Article

A Digital Auto-Zeroing Circuit to Reduce Offset in Sub-Threshold Sense Amplifiers

Peter Beshay [1,*], Joseph F. Ryan [2] and Benton H. Calhoun [1]

[1] The Charles L. Brown Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904, USA; E-Mail: bcalhoun@virginia.edu
[2] Intel Corporation, Hillsboro, OR 97124, USA; E-Mail: joseph.f.ryan@intel.com

* Author to whom correspondence should be addressed; E-Mail: plb3qt@virginia.edu; Tel.: +1-435-938-1659; Fax: +1-434-924-8818.

Abstract: Device variability in modern processes has become a major concern in SRAM design leading to degradation of both performance and yield. Variation induced offset in the sense amplifiers requires a larger bitline differential, which slows down SRAM access times and causes increased power consumption. The effect aggravated in the sub-threshold region. In this paper, we propose a circuit that reduces the sense amp offset using an auto-zeroing scheme with automatic temperature, voltage, and aging tracking. The circuit enables flexible tuning of the offset voltage. Measurements taken from a 45 nm test chip show the circuit is able to limit the offset to 20 mV. A 16kB SRAM is designed using the auto-zeroing circuit for the sense amps. The reduction in the total read energy and delay is reported for various configurations of the memory.

Keywords: offset compensation; SRAM; sense amplifier; auto-zeroing

1. Introduction

Variation induced offset in the sense amplifiers requires a larger bit-line differential, which slows down SRAM access times and causes increased power consumption. In the sub-threshold (sub-V_T) region of operation, in particular, the effect is more dominant because the threshold voltage (V_T)

variation has an exponential effect on the drive current. It is shown in [1] that the offset gets worse in sub-threshold relative to strong inversion as technology scales. Furthermore, increasing the sizes of devices in the sense amplifier does not yield the reduction of input referred offset according to $1/(WL)^{0.5}$ that is achieved for strong inversion operation. In addition it causes increased cell instability and a severely degraded read-current [1,2]. Several attempts have been made before to tackle the problem of offset voltage in sense amplifiers (SAs) including redundancy [3], transistor upsizing [4], digitally controlled compensation [5] and dynamic compensation [6]. Our approach to eliminating offset is a digital auto-zeroing (DAZ) scheme inspired by analog amplifier offset correction [7]. The main advantages of the approach are the near-zero offset after cancellation, offset tuning, and the automatic temperature, voltage, and aging tracking achievable using a repeated offset calibration phase, which makes the design useful in the sub-threshold region due to the high offset voltage sensitivity to supply voltage and temperature variations. In Verma and Chandrakasan [3], SA redundancy is used. It requires the SRAM bit-line from each column be connected to N different SAs. One SA will be selected whose offset is bound by the high and low logic levels of bit-line. This scheme statistically reduces the reliance on SAs with high offset, but it adds area and test time calibration to select the SA to use from each set. In Pileggi [4], transistor upsizing is used. A linear response surface model was developed to relate the SA offset voltage to the threshold voltages variations. The model is used to determine the statistical optimized transistors sizing. The optimized design resulted in a 25% decrease in the standard deviation of the offset voltage at a cost of 10% increase in active area. In Ryan and Calhoun [1], a methodology is proposed for sizing sub-V_T SAs to minimize offset using SAs for ultra-low power operation. In Bhargava [5] a digitally controlled compensation is used. The scheme is applied to a latch-style and Strong-ARM SA topologies. Measured results from a 45nm test-chip show that the standard deviation of the offset is reduced by 5X. In Sachdev [6], dynamic compensation is used. A group of transistors are selectively coupled to high and low voltage levels via multi-phase timing. This results in a voltage level on nodes of interest that is a function of transistor mismatch. The voltage levels act to compensate for the transistor mismatch. This scheme is similar to the auto-zeroing scheme presented in this work. However the presented scheme uses a compensation capacitor, charge pump and feed-back circuit. Hence, the calibration phase is not necessarily needed prior to every sensing cycle. This improves the SA power consumption as will be illustrated in Section 5.

Section 2 describes the implementation of the auto-zeroing offset compensation scheme. Section 3 illustrates the voltage, temperature and aging tracking. Section 4 illustrates the offset tuning. Section 5 provides details of the power consumption and methods proposed to minimize it. Section 6 illustrates the offset sensitivity. Section 7 illustrates the offset compensation across technologies. Section 8 shows the improvements in read energy and delay gained by utilizing the DAZ SA in a 16kB SRAM. Section 9 provides a comparison to other offset compensation schemes. Section 10 provides measurements of the DAZ SA offset voltage from a 45 nm bulk CMOS test chip.

2. Mismatch Compensation using Auto- Zeroing Circuitry

Our auto-zeroing scheme uses a split-phase clock and charge pump feedback circuit. Figure 1a shows a conventional latch-based sense amp with PMOS inputs (e.g., to support near-V_{SS} sensing on a

low swing bus). Figure 1b shows the auto-zeroing circuit attached to the sense amp. The same scheme can apply to a SA with NMOS inputs in an SRAM. The charge pump circuit is shown in Figure 2. ENI and ENO are the input voltage differential and offset tuning phases respectively. ENR1 and ENR2 are reset phases. During ENR1, a zero differential input is applied to the sense amp. The ENO phase then occurs, and the SA resolves based on its intrinsic offset.

Figure 1. (**a**) Latch-based sense amp for near-V_{SS} inputs; (**b**) Auto-zeroing circuit attached to the sense amp [7].

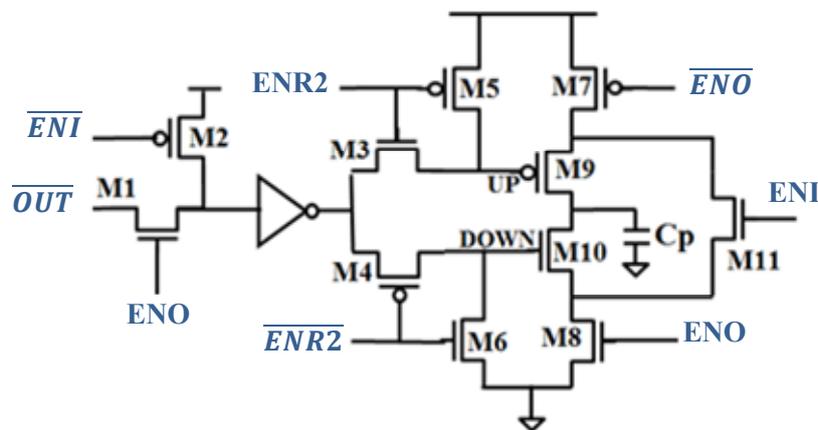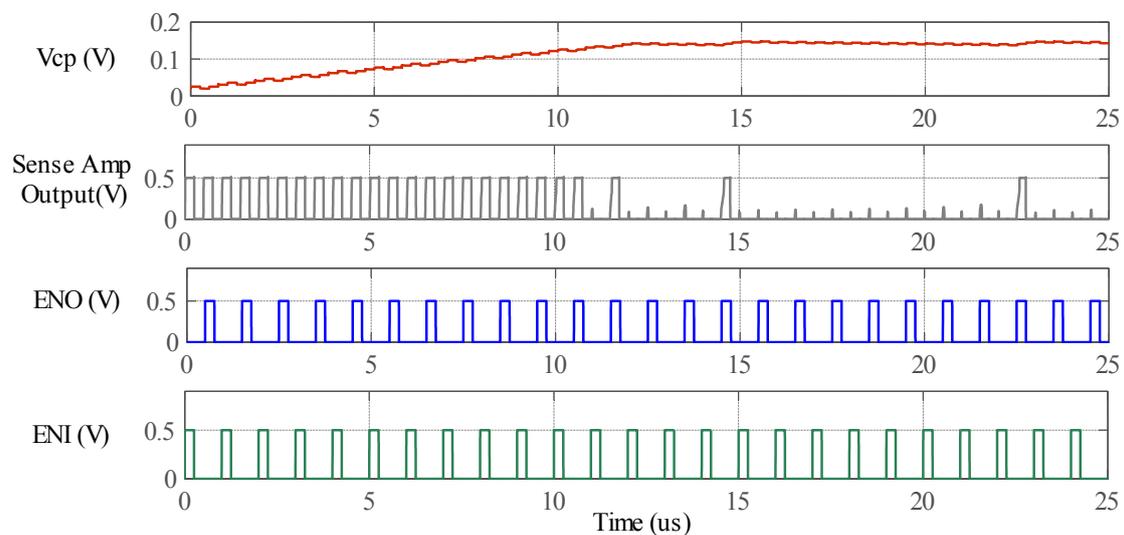(**a**)   (**b**)

Figure 2. Charge pump circuit for adjusting the voltage on Cp. [7].

The sense amp output is fed to the charge pump circuit that charges the capacitor, Cp, up or down. During ENR2, the differential input is applied to the sense amp. ENI then occurs, and the SA resolves based on the differential input. Note that phases ENR1 and ENO can be omitted or included based on how often re-calibration is needed. Transistors MC1 and MC2 control the drive strength of the right side of the sense amp to compensate for the offset.
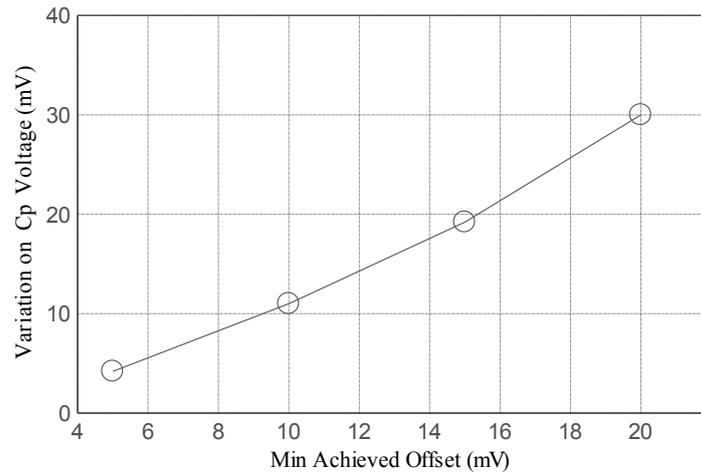
The charge pump controls the drive current in both transistors to equalize the strength of the SA right and left sides to reduce the offset. The offset is compensated with minimal capacitive loading at the output and is independent of input DC bias (V_{INDC}). A supply voltage and clock frequency of 0.5 V and 1MHz are used in the simulations. The output voltage of the sense amp and the voltage on Cp are illustrated in Figure 3 for an input differential of −10 mV. The initial voltage on Cp is zero. This causes an intrinsic positive offset voltage that set the SA output voltage to 1. Simulations indicate that the voltage on Cp required for a zero offset is 142 mV. For a 10 mV offset, the voltage on Cp can vary within ±12 mV.

Figure 3. Simulated output voltage of the sense amp and Cp voltage for a −10 mV differential input voltage at 0.5 V and 1MHz in 45 nm CMOS.

This imposes a minimum and maximum limit on Cp voltage to 130 mV and 154 mV in order to maintain an offset less than 10 mV. The deviation of Cp voltage from to the value corresponding to zero offset (142 mV in this case) is plotted in Figure 4 for desired final offset voltages of 5 mV, 10 mV, 15 mV, and 20 mV. Low offset voltages are usually realized using a higher value of Cp. In Figure 3, the offset compensation is completed when the voltage on Cp settles to its final value within the 130 mV to 154 mV range. The sense amp then resolves its output correctly to 0 during the input phase. A zero differential voltage is applied to the SA input during the offset phase. This sets the SA output to "1" when Cp voltage drops below 142mV and "0" otherwise. In this design, rate at which Cp charges up is higher than its charge down rate. This helps to minimize the power consumption as will be discussed in Section 5.
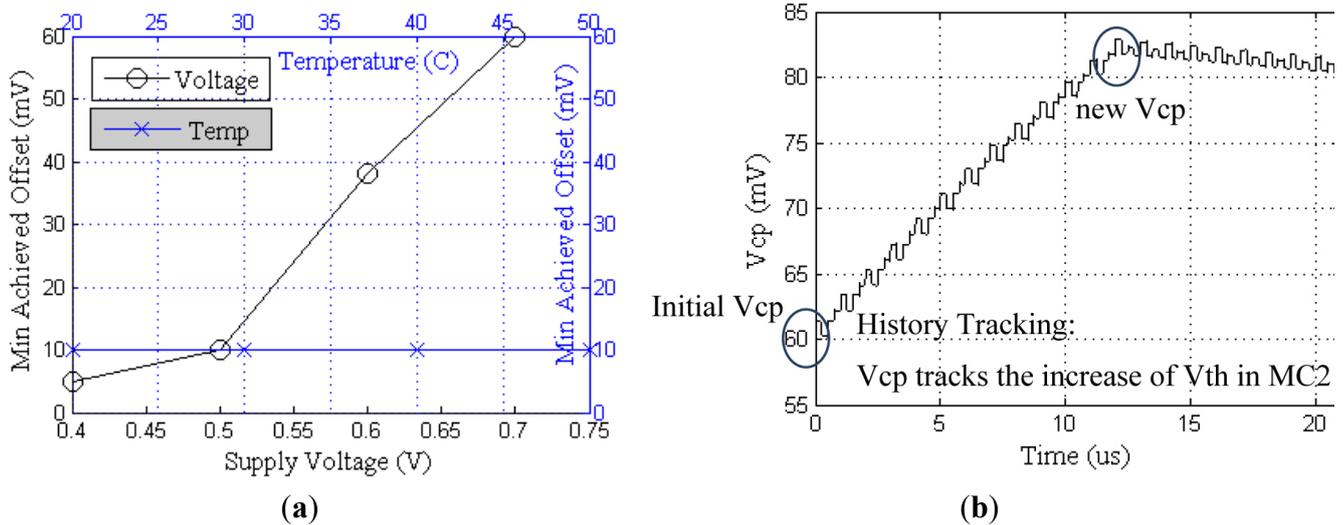
Figure 4. Variation on Cp voltage *vs.* Minimum Achieved Offset.

3. Voltage, Temperature, and Aging Tracking

To demonstrate temperature, voltage, and aging tracking, the offset voltage that remains after compensation is calculated for various voltages and temperatures as shown in Figure 5a. Simulations in a commercial 45 nm process show that the circuit maintains a constant offset across temperature. The accuracy of voltage tracking depends on the supply voltage. Higher supply voltage causes more charge to be pumped to Cp during each offset calibration cycle, and this larger change in charge leads to a coarser resolution, as Figure 5a illustrates.

Figure 5. (**a**) Voltage and temperature tracking; (**b**) Aging tracking [7].

(**a**)	(**b**)

The auto-zeroing scheme also has the ability to compensate for any changes in device characteristics after circuit deployment. One common cause for such changes is effective threshold voltage shifting due to Bias Temperature Instability (BTI), hot carrier injection, or other aging effects. To demonstrate how this circuit can compensate for such changes, Figure 5b shows the capacitor voltage after an abrupt increase in the threshold voltage of MC2, to emulate an aging effect. The

charge pump boosts the voltage on Cp to decrease the drive strength of MC2 in response and rapidly restores the compensated offset voltage.

4. Offset Tuning

We define settling time as the difference between the time when the zero differential-input is first applied and the time when the voltage of the output capacitor settles as shown in Figure 6. Changing the size of the output capacitor (Cp) affects the amount of charge added during the offset compensation phase (ENO) and so controls both the offset and the settling time. Figure 7 demonstrates the trade-off between accuracy (min achieved offset) and settling time using different values of output capacitors.

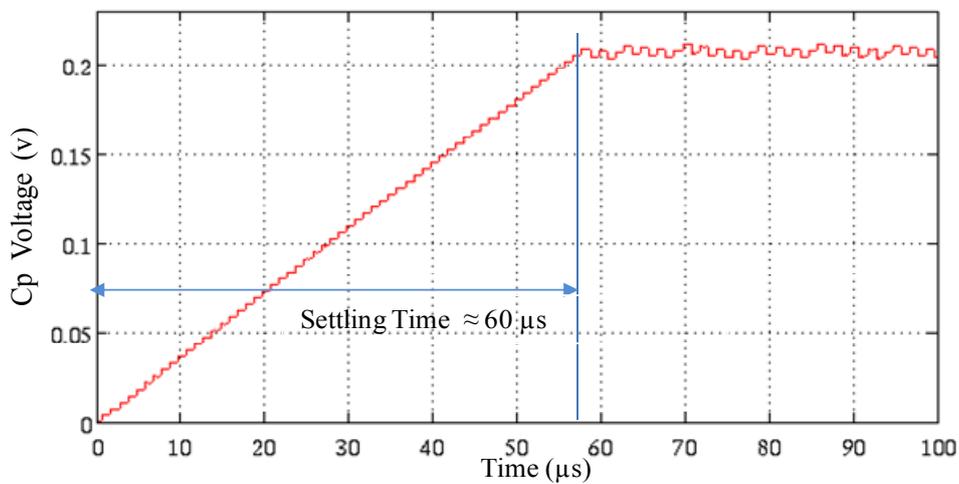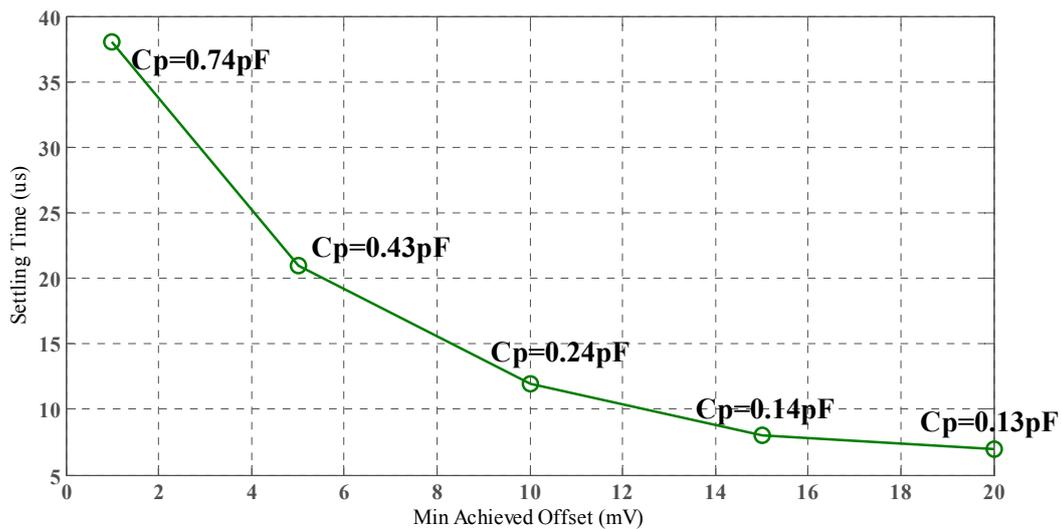**Figure 6.** Settling time simulation [7]

Figure 7. Min achieved offset *vs.* settling time for different values of output capacitor. [7]
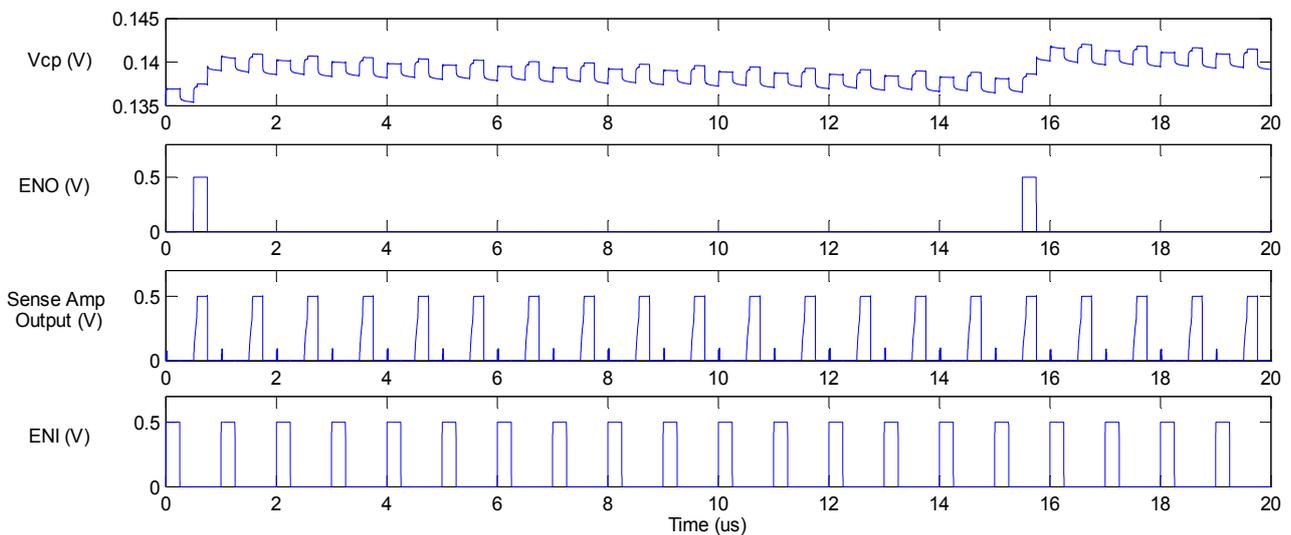

5. Power Consumption

The main contribution to the power consumed by the DAZ SA comes from the continuous calibration. Decreasing the number of cycles of calibration phase (ENI and ENO) relative to the input phase decreases the switching power of the feedback circuit and the power consumed in charging and

discharging (Cp) but is limited by the leakage at the output capacitor (Cp). The overhead area of the scheme includes the area of the timing circuit, the charge pump circuit, and the output capacitance (Cp). For an offset voltage of 1mV, a 0.74pF output capacitance is needed. In this case, the area of Cp can dominate the total area overhead. In Figure 8, the offset calibration phase occurs once every 15 clock cycles. The maximum calibration period or the minimum number of offset calibration cycles needed is limited by the leakage on Cp. Simulation results indicated a maximum calibration period of 200 µs. This high period makes the difference in power consumption between the DAZ SA and the Latch SA insignificant. The total power consumption of the DAZ SA and the Latch SA is 2.02 nW and 2 nW respectively. The minimum number of offset calibrations is independent of the required offset or the value of Cp, but it depends on the charge pump current. In this design the charging current is 0.5 µA. High current allows fewer number of calibration cycles. Increasing the charge pump current however increases the dynamic power consumption.

Figure 8. Offset compensation clock period for a 10 mV offset voltage.

Shorting the output virtual nodes of the charge pump through M11 can decrease the leakage by reducing V_{DS} of the switches and improve settling time as shown in Figure 9. The switching power can also be decreased by strengthening M9 in the charge pump circuit relative to M10 to avoid the continuous toggling of the sense amp output during offset compensation phase (ENO) after settling as shown in Figure 10. Strengthening M9 has the downside effect of increasing the settling time when Cp is moving to lower voltages; the time Cp takes to discharge will increase. However, the compensation usually starts with zero-initial voltage on the capacitor Cp that makes the settling time mostly dependent on the charging rate.

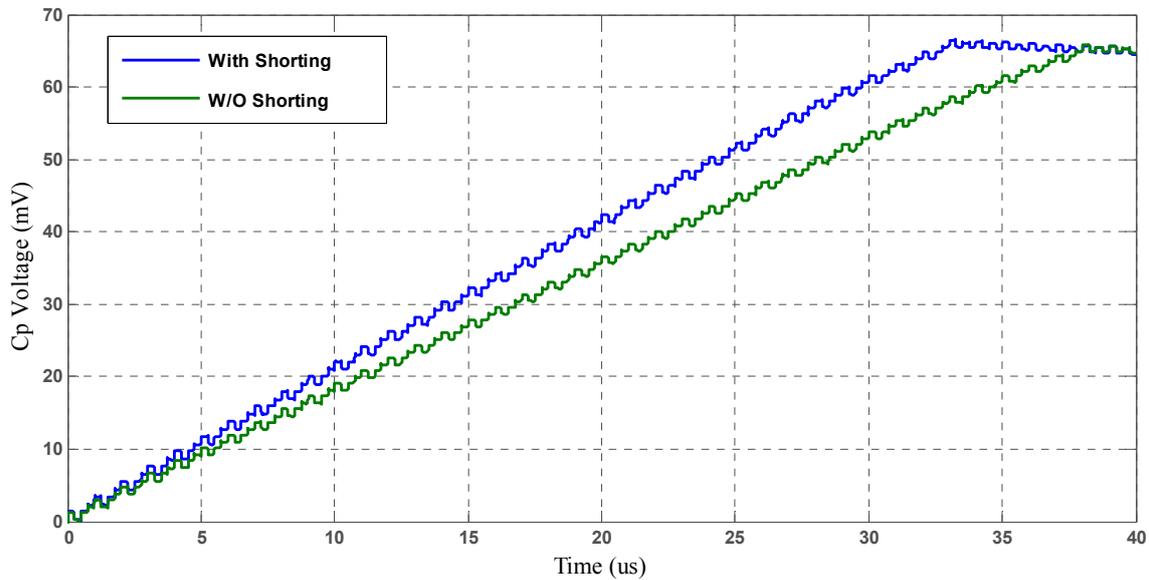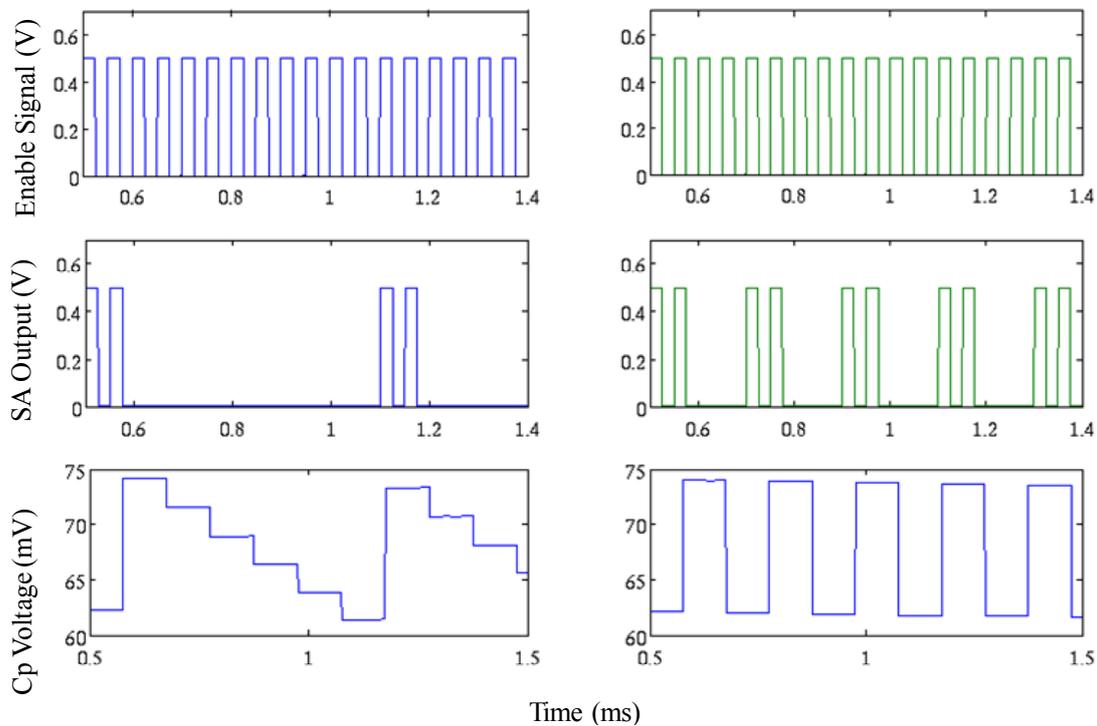**Figure 9.** Voltage on Cp with and without shorting the virtual supply nodes of the charge pump [7].

Figure 10. Strengthening the pull up transistor of the charge pump (left column of sims) reduces the rate at which the SA output switches high relative to equal strength devices (right column). This reduces power consumption [7]**.**
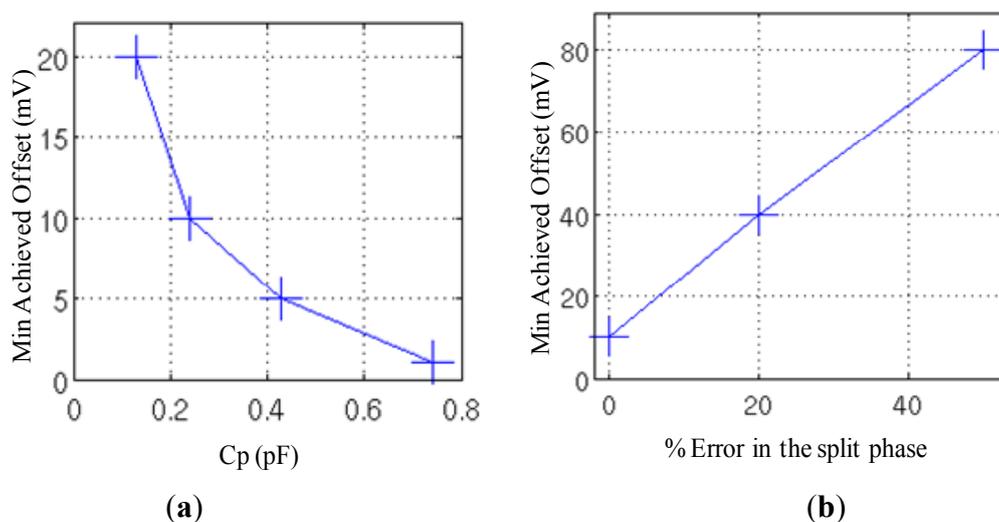

6. Offset Sensitivity

The sensitivity of the offset compensation depends on the split phases, charge pump circuit, and the output capacitance. The accuracy of the split phases has the dominant influence on the resolution. A small overlap between ENO and ENR2 phases can dramatically degrade the accuracy by

connecting M1 and M2 to the supply rails during charging. That leads to a significant increase in the charge pump rate degrading the accuracy as shown in Figure 11b, where the min achieved offset is plotted against the error in split phase timing, measured as the percentage of time overlap between ENO and ENR2. The scheme is also sensitive to variations in the M9 and M10 transistors in the charge pump circuit. They are responsible for charging/discharging Cp, and so the one with more drive strength determines the final offset value. Figure 11a shows the sensitivity of the offset voltage to the output capacitance Cp.
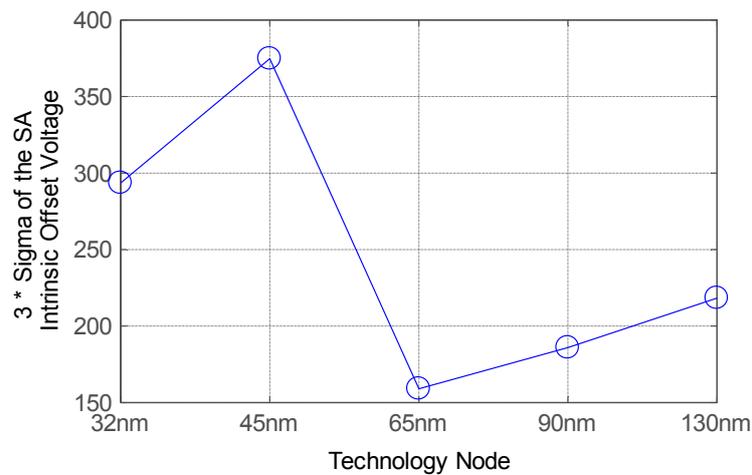
Figure 11. (**a**) Offset voltage sensitivity to output capacitance; (**b**) sensitivity to split phases [7]**.**

(**a**)   (**b**)

The offset voltage is also sensitive to the frequency of the split phase. The increase in the split phase frequency increases the enable signal switching and degrades the compensated offset voltage.

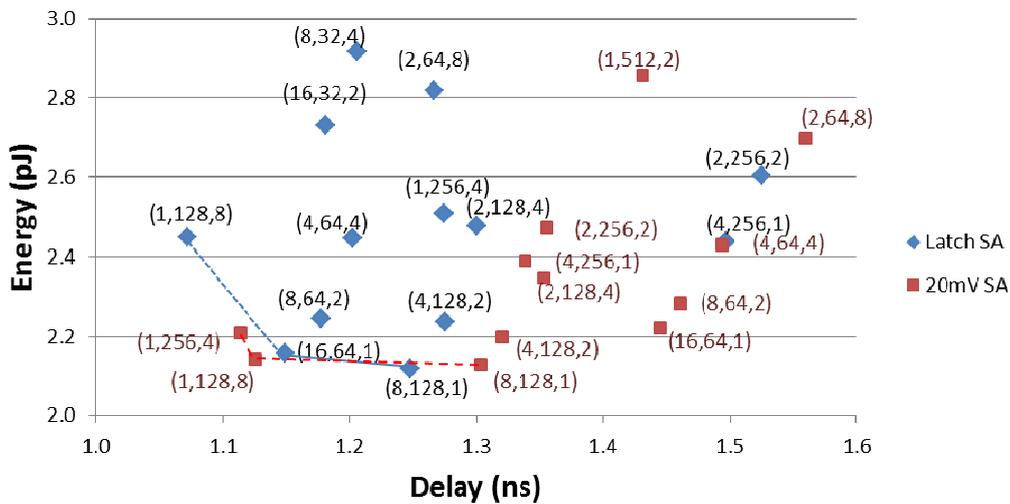## 7. Offset Compensation across Technology Nodes

The intrinsic offset voltage of the sense amplifier relies on the technology of the design. The effect of process variability on the final offset of the auto-zeroing circuit is marginal as explained in Section 1 due to the repeated compensation. To illustrate the variation of the SAs offset across technologies, 1000 Monte-Carlo simulations were performed to evaluate the offset voltage of a latch-based SA using 45, 65, 90 and 130 nm commercial technology models. Figure 12 illustrates the 3σ value of the SA offset voltage across technologies. The results indicate that the largest offset voltage belongs to 45 nm technology, followed by 32 nm, 65 nm, 90 nm and 130 nm respectively. Although the offset behavior is not monotonically increasing with technology scaling, the plot indicates a trend of increased offset in emerging technologies, *i.e.*, 32, 45 nm. Since offset voltage is increasing with newer technologies, compensation becomes more essential as technology scales. The next section demonstrates the benefits gained from using the DAZ SA in a 16kB SRAM. The total energy and delay of the SRAM is calculated and compared to the uncompensated SA case.

Figure 12. 3σ of the intrinsic voltage of the SA across technology nodes.

8. 16 kB SRAM Design

In this section, we investigate the effect of utilizing the DAZ SA in a 16kB SRAM memory. The power consumption of the DAZ SA is higher than the uncompensated one due to the clock generator, charge pump and the buffer stages needed for the non-overlapping clock. The sense amplifier delay is also higher due to the high capacitive loading. Reducing the sense amp offset reduces the necessary bit-line swing, which decreases both the precharge and bitcell energy and delay during the read operation. The reduction in the read energy and delay depends on the number of banks, rows and words per row of the memory. The energy and delay of the 16 kB SRAM is calculated using a 20 mV DAZ SA for all possible configurations and plotted in Figure 13. Each point is annotated with (B, R, W) where B is the number of banks, R is the number of rows, and W is the number of words per row. The results indicate a significant improvement in both the energy and delay for cases with large numbers of rows and small improvement or degradation for cases with small number of rows. The design point of 1 bank, 512 rows and 2 words per row has the biggest improvement of 10% in energy and 24% in delay. The design point of 4 banks, 32 rows and 8 words per row has the biggest degradation of 6% in energy and 5% in delay. The DAZ SA created 3 new optimal design points (1, 256, 4), (1, 128, 8) and (8, 128, 1). The improvement in energy and delay for (1, 256, 4) is 12% and 13% respectively. For (1, 128, 8) it improves the energy by 13% and degrades the delay by 5%. The energy of (1, 128, 8) with DAZ SA is the lowest. This could not be achieved using a Latch SA. The system level parameters of the SRAM that satisfy the energy/delay requirements changed (*i.e.*, design point (16, 64, 1) is the minimum energy/delay point with uncompensated SA. Using DAZ SA, the min delay/energy design changed to (1, 128, 8).

Figure 13. Design space of 16kB SRAM Memory with uncompensated and 20 mV digital auto-zeroing (DAZ) SA.

9. Comparison to other Offset Compensation Schemes

The main advantages of the scheme are the continuous calibration that makes it specifically useful for sub-threshold operation and the flexibility to tune the offset voltage. The latter provides different design options that can be utilized in the SRAM design process. Approaches like redundancy [3], transistor upsizing [4], and digitally controlled compensation [5] do not support continuous calibration and hence would not be tolerant to voltage and temperature variation. The approach in [6] provides continuous calibration. The power consumed in this approach is only for compensation clock phase generation. There is no charge pump or additional circuitry. However this approach requires the calibration phase to essentially occur before every sensing cycle. As explained in Section 5, the DAZ SA can perform compensation every N number of cycles. High charge pump current can be used to increase N at a cost of higher dynamic and leakage power of the charge pump circuit. The power consumption of [6] is compared to that of DAZ SA with the offset calibration phase occurring every cycle (cycle period = 1 μs) and every 200μs with a controllable offset phase. The controllable offset phase logic is employed to force calibration every cycle at the beginning. The logic then enables calibration every N cycles once the voltage on Cp settled to its final value. The results are shown in Table 1. The settling time of both schemes is compared. The DAZ SA with controllable offset phase consumes the lowest power with a 12 μs settling time.

Table 1. Power consumption of dynamic offset compensation and auto-zeroing circuit.

Offset Compensation Scheme	Power Consumption	Settling time
DAZ SA	6 nW	12 μs
Dynamic Compensation [6]	4 nW	0.5 μs
DAZ SA with controllable offset phase	2.5 nW	12 μs

10. 45 nm Test Chip Measurements

A test chip fabricated in 45 nm technology is used to verify the scheme. The chip contains one regular SA array for benchmarking and another array that uses SAs with the auto-zeroing circuitry, with Cp equal to 32fF. The chip micrograph is shown in Figure 14. Figure 15 shows the layout of a single DAZ sense amplifier. The layout of the output capacitor is shown on the left side of Figure 15, consuming an area of 2.97 μm × 3.9 μm. The sense amplifier and the charge pump layout are shown on the right side of Figure 15, consuming an area of 4.39 μm × 5.29 μm. The supply voltage is set to 0.6V during measurements to mitigate the effect of noise on the measured results. The control signals are supplied to the auto-zeroing circuit at 1 MHz. Figure 16 shows the measured offset distribution of both banks. The positive terminal of the SAs is connected to 0.45 V. The negative terminal of the SAs is swept from 0.3 V to 0.6 V in increments of 5 mV. The SAs are enabled during each increment, and measurements of the SAs outputs are recorded. This information is then used to construct the SAs offset distribution in Figure 16. The measured mean (μ) and standard deviation (σ) of the uncompensated SA banks is −31 mV and 45 mV respectively. The auto-zeroing circuitry reduced the value of μ to −13mV and lowered σ to 9.3 mV. This indicates an 80% improvement in σ. The scheme limits the absolute value of the maximum offset to 50 mV.
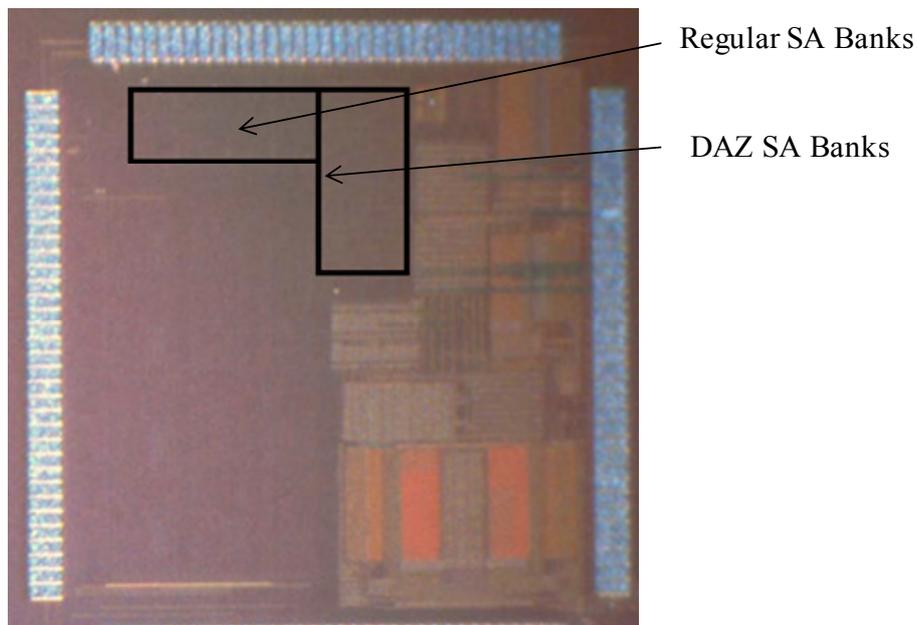
Figure 14. Chip Micrograph.
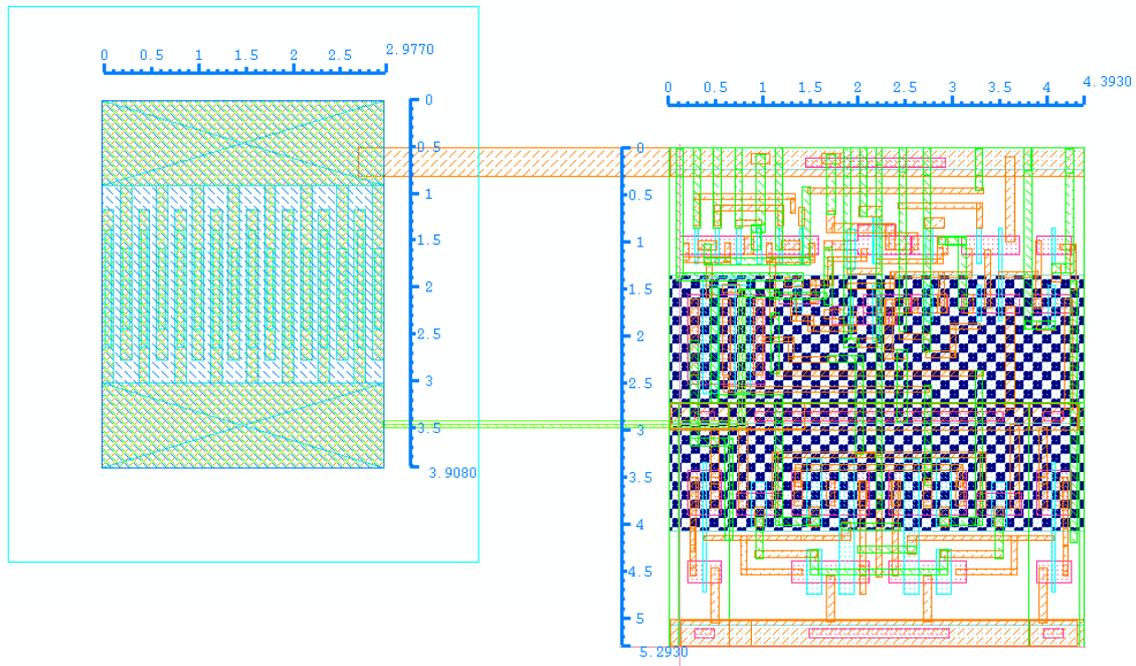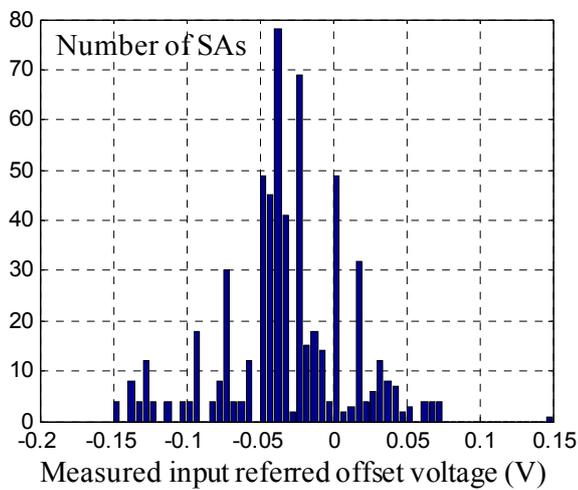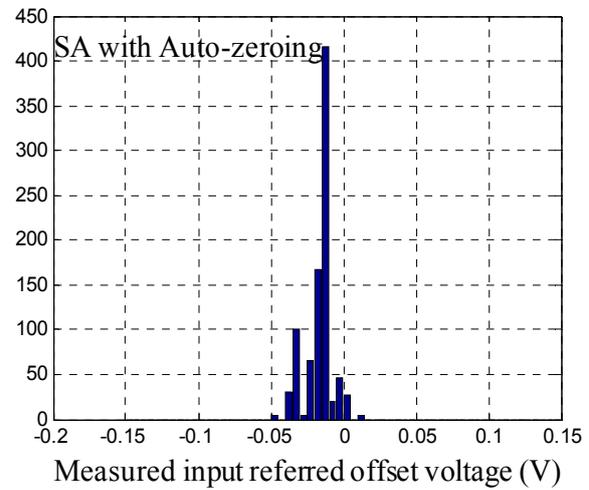
Figure 15. DAZ Sense Amplifier Layout.

Figure 16. Measured offset voltage distribution (**a**) regular SA; (**b**) SA with auto-zeroing circuitry [7].
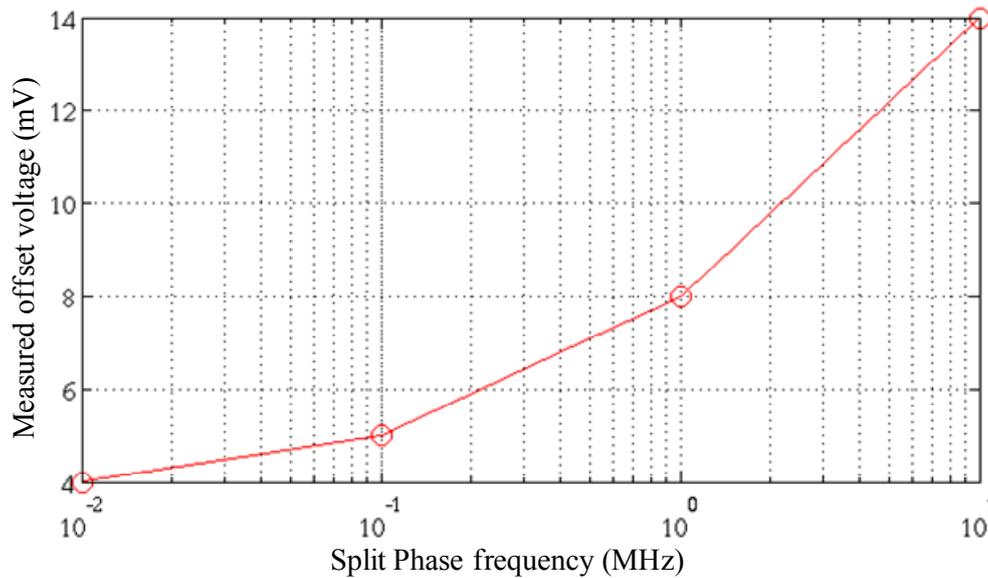

(**a**)

(**b**)

To verify the offset sensitivity to split phases, the offset of a sample DAZ SA is measured for different split phase frequencies. Figure 17 shows the offset voltage values for different split phase frequency.

Figure 17. Measured offset voltage *vs.* split phase frequency [7].

11. Conclusions

We proposed a circuit that is capable of improving sense-amp offset to near zero, which is valuable for sub-threshold operation due to the heightened effect of mismatch. Simulations of the design (0.5 V, 1 MHz) show a compensated offset voltage of 1mV, settling time of 37 μs, and total power consumption of 12 nW. Measurements from a test chip fabricated in 45 nm technology showed the circuit's ability to improve σ of the offset voltage by 80% and limited the absolute maximum value of the offset voltage to 50 mV using a 1 MHz split phase frequency and 32fF output capacitance. Using the circuit in a 16 kB SRAM showed a reduction in the total energy and delay of 10% and 15% respectively. The trade-off between the sense amp compensated offset and power consumption is demonstrated. This makes the circuit able to provide the offset/power values that can generate the optimal SRAM design.

References

1. Ryan, J.F.; Calhoun, B.H. Minimizing Offset for Latching Voltage-Mode Sense Amplifiers for Sub-Threshold Operation. In Proceedings of the 9th International Symposium on Quality Electronic Design, San Jose, California, USA, 17–19 March 2008; pp. 127–132.
2. Wang, A.; Chandrakasan, A.P.; Kosonocky, S.V. Optimal Supply and Threshold Scaling for Subthreshold CMOS Circuits. In Proceedings of the IEEE Computer Society Annual Symposium on VLSI, Pittsburgh, PA, USA, 25–26 April 2002; pp. 5–9.
3. Verma, N.; Chandrakasan, A.P. A 256 kb 65 nm 8T subthreshold SRAM employing sense-amplifier redundancy. *IEEE J. Solid State Circ.* 2008, *43*, 141–149.
4. Pileggi, L.; Keskin, G.; Li, X.; Mai, K.; Proesel, J. Mismatch Analysis and Statistical Design at 65 nm and Below. In Proceedings of the IEEE Custom Integrated Circuits Conference, San Jose, California, USA, 21–24 September 2008; pp. 9–12.

5. Bhargava, M.; McCartney, M.P.; Hoefler, A.; Mai, K. Low-overhead, Digital Offset Compensated, SRAM Sense Amplifiers. In Proceedings of the IEEE Custom Integrated Circuits Conference, San Jose, California, USA, 13–16 September 2009; pp. 705–708.

6. Sachdev, M.; Sharifkhani, M.; Shah, J.S.; Rennie, D. Sense-amplification with Offset Cancellation for Static Random Access Memories. U.S. Patent Application 12/757,033, 8 April 2010.

7. Beshay, P.; Calhoun, B.H.; Ryan, J.F. Sub-threshold Sense Amplifier Compensation Using Auto-zeroing Circuitry. In Proceedings of the 2012 IEEE Subthreshold Microelectronics Conference, Waltham, Massachusetts, USA, 9–10 October 2012; pp. 1–3.