A Double-Tail Sense Amplifier for Low-Voltage SRAM in 28nm Technology

Pi-Feng Chiu, Brian Zimmer, Borivoje Nikolić
Email: pfchiu@eecs.berkeley.edu
Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA

***Abstract*—A double-tail sense amplifier (DTSA) is designed as a drop-in replacement for a conventional SRAM sense amplifier (SA), to enable a robust read operation at low voltages. A pre-amplification stage helps reduce the offset voltage of the sense amplifier by magnifying the input of the regeneration stage. The self-timed regenerative latch simplifies the timing logic so the DTSA can replace the SA with no area overhead. A test chip in 28nm technology achieves 56% error rate reduction at 0.44V. The proposed scheme achieves 50mV of VDDmin reduction compared to commercial SRAM with a faster timing option that demonstrates a smaller bitline swing.**

I. Introduction

Variations associated with technology scaling together with increased on-die memory capacities cause a wide distribution of bitline swings (ΔV), requiring a tight control of the sense amplifier offset voltage in SRAMs. To achieve the targeted sensing yield, the conventional sensing scheme margins the bitline swing with a longer sensing time (CLK-SAE delay), which degrades the performance and consumes more energy. Figure 1 shows the Monte Carlo simulation results of ΔV in a commercial SRAM in a 28nm process, indicating that the distribution of ΔV at 0.6V is twice as wide as the distribution at 0.9V. Therefore, a more robust sensing scheme is needed at low voltage.

The conventional latch-type sense amplifier [1] shown in Figure 2(a) is widely used for SRAM read due to its compact structure, but the offset voltage increases with decreasing voltage, limiting robustness at low voltage [2]. Various sensing techniques have been proposed to improve SA robustness, but proposed techniques require additional complexity, such as matching capacitors for offset cancellation [3], extra configuration steps for SA redundancy [4] [5] and microarchitectural modifications in error detection and correction [6].

II. Double-Tail Sense Amplifier

The schematic of the DTSA is shown in Figure 2(b) [7] [8]. It consists of two stages. The dynamic pre-amplification stage provides a small gain through the input pair, $M_{IN1,\ IN2}$ and the regeneration stage, consisting of cross-coupled inverters ($M_{U1,U2,D1,D1}$) completes the comparison. Figure 3 shows the simulated waveforms of the DTSA at 0.6V VDD and 50mV ΔV. Vinp and Vinn are set to VDD and VDD-ΔV, respectively. During the reset phase, DN and DP are precharged to VDD by M_{PRE1} and M_{PRE2}. All the internal nodes in the second stage are discharged to 0 by $M_{N1\text{-}N4}$. After the first stage is activated by SAE, DN and DP are discharged at different rates depending on the input voltage levels (Vinn, Vinp), the parasitic capacitance, and the tail current. The output signal of the first stage, DN and DP, serves as both the enabling and the input signal for the second stage. As the voltage of DN and DP drop below VDD-V_{TP}, the second stage is turned on by $M_{H1,2}$. There is a short period of time in which only one side of the latch is enabled so that the side that is pulled high can develop the signal without contention with the other branch. Three pairs of input transistors, $M_{H1,2}$ and $M_{N1\text{-}4}$, increase the pre-amplification gain and reduce the input-referred offset voltage. Finally, the latch completes the signal regeneration and the output signals (OP, ON) are fed to an SR latch.

Figure 4 shows the simulated error rate of the DTSA and the conventional SA at different ΔV and supply voltages. The error rate of the DTSA is slightly worse than that of the conventional SA at 0.9V, but it is more robust at lower supply

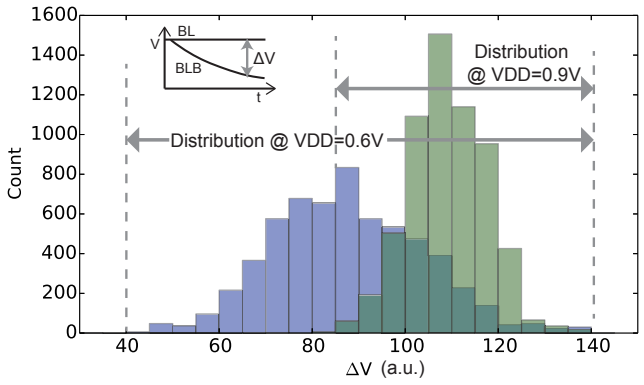
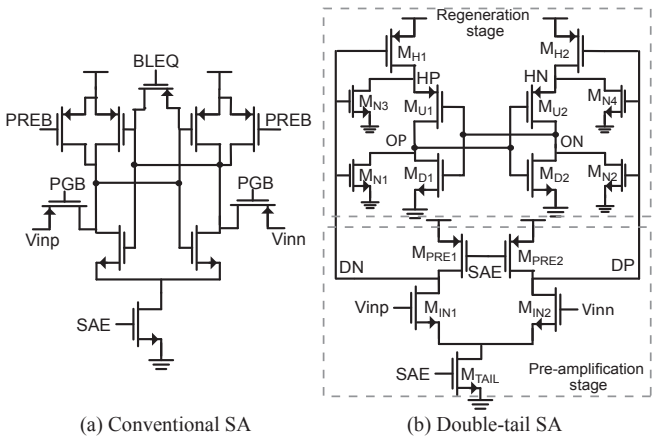


Fig. 1. Histogram of bitline swing (ΔV) at 0.9V and 0.6V.



(a) Conventional SA (b) Double-tail SA

Fig. 2. Circuit diagrams of (a) the conventional SA and (b) the DTSA.

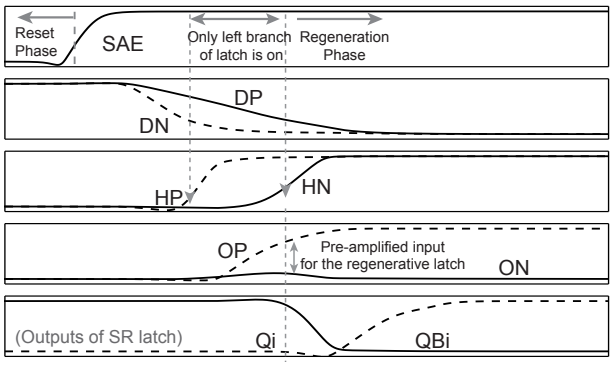


Fig. 3. Operation waveform of the DTSA at 0.6V and ΔV = 50mV.

voltages because of the integration in the pre-amplification stage. The error rate is reduced by 83% at 0.45V with a ΔV of 20mV and the offset voltage under the iso-robustness condition (99.9% yield) is 22% lower than the conventional SA. The conventional SA shows the worst case at SF corner with 8.1x larger error rate than that at TT corner, while a mix of PMOS and NMOS input pairs make the DTSA equally resilient across all corners, as shown in Figure 5.

The wider ΔV distribution at low VDD requires a sense amplifier that has lower offset at low voltage to compensate for the low read current in the weakest bitcells on the chip. The current ratio of M_{IN1} and M_{IN2} in the DTSA under the same ΔV becomes larger as VDD is decreased. A higher gain for the pre-amplification stage in the DTSA at low voltage is achieved due to a larger current ratio in the input pairs and a longer integration time. The Monte Carlo simulation result in Figure 6 shows that the offset voltage of the conventional SA increases for lower VDD, while the offset voltage of DTSA actually decreases at low voltages. The DTSA shows better robustness when VDD is lower than 0.7V. Therefore, the DTSA is favored for low-voltage applications.

The offset voltage is largely determined by the sizing of the input pairs (M_{PRE1} and M_{PRE2}) and the tail transistor (M_{TAIL}). The sizing of the regeneration stage is relaxed because of the pre-amplified signal. The offset voltage is reduced by using a longer gate length for the input pairs, which also increases the

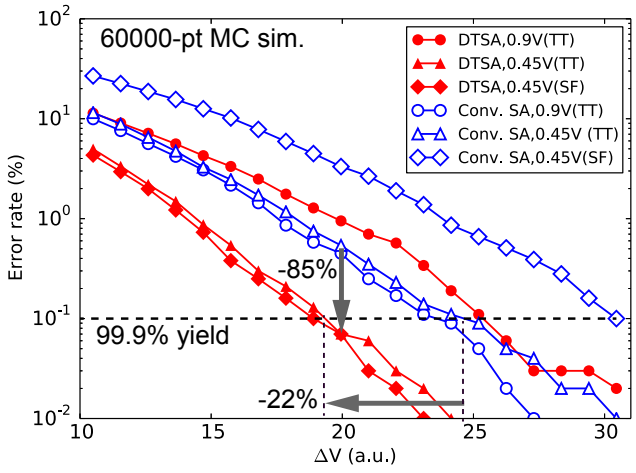


Fig. 4. Simulated error rate with various ΔV for the conventional SA and the DTSA.

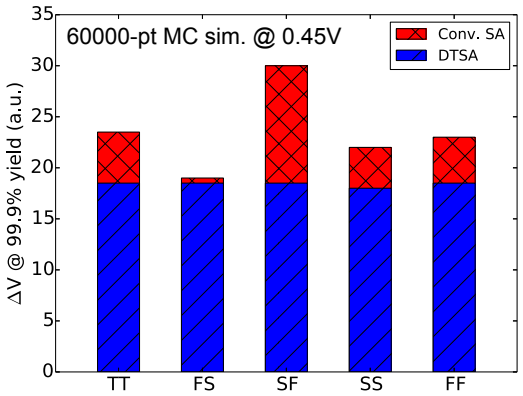


Fig. 5. Offset voltages at different process corners (VDD=0.45V).

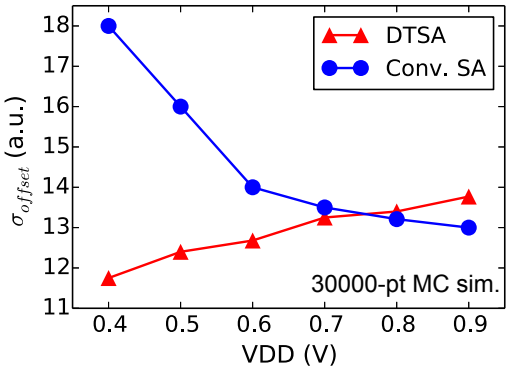


Fig. 6. The offset voltage of the conventional SA and the DTSA at different supply voltages.

area and the load capacitance of the first stage. By decreasing the size of the tail transistor, the integration time for the pre-amplification stage becomes longer, which results in a larger ΔV for the regeneration stage. However, this change degrades the sensing time. Figure 7 shows the trade-off between speed and offset voltage. The size of M_{TAIL} is chosen to have a low offset voltage with acceptable speed.

Figure 8 illustrates the dimensions of the conventional SA and the DTSA with the timing logic and the output latch. The DTSA has 50% more transistors than the conventional SA, resulting in a 23% area overhead. However, the sense amplifier itself occupies only half of the area of the read-out circuit in the traditional design, as timing logic is required to generate the PREB, PGB, and BLEQ inputs. Since the regenerative latch is self-timed, it only requires a single-phase clock, reducing the area required to generate the timing logic. As a result, although the DTSA consists of more transistors, the entire read-out circuit fits in the same area footprint. Similarly, although the additional pre-amplification stage consumes 74% more power than the conventional SA, the overall power of the memory array is reduced by the lower bitline swing.

III. Silicon Measurement Results

Four 72kb SRAM macros were fabricated in the TSMC 28nm HPM process, two with the conventional SA and two with the DTSA. The die photo and the measurement setup are shown in Figure 9. The data from the arrays are read by running built-in self-test (BIST) to detect errors while sweeping voltages and frequencies. To generate different ΔV,

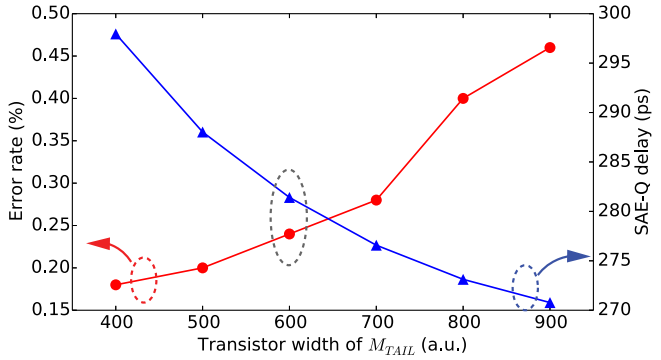


Fig. 7. Trade-off between error rate and SAE-Q delay for various sizings of M_{TAIL}.

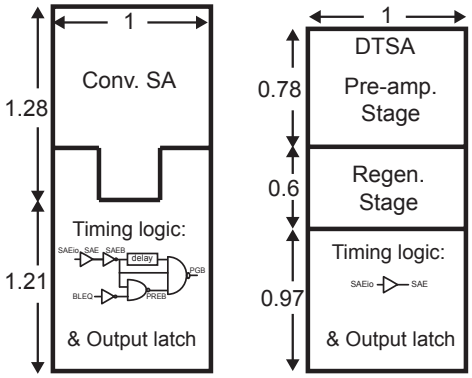


Fig. 8. The dimensions of the conventional SA and the DTSA (normalized to width).

two of the macros are designed with tunable timing options. The tunable timing circuit is constructed by a 3-bit delay chain. The tunable timing code (TTC) is set to a lower value to enable early SAE triggering for a smaller ΔV.

The measured error rate of the SRAM with the DTSA and different tunable timing codes are shown in Figure 10. The error rate is averaged over 6 chips with 72 SAs per chip. The SRAM with the DTSA achieves a 56% error rate reduction compared to the original timing circuit of the commercial SRAM. It also achieves comparable error rate to the commercial SRAM around tunable timing code 3, which corresponds to 41% shorter sensing time (CLK-SAE delay)

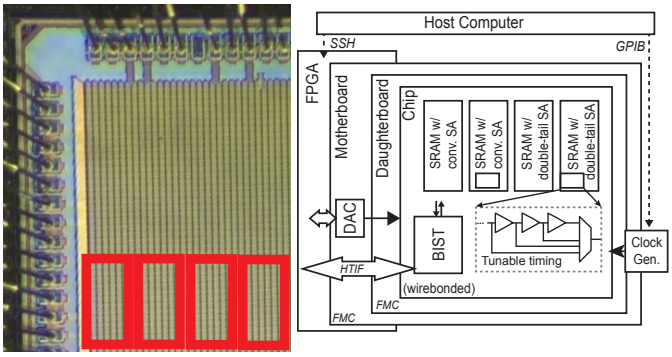


Fig. 9. Die photo and the measurement setup.

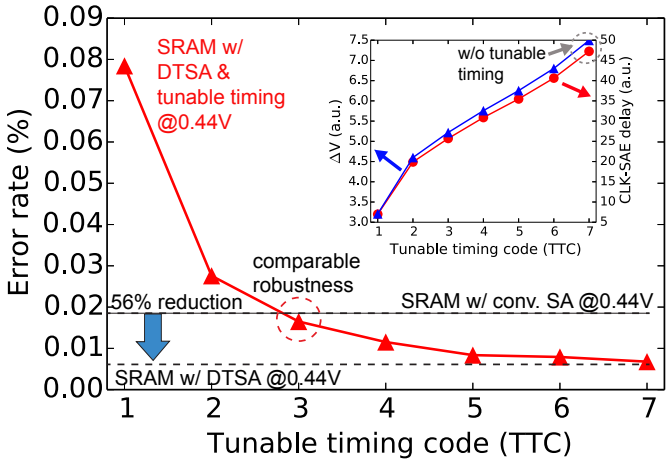


Fig. 10. Measured error rate for various tunable timing settings at 0.44V (averaged across 6 chips).

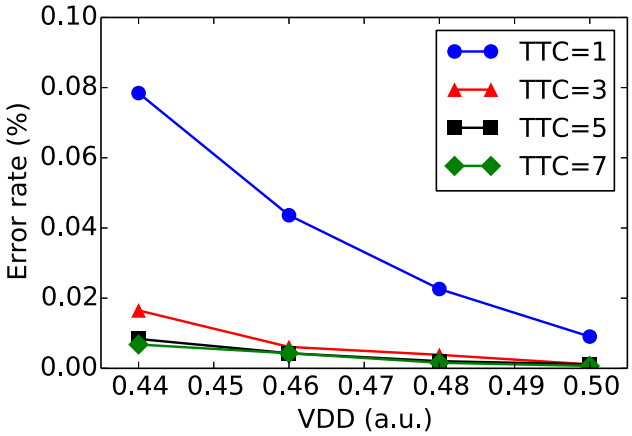


Fig. 11. Measured error rate at different VDD and tunable timing settings (averaged across 6 chips).

and 46% smaller ΔV in post-layout simulation. However, the distribution for 432 SAs does not capture the worst-case process variation. Therefore, to guarantee high yields, extra margin should be added to cover the worst case. If a 10mV margin is added based on the worst-case corner (SF corner) simulations shown in Figure 4, the ΔV is still improved by 30%.

Adjusting the tunable timing effectively shifts the ΔV distribution. Figure 11 shows the error rate at different VDD and CLK-SAE delay. A longer CLK-SAE delay (smaller TTC) ensures enough read margin, reducing the error rate at the expense performance and energy efficiency. With more aggressive timing settings, the error rate at lower voltages increases rapidly because of the long tail in the ΔV distribution.

Figure 12 shows shmoo plots of the SRAM with the conventional SA and the DTSA, respectively. Even with 46% smaller ΔV (tunable timing code 5), the SRAM with the DTSA achieves 50mV of VDDmin reduction compared to the commercial SRAM. The CLK-SAE delay reduction allows the SRAM with the DTSA to operate at a higher frequency and a lower voltage.

TABLE I. COMPARISON BETWEEN DIFFERENT SAS

Source	Conventional	DTSA (This work)	VTS-SA [3]	Reconfigurable SA [4]	RazorSRAM [6]
Technology	28 nm	28 nm	28 nm	28 nm FDSOI	28nm FDSOI
Circuit area	1X	1X	1X + MOM cap	1X + fuse	–
Features	–	Dynamic preamplifier, Self-enabled latch	Reconfiguration, Static preamplifier, Offset calibration	Reconfigurable SA redundancy	High throughput, Error correction and detection
Design effort overhead	–	None	Extra capacitors Complex timing	Run BIST Test, Configure with fuses	Error control in memory controller
Sensing speed improvement	–	13.3% @0.6V	34% @1V	–	79% @0.6V
Offset voltage reduction	–	22% @0.45V	–	58.8%	–

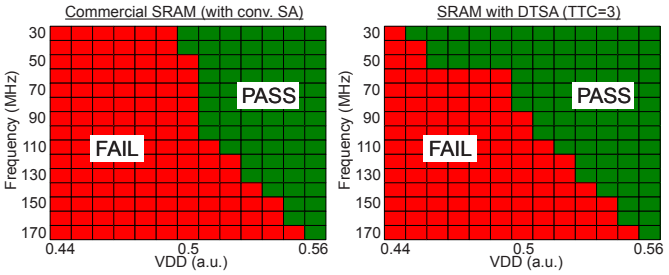


Fig. 12. Shmoo plots of the SRAM with the conventional SA and the DTSA.

IV. COMPARISON

Table I compares the conventional SA and the DTSA with previously published SAs. VTS-SA [3] utilizes the same cross-coupled latch for the static pre-amplification and offset compensation with reconfiguration. It induces short-circuit current in the pre-amplification phase, requires extra auto-zeroing capacitors and precise array-scale timing circuits in the offset sampling phase. The Reconfigurable SA [4] is able to combine the offset voltages of sub-unit SAs and choose the best configuration. The hidden cost of the approach is the time/area/energy of programming the fuses for the configuration. RazorSRAM [6] improves the throughput by eliminating the guard band of the sensing margin and uses error detection and correction to fix the failing bits. Microarchitectural modification is required to realize this scheme.

All approaches report large improvements in robustness, performance, or offset voltage reduction. However, the hidden costs of design complexity make adoption of these approaches cumbersome. The DTSA improves the offset voltage and the sensing speed at low voltage by changing the circuit topology with minimum design effort. The comparable area and compatible timing signals make the DTSA attractive as a drop-in replacement of the conventional SA in commercial memories.

V. CONCLUSION

The DTSA enables robust SRAM sensing at low voltages, which allows for smaller bitline swings with early SAE timing. Preventing excessive bitline discharge improves both the performance and the energy efficiency of SRAM. Although small ΔV leads to a longer SAE-Q delay, the early SAE timing still contributes to a 13.3% reduction of the overall CLK-Q delay in post-layout simulation at 0.6V. The silicon measurement results show 56% error rate reduction at 0.44V and 50mV VDDmin reduction by replacing the conventional SA in commercial SRAM with the DTSA design. This design is a direct drop-in replacement for the existing sensing circuit in the commercial SRAM.

ACKNOWLEDGMENT

The authors would like to thank Vladimir Milovanović, Krste Asanović, Brian Richards, James Dunn, Scott Liao, and Jonathan Chang for their contributions. This work was funded in part by BWRC, ASPIRE, DARPA PERFECT Award Number HR0011-12-2-0016, Semiconductor Research Corporation (SRC) and fabrication donation by TSMC.

REFERENCES

[1] A. Abidi and H. Xu, "Understanding the regenerative comparator circuit," in *Proc. IEEE Custom Integrated Circuits Conference*, Sept 2014, pp. 1–8.

[2] M. Abu-Rahma, Y. Chen, W. Sy, W. L. Ong, L. Y. Ting, S. S. Yoon, M. Han, and E. Terzioglu, "Characterization of sram sense amplifier input offset for yield prediction in 28nm cmos," in *Proc. IEEE Custom Integrated Circuits Conference*, Sept 2011, pp. 1–4.

[3] B. Giridhar, N. Pinckney, D. Sylvester, and D. Blaauw, "A reconfigurable sense amplifier with auto-zero calibration and pre-amplification in 28nm cmos," in *Proc. IEEE International Solid-State Circuits Conference*, Feb 2014, pp. 242–243.

[4] M. Khayatzadeh, F. Frustaci, D. Blaauw, D. Sylvester, and M. Alioto, "A reconfigurable sense amplifier with 3x offset reduction in 28nm fdsoi cmos," in *Proc. IEEE Symposium on VLSI Circuits*, June 2015, pp. 270–271.

[5] N. Verma and A. P. Chandrakasan, "A 256 kb 65 nm 8t subthreshold sram employing sense-amplifier redundancy," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 1, pp. 141–149, Jan 2008.

[6] M. Khayatzadeh, M. Saligane, J. Wang, M. Alioto, D. Blaauw, and D. Sylvester, "A reconfigurable dual-port memory with error detection and correction in 28nm fdsoi," in *Proc. IEEE International Solid-State Circuits Conference*, Jan 2016, pp. 310–312.

[7] D. Schinkel, E. Mensink, E. Klumperink, E. van Tuijl, and B. Nauta, "A double-tail latch-type voltage sense amplifier with 18ps setup+hold time," in *Proc. IEEE International Solid-State Circuits Conference*, Feb 2007, pp. 314–605.

[8] M. Miyahara, Y. Asada, D. Paik, and A. Matsuzawa, "A low-noise self-calibrating dynamic comparator for high-speed adcs," in *IEEE Asian Solid-State Circuits Conference*, Nov 2008, pp. 269–272.