21.3 A High-Density 45nm SRAM Using Small-Signal Non-Strobed Regenerative Sensing

Naveen Verma, Anantha P. Chandrakasan

Massachusetts Institute of Technology, Cambridge, MA

High-density SRAMs are a primary contributor to the dramatic cost reductions and expanding features of ICs every technology node. Unfortunately, their small bit-cell devices have large variation, and the ensuing degradation in both I_{READ} and read SNM must be addressed simultaneously. In addition, variation severely affects sense-amplifier (SA) performance, as offset voltage and strobe timing uncertainty are dominating limitations. In this work, an SRAM in low-power 45nm CMOS is composed of 0.25μm^2 bit-cells and uses an offset-compensating non-strobed regenerative sense-amplifier (NSR-SA); this eliminates the need to insert timing margin for variation and tracking errors in the strobe signal, and it achieves higher sensitivity than conventional SAs, allowing read SNM to be optimized at the cost of I_{READ}.

Figure 21.3.1 underlines the trade-offs plaguing high-density SRAMs. The scatter plot shows an inverse correlation between I_{READ} and read SNM for the 6T cell used. Extreme variation in 0.25μm^2 cells causes both to be severely degraded, but the foremost concern of vanishing read SNM requires aggressive stability assists, usually reducing I_{READ} even further [1]. Accordingly, BL/BLB discharge, to exceed the SA offset, which itself is prominently affected by variation, dominates the access time. Further, timing variation in the strobe delay circuit, and its poor tracking of the array read-path across operating corners [2], is emerging as a significant source of uncertainty requiring 100 to 200ps of additional margin. Although full-swing sensing avoids offset and timing limitations [3], it is impractical for high-density SRAMs, where I_{READ} is very low and BL/BLB capacitance is very high—with up to 256 cells/column.

The NSR-SA of Fig. 21.3.2 overcomes these challenges giving a measured reduction of up to 34% in read access-time. The NSR-SA has the following features: 1) simple offset compensation, imposing negligible loading on high-speed nodes while reducing sensitivity to variation; 2) continuous sensing after reset, eliminating the strobe signal and its associated timing uncertainty; 3) single-sided regeneration, efficiently providing high sensitivity [4] but responding only to BL discharge, ensuring robustness against noise causing false regeneration; and, 4) single-ended operation, allowing its use with asymmetric cells (8T, 6T), whose improved stability is gaining popularity at advanced CMOS nodes [5,6]. Importantly, the NSR-SA uses nearly minimum-sized inverters, so its delay and area scaling follow logic trends more closely than conventional SAs.

During BL/BLB precharge, RST, in Fig. 21.3.2, is asserted, and the inverters, formed by M1/M2 and M3/M4, are reset to their high-gain regions. Subsequently, as the waveforms of Fig. 21.3.3 show, when RST is de-asserted, nodes X/Y remain at approximately equal voltages. Accordingly, the V_{GS} of the positive feedback device, M7, is less than or nearly zero, and the device is off. After WL assertion, if no BL discharge is detected, all node voltages remain unchanged. Alternatively, small discharge causes rapid increase in the voltage of node X and even more rapid decrease in the voltage of node Y due to the inverter gains. Eventually, M7 is turned on by its increasing V_{GS}, triggering positive feedback; the first inverter input is actively pulled low, causing X/Y to sharply regenerate (in <100ps), as shown in Fig. 21.3.3. The strong resulting overdrive on M5 quickly causes the output state, QB, to change. As noted, regeneration is triggered by the input signal itself, not an explicit strobe signal.

The goal of offset compensation is to set the amount of BL discharge required to flip the output state and diminish its sensitivity to variation. The reset X/Y voltages set $V_{GS,\,M7}$ after reset and therefore determine how much additional BL discharge triggers regeneration. Their values are chosen, based on speed and noise-rejection considerations, by setting the M1/M2 and M3/M4 strength ratios (i.e. inverter trip-points). In practice, the actual voltages get skewed by variation, which can be modeled as a voltage error in series with the input; however, during reset, negative feedback forces each inverter to its nominal trip-point minus that input offset. This voltage is stored on C1/C2, and, now, since the negative of each offset effectively appears in series with the true input offset, the transfer functions from IN to X and from X to Y are nearly ideal. Only the offset of M7 remains; however, when input-referred, its effect is reduced by a factor of $[g_m r_o + (g_m r_o)^2]$, the input gain to $V_{GS,M7}$ (g_m and r_o are the transconductance and output-resistance of transistors M1 to M4). A residual source of error is false regeneration due to charge-injection from the reset switches. However, the NSR-SA, exploits the fact that it must respond only to BL discharge, not charge-up. Specifically, it regenerates only when node X increases and node Y decreases. So, the reset switches are implemented as indicated in Fig. 21.3.2, where PMOS switch charge-injection causes the M1/M2 gate voltage to increase, while NMOS switch charge-injection causes the M3/M4 gate voltage to decrease. The resulting voltage errors cause X/Y to oppose regeneration (see case where BL does not discharge in Fig. 21.3.3).

The benefit of offset compensation and non-strobed sensing is shown in the simulated distributions of Fig. 21.3.4. Being differential, the conventional SA (schematic shown) nominally requires very little BL/BLB discharge and, therefore, achieves good mean/mode access-time. With variation, however, the NSR-SA has superior sigma, and it is free from the 150ps of strobe timing margin that, in simulations, the conventional SA requires. Its 3σ access time, for the considered 256×256 array with mean bit-cells, is 486ps compared to 610ps for the conventional SA, representing a speed-up of over 20%.

The test-chip architecture is shown in Fig. 21.3.5. Two 256×256 arrays of high-density, low-power 0.25μm^2 cells allow comparison of the NSR-SA against the conventional SA of Fig. 21.3.4. Their timing paths for WLE/CLKIN are carefully matched to ensure minimum relative measurement errors. To evaluate the trade-off between sensitivity and noise rejection, which is fundamental to single-ended sensing, a circuit that injects a controlled voltage noise on BL/BLB via capacitive coupling is also incorporated, along with parallel current-source devices that manually adjust sensitivity by trimming the reset X/Y voltages.

The die micrograph is shown in Fig. 21.3.7. Each NSR-SA occupies 19μm^2, approximately equal to the conventional SA. The measured access-time (WLE to CLKIN) distributions from 53 chips are shown in Fig. 21.3.6. Since the strobe signal is generated off-chip, its variation is not considered. Still, the access-time differences show that the NSR-SA offers a speed-up of up to 34%, and a mean speed-up of 20%, even without the additional timing margin required by the conventional SA. Further, as expected, the NSR-SA has lower sigma, confirming the benefit of offset-compensation. Fig. 21.3.6 also shows measurements of how the NSR-SA's speed can be increased at the cost of BL noise rejection. Finally, the power of the NSR-SAs is measured to be 19μW each in reset, and they increase the total array power by 7% when operating at 100MHz.

Acknowledgements:
N. Verma is supported by the Intel Foundation Ph.D. fellowship program and IC fabrication is provided by Texas Instruments. We thank M. Clinton, X. Deng, T. Houston, and W.-K. Loh for their support and feedback.

References:
[1] M. Yabuuchi, K. Nii, Y. Tsukamoto et al., “A 45nm Low-Standby-Power Embedded SRAM with Improved Immunity Against Process and Temperature Variations,” *ISSCC Dig. Tech. Papers*, pp. 326-327, Feb. 2007.
[2] K. Sohan, N. Cho, H. Kim et al., “An Autonomous SRAM with On-Chip Sensors in an 80nm Double Stacked Cell Technology,” *Dig. Symp. VLSI Circuits*, pp. 232-235, Jun. 2005.
[3] K. Zhang, K. Hose, V. de and B. Senyk, “The Scaling of Data Sensing Techniques for High-Speed Cache Design in Sub-0.18μm technologies,” *Dig. Symp. VLSI Circuits*, pp. 226-227, Jun. 2000.
[4] J.-T. Wu and B. Wooley, “A 100MHz Pipelined CMOS Comparator,” *IEEE J. Solid-State Circuits*, vol. 23, no. 6, pp. 1379–1385, Dec. 1988.
[5] Y. Morita, H. Fujiwara, H. Noguchi et al., “An Area-Conscious Low-Voltage-Oriented 8T-SRAM Design Under DVS Environment,” *Dig. Symp. VLSI Circuits*, pp. 256-257, Jun. 2007.
[6] R. Joshi, R. Houle, K. Batson et al., “6.6+ GHz Low Vmin, Read and Half Select Disturb-Free 1.2Mb SRAM,” *Dig. Symp. VLSI Circuits*, pp. 250-251, Jun. 2007.

 978-1-4244-2010-0/08/$25.00 ©2008 IEEE

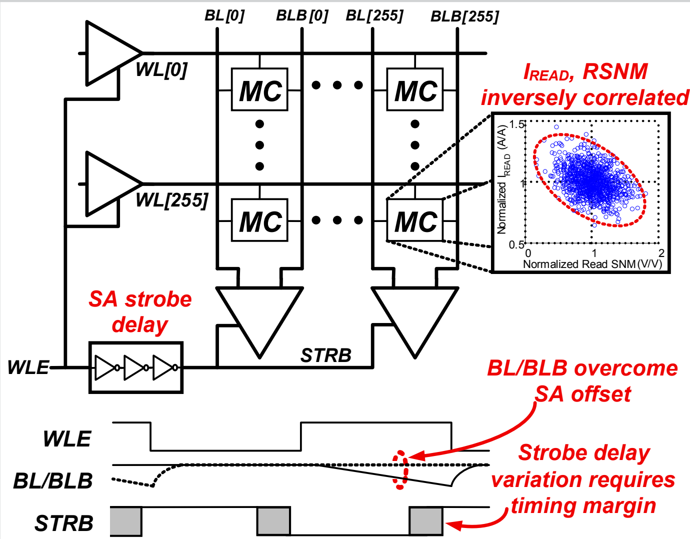

Figure 21.3.1: Conventional SRAM read-access path showing I_{READ}-read SNM limitation and SA-strobe timing uncertainty.

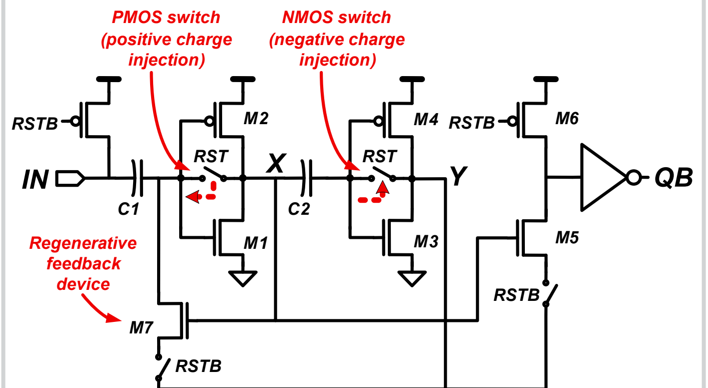

Figure 21.3.2: Non-strobed regenerative sense-amplifier (NSR-SA) circuit schematic.

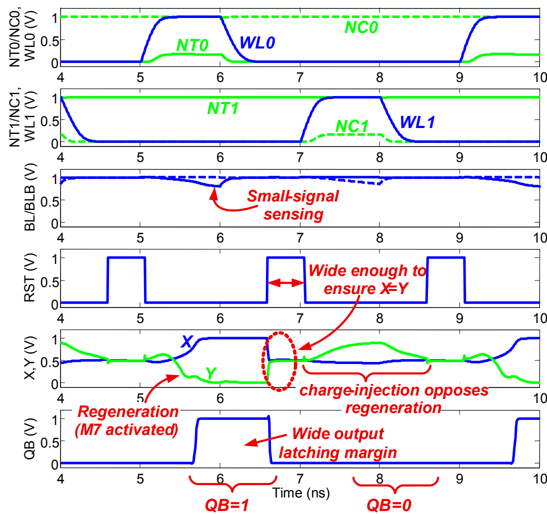

Figure 21.3.3: NSR-SA simulated waveforms showing sensing of both data states (slow process, low voltage condition); NT/NC represent bit-cell storage nodes.

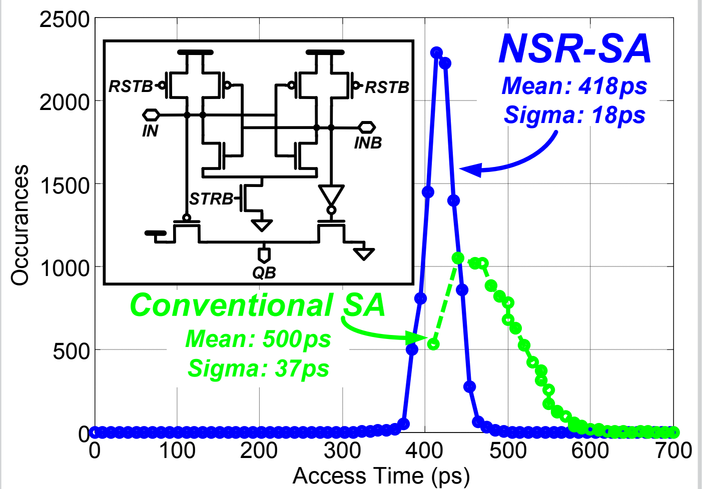

Figure 21.3.4: Simulated access-time distributions for conventional SA (schematic shown) and NSR-SA.

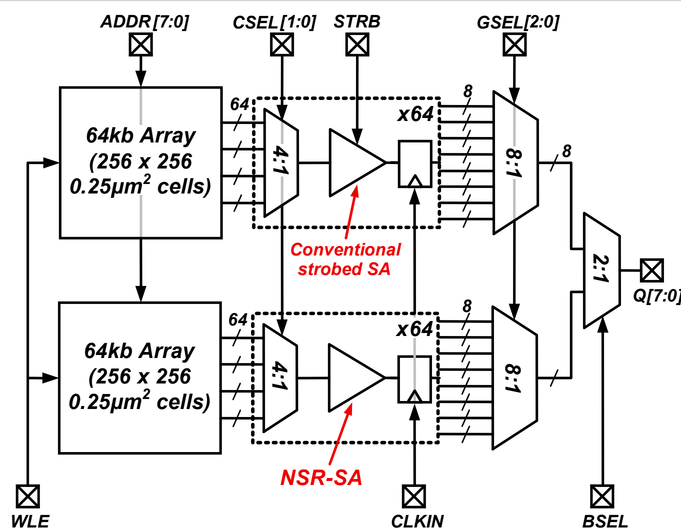

Figure 21.3.5: Block diagram of test-chip designed to measure NSR-SA access time speed-up compared to conventional SA.

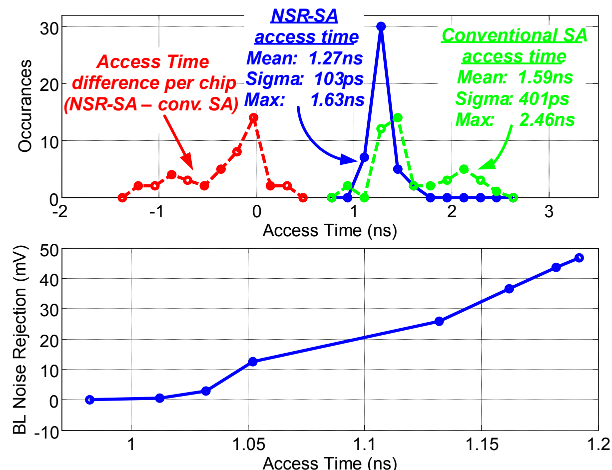

Figure 21.3.6: Measured access times (53 chips) and noise rejection. Worst-case NSR-SA and conventional SA access-time are 1.63ns and 2.46ns respectively. NSR-SA speed increases at the cost of BL noise-rejection (injected noise is estimated from nominal ratio of on-chip capacitor divider).

Continued on Page 621

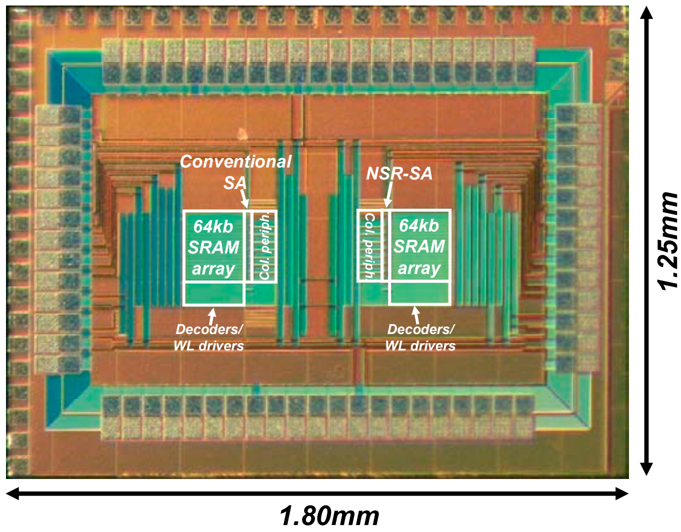

Figure 21.3.7: Die micrograph of 45nm SRAM test-chip with two 64kb arrays.

 978-1-4244-2010-0/08/$25.00 ©2008 IEEE