A Single-Ended Offset-Canceling Sense Amplifier Enabling Wide-Voltage Operations

Shan Shen, *Member, IEEE*, Hao Xu, Yongliang Zhou, and Wenjian Yu, *Senior Member, IEEE*

Abstract—The input-referred offset voltage (V_{OS}) of the sense amplifier (SA) often dictates the minimum required differential voltage swing on bit-lines and is an important indicator that determines the access delay and the energy dissipation of low-power SRAMs. This brief presents a single-ended offset-canceling sense amplifier (SOSA), where a capacitor stores the offset voltage and another capacitor couples the bit-line voltage swing to the internal sensing node. SOSA can be operated in a wide range of V_{DD} and is more compatible with low-power 8T SRAMs with large array configurations. According to the simulation results under TSMC 28nm CMOS technology, the average standard deviation of V_{OS} (σ_{OS}) and the average sensing delay of SOSA across 0.2V to 0.9V V_{DD} are only 5.83% and 15% of the baseline respectively. And compared to the state-of-the-art, it can reduce σ_{OS} by 2.2X.

Index Terms—Sense amplifier, SRAM, wide voltage.

I. INTRODUCTION

SRAMS with a wide range of supply voltages are demanded to achieve high performance during normal operation modes while minimizing power consumption during low voltage modes [1]. In the sub/near-threshold region, the current flowing through transistors is extremely sensitive to process, voltage, and temperature (PVT) variations, which will have a significant effect on circuit performance in different process corners. This could incur speed degradation by two orders of magnitude, compared with the operation at nominal supply voltages [2].

A sense amplifier (SA) circuitry is an important component of SRAM that is responsible for reading and amplifying the data stored in a cell. Its characteristics influence several important performance metrics of SRAM, such as read frequency, minimum operating voltage, and power consumption [3]. This is because an SA has an input-referred offset voltage (V_{OS}) and requires a minimum differential input signal (ΔV_{BL_min}) larger than V_{OS} to make a reliable decision [4], [5]. A larger

Shan Shen and Wenjian Yu are with the Department of Computer Science and Technology, BNRist, Tsinghua University, Beijing 100084, China (e-mail: shanshen@tsinghua.edu.cn; yu-wj@tsinghua.edu.cn).

Hao Xu is with the Nation ASIC System Engineering Technology Research Center, Southeast University, Nanjing 210096, China (e-mail: 220205865@seu.edu.cn).

Yongliang Zhou is with the School of Integrated Circuits, Anhui University, Hefei 230601, China (e-mail: zhouyongliang@ahu.edu.cn).

standard deviation of offset voltage distributions (σ_{OS}) has a huge negative effect on read speed and read yield (Y_{READ}) reported by Abu-Rahma et al. [6]. According to their analysis, every 1 mV increase in σ_{OS} of the SA forces a 10 mV increase in the ΔV_{BL_min} for a 28 nm-technology 16 Mb SRAM with $Y_{READ} = 97\%$. Increased ΔV_{BL_min} results in more energy consumption per access and more time to develop a larger differential voltage on the bitlines with large loads.

However, the lower bound of ΔV_{BL_min} is the SA's V_{OS}, which is arising from several mismatches of circuits including the gain factor, the drain current, the threshold voltage (V_{TH}), and the layout of the devices [7], [8]. Among these, Mohammad et al. [9] found that the V_{TH} mismatch is the most dominant contributing factor to V_{OS}. The voltage latch SA (VLSA) and the current latch SA (CLSA) are two popular topologies. The VLSA schemes usually show lower area occupation at the cost of higher power consumption. It is shown that VLSA is more robust and less sensitive to process and temperature variation than CLSA [9]. With the same area budget, CLSA exhibits ~3X wider σ_{OS} compared to VLSA [6]. Research works [10] and [11] summarized that the V_{TH} mismatch between the NMOS sensing pair dominates the V_{OS} of VLSA and CLSA, respectively. Unfortunately, aggressive device scaling of cutting-edge technology further increases device variations and contributes to larger V_{OS} in SAs [12], [13].

This brief aims to investigate and compare the topologies of offset-tolerant low-voltage SAs that can enable SRAMs to support wide-voltage operations. However, in the near/subthreshold region, conventional 6T SRAMs fail to deliver the yield requirements due to the reduced read static noise margin, poor writability, etc. The decoupled 8T cell (Fig. 4) is the most popular solution due to its reasonable area overhead where the read and write ports are separated to deliver good read stability.

Therefore, in this brief, we propose a single-ended offset-canceling SA (SOSA) that is more compatible with 8T SRAMs with large array configurations and can be operated from the nominal supply down to the subthreshold supply range. The basic principle of SOSA, derived from the basic VLSA, is to consecutively amplify a small signal twice by the 2 inverters with the maximum gain. The average σ_{OS} of SOSA is only 5.83% of the baseline CLSA [14] with 15% sensing delay. Compared to the state-of-the-art [15], [16], [17], [18], [19], the maximum and minimum σ_{OS} improvements of SOSA are 9.9X and 2.2X. In addition, it can also be integrated into the computing-in-memory structure as a high-resolution AD converter.

II. BACKGROUND & RELATED WORKS

Conventional VLSA [10] is depicted in Fig. 1 (a). Due to V_{TH} variations, the trip points (V_{TRIP}) within the

Fig. 1. (a) Conventional VLSA. (b) Correct and (c) failure read due to the V_{TRIP} mismatch of VLSA.

Fig. 2. (a) Overall circuits of SOSA. (b) Precharge phase, (c) voltage sampling phase, and (d) amplifying phase in a read operation.

cross-coupled inverters (INV1 and INV2) are mismatched and further result in the offset voltage of a SA. Fig. 1 (c) shows a transient waveform of a read failure ($V_{TRIP1} > V_{TRIP2}$). The failure occurs when the difference of V_{TRIP} triggers the regeneration of V_{INR} earlier than that of V_{INL}. This situation would be frequent with PVT variations under nanoscale technology. Although the offset in V_{TRIP} can be reduced by increasing transistor sizes [13], the corresponding drawbacks are obvious such as increased area and power consumption. An additional shortcoming of this design strategy is that the increased capacitance, due to the use of larger devices, slows the latching speed of the VLSA.

There are several advanced offset compensation schemes proposed recently. The variation-tolerant small-signal differential sensing SA (VTSSA) was presented by Giridhar et al. [15] with a capacitive offset compensation strategy where the SA inverter pair were reconfigured as amplifiers to the pre-amplification of bitline differential and then resembled to its conventional cross-coupled latch-type structure. Similarly, Fragasse et al. [16] proposed an SA topology implemented in a current-mode latch-type SA (CSA) with capacitive-offset correction. The operation of this SA relies on storing V_{TRIP} across an offset correction capacitor, then biasing the cross-coupled latch at each inverter's respective V_{TRIP} just before sensing. This allows offset-free sensing to effectively occur when SAE is asserted. However, the above two designs do not perform well with a small V_{OS}-storing capacitor. The bit-line voltage sense amplifier (VSA) presented by Licciardo et al. [17] combines offset compensation and cancellation features in four phases, in which two amplifiers are pre-charged in the high voltage gain region in order to reduce the effects of the offset. The drawback of VSA is the requirement of 2 large offset compensation inverters. Patel et al. [18] proposed a hybrid latch-type sense amplifier (HYSA) that integrated both current and voltage latch-type SA features by applying the differential input signal to multiple sensing nodes to significantly reduce ΔV_{BL_min}. The transistor sizes of HYSA need to be fine-tuned to get a balanced performance. A novel reconfigurable sensing amplifier (RSA) for computing SRAM presented by Chen et al. [19] can dynamically reconfigure its circuits in a compute access or in normal read access, to optimize its access delay in both modes. However, its σ_{OS} is still large in the normal mode due to the simple structure.

III. SINGLE-ENDED OFFSET-CANCELING SENSE AMPLIFIER

A. Principle of SOSA

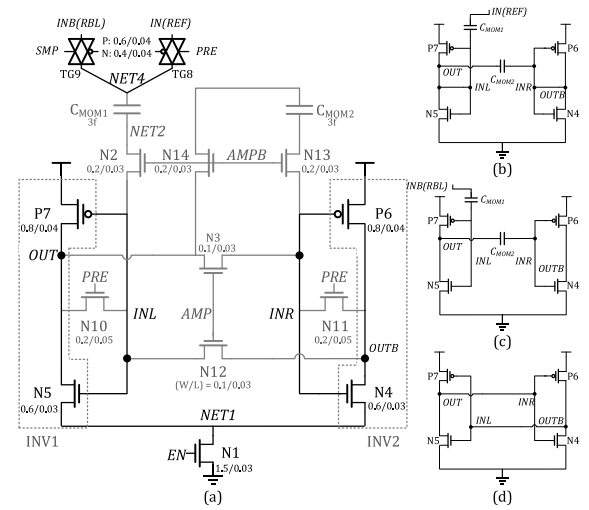The overall circuitry of SOSA is shown in Fig. 2 (a). SOSA is constructed from a basic voltage latch-type SA structure (including INV1 and INV2). In addition, it is equipped with 2 capacitors, C_{MOM1}, to couple the input voltage swing into the internal sensing nodes, and C_{MOM2}, to store the V_{TRIP} mismatch between the 2 inverters. Different from other symmetric SAs, the basic principle of SOSA is transmitting a small signal from the left to the right inverter sequentially. Thereby, it is very sensitive to the input voltage change and has a very small V_{OS} across a wide voltage range. The operation of SOSA is comprised of 3 phases: precharge, voltage sampling, and amplifying. In the precharge phase [Fig. 2 (b)], TG8 connects NET4 to a voltage reference, N2, N10, N11, N13, and N14 are switched on by PRE and AMPB. This makes $V_{INL} = V_{TRIP1}$, and $V_{INR} = V_{TRIP2}$. The voltage difference between IN and INL is stored by C_{MOM1}, meanwhile, the difference between INL/OUT and INR is stored by C_{MOM2}. In the voltage sampling phase [Fig. 2 (c)], TG8 is off, TG9 is turned on by SMP, N2, N13, and N14 remain in their active states. The voltage difference between RBL and the reference is passed through C_{MOM1} to INV1, then to INV2 through C_{MOM2}. In the amplifying phase [Fig. 2 (d)], only N3 and N12 are activated by AMP to configure SOSA to latch mode.

Due to the precharge operation, the 2 inverters are all in the metastable state with the largest gain factors (defined as $|\frac{dV_{OUT}}{dV_{IN}}|$) so that any small stimulation at the node INL and INR can be amplified quickly. This feature is leveraged by the voltage sampling phase. Assuming the SA reads a "1", the voltage increment Δv at NET4 is coupling to INL after the input node switches to INB from IN. This increment is first amplified by INV1 with its maximum gain factor a_1 ($a_1 > 1$) making V_{OUT} decrease to

$$V_{OUT} = V_{TRIP1} - a_1 \Delta v \qquad (1)$$

The coupling capacitor C_{MOM2} brings the abrupt change to the input of INV2

$$V_{INR} = V_{TRIP2} - a_1 \Delta v \qquad (2)$$

Then the INV2 again amplifies the voltage difference with the gain factor a_2 ($a_2 > 1$) and the voltage at OUTB increases to
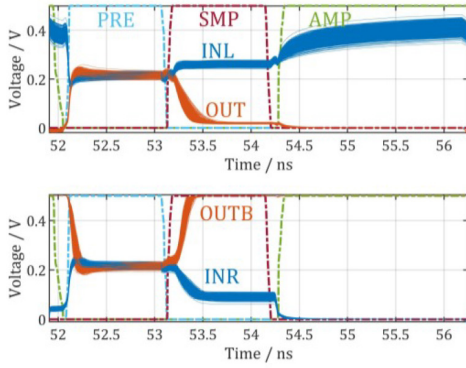
$$V_{OUTB} = V_{TRIP2} + a_1 a_2 \Delta v \qquad (3)$$

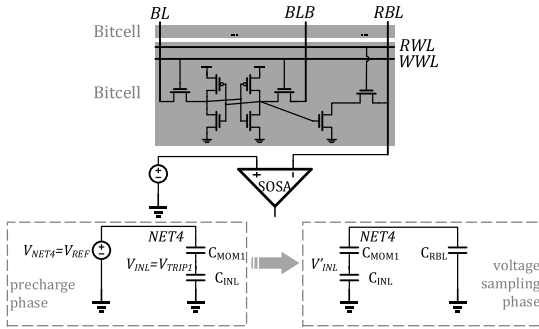Fig. 3. Timing diagram for SOSA, assuming $V_{IN} < V_{INB}$.

Fig. 4. Charge-sharing process in the voltage sampling phase.

Note that the maximum gain factors a_1 and a_2 of inverters should be larger than 1 in a balanced design. Hence, this twice-amplified voltage difference $a_1 a_2 \Delta v$ can greatly reduce the setup time of the SOSA for establishing the positive feedback path through N3 and N12 in the amplifying phase. The footer N1 is always turned on by EN during the entire operation.

The prime advantage of SOSA is providing a very small V_{OS} across a wide voltage range. This is accomplished by the V_{OS}-storing and the input-coupling capacitors that make SOSA very sensitive to the input voltage change. Another advantage is that the delay of the precharge and sampling phases can be overlapped with the RBL precharge/discharge time when SOSA is integrated into 8T SRAM. The timing diagram collected from 1K Monte Carlo (MC) sweeps is shown in Fig. 3.

B. Design Tradeoffs

There are some design trade-offs in SOSA (all sizes of devices are listed in Fig. 2 (a)). Firstly, the switches should be in the minimum size to reduce the internal node capacitance. All other transistors are ultra-low threshold devices to enhance the sub-threshold operations of SOSA. The gate lengths of N10 and N11 are longer to reduce leakage. Secondly, the key transistors, N1, N4, N5, P6, and P7, determine the overall sensing speed and power dissipation. Therefore, we carefully balance their sizes in this design. Thirdly, the capacitance constraints of C_{MOM1} and C_{MOM2} need to be analyzed to make SOSA more robust. The transition from the precharge phase to the sampling phase is a charge-sharing process, shown in Fig. 4, with the initial state described as (assuming SA reads a "1")

$$Q = (V_{REF} - V_{TRIP1})C_{MOM1} + V_{TRIP1}C_{INL} + V_{RBL}C_{RBL} \quad (4)$$

where Q is the total charges comprised of the charges stored by C_{MOM1} (the first term), C_{INL} (the second term) in the precharge phase, and the capacitive RBL (the last term). After charge sharing, the voltage at NET4 becomes V'_{NET4} and the increment is

$$\Delta V_{NET4} = V'_{NET4} - V_{NET4} = \frac{Q}{C_{RBL} + C_{EQ}} - V_{REF}$$

$$= \frac{C_{RBL}}{C_{RBL} + C_{EQ}}V_{RBL} - V_{REF} + r,$$

$$r = \frac{C_{INL} - C_{MOM1}}{C_{RBL} + C_{EQ}}V_{TRIP1} + \frac{C_{MOM1}}{C_{RBL} + C_{EQ}}V_{REF} \quad (5)$$

where C_{EQ} is the equivalent capacitance of C_{INL} and C_{MOM1} connected in series

$$C_{EQ} = \frac{C_{INL}C_{MOM1}}{C_{INL} + C_{MOM1}}. \quad (6)$$

According to (5), to get a full voltage swing at NET4, the condition, $C_{EQ} << C_{RBL}$, should be satisfied and C_{INL} should be as small as possible to make $r \approx 0$. This is naturally accomplished since the C_{INL} is around 1.1fF and C_{RBL} is usually larger than 25fF for a 256-bit depth SRAM column. Thus, SOSA is more compatible with large SRAM arrays. The voltage at INL becomes V'_{INL} in the sampling phase, and the increment equals to

$$\Delta V_{INL} = V'_{INL} - V_{INL} = \frac{C_{MOM1}}{C_{INL} + C_{MOM1}}V'_{NET4} - V_{TRIP1} \quad (7)$$

which means that C_{MOM1} should be several times larger than C_{INL} to get a maximum voltage increase at the input of INV1. A similar analysis can be reproduced for C_{MOM2}. Fortunately, the performance (σ_{OS}) of SOSA is not significantly affected by C_{MOM1} and C_{MOM2} when the capacitance is larger than 1fF. This shows another advantage of SOSA compared to the designs with V_{OS}-storing capacitors, such as [15], [16]. Considering the area budget and pitch of SRAM columns, we use two 3fF metal-oxide-metal (MOM) capacitors stacked on transistors (Fig. 11).

IV. EXPERIMENTAL RESULTS

All aforementioned designs are validated in the commercial TSMC 28nm technology. All performance metrics are collected from 1K MC sweeps at the typical-NMOS typical-PMOS corner (TTG) 25°C with local variation turned on. An implementation of an accurate sub-threshold reference is really challenging and out of the scope of this brief, so we leave it to our future work and use ideal voltage sources connecting to the SA inputs in simulations. To mimic the variation of the control signal, we collect the mean of delay and rise/fall time of an inverter chain with the length of 50 via MC sweeps at different PVT conditions. These values are used to configure the slew and assertion time of control signals. Besides, the signal overlapping and mismatch between the pulse time and the read time of SRAM cell can be calibrated by the digitized timing with replica-bitline tracking circuits [14], [16]. As a result, we do not consider these non-ideal factors here.

A. Results of 8T SRAM With SOSA Integrated

For simplicity, SOSA is integrated into a 256-bit depth 8T SRAM column to demonstrate the performance improvement of SRAM. The bitcells are 8T dual-port provided by the foundry (Fig. 4). In our simulations, a read operation is
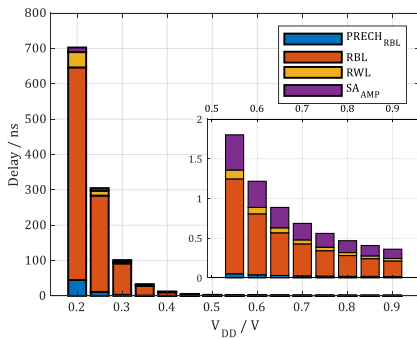
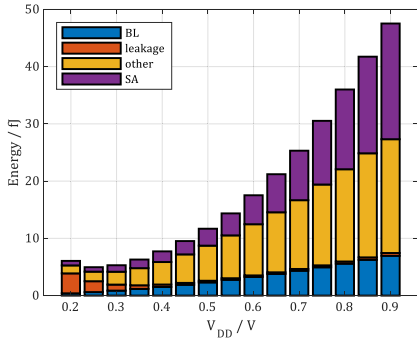Fig. 5. Delay breakdowns of SRAM read at different supply voltages.

Fig. 6. Energy decomposition of SRAM read at different supply voltages.
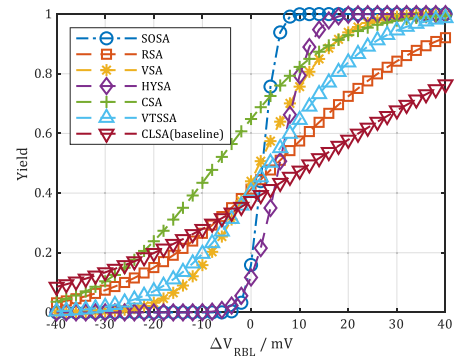

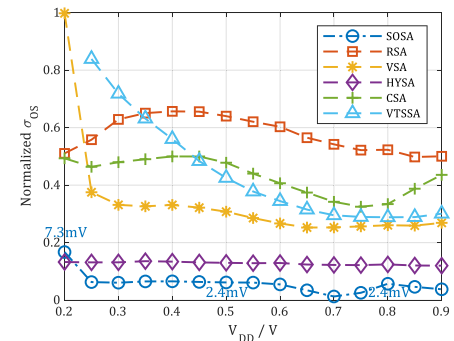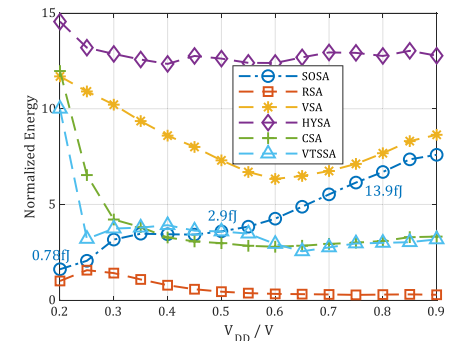Fig. 7. Yield changes with different input voltages for different SA designs.

Fig. 8. Normalized standard deviation of V_{OS} at different supply voltages.

Fig. 9. Normalized energy consumption of SAs at different supply voltages.

comprised of RBL precharging from $V_{DD}/2$ to V_{DD}, RWL assertion, RBL discharging to $V_{DD}/2$, and SA enabling. The delay and energy consumption of decoding and buffering are not considered. SA delay is defined as the time duration from AMP signal assertion to the slowest internal node (OUT) reaching 90% of its steady-state voltage. Fig. 5 depicts delay breakdowns of 8T SRAM. The total delay increases exponentially as the V_{DD} scales down. The RBL discharge time dominates the overall read delay. Compared to the SRAM delay, the delay of SOSA can be ignored, especially at low supply voltages ($V_{DD} \leq 0.5V$).

Fig. 6 also shows the energy decomposition of an 8T SRAM read. SOSA has similar energy consumption with RBL discharge of SRAM when $V_{DD} \leq 0.6V$. However, energy grows fast as the supply voltage goes to the nominal value. This is caused by the meta-stable state of the 2 inverters forming 2 DC paths in the precharge phase (Fig. 2(b)). It can be alleviated by a fine-grain timing control or by using column multiplexers to reduce the total number of SAs in an SRAM array.

B. Comparisons With Other SAs

SOSA is compared with several state-of-the-art SA designs with offset compensation strategies. To ensure the correct functionality under a wide range of V_{DD}, we use *ulvt* devices for all schemes. The baseline is a CLSA from [14]. Fig. 7 plots the yield changes as the input voltage increases. The proposed scheme has the steepest yield curve, which indicates a small RBL swing is sufficient to make SOSA sense correctly. Fig. 8 shows the normalized standard deviation of V_{OS} (σ_{OS}) w.r.t. that of the baseline. SOSA has the smallest σ_{OS}, only 5.83% of the baseline on average. It improves σ_{OS} by 9.9X and 2.2X compared to RSA [19] and HYSA [18] respectively. The performance of VTSSA [15] and VSA [17] deteriorates with the reduced supply voltage. To take the input variations

caused by memory cells into account, we re-run the MC simulations where an SRAM column is connected to the input of SOSA (an ideal voltage reference is still used), and a slow and steady voltage drop on RBL is contributed by the leakage current of that column. The results show that the input voltage variations make σ_{OS} of SOSA 1.24X larger than that in the ideal situation.

Fig. 9 shows the normalized energy consumption of all SAs. RSA and HYSA achieve the minimum and maximum energy consumption respectively. The average energy of SOSA is 4.4X of the baseline. Note that SOSA consumes more energy as V_{DD} increases. This is mainly caused by the short-cut current in the precharge phase. Fig. 10 depicts the normalized sensing delay of SAs. The time delay of precharge and voltage sampling phases can be overlapped with RBL precharge or discharge, thus, we only compare the delay of amplifying phase in this section (it is the total delay for RSA and HYSA). All SAs aggressively improve the sensing delay compared to the baseline in the range of $V_{DD} > 0.3V$. The slight performance improvement of the baseline as V_{DD} increases to the nominal
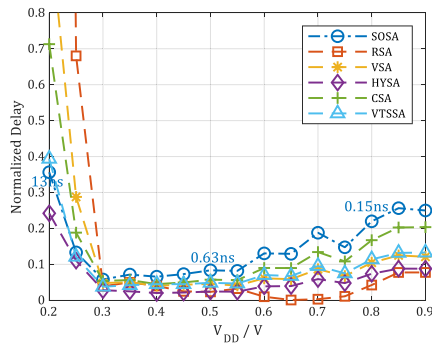
Fig. 10. Normalized delay of SAs at different supply voltages.

TABLE I
COMPARISON WITH OTHER SAs

	CLSA [14]	VTS-SA[15]	CSA [16]	HYSA [18]	VSA [17]	RSA [19]	ACSA [20]	**SOSA**
Tech./ nm	28	28	130	65	180	55	45	**28**
Type*	C	V	C	V	V	V	C	**V**
Proj. Area+	0.60	0.70	0.94	2.4	2.37	0.60	1.68	**1.39**
# of Devices	9T	15T+2 MOM	18T+1 MOM	15T	15T	9T	6T+Mo sCap	**16T+2 MOM**
Sram Type	6T	6T	6T	6T	6T	6T	8T	**8T**

* C for the current-latch SA, V for the voltage-latch SA.
+ Projected area is calculated from the designs under the same technology where the device sizes are tuned manually according to the corresponding published works.
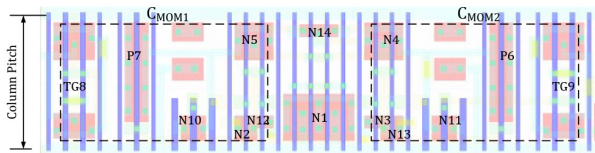

Fig. 11. Layout of SOSA.

TABLE II
POST-LAYOUT SIMULATION RESULTS OF SOSA

V_{DD} /V	SS			TT			FF		
	σ_{OS} /mV	t_{AMP} /ns	E_{TOT} /fJ	σ_{OS} /mV	t_{AMP} /ns	E_{TOT} /fJ	σ_{OS} /mV	t_{AMP} /ns	E_{TOT} /fJ
0.2	11.4	31.2	3.1	10.1	28.7	2.37	10.0	26.5	2.09
0.5	5.4	1.0	4.72	5.9	0.82	5.76	7.1	0.71	6.13
0.8	6.1	0.2	29.1	6.0	0.19	27.5	5.8	0.17	26.2

value makes the normalized delay curves rise. Averagely, SOSA has 15% sensing delay of the baseline. When the supply enters the sub-threshold regime, sensing speeds of RSA, VSA, as well as CSA [16] slow down dramatically. Table I further compares the projected area/SA type/SRAM type with several designs. Although [15], [16], [17], [18], [19] are based on 6T-SRAM, they can still work in 8T with additional voltage references, the same as [21].

C. Layout and Post-Layout Simulations

Fig. 11 shows the layout of SOSA. The entire layout is symmetric about N1 and N14, the 2 MOM capacitors are stacked on the transistors to save area ($4.9\mu m^2$) while satisfying the SRAM column pitch. The detailed performance metrics of SOSA collected from post-layout simulations are listed in Table II, where σ_{OS} is always below 11.4mV. The delay of SOSA is maximum at SSG due to the slowest PMOS and NMOS. The lengthened operation time further makes the energy at SSG larger. Oppositely, SOSA has the best delay/energy at FFG. We found that the precharge phase consumes nearly half of the total energy of SOSA in the simulation. Fortunately, it can be alleviated by using a better timing controller or multiplexers.

V. CONCLUSION

The offset voltage of SA determines the access delay and the energy dissipation of SRAMs. This brief presents SOSA that achieves the smallest V_{OS} compared to other SA structures. It can be leveraged by low-power 8T SRAM to enable wide-voltage operations while satisfying a stringent yield constraint.

REFERENCES

[1] L. Chang et al., "A 5.3 GHz 8T-SRAM with operation down to 0.41 V in 65 nm CMOS," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2007, pp. 252–253.

[2] H. Reyserhove and W. Dehaene, "A differential transmission gate design flow for minimum energy sub-10-pJ/cycle ARM cortex-M0 MCUs," *IEEE J. Solid-State Circuits*, vol. 52, no. 7, pp. 1904–1914, Apr. 2017.

[3] K. Zhang, K. Hose, V. De, and B. Senyk, "The scaling of data sensing schemes for high speed cache design in sub-0.18 μ/m technologies," in *Proc. IEEE Symp. VLSI Circuits*, Honolulu, HI, USA, Jun. 2000, pp. 226–227.

[4] S. Shen, L. Pang, T. Shao, M. Ling, X. Shi, and L. Shi, "TYMER: A yield-based performance model for timing-speculation SRAM," in *Proc. 57th DAC*, San Francisco, CA, USA, Jul. 2020, pp. 1–6.

[5] S. Shen, P. Cao, M. Ling, and L. Shi, "A timing yield model for SRAM cells in sub/near-threshold voltages based on a compact drain current model," 2022, *arXiv:2202.11941*.

[6] M. H. Abu-Rahma et al., "Characterization of SRAM sense amplifier input offset for yield prediction in 28 nm CMOS," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, San Jose, CA, USA, Sep. 2011, pp. 1–4.

[7] S. J. Lovett, G. A. Gibbs, and A. Pancholy, "Yield and matching implications for static RAM memory array sense-amplifier design," *IEEE J. Solid-State Circuits*, vol. 35, no. 8, pp. 1200–1204, Aug. 2000.

[8] D. G. Laurent, "Sense amplifier signal margins and process sensitivities [DRAM]," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 3, pp. 269–275, Mar. 2002.

[9] B. Mohammad, P. Dadabhoy, K. Lin, and P. Bassett, "Comparative study of current mode and voltage mode sense amplifier used for 28 nm SRAM," in *Proc. 24th Int. Conf. Microelectron. (ICM)*, Algiers, Algeria, Dec. 2012, pp. 1–6.

[10] L. Pileggi, G. Keskin, X. Li, K. Mai, and J. Proesel, "Mismatch analysis and statistical design at 65 nm and below," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, San Jose, CA, USA, Sep. 2008, pp. 9–12.

[11] J. S. Shah, D. Nairn, and M. Sachdev, "An energy-efficient offset-cancelling sense amplifier," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 60, no. 8, pp. 477–481, Aug. 2013.

[12] J. F. Ryan and B. H. Calhoun, "Minimizing offset for latching voltage-mode sense amplifiers for sub-threshold operation," in *Proc. 9th Int. Symp. Quality Electron. Design (ISQED)*, San Jose, CA, USA, Mar. 2008, pp. 127–132.

[13] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989.

[14] S. Shen et al., "TS cache: A fast cache with timing-speculation mechanism under low supply voltages," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 28, no. 1, pp. 252–262, Jan. 2020.

[15] B. Giridhar, N. Pinckney, D. Sylvester, and D. Blaauw, "13.7 A reconfigurable sense amplifier with auto-zero calibration and pre-amplification in 28nm CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf (ISSCC)*, San Francisco, CA, USA, Feb. 2014, pp. 242–243.

[16] R. Fragasse et al., "Analysis of SRAM enhancements through sense amplifier capacitive offset correction and replica self-timing," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 6, pp. 2037–2050, Jun. 2019.

[17] G. D. Licciardo, L. D. Benedetto, A. De Vita, A. Rubino, and A. Femia, "A bit-line voltage sensing circuit with fused offset compensation and cancellation scheme," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 10, pp. 1633–1637, Oct. 2019.

[18] D. Patel, A. Neale, D. Wright, and M. Sachdev, "Hybrid latch-type offset tolerant sense amplifier for low-voltage SRAMs," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 7, pp. 2519–2532, Jul. 2019.

[19] J. Chen, W. Zhao, Y. Wang, and Y. Ha, "Analysis and design of reconfigurable sense amplifier for compute SRAM with high-speed compute and normal read access," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 12, pp. 3503–3507, Dec. 2021.

[20] M. Qazi, K. Stawiasz, L. Chang, and A. P. Chandrakasan, "A 512kb 8T SRAM macro operating down to 0.57V with an AC-coupled sense amplifier and embedded data-retention-voltage sensor in 45nm SOI CMOS," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 85–96, Jan. 2011.

[21] L. Wen, X. Cheng, K. Zhou, S. Tian, and X. Zeng, "Bit-interleaving-enabled 8T SRAM with shared data-aware write and reference-based sense amplifier," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 63, no. 7, pp. 643–647, Jul. 2016.