KB nvSRAM configuration, while 20ns is achieved under a 16B nvSRAM configuration. The Shmoo figure of the nvSRAM shows it operates at a wide supply range of 0.4-to-1V. The ReRAM-based NVP is 4× faster than state-of-the-art NVPs benefiting from the 1-macro hybrid memory architecture. The adaptive data retention scheme achieves 6× speedup and >6000× energy reduction in restore operations. Figure 4.7.7 shows the die photograph and summary table.

Acknowledgements:
Supported in part by 863 Project 2013AA01320, YETP0102, Beijing Advanced Innovation Center for Chip Excellence, and NSFC Grant #61271269.

References:
[1] Y. Wang et al., "A 3us Wake-up Time Nonvolatile Processor Based on Ferroelectric Flip-Flops," *European Solid-States Circuits Conf.,* pp. 149-152, 2012.
[2] S. Bartling et al., "An 8MHz 75μA/MHz Zero-Leakage Non-Volatile Logic-Based Cortex-M0 MCU SoC Exhibiting 100% Digital State Retention at VDD=0V with <400ns Wakeup and Sleep Transitions," *ISSCC Dig. Tech. Papers*, pp. 432-433, 2013.
[3] N. Sakimura et al., "A 90nm 20MHz Fully Nonvolatile Microcontroller for Standby- Power-Critical Applications," *ISSCC Dig. Tech. Papers*, pp.184-185, 2014.
[4] 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." *IEEE Symp. VLSI Circuits*, pp. C76-C77, 2015.
[5] M.-F.Chang et al., "Embedded 1Mb ReRAM in 28nm CMOS with 0.27-to-1V Read Using Swing-Sample-and-Couple Sense Amplifier and Self-Boost-Write-Termination Scheme." *ISSCC Dig. Tech. Papers*, pp. 332-333, 2014.
[6] M.-F. Chang et al., "A 0.5V 4Mb Logic-Process Compatible Embedded Resistive RAM (ReRAM) in 65nm CMOS Using Low Voltage Current-Mode Sensing Scheme with 45ns Random Read Time," *ISSCC Dig. Tech. Papers*, pp. 434-435, 2012.

 978-1-4673-9467-3/16/$31.00

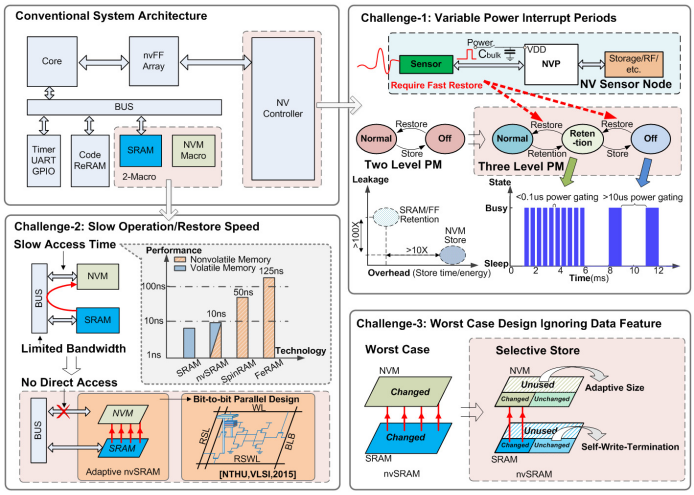


Figure 4.7.1: Challenges of a conventional NVP.

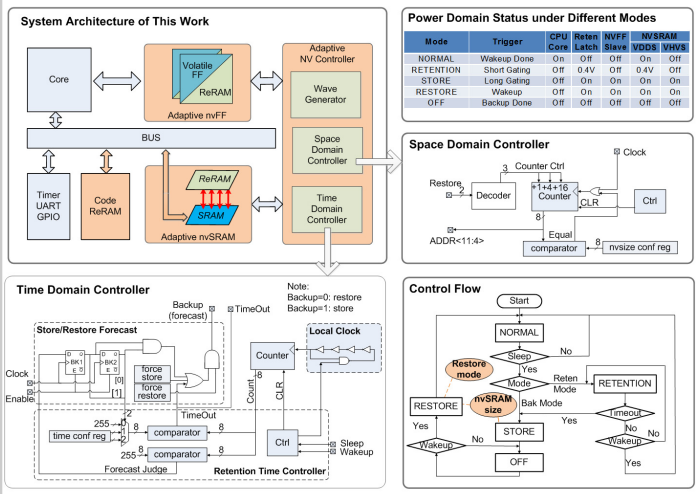


Figure 4.7.2: The fabricated NVP with adaptive data retention.

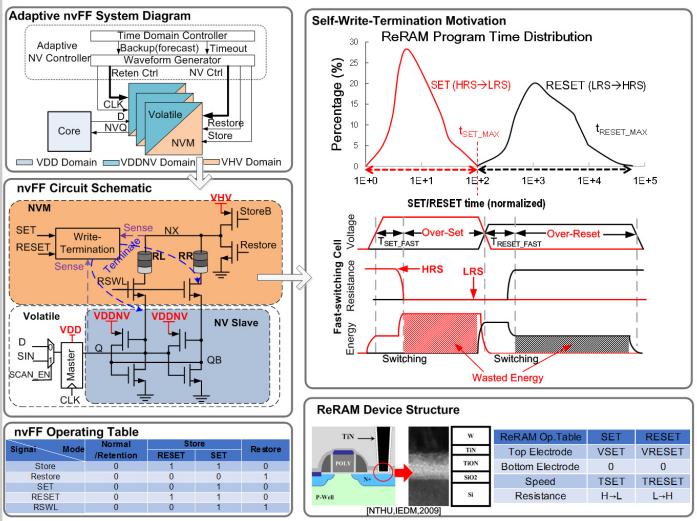


Figure 4.7.3: Adaptive nvFF with data retention and self-write-termination (SWT).

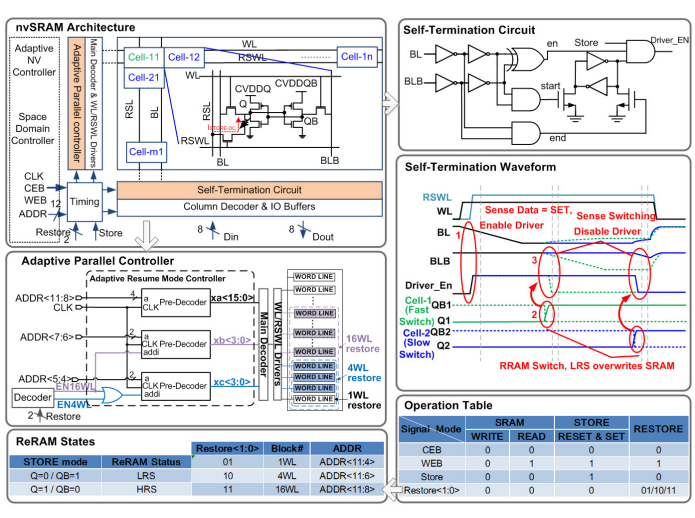


Figure 4.7.4: Adaptive nvSRAM with reconfigurable size and self-write-termination (SWT).

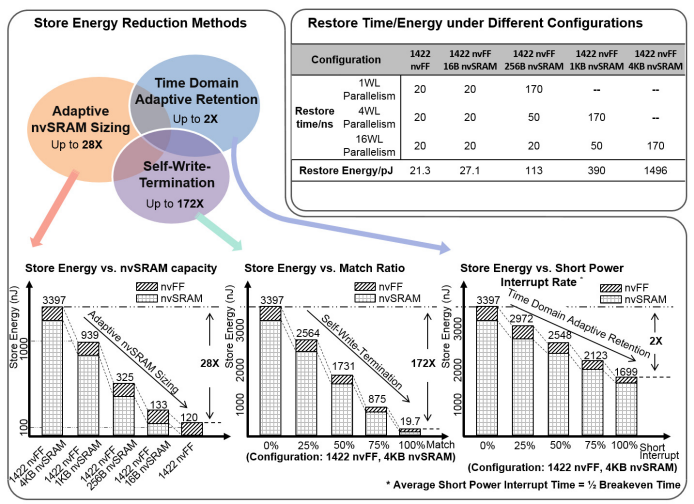


Figure 4.7.5: Restore/store time/energy of the adaptive NVP.

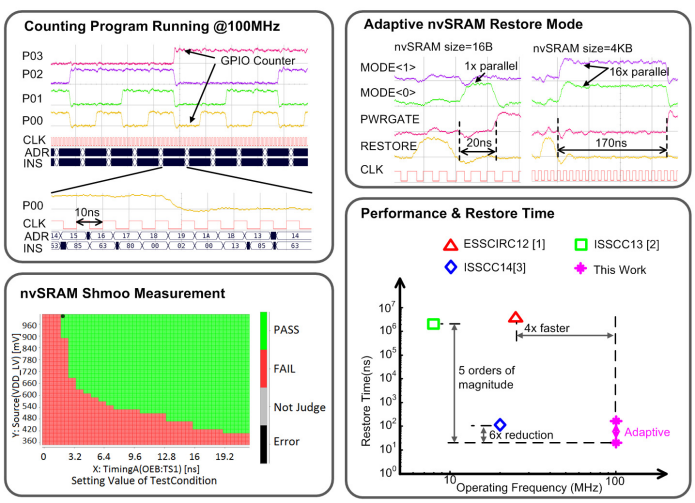
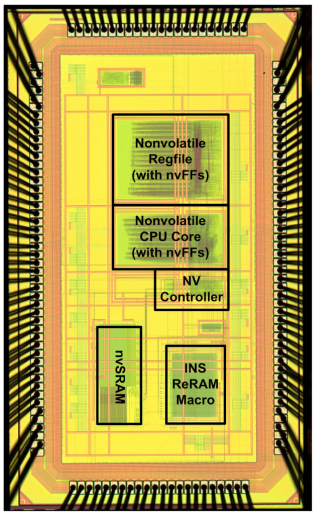


Figure 4.7.6: Measurement results of the adaptive NVP.



Affiliation	THU/NTHU
Technology	65nm SVT CMOS +ReRAM
ISA	8051
NVM	8KB ReRAM 4KB nvSRAM 1422bits nvFF
Supply	0.8V(Core), 3V(HV)
Chip Area	1560 x 2860 um^2
Nonvolatile Area	nvFFs: 5.98% nvSRAM: 4.71% ReRAM Macro: 5.84%
Fequency	>100 MHz
Active Power	33 uW/MHz@100MHz
System Restore Time	20 - 170 ns
System Restore Energy	0.45nJ (Avg)
System Store Time	4us- 1.02 ms
System Store Energy	0.40uJ (Avg)

Figure 4.7.7: Die photo and metric table.

978-1-4673-9467-3/16/$31.00 ©2016 IEEE