Hybrid Latch-Type Offset Tolerant Sense Amplifier for Low-Voltage SRAMs

Dhruv Patel, *Student Member, IEEE*, Adam Neale, *Member, IEEE*, Derek Wright, *Member, IEEE*, and Manoj Sachdev, *Fellow Member, IEEE*

Abstract—Sense Amplifier (SA) input-referred offset often dictates the minimum required differential input ($\Delta V_{BL\text{-}min}$) and is an important factor in realizing low-voltage SRAMs. This work presents a HYbrid Latch-Type Sense Amplifier (HYSA-QZ), where the bitline signals are supplied to multiple internal nodes to significantly reduce $\Delta V_{BL\text{-}min}$. A 65nm CMOS test chip with arrays of HYSA-QZ, two intermediate formulations of HYSA-QZ, conventional Current Latch SA (CLSA) and conventional Voltage Latch SA (VLSA) were fabricated. Measurements over 5120 SAs of each type show that the HYSA-QZ implemented with regular-V_T transistors require 50.0%, and 22.8% lower $\Delta V_{BL\text{-}min}$ with 6.5% (or 4.5%) and 30.7% (or 18.8%) of total gate (or layout) area overhead compared to CLSA and VLSA at 0.4 V, respectively. Iso-gate-area offset improvement was substantiated with Pelgrom's mismatch model where HYSA-QZ with regular-V_T transistors showed 46.6% and 7.7% improvements in measured standard deviation of offset distribution compared to CLSA and VLSA, respectively. Measured $\Delta V_{BL\text{-}min}$ for HYSA-QZ remains stable and low over a temperature range from 0 °C to 75 °C at 0.4 V. Moreover, an additional 13.0% reduction in $\Delta V_{BL\text{-}min}$ was measured in HYSA-QZ when using Low-V_T transistors. Finally, HYSA-QZ operates reliably at $V_{DD\text{-}min}$ of 260 mV in 25 °C.

Index Terms—Comparator, Offset Tolerant Latch, Offset Cancellation, Sense Amplifier, Variation Tolerant Circuits, SRAM

I. Introduction

Static Random Access Memories (SRAMs) often occupy a significantly large area of System on Chips (SOCs) and consequently affect their energy consumption, yield, and reliability. A sense amplifier (SA) is an important circuit that reads and amplifies the data stored in an SRAM cell. The characteristics of an SA determine several important SRAM metrics, including minimum operating voltage, maximum read frequency, and power/energy consumption [1]. Among all these related characteristics, the SA sensing delay, minimum required differential input voltage, $\Delta V_{BL\text{-}min}$ (resolution), and energy of the read/write operations are the most important [2].

II. Background

Two popular SA topologies are the Voltage Latch SA (VLSA) and the Current Latch SA (CLSA). The choice of specific topology is dependent on the technology used [3]. To make a reliable decision, an SA requires a minimum worst-case differential signal ($\Delta V_{BL\text{-}min}$), which should be greater than the SA's input-referred offset voltage (V_{OS}) [4]. The SA's V_{OS} is determined by the V_T mismatches of the sensing and input

transistors [5]. Abu Rahma et al. reported that larger standard deviation of input referred offset distribution (σ_{OS}) has a significant negative impact on SRAM speed and read access yield, Y_{read} [6]. Analysis carried out by the authors shows that for a 28 nm, 16 Mb SRAM with Y_{read} of 97%, every 1 mV increase in σ_{OS} of the SA requires a 10 mV increase in the $\Delta V_{BL\text{-}min}$ [6]. Minimizing $\Delta V_{BL\text{-}min}$ improves energy consumption as it takes less time to develop a smaller differential voltage on the highly capacitive bitlines, which experience less discharge per read access. The SA's V_{OS} arises from the mismatches in the gain factor, the drain current, the threshold voltage V_T, and the layout of the devices used [7, 8]. Among these, V_T mismatch has been identified as the dominant contributing factor to V_{OS}. Shah [9] concluded that the V_T mismatch between the NMOS sensing pair mostly determines the CLSA's V_{OS}. Similarly, the work on VLSA [10] reported that the majority of V_{OS} is contributed by the V_T mismatch in NMOS pair when bitlines are precharged to V_{DD}. Unfortunately, aggressive device scaling has resulted in increased device variations and contributed to larger V_{OS} in SAs [11, 12].

The main target for this work is to investigate into offset tolerant low-voltage SA topologies that can leverage low-voltage SRAMs to enable wide range of battery-operated mobile, sensory and implantable SOCs with stringent energy constraints. In this paper, we propose a differential SA architecture that reduces the required $\Delta V_{BL\text{-}min}$ while offering reliable operation from the nominal supply down to the subthreshold supply range. The proposed SA combines both CLSA and VLSA features by applying the differential input signal to multiple sensing nodes, and thus referred to it as the HYbrid SA (HYSA). This also makes our proposed differential HYSA compatible with most of the symmetrical fully differential SRAM cells i.e. 6T, 8T, 10T etc.

The rest of this paper is outlined as follows. Section III introduces the proposed SA, HYSA-QZ as a succession from the conventional topologies with their description, and simplified circuit analysis and proving the improvement in offset tolerance. Section IV shows the simulations on transient waveforms, and offset tolerance supporting the hypothesis from the prior section. It also analyzes each SA's sensing delay and dynamic power consumption. Section V contains measured offset statistics and shmoo plots. Section VI compares proposed HYSA-QZ with other SAs considered in this work as well as previously published works. Finally, Section VII concludes the work with summarized results.

III. Hybrid SRAM sense amplifier

Abu Rahma et al. reported that the CLSA exhibits ~3x wider σ_{OS} compared to a VLSA of the same total gate area [6]. However, the CLSA having access to multiple internal nodes over a regenerative signal path allowed us to apply the bitline signals to those internal nodes to preset the regenerative mechanism's converging direction with higher gain factor against mismatches. As a result, the HYSA-QZ was developed where its schematic and layout are shown in Fig. 1. Similar to the operation of VLSA and CLSA, the operation of the proposed HYSA-QZ can be divided into two phases: (I) Precharge with BL/BLB (II) Enable Latch, Regenerate and Resolve. In phase (I), SAE is kept LOW which turns ON all the P3-P6 access PMOS switches allowing differential BL/BLB to precharge Z/ZB and Q/QB nodes with small voltage difference. In phase (II), SAE is asserted HIGH which turns off all the P3-P6 PMOS access switches and turns on N5. This allows differential current driving transistors, N3/N4 driven by BL/BLB to further amplify the voltage difference in Z/ZB and Q/QB nodes. This ultimately helps start the regeneration process in the latching element formed by the P1/P2 and N1/N2 transistors. Finally, the latching element resolves one of Q/QB nodes to V_{DD} and another to GND. The resulting output on Q/QB nodes is later buffered through an inverter as OUTB/OUT.
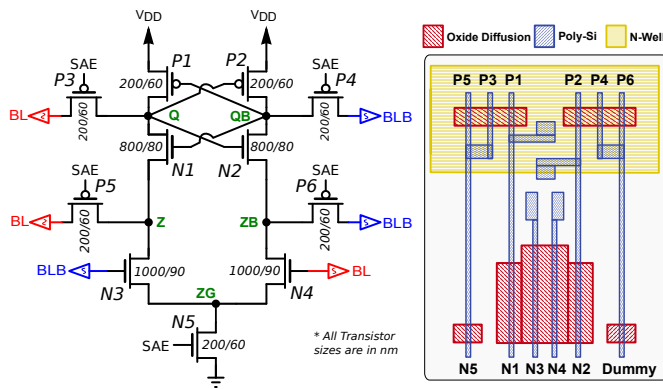

Fig. 1. Proposed offset tolerant HYbrid Latch-type Sense Amplifier (HYSA-QZ) schematic and layout.

Fig. 2 (a) illustrates the proposed HYSA-QZ and its associated mode selection transistors. The mode selection multiplexors and transistors are for investigating and understanding the effectiveness of applying the differential signal to multiple SA locations to mitigate the V_T mismatch between otherwise matched transistor pairs. These are shown in gray to distinguish them from the HYSA-QZ transistors. In the most preferred configuration, **HYSA-QZ**, the differential bitline signals are applied to all three following pairs of nodes: (i) Q and QB nodes; (ii) Z and ZB nodes; and (iii) the gates of transistors N3 and N4. Transistors P1, N1, P2, and N2 make a latching element in the HYSA-QZ and, together with P3 and P4, behave like a VLSA as shown in Fig. 2 (b). Transistors N3 and N4 with their gates connected to BLB and BL, respectively, mimic a CLSA.

Table I shows seven different topologies with their respective modes implemented on the test chip along with their relative total gate area (calculated sum of width x length of transistors) and physical layout area (both excludes the area of selection multiplexors/transistors). The first letter in the first column signifies Regular or Low V_T (R-V_T or L-V_T) transistors
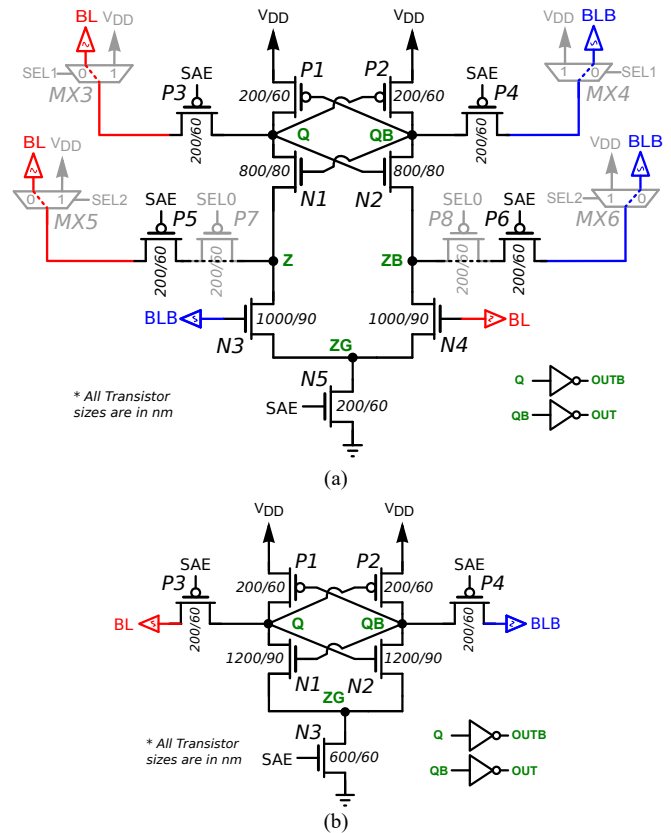

(a)

(b)

Fig. 2. SAs fabricated in 65nm test Chip – (a) Hybrid and Current Latch-Type SRAM SAs with configurable test modes (b) Conventional VLSA.

in the SA. Subsequent letters signify the types of SA, e.g., HYSA, CLSA or VLSA. The last set of letters signify the applied differential input signal (BL/BLB) locations. For example, L-HYSA-QZ signifies it to be an L-V_T HYSA with BL/BLB signals applied to Q/QB and Z/ZB nodes. As apparent from Fig. 2 (a), an HYSA with appropriate SEL signals can be reduced to a CLSA. In all CLSA and HYSA configurations, the BL/BLB are always applied to the gates of N3/N4 transistors. In literature, the SA proposed in the work of [13] is similar to the L-CLSA-Z which additionally applies BL/BLB signals to Z/ZB nodes. On the other hand, the proposed SAs in the works of [14], [15] and [16] are similar to the L-HYSA-Q which additionally applies BL/BLB signals to Q/QB nodes, but not to Z/ZB nodes. In our proposed R/L-HYSA-QZ, we apply BL/BLB signals to both Q/QB and Z/ZB nodes simultaneously to further improve the offset tolerance and in fact, the best

TABLE I. Test Modes Description

Mode	Mode description	SEL [2:0]	Rel. Total Gate Area	Rel. Total Layout Area
R-VLSA	R-V_T VLSA (Conventional)	-	1	1
R-CLSA	R-V_T CLSA (Conventional)	111	1.23	1.14
L-CLSA	L-V_T CLSA (Conventional)	111	1.23	1.14
L-CLSA-Z	L-V_T CLSA with BL/BLB to Z/ZB (Intermediate-I)	010	1.31	1.19
L-HYSA-Q	L-V_T HYSA with BL/BLB to Q/QB (Intermediate-II)	101	1.23	1.14
R-HYSA-QZ	R-V_T HYSA BL/BLB to Q/QB and Z/ZB (Proposed)	000	1.31	1.19
L-HYSA-QZ	L-V_T HYSA BL/BLB to Q/QB and Z/ZB (Proposed)	000	1.31	1.19

possible offset tolerance among all analyzed topologies in this work. The significance and the effect of applying BL/BLB signals to both Q/QB and Z/ZB nodes will be shown throughout this paper by the means of offset tolerance analysis, simulations and detailed on-chip measurements.

The rest of this section provides a substantiation with simplified circuit analysis showing HYSA-QZ having higher offset tolerance compared to other topologies. The circuit analysis shown in Fig. 3 evaluates differential overdrive voltages (ΔV_{OV}) and differential drain-to-source voltages (ΔV_{DS}) on the NMOS transistor pairs, N1/N2 and N3/N4 at the instant when the SAE signal makes a 0→1 transition. ΔV_{OV} and ΔV_{DS} of NMOS transistor pairs are analyzed as they are the key parameters for creating differential transconductance (Δg_m) and differential impedance ($\Delta\Omega$), respectively giving rise to differential current (ΔI) in the left and right branches of a given latch-type SA. The polarity of ΔI set by the applied ΔV_{BL} and other preset conditions is of importance in making a correct decision at the moment when the SAE is asserted. Hence, ΔV_{OV} and ΔV_{DS} are analyzed at this instant.

Assumptions and definitions used in this analysis in Fig. 3 are shown in the first row. To impose worst-case disadvantageous mismatch conditions, V_T mismatches (ΔV_T) are assumed such that $V_{T2} > V_{T1}$ and $V_{T4} > V_{T3}$ inducing ΔI in undesirable opposite polarities. Given these mismatch assumptions and applied ΔV_{BL} according to the topology configurations (see third and fourth rows), $\Delta V_{OV:2-1}/\Delta V_{OV:4-3}$ and $\Delta V_{DS:2-1}/\Delta V_{DS:4-3}$ are calculated as shown in fifth and sixth rows, respectively. Based on the applied BL/BLB in all topologies, the expected value through the left branches (Q) is '1' (V_{DD}) and through the right branches (QB) is '0' (GND). Hence, favorably, left branches should exhibit higher impedance to GND trying to hold to V_{DD} and should drain less current compared to right branches. On the other hand, right branches should exhibit lower impedance to GND trying to drain QB to GND and should drain more current compared to left branches. Therefore, the preset or pre-charge conditions where the transistors in the right branches have higher V_{OV} (indicating higher current draining capability) and lower V_{DS} (indicating lower impedance to GND) compared to the transistors in the left branches is desirable. This dictates that higher magnitudes of $\Delta V_{OV:2-1}/\Delta V_{OV:4-3}$ with +ve

polarity and higher magnitudes of $\Delta V_{DS:2-1}/\Delta V_{DS:4-3}$ with –ve polarity is desired resulting in higher ΔI with desirable polarity working against mismatches; this ultimately dictates topology's ability to tolerate disadvantageous mismatch conditions.

Calculated ΔV_{OV} and ΔV_{DS} in Fig. 3 are put into three categories: (1) working with mismatch (Bad), (2) working against mismatch to neutralize the mismatch effect (Neutral) and (3) Compensating mismatch to 0 and/or giving additional offset tolerance (Good). **VLSA:** The required ΔV_{BL-min} must be at least equal to $\Delta V_{T:2-1}$ to neutralize the mismatch effect with some aid from $\Delta V_{DS:2-1}$ pre-setting $\Delta\Omega_{2-1}$ in favor of the desired decision. **CLSA:** It has $\Delta V_{OV:2-1}$ compensated to 0 where the mismatch due to $\Delta V_{T:2-1}$ works in favor of presetting $\Delta\Omega_{4-3}$ but works against presetting $\Delta\Omega_{2-1}$. Coincidental advantages due to $\Delta V_{T:2-1}$ are bound to randomness and therefore not assured. CLSA will typically require ΔV_{BL-min} of at least $\Delta V_{T:4-3}$ for making a correct decision, requiring it to neutralize $\Delta V_{OV:4-3}$. **CLSA-Z:** It has applied ΔV_{BL} working in favor of presetting $\Delta\Omega_{4-3}$ giving assured aid in desirable initial amplification phase, but works against presetting $\Delta\Omega_{2-1}$ which is relatively less of a concern after the sufficient signal amplification on Z/ZB is achieved [17]. It will require ΔV_{BL-min} of around $\Delta V_{T:4-3}$ to neutralize $\Delta V_{OV:4-3}$ or little larger to overcome worst-case $\Delta V_{T:2-1}$ mismatch. **HYSA-Q:** It has coincidental aid from $\Delta V_{T:2-1}$ to pre-set $\Delta\Omega_{4-3}$ in the desired polarity but is diminished by the intentionally applied ΔV_{BL}. However, it has its $\Delta V_{OV:2-1}$ compensated to 0 and has 2x impact of ΔV_{BL} to fight against $\Delta V_{T:2-1}$ mismatch to neutralize $\Delta V_{DS:2:1}$; which consequently reduces ΔV_{BL-min} requirement to neutralize $\Delta V_{OV:4-3}$. **HYSA-QZ:** This proposed topology has assured ΔV_{BL} presetting $\Delta\Omega_{4-3}$ in the desired polarity during initial amplification phase and has $\Delta V_{DS:2-1}$ compensated to 0. Moreover, its $\Delta V_{OV:2-1}$ is benefitted by 2x effect of ΔV_{BL} to fight against $\Delta V_{T:2-1}$ mismatch; this further helps in reducing the ΔV_{BL-min} requirement to fight $\Delta V_{T:4-3}$ to neutralize $\Delta V_{OV:4-3}$. All these benefitting attributes indicate HYSA-QZ has higher offset tolerance than topologies discussed before. Overall, applying BL/BLB to more internal nodes bring consistency in preset conditions and relaxes ΔV_{BL-min} for neutralizing mismatches or disadvantageous preset conditions. The simulations supporting this hypothesis is shown in the next section followed by supporting measurement results in Section V.
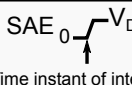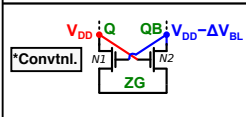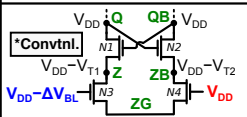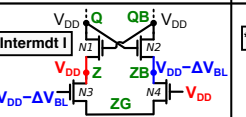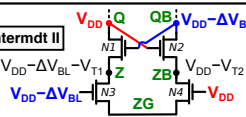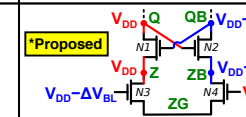
Circuit Analysis Assumptions & Definitions:	SAE 0→V_{DD} Time instant of interest	$\Delta V_{T:2-1} = V_{T2} - V_{T1}$ $\Delta V_{T:4-3} = V_{T4} - V_{T3}$	$V_{OV} = V_{GS} - V_T$ $\Delta V_{OV:2-1} = V_{OV2} - V_{OV1}$ $\Delta V_{OV:4-3} = V_{OV4} - V_{OV3}$	$V_{DS} = V_D - V_S$ $\Delta V_{DS:2-1} = V_{DS2} - V_{DS1}$ $\Delta V_{DS:4-3} = V_{DS4} - V_{DS3}$	Worst Case random local mismatch condition based on applied BL/BLB signals → $V_{T2} > V_{T1}$, $V_{T4} > V_{T3}$

	VLSA	CLSA	CLSA-Z	HYSA-Q	HYSA-QZ
Q/QB: Pre-charge	Pre-charged to BL/BLB	Pre-charged to V_{DD}	Pre-charged to V_{DD}	Pre-charged to BL/BLB	Pre-charged to BL/BLB
Z/ZB:	—	Floated	Pre-charged to BL/BLB	Floated	Pre-charged to BL/BLB
Circuit	*Convtnl.	*Convtnl.	*Intermdt I	*Intermdt II	*Proposed
$\Delta V_{OV:2-1}$	$\Delta V_{OV:2-1} = \Delta V_{BL} - \Delta V_{T:2-1}$	$\Delta V_{OV:2-1} = 0$	$\Delta V_{OV:2-1} = \Delta V_{BL} - \Delta V_{T:2-1}$	$\Delta V_{OV:2-1} = 0$	$\Delta V_{OV:2-1} = \Delta V_{BL} + \Delta V_{BL} - \Delta V_{T:2-1}$
$\Delta V_{OV:4-3}$		$\Delta V_{OV:4-3} = \Delta V_{BL} - \Delta V_{T:4-3}$	$\Delta V_{OV:4-3} = \Delta V_{BL} - \Delta V_{T:4-3}$	$\Delta V_{OV:4-3} = \Delta V_{BL} - \Delta V_{T:4-3}$	$\Delta V_{OV:4-3} = \Delta V_{BL} - \Delta V_{T:4-3}$
$\Delta V_{DS:2-1}$	$\Delta V_{DS:2-1} = -\Delta V_{BL}$	$\Delta V_{DS:2-1} = +\Delta V_{T:2-1}$	$\Delta V_{DS:2-1} = +\Delta V_{BL}$	$\Delta V_{DS:2-1} = -\Delta V_{BL} - \Delta V_{BL} + \Delta V_{T:2-1}$	$\Delta V_{DS:2-1} = 0$
$\Delta V_{DS:4-3}$		$\Delta V_{DS:4-3} = -\Delta V_{T:2-1}$	$\Delta V_{DS:4-3} = -\Delta V_{BL}$	$\Delta V_{DS:4-3} = -\Delta V_{T:2-1} + \Delta V_{BL}$	$\Delta V_{DS:4-3} = -\Delta V_{BL}$

Bad: Working **With** mismatch **Neutral:** Working **against** mismatch to **neutralize** the mismatch effect **Good:** **Compensating** mismatch to 0 and/or beyond giving additional offset tolerance

Fig. 3. Simplified circuit analysis showing how applying bitlines to various internal nodes over each of the topologies studied in this work is effectively overcoming the V_T mismatches. The analysis consists of comparing resulted differential over drive voltages (ΔV_{OV}) and differential drain-to-source voltages (ΔV_{DS}) for an applied ΔV_{BL} on a given SA topology. The analysis was performed just at the moment when the SAE is asserted high (0 → 1 transition).

From the previous analysis it was clear that all topologies do not have any mismatch compensation happening on N4/N3 pair ($\Delta V_{OV:4-3}$ terms in Fig. 3). Hence, this led to spending more gate area budget on N4/N3 than on N2/N1 reducing the $\Delta V_{T:4-3}$ mismatch. The sizing of the R-VLSA in Fig. 2 (b) was selected as a tradeoff between V_{OS}, energy consumption, and the overall area. Furthermore, the PMOS access switches (P3-P6) in all topologies exhibit mismatch and coupling effect from the SAE signal on internal nodes. Since the common mode on bitlines is assumed to be close to V_{DD}, PMOS transistors were selected for these access switches to avoid V_T drop across them while pre-charging internal nodes. By giving sufficient time to pre-charge internal nodes, the impact of I_d variations due to mismatch effect could be reduced. Moreover, these switches were sized near-

minimum to avoid additional bitline loading and the coupling effect through the overlap capacitances of the access switches which are relatively small fraction of their gate capacitances (~10% of C_{gate}).

IV. SIMULATION RESULTS

As this work is mainly targeted towards low-voltage operation, the comparison among each SA's output (Q/QB) transient response with $V_{DD} = 0.2$ V and $\Delta V_{BL} = -40$mV at TT/25 °C corner is shown in Fig. 4. It shows 1k Monte Carlo (MC) simulations with the decision failures associated with each SA. SAs with Q/QB precharged with bitlines (R-VLSA, L-HYSA-Q, R/L-HYSA-QZ) were able to perform successfully at such low supply voltage without any failures. Also, similar results were achieved with $\Delta V_{BL} = +40$mV. Despite having four stacking transistors, L-HYSA-Q and proposed R/L-HYSA-QZ were capable of operating without any decision failures owing to their precharge conditions and high offset tolerance. Hence, they are suitable for low-voltage operations. The benefits of applying bitlines at multiple locations is further analyzed in detail later in this section.
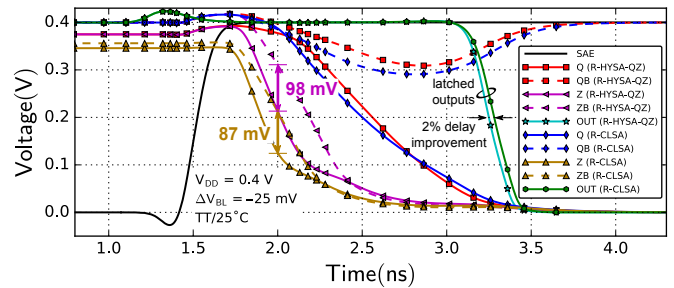
Fig. 5. Transient simulations comparison on pre-amplification of Z/ZB nodes on R-HYSA-QZ and R-CLSA just before the Q/QB peels away.

Fig. 5 shows the transient simulation comparison between R-HYSA-QZ and R-CLSA with $V_{DD} = 0.4$ V and $\Delta V_{BL} = -25$ mV at TT/25°C corner. The purpose of this simulation is to compare the initial amplification of the differential signal in these two topologies before the regenerative phase starts (Q/QB peeling away). As analyzed in Fig. 3, pre-charging Q/QB and Z/ZB with bitline signals in R-HYSA-QZ help increase the effective ΔV_{BL} applied to the SA. For example, Z/ZB in R-HYSA-QZ has 11 mV (98 mV–87 mV) more differential amplitude developed compared to R-CLSA showing the increased effectiveness of ΔV_{BL} and ultimately increasing offset tolerance [17] [18]. Notice that Z/ZB in R-HYSA-QZ starts discharging from a relatively higher voltage compared to Z/ZB in R-CLSA. Despite that, R-HYSA-QZ could resolve in 2% less time owing to higher preamplification as a result of applied bitlines at Q/QB and Z/ZB nodes.

An SA with offset mitigation features should be able to tolerate offset and still make the correct decision. Simulated offset tolerance for each transistor pair in different topologies is shown in Fig. 6. It signifies how much V_T offset in a given transistor pair can a given SA tolerate while still making a correct decision with applied value of ΔV_{BL}. The analysis was performed across different values of ΔV_{BL} with $V_{DD} = 0.4$ V at TT/25°C corner. $V_{T:1,2,3,4}$ values were modified from their nominal values (i.e. $V_{T:2} = V_{T-nom} + \Delta V_{T:2-1}/2$, $V_{T:1} = V_{T-nom} - \Delta V_{T:2-1}/2$) in the transistor model files to emulate $\Delta V_{T:4-3}$ and $\Delta V_{T:2-1}$ in undesirable direction. Fig. 6 (a) and (b) shows the
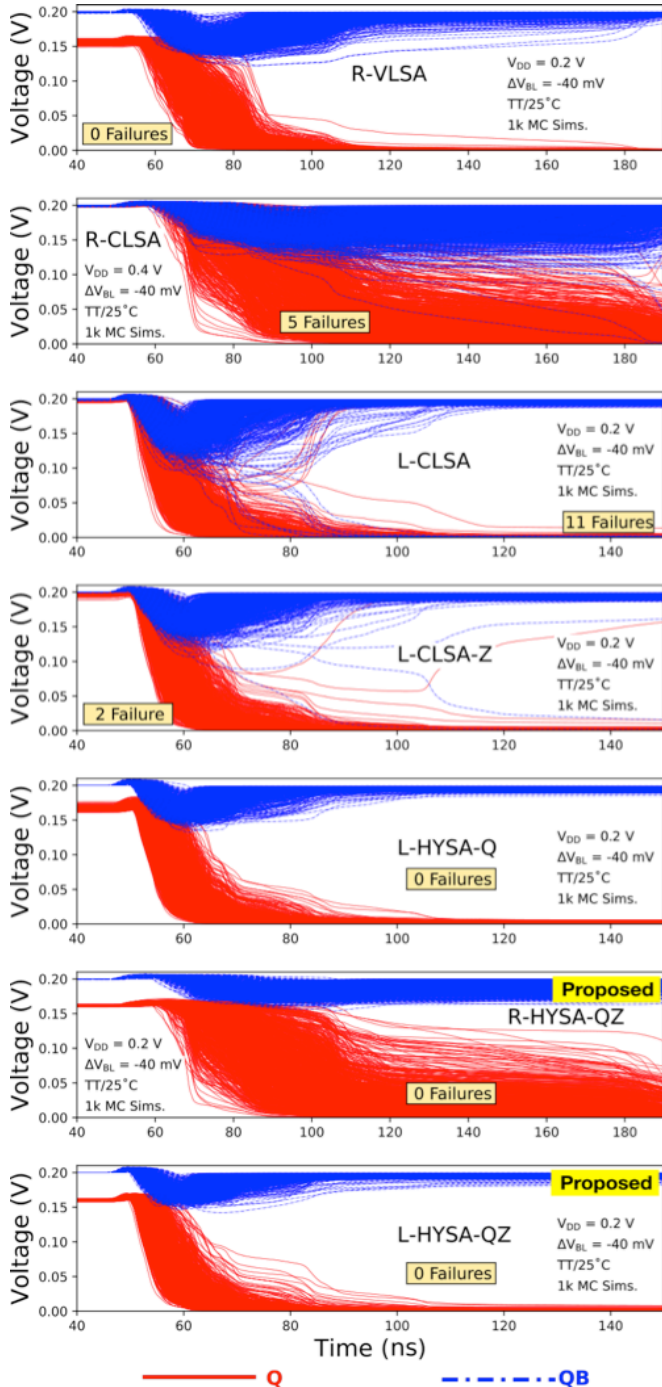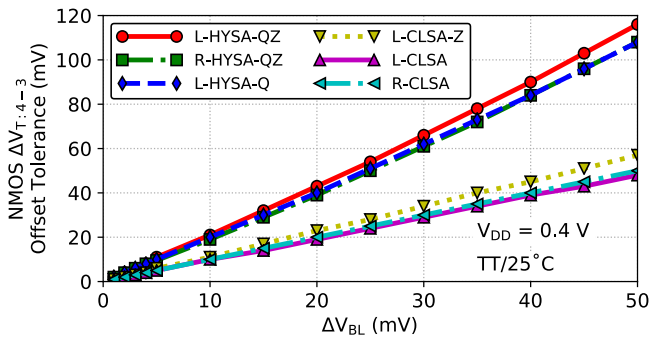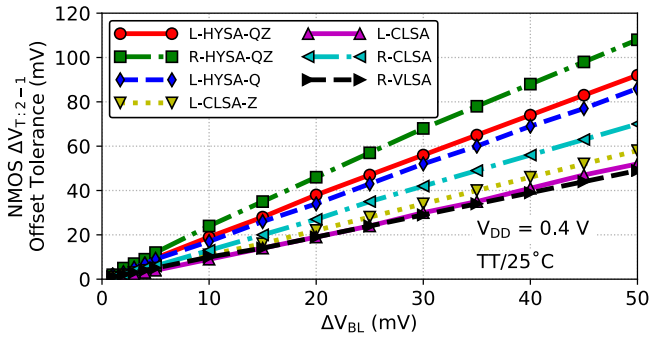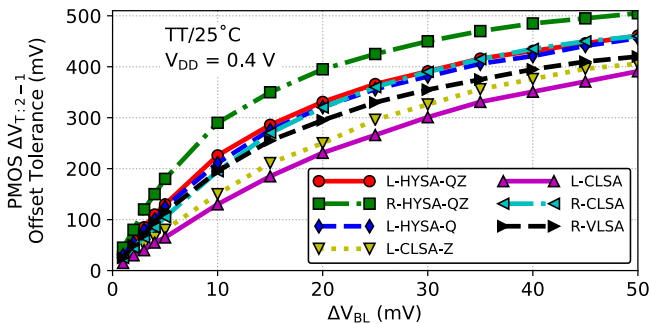

Fig. 4. Transient response (1k Monte Carlo simulations) of SAs analyzed in this work. Simulated with $V_{DD} = 0.2$ V and $\Delta V_{BL} = -40$mV at 25° C.

offset tolerance for NMOS pairs N4/N3 and N2/N1 one at a time, respectively. For example, R-HYSA-QZ with applied ΔV_{BL} of 20 mV is capable of tolerating 44 mV of $\Delta V_{T:2-1}$ offset (considering other transistor pairs in perfect matching conditions, i.e $\Delta V_{T:4-3} = 0$) and still make the correct decision. This can be explained by inspecting the simplified offset tolerance analysis from Fig. 3 in a following manner. Let us substitute the known values such as $\Delta V_{BL} = 20$ mV and $\Delta V_{T:4-3} = 0$ into the column associated with R-HYSA-QZ in Fig. 3. Now, this results in $\Delta V_{OV:2-1} = +40$ mV$-\Delta V_{T:2-1}$, meaning $\Delta V_{T:2-1}$ has to be greater than 40 mV to flip $\Delta V_{OV:2-1}$ into $-$ve (undesirable) polarity to make an incorrect decision. Not only that but with the resulting $\Delta V_{OV:4-3}$ of +20 mV, $\Delta V_{DS:2-1}$ of 0 mV and $\Delta V_{DS:4-3}$ of -20 mV, with all terms either being with desired polarity or compensated to 0, provides additional pull to maintain latching inertia into correct direction. Hence requiring $\Delta V_{T:2-1}$ of 44 mV to overcome effective desirable impact of precharging conditions. Overall, all SAs' offset tolerance associated with NMOS pairs show linear scaling with applied ΔV_{BL}.
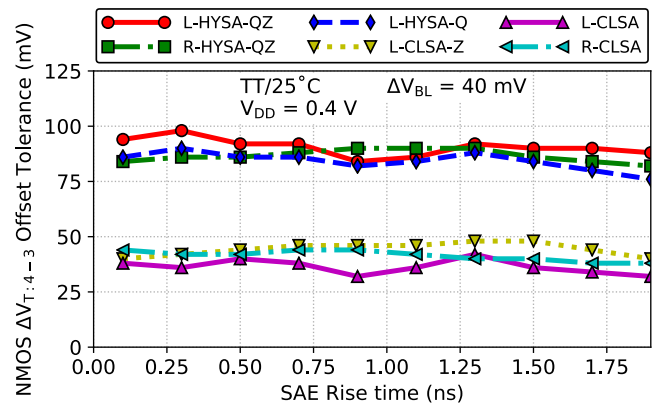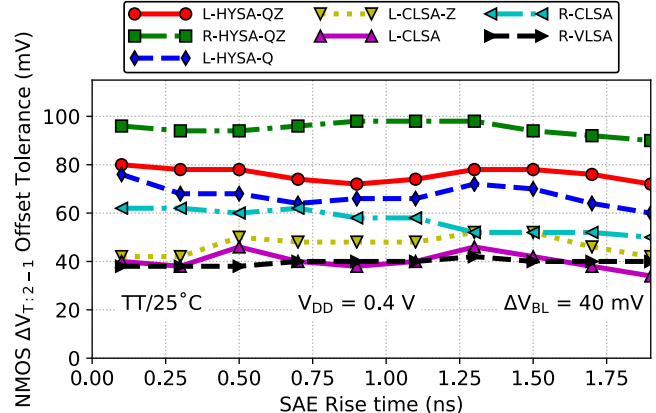
Furthermore, Fig. 6 (c) shows a similar analysis on the P1/P2 pair. All topologies can tolerate relatively much higher PMOS mismatches compared to NMOS mismatches. Thus, sizing PMOS pair to near-minimum size poses almost no harm to V_{OS} and allowed us to spend most of the gate area budget on the NMOS transistors where the V_T mismatch matters the most. In all three transistor pairs (N4/N3, N2/N1 and P1/P2) and within their respective V_T flavors, offset tolerance is increased as more internal nodes are precharged with bitlines where R/L-HYSA-QZ have the highest offset tolerance compared to other SAs. From these offset tolerance simulations, the general trend of the offset tolerance of HYSA-QZ > HYSA-Q > CLSA-Z > VLSA $\approx$ CLSA, which supports the circuit analysis performed in the previous section. Also, applying bitlines at Q/QB is relatively more effective compared to applying bitlines at Z/ZB as the offset tolerance improvement from L-CLSA-Z to L-HYSA-Q is relatively much higher than the offset tolerance improvement from L-CLSA to L-CLSA-Z.

(a)

(b)

(c)

Fig. 6. Simulated offset tolerance comparison across ΔV_{BL} with $V_{DD} = 0.4$ V and $\Delta V_{BL} = 20$ mV at 25°C. (a) Offset tolerance due to V_T mismatches of N4/N3 ($\Delta V_{T:4-3}$) only. (b) Offset tolerance due to N1/N2 V_T mismatches ($\Delta V_{T:2-1}$) only. (c) Offset tolerance due to P1/P2 V_T mismatches ($\Delta V_{T:2-1}$) only.

(a)

(b)

Fig. 7. Simulated Offset Tolerance comparison across SAE Rise time with $V_{DD} = 0.4$ V and $\Delta V_{BL} = 40$ mV at TT/25°C. Simulations were performed with transistor noise and $V_{RMS-noise} = 5$ mV on all V_{DD}, bitlines and SAE signals. (a) Offset tolerance due to V_T mismatches of N4/N3 ($\Delta V_{T:4-3}$) only. (b) Offset tolerance due to N1/N2 V_T mismatches ($\Delta V_{T:2-1}$) only.

The SAE signal responsible for enabling footer transistor and disabling access switches leads to charge injection and coupling effects on to the critical internal nodes of the SA. Also, the SAE on/off timing determines the final considered output voltages developed on Q/QB nodes which are typically taken as an input by another latch following the SA. Hence, it is critical to analyze the sensitivity of the offset tolerance due to the dynamics of the SAE signal such as variations in SAE rise time

(Fig. 7) and SAE pulse width (Fig. 8). These simulations were performed with $V_{DD} = 0.4$ V and $\Delta V_{BL} = 40$ mV at TT/25 °C corner. To add additional rigor, the transistor noise was enabled and $V_{RMS\text{-}noise} = 5$ mV was added to all V_{DD}, bitlines and SAE signals during these simulations. Fig. 7 (a) and (b) shows the $\Delta V_{T:4\text{-}3}$ and $\Delta V_{T:2\text{-}1}$ offset tolerance across varied SAE rise time. Both plots have their offset tolerance relatively stable across SAE rise time and hence, sufficient to conclude that offset tolerance is insensitive to the SAE rise times within the range of 0.1 ns – 2 ns. Fig. 8 (a) and (b) shows the $\Delta V_{T:4\text{-}3}$ and $\Delta V_{T:2\text{-}1}$ offset tolerance across varied SAE pulse width. For all topologies, the offset tolerance starts off with lower value due to the pulse width not meeting the finite sensing delay associated with 40 mV of applied ΔV_{BL}. After the pulse width exceeds the SA's sensing delay, the offset tolerance starts to stabilize and converge to the offset tolerance value found from Fig. 6 (a) and (b). SAs in L-V_T flavor has much lower sensing delay and therefore requires relatively much less SAE pulse width before converging to their peak offset tolerance compared to SAs in R-V_T. For example, L-HYSA-QZ requires pulse width of ~2 ns whereas R-HYSA-QZ requires ~4.5 ns to converge to their peak offset tolerance values. To highlight, in both analysis from Fig. 7 and Fig. 8, proposed R/L-HYSA-QZ is consistent in offering highest offset tolerance among other topologies.

Possible offset improvement due to the use of transistors with different V_T flavors is also analyzed. Fig. 9 shows the V_T statistics of 1k MC simulations on an NMOS transistor ($W_n = $

1 μm, $L_n = 90$ nm) with L-V_T and R-V_T flavors under the saturation region at 25 °C in TT corner. Only an NMOS transistor with the size similar to the ones used in the implemented topologies was analyzed as the NMOS pairs have predominant impact on offset compared to PMOS pairs. Both μ_{V_T} and σ_{V_T} are reduced from R-V_T to L-V_T by 22% and 12%, respectively. Lower σ_{V_T} is an indication of reduced mismatch effect and ultimately lower V_{OS} for a given SA. Lower μ_{V_T} is an indication for a lower sensing delay with the cost of increased total dynamic power consumption of a given SA ($P_{dy\text{-}SA}$).
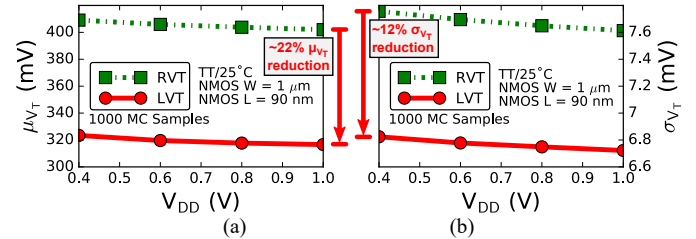

Fig. 9. NMOS (1μm/90nm) V_T statistics for 1k MC simulations at TT/25°C: (a) V_T average (μ_{V_T}) (b) V_T standard deviation (σ_{V_T}).

Fig. 10. (a) Simulated sensing delay across V_{DD}. (b) Normalized sensing delay w.r.t R-VLSA at a given V_{DD}. Performed at TT/25°C with $\Delta V_{BL} = -40$ mV.

Fig. 10 (a) shows the simulated sensing delays across V_{DD} on a semi-log scale with perfect matching condition at 25 °C while keeping $\Delta V_{BL} = -40$ mV. The sensing delays were extracted from $SAE_{50\%\text{-}rise}$ to the $OUT_{50\%\text{-}fall}$, where the OUT is the SAs' buffered outputs with an inverter depicted in Fig. 2. Fig. 10 (b) shows the relative sensing delay with respect to R-VLSA's sensing delay at a given V_{DD}. The sensing delay is reduced from R-V_T to L-V_T, with the difference increasing at lower V_{DD}. The R-VLSA has lower sensing delay compared to R-CLSA and R-HYSA-QZ due to fewer transistor stacking and larger footer transistor (N3) width. The L-HYSAs and L-CLSAs
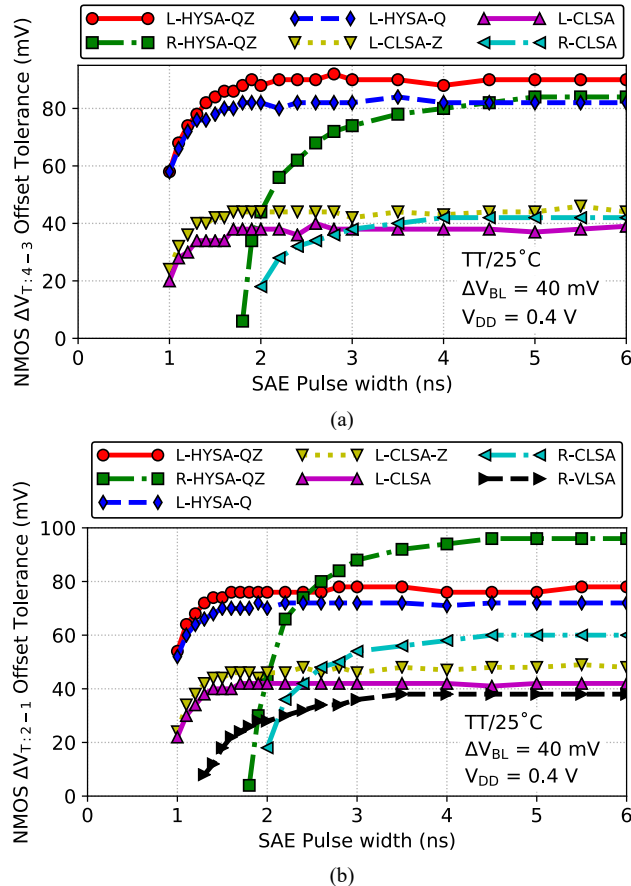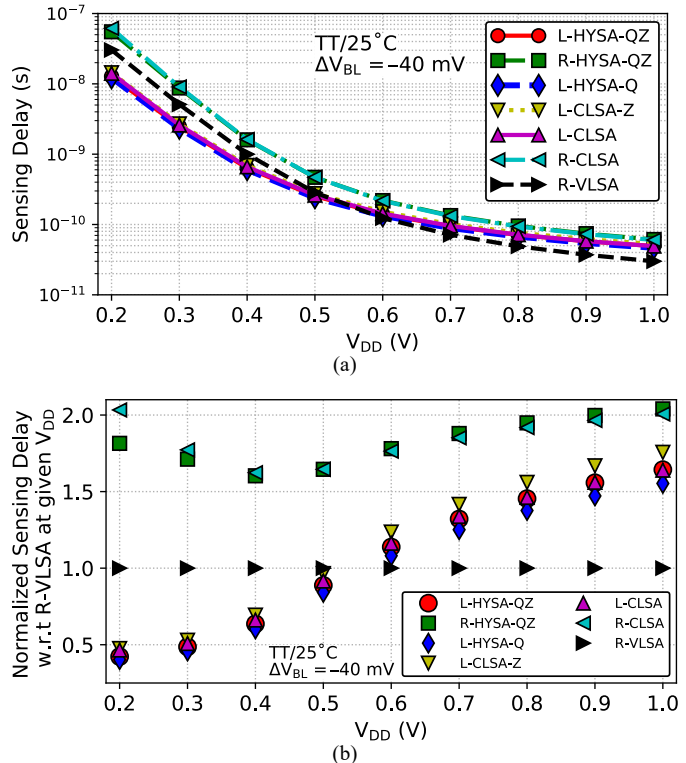

Fig. 8. Simulated Offset Tolerance comparison across SAE Pulse width with $V_{DD} = 0.4$ V and $\Delta V_{BL} = 40$ mV at TT/25°C. Simulations were performed with transistor noise and $V_{RMS\text{-}noise} = 5$ mV on all V_{DD}, bitlines and SAE signals. (a) Offset tolerance due to V_T mismatches of N4/N3 ($\Delta V_{T:4\text{-}3}$) only. (b) Offset tolerance due to N1/N2 V_T mismatches ($\Delta V_{T:2\text{-}1}$) only.

have lower sensing delay than R-VLSA up to 0.5 V owing to the combination of pre-charge conditions and lower V_T. Nevertheless, sensing delay of an SA is just a small fraction of the entire read path delay. Importantly, lower SA V_{OS} resulting in smaller $\Delta V_{BL\text{-}min}$ development on highly capacitive bitlines would ultimately help reduce the overall read path delay.
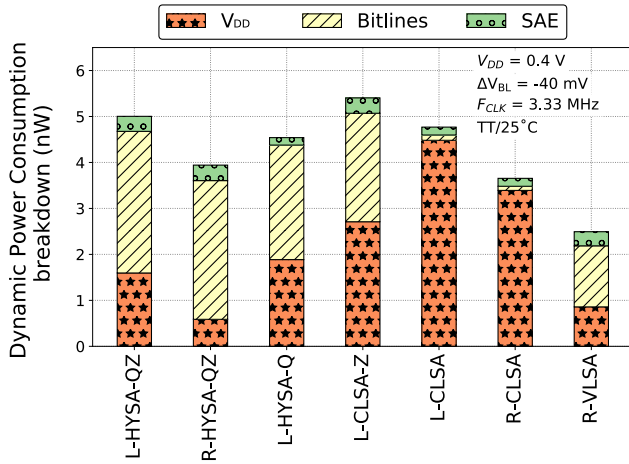

Fig. 11. Simulated $P_{dy\text{-}SA}$ breakdown into V_{DD}, bitlines and SAE. (Simulated with $V_{DD} = 0.4$ V, $\Delta V_{BL} = -40$mV, and $F_{CLK} = 3.33$ MHz at TT/25˚C corner).

The $P_{dy\text{-}SA}$ was simulated to realize the cost of precharging internal nodes. The $P_{dy\text{-}SA}$ breakdown of each SA is shown in Fig. 11 where the simulations were performed with $V_{DD} = 0.4$ V and $F_{CLK} = 3.33$ MHz at TT/25 ˚C corner. The simulated $P_{dy\text{-}SA}$ is divided into three key components: (1) Power drawn from the SAs' V_{DD} source during regeneration phase and in the case of R/L-CLSA and L-CLSA-Z, precharging their Q/QB nodes. (2) Portion of the power drawn from the bitlines loading the gates of the current sensing transistors (not applicable for R-VLSA) and in the case of R/L-HYSA-QZ, L-HYSA-Q, L-CLSA-Z and R-VLSA, precharging their Q/QB and/or Z/ZB nodes. (3) Power drawn from the SAE driver loading the gates of the SA enable footer transistor and the PMOS access switches (P3-P6). From Fig. 11, it is apparent that R/L-CLSA draws relatively much less power from bitlines as they are only applied to the gates of the current sensing transistors (N4/N3). On the other hand, R/L-HYSA-QZ, L-HYSA-Q, L-CLSA-Z, and R-VLSA has significant portion of their $P_{dy\text{-}SA}$ contributed by the bitlines as the internal nodes are being precharged by them. Also, the power contributed by the SAE signal is relatively low and as one would expect, it increases in proportion with the total gate area driven by the SAE signal (i.e. SAE power increases in topologies with higher number of access switches). The inverters used for buffering SA outputs had the same sizes for all topologies and hence, the power spent on loading those buffers (embedded in power contribution from V_{DD} and bitlines) is equal in all topologies. Overall, R-VLSA consumes the lowest $P_{dy\text{-}SA}$ as it has the lowest relative total gate area compared to other SAs. Also, note that despite having the same total gate area, L-CLSA-QZ has lower $P_{dy\text{-}SA}$ than L-HYSA-Z as latter precharges both Q and QB nodes to V_{DD} whereas former precharges only Q node to V_{DD} but prechages QB node to $V_{DD} - \Delta V_{BL}$, which is lower than V_{DD}. The increased $P_{dy\text{-}SA}$ cost for proposed R/L-HYSA-QZ compared to other SAs (excluding L-CLSA-Z) within same V_T flavor mainly comes from supplying bitlines to multiple internal capacitances as well as driving more number of access switches.
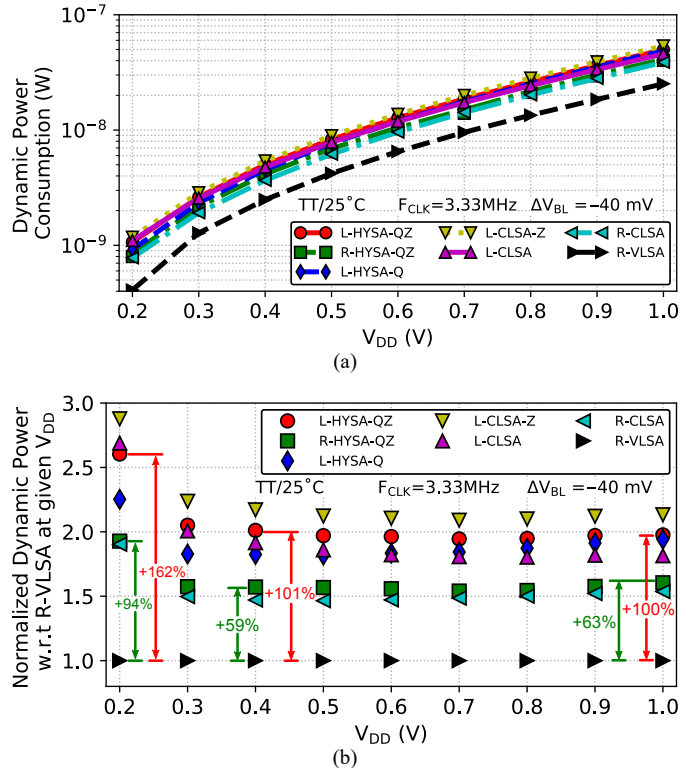

Fig. 12. (a) Simulated dynamic power (b) Normalized dynamic power w.r.t R-VLSA at a given V_{DD}. Performed at $\Delta V_{BL} = -40$mV, $F_{CLK} = 3.33$ MHz, TT/25˚C.

The $P_{dy\text{-}SA}$ comparison was further analyzed in Fig. 12 (a) showing the $P_{dy\text{-}SA}$ simulated across V_{DD} with $\Delta V_{BL} = -40$ mV and $F_{CLK} = 3.33$ MHz at TT/25˚C corner. Fig. 12 (b) shows the relative $P_{dy\text{-}SA}$ where R-VLSA was set as a reference for each V_{DD}. For R-HYSA-QZ and L-HYSA-QZ, the $P_{dy\text{-}SA}$ cost relative to R-VLSA is increased from 63% to 94% and from 100% to 162% while reducing V_{DD} from 1 V to 0.2 V, respectively. Despite the fact that the relative $P_{dy\text{-}SA}$ is higher for HYSA-QZs, with improved offset tolerance, they would lower highly capacitive bitline swing requirement reducing the overall SRAM power consumption and energy [19] [20]. Moreover, the $P_{dy\text{-}SA}$ is a much smaller fraction of SRAM's overall read access dynamic power consumption ($P_{dy\text{-}read}$) as reported by the works of [21] and [22]. Using the measured σ_{OS} values of SAs, the net benefit in $P_{dy\text{-}read}$ with the proposed R/L-HYSA-QZ topologies is justified in Section VI.

V. MEASUREMENT RESULTS

SA arrays of L-HYSAs/CLSAs, R-HYSAs/CLSAs and R-VLSA cells with different configurable modes were implemented in a 65nm-GP CMOS process shown in Fig. 13 with the test chip and the test bench setup. Similar to the works of [6, 10], each SA array was organized with 64 rows, and 8 columns containing total 512 cells that are individually addressable with a row and a column decoder. The test chip architecture for characterizing SA V_{OS} is shown in Fig. 14. Each row slice contains SAs, pre-charge circuitry, pull-down NMOS transistors, and a level shifter with a latch. BL and BLB signals are routed vertically throughout the array. All the control signals for the mode selection transistors and the multiplexers were driven by 1.0 V supply to minimize the timing uncertainty of control signals. The measurements consist of two key aspects of characterization: (1) Directly measured VOS statistics across

V_{DD} and temperatures. (2) Comparison of different topologies on reliability/reproducible results over clock frequency and the operating temperature at a given V_{DD}.
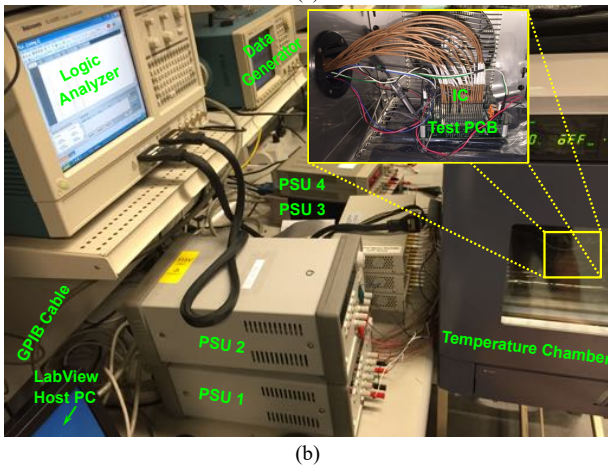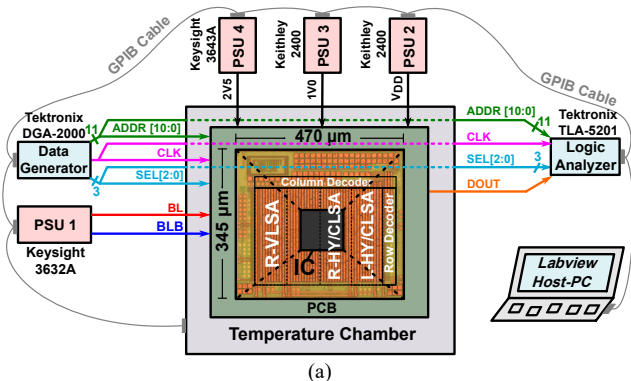

Fig. 13. (a) 65-nm Sense amplifier array test chip (b) test bench setup.
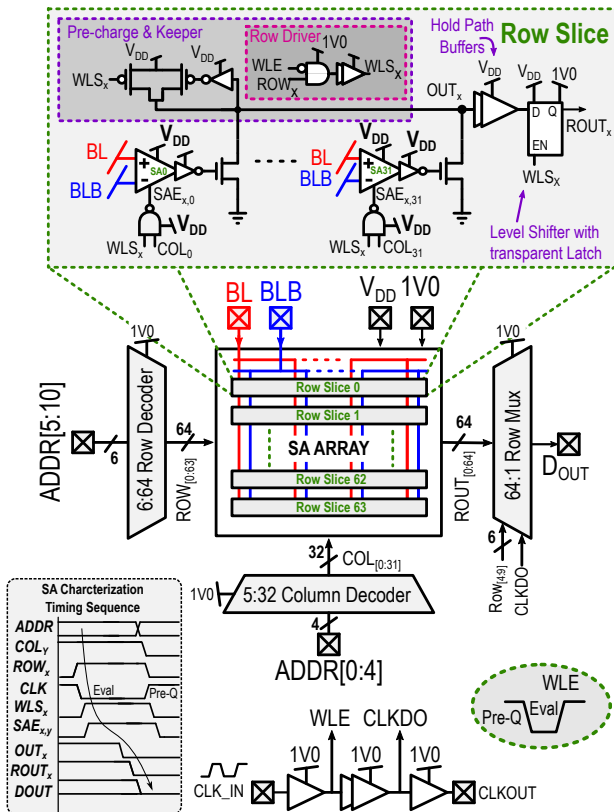
Fig. 14. Test chip implementation for SA offset characterization
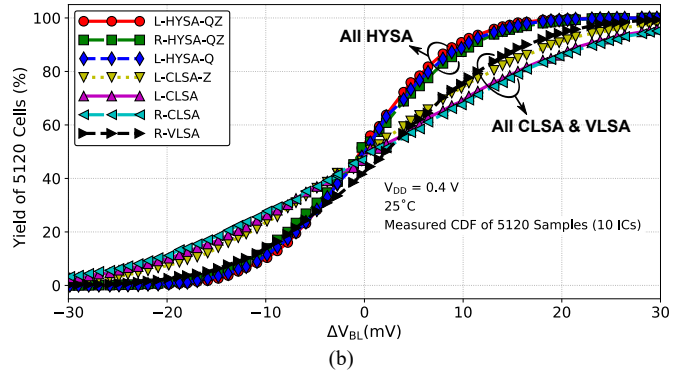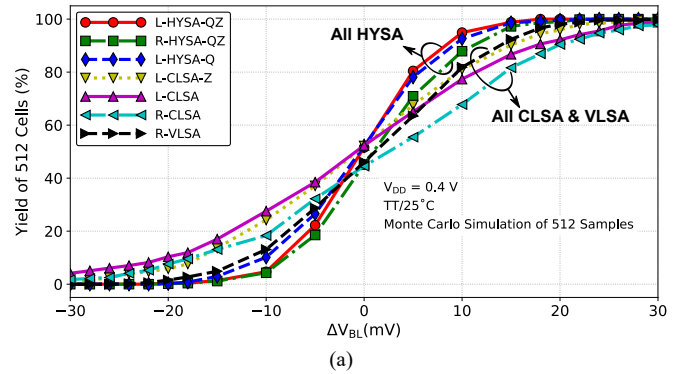
Fig. 15. (a) Simulated Cumulative Distribution of 512 SAs (b) Measured Cumulative Distribution of 5120 SAs across 10 ICs at 1 mV of ΔV_{BL} step resolution.
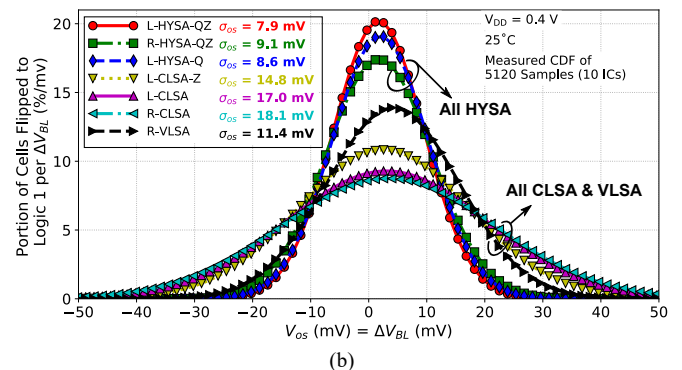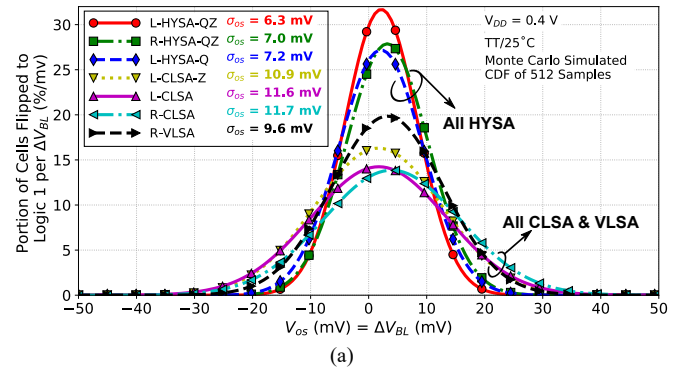
Fig. 16. (a) Simulated Probability Density of V_{OS} across 512 SAs (b) Measured Probability Density of V_{OS} across 5120 SAs (10 ICs) at 1 mV of ΔV_{BL} step resolution.

Fig. 15 (a) and (b) show simulated and measured Cumulative Distributions (CDF) of topologies analyzed in this work at 0.4 V at 25 °C, respectively. MC simulations and measured CDF plots across 10 ICs with a ΔV_{BL} step resolution

of 1 mV show that the ΔV_{OS} distribution is tighter for proposed topologies compared to others with same V_T flavor. Simulated and measured σ_{OS} values in Fig. 16 (a) and (b) were obtained by curve fitting (non-linear least-square minimization method) probability density function curves derived from their respective CDF curves illustrated in Fig. 15 (a) and (b), respectively. For example, 512 MC simulations in Fig. 16 (a) show that R-HYSA-QZ has σ_{OS} of 7.0 mV compared to R-CLSA of 11.7 mV, or R-VLSA of 9.6 mV. A similar trend is observed in Fig. 16 (b) over 5120 measured samples from 10 ICs with σ_{OS} of 9.1 mV, 18.1 mV and 11.4 mV for R-HYSA-QZ, R-CLSA and R-VLSA, respectively. The differences between measurements and simulations could be explained by the differences in the sample size, excluded layout parasitics and the inaccuracy of the mismatch model used in MC simulations. Fig. 17 (a) and (b) show the measured σ_{OS} and the mean of offset distribution (μ_{OS}) of each individual ICs with $V_{DD} = 0.4$ V at 25 ˚C, respectively, distinguishing the impact of within-die and inter-die variations. It was evident that σ_{OS} of the proposed R/L-HYSA-QZ remained relatively low and stable (deviated less than $|\pm 1$ mV$|$) across each of the 10 ICs. The μ_{OS} of each topologies remained within 0 mV – 6 mV across all 10 ICs.
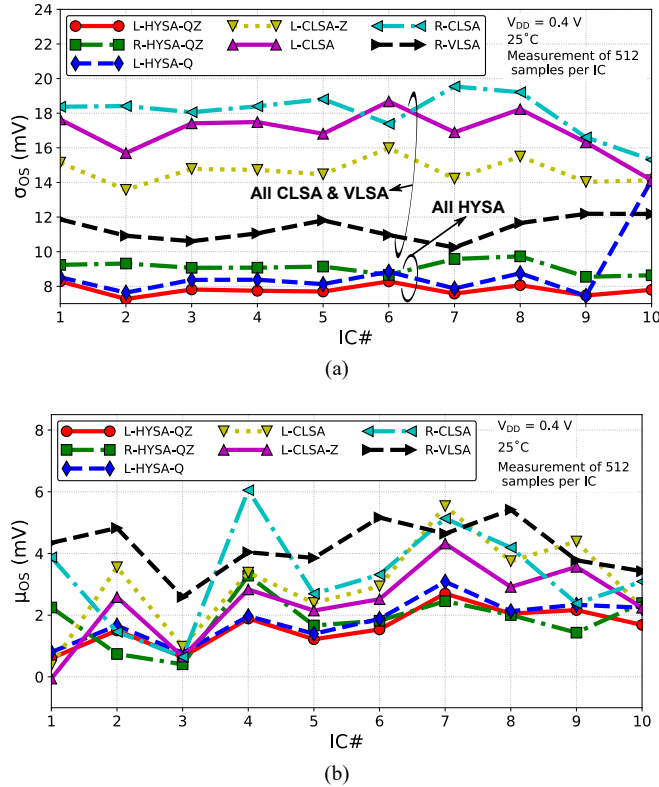

(a)

(b)

Fig. 17. Statistics of measured within-die and inter-die offset distribution (a) Measured σ_{OS} across each of the 10 ICs. (b) Measured μ_{OS} across each of the 10 ICs. (512 Samples per IC).

A similar measured statistical analysis was carried out over V_{DD} between 0.4 V to 1.0 V at 10 MHz. For each topology on each IC, worst case ΔV_{BL} (ΔV_{BL-min}) required for 100% yield (0 failures in both logic '1' and '0' from 512 SAs) was extracted. Subsequently, the average of ΔV_{BL-min} from 10 ICs, $\mu_{\Delta V_{BL-min}}$ was computed for each V_{DD}. Fig. 18 (a) summarizes these measurements, and each data point in the figure represents this value. For example, for R-HYSA-QZ, the $\mu_{\Delta V_{BL-min}}$ at

(a)

(b)

(c)
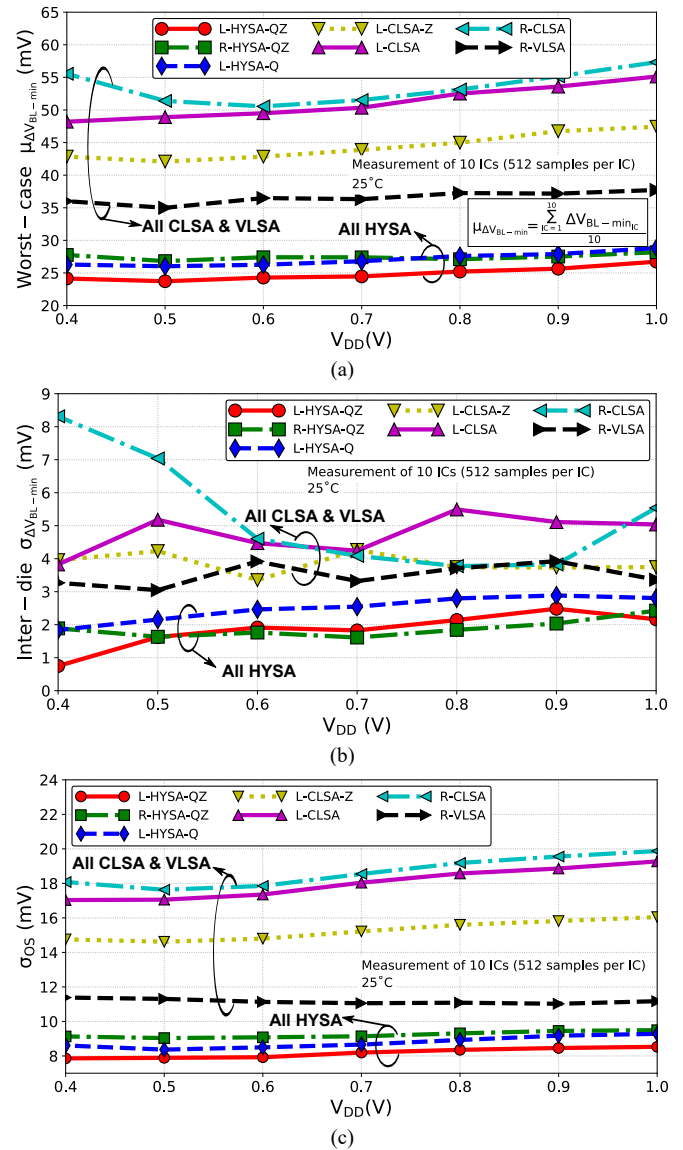
Fig. 18. (a) Measured avg. of V_{BL-min} ($\mu_{\Delta V_{BL-min}}$) across 10 ICs. (b) Measured Inter-die standard deviation of ΔV_{BL-min} ($\sigma_{\Delta V_{BL-min}}$) across 10 ICs. (c) Measured σ_{OS} across 5120 SAs (10 ICs).

0.4 V is 27.8 mV as compared to 56.0 mV and 36.0 mV of R-CLSA and R-VLSA, respectively. Based on these measured results we conclude that R-HYSA-QZ requires 50.0% and 22.8%, smaller ΔV_{BL-min} at the cost of 6.5% and 30.7% gate area overhead compared to R-CLSA and R-VLSA, respectively. Also, using L-V_T transistors in L-HYSA-QZ lowered $\mu_{\Delta V_{BL-min}}$ by 13.0% (56.5% w.r.t R-CLSA & 32.9% w.r.t. R-VLSA) compared to using R-V_T transistors in R-HYSA-QZ. To validate the inter-die consistency of ΔV_{BL-min}, standard deviation of ΔV_{BL-min} ($\sigma_{\Delta V_{BL-min}}$) across 10 ICs was computed and shown in Fig. 18 (b). Again, the proposed HYSA-QZs having relatively low $\sigma_{\Delta V_{BL-min}}$ confirms inter-die consistency in offset tolerance. Also, as shown in Fig. 18 (c), σ_{OS} were extracted across V_{DD} from their respective PDF curves to understand the improvement over gate area overhead explained in the next paragraph. ΔV_{BL-min} characteristics of the SAs were also characterized across temperature between 0 ˚C – 75 ˚C at 0.4 V at 10 MHz. As depicted in Fig. 19, measured ΔV_{BL-min}

shows that all HYSA topologies have relatively more stable and lower ΔV_{BL-min} across temperature compared to CLSA and VLSA. Overall, the measured relative trends were in good agreement with the circuit theory predictions, the offset tolerance and the MC simulations of V_T. They showed that the HYSA topologies had increased offset tolerance as more internal nodes get precharged with bitlines and increased even further by using L-V_T transistors. As was also predicted in Section IV, applying bitlines at Q/QB nodes is far more effective (~4.5x at 0.4 V) than applying bitlines at Z/ZB nodes. R-VLSA had relatively more gate area spent on its critical NMOS pair resulting in lower mismatch compared to other SAs. Therefore, despite having low offset tolerance as shown in Figs. 6-8, R-VLSA achieves lower ΔV_{BL-min} and lower σ_{OS} compared to, for example, L-CLSA-Z, where it provides only comparable offset tolerance as R-VLSA.
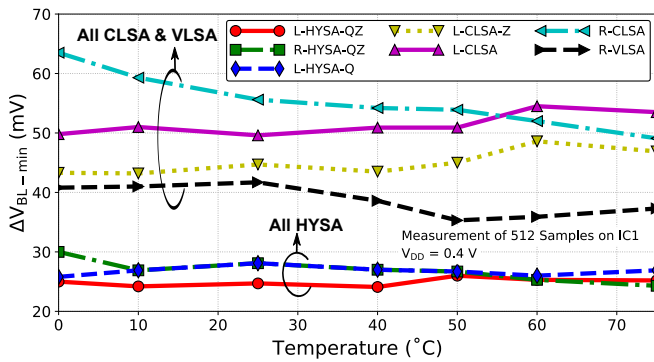

Fig. 19. Measured ΔV_{BL-min} stability across temperature range between 0 ˚C to 75 ˚C.

Due to the gate area penalty in the proposed HYSA-QZ, we analyze measured σ_{OS} from Fig. 18 (c) while substantiating iso-gate area improvement. Pileggi et. al. [10] reported that

VLSA σ_{OS} accurately follows Pelgrom's model [12] described by $\sigma(\Delta V_T) = \frac{A_{T0}}{\sqrt{WL}}$, where A_{T0}, W and L are the area proportionality constant, and width and length of two devices with close proximity, respectively. Considering this, adding 30.7% extra area to R-VLSA's NMOS pair would give maximum of $\left(1 - \frac{1}{\sqrt{1.307}}\right) \approx 12.5\%$ σ_{OS} improvement in R-VLSA whereas R-HYSA-QZ improved measured σ_{OS} by 20.2%; therefore resulting in net iso-gate-area improvement of 7.7% at 0.4 V. However, at super-threshold supply, i.e. 1 V, this improvement reduces to 2.5%. With a similar analysis for R-CLSA, adding 6.5% extra area in R-CLSA's input NMOS sensing pair (best possible case) would give maximum improvement of $\left(1 - \frac{1}{\sqrt{1.065}}\right) \approx 3.1\%$ which is far less than 49.7% measured σ_{OS} improvement from R-HYSA-QZ at 0.4 V. Moreover, comparing the proposed L-HYSA-QZ with L-HYSA-Q, where the latter topology is similar to the one proposed by [14, 15], the former achieves 8.1% σ_{OS} improvement with 6.5% total gate area penalty. This still results in 5.0% iso-gate-area improvement in σ_{OS} at 0.4 V. At 1 V, improvement in measured σ_{OS} slightly increases to 5.5%.

Fig. 20 (a)-(e) and Fig. 20 (f)-(j) shows the measured Frequency-V_{DD} and Temperature-V_{DD} shmoo plots, respectively. Conventional and intermediate topologies were compared with proposed HYSA-QZ within same V_T flavor as shown in Fig. 20 (a-d, f-i). Also, improvement achieved by changing proposed HYSA-QZ's V_T flavor from R-V_T to L-V_T is shown in Fig. 20 (e, j) shmoo plots. All shmoo plots were measured at ΔV_{BL} of ± 40 mV ($\pm$ for testing both logic '1' and '0') while changing the operating parameters, i.e. frequency, temperature and V_{DD}. For a given operating condition, worst case yield (considering both '1' and '0' yields) for a typical die (512 samples) was obtained. The error rate less than 0.8% was considered a pass, and a fail otherwise.

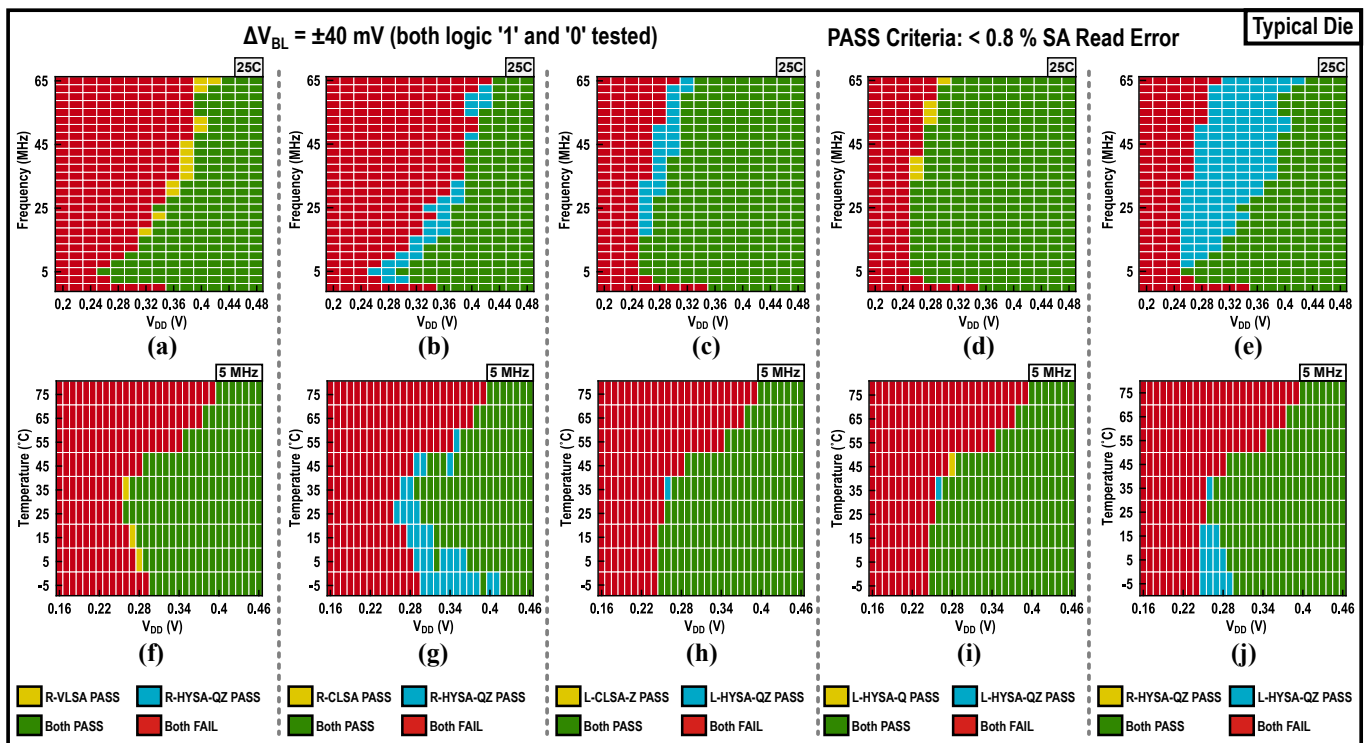Fig. 20. **Measured Frequency vs V$_{DD}$ shmoo plots (a-e): (a)** R-VLSA vs R-HYSA-QZ **(b)** R-CLSA vs R-HYSA-QZ **(c)** L-CLSA-Z vs L-HYSA-QZ **(d)** L-HYSA-Q vs L-HYSA-QZ **(e)** R-HYSA-QZ vs L-HYSA-QZ. **Measured Temperature vs V$_{DD}$ shmoo plots (f-j): (f)** R-VLSA vs R-HYSA-QZ **(g)** R-CLSA vs R-HYSA-QZ **(h)** L-CLSA-Z vs L-HYSA-QZ **(i)** L-HYSA-Q vs L-HYSA-QZ **(j)** R-HYSA-QZ vs L-HYSA-Q.

Relative trends in Frequency-V_{DD} shmoo plots reflect their relative sensing delays where a topology with lower sensing delay would be capable of operating at a relatively higher frequency under the same ΔV_{BL} and other operating conditions. Since these measured shmoo plots include the impact of mismatch, they do not perfectly resemble relative sensing delay trends from Fig. 10 (b) which were simulated in perfect matching conditions. For example, R-VLSA is capable of operating at ~1.5x−1.7x higher frequency than R-HYSA-QZ as seen in sensing delay analysis from Fig. 10 (b). However, from Fig. 20 (a), their relative difference is reduced due to higher offset tolerance in R-HYSA-QZ. In fact, below 0.32 V, R-HYSA-QZ offers identical performance as R-VLSA. Furthermore, Fig. 20 (b) shows R-HYSA-QZ performing better than conventional R-CLSA and Fig. 20 (c) shows L-HYSA-QZ further improving from intermediate topology, L-HYSA-Z; all owing to higher offset tolerance. Fig. 20 (d), however, illustrates L-HYSA-QZ operating slightly slower than L-HYSA-Q. This is because, at $\Delta V_{BL} = \pm 40$ mV, both topologies have more than enough offset tolerance to overcome the impact of mismatches (as shown in Fig. 18 (a)) and since L-HYSA-QZ has its Z/ZB nodes starting to discharge from relatively higher voltages than the floating Z/ZB in L-HYSA-Q, it marginally takes more time to resolve compared to L-HYSA-Q (also observed in Fig. 10 (b)). Finally, Fig. 20 (e) reflects predicted results from Section IV that L-HYSA-QZ is capable of operating much faster than R-HYSA-QZ due to lower μ_{V_T} and σ_{V_T}.

From the Temperature-V_{DD} shmoo plots, Fig. 20 (f)-(j), measured at 5 MHz, it is clear that R/L-HYSA-QZs have similar

or improved reliability coverage across temperature compared to conventional and intermediate topologies. Importantly, comparing R-HYSA-QZ with L-HYSA-QZ in Fig. 20 (j), latter offers improved coverage at lower temperatures and lower V_{DD}. Fig. 21 (a) summarizes the V_{DD-min} extracted from the Temperature-V_{DD} shmoo plots in Fig. 20 (f)-(j) for each topology across −5 °C, 25 °C and 75 °C. At 25 °C, R-HYSA-QZ equals R-VLSA with the V_{DD-min} of 260 mV whereas R-CLSA requires V_{DD-min} of 300 mV. Similarly, at 25 °C, L-HYSA-QZ, L-HYSA-Q and L-CLSA-Z achieved V_{DD-min} of 260 mV whereas V_{DD-min} for L-CLSA was limited to 320 mV. All topologies offered reliable operation at and above 400 mV for the temperature range between −5 °C to 75 °C (with a minor exception for R-CLSA with $V_{DD-min} = 420$ mV at −5 °C). The V_{DD-min} for most topologies was mainly dictated by higher temperature as it reduces V_T (see Fig. 21 (b)) further increasing leakage and hence deteriorating the yield. However, in general terms, the V_{DD-min} trend across temperature at a given F_{CLK} for SAs in R-V_T and L-V_T flavors is dictated by the balance between yield affected by both leakage at higher temperatures and reduced sensing delay due to increased V_T at lower temperatures (see Fig. 21 (b)).

VI. SENSE AMPLIFIERS COMPARISON

The resulting $P_{dy-read}$ with the consideration of increased P_{dy-SA} in the proposed R/L-HYSA-QZ topologies needs to be analyzed and compared with the resulting $P_{dy-read}$ with other SAs. The $P_{dy-read}$ can be quantified as shown in eq. (1) where the $P_{dy-BL/BLB}$ is the power consumption due to the bitline discharge defined by eq. (2) and the P_{dy-SA} is the SA dynamic power analyzed in Section IV and also defined by eq. (3). Note that other dynamic power components of the read access such as word-line drivers and control/timing are ignored in this comparative analysis as they would remain constant regardless of the choice of an SA. As emphasized by the work of [6], the amount of ΔV_{BL} discharge in eq. (4) is determined by the combination of σ_{OS} of a given SA and the ζ factor defined in eq. (5). The ζ factor is indicated by the amount of ΔV_{BL-min} required per σ_{OS} to meet certain Y_{read} target for particular SRAM size in a given technology. For example, in the work of [6], for 28-nm CMOS and 16 Mb SRAM, the ζ factor was found to be ~10mV/mV for 97 % yield target.

Since the $P_{dy-read}$ improvement is a function of bitline capacitance (C_{BL}), the $P_{dy-read}$ improvement is analyzed across C_{BL} comparing R-HYSA-QZ with R-VLSA and R-CLSA, and comparing L-HYSA-QZ with R-HYSA-QZ, L-HYSA-Q, L-HYSA-Z and L-CLSA. For the calculation of $P_{dy-read}$, $V_{DD} = 0.4$ V, $F_{CLK} = 3.33$ MHz, measured σ_{OS} from Fig. 18 (c) and simulated values of P_{dy-SA} with appropriate ΔV_{BL-min} calculated from eq. (4) were used. For example, for the ζ factor of 8 mV/mV, Fig. 22 (a) shows the predicted $P_{dy-read}$ improvement with proposed R/L-HYSA-QZ. It shows the $P_{dy-read}$ improvement across different values of C_{BL} where the 0% crossover point is annotated to highlight the minimum C_{BL} requirement for $P_{dy-read}$ improvement > 0%. At C_{BL} of 150 fF, R-HYSA-QZ results in $P_{dy-read}$ improvement of 11% and 45% compared to R-VLSA and R-CLSA, respectively. Also, L-HYSA-QZ results in $P_{dy-read}$ improvement of 5%, 3.5%, 38%, and 46% compared to R-HYSA-QZ, L-HYSA-Q, L-CLSA-Z and L-CLSA, respectively. The ζ factor is subject to vary depending on the SRAM memory size, target Y_{read} and the
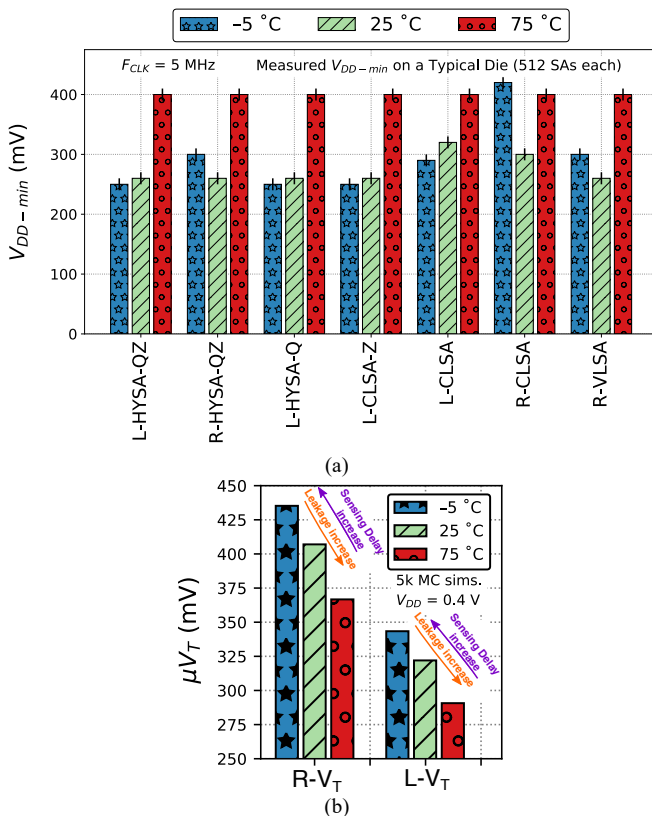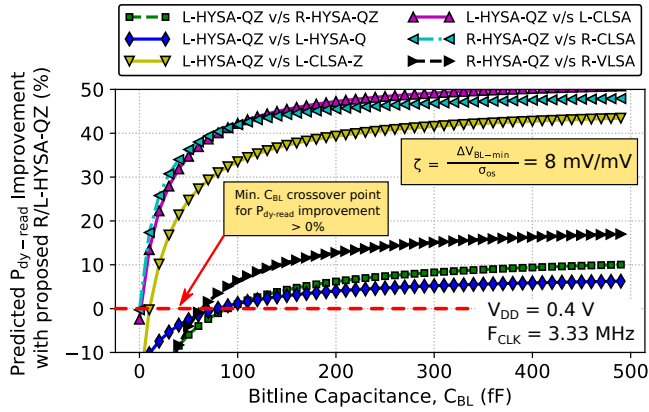

Fig. 21. (a) V_{DD-min} at $F_{CLK} = 5$ MHz from the shmoo plots in Fig. 20. (Extracted from the measured shmoo plots.) (b) 5k MC simulations of μ_{V_T} (average threshold) across temperature for critical NMOS Transistor in R-V_T and L-V_T flavors.
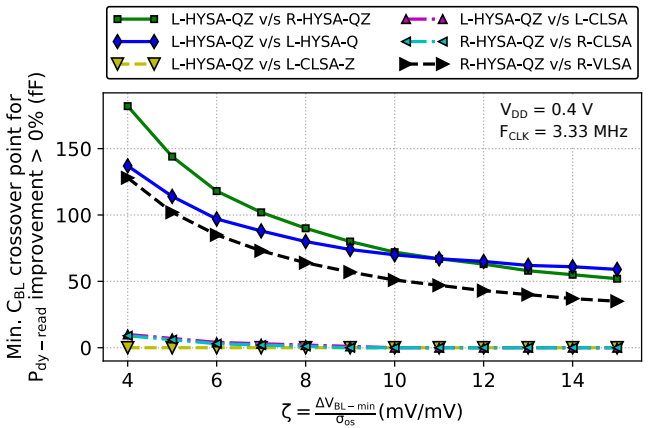
technology related effects such as proximity, stress effects, V_T variations, and active area rounding [6]. Therefore, the minimum C_{BL} crossover points were recorded as shown in Fig. 22 (b) for varied values of ζ between 4 mV/mV to 15 mV/mV. The values for minimum C_{BL} crossover points reduces with the increased values of ζ, and for the lower values of ζ, minimum C_{BL} crossover points are within typical C_{BL} range for an SRAM macro. This assures $P_{dy-read}$ savings with proposed R/L-HYSA-QZ topologies around threshold region. This analysis also highlights that utilizing L-V_T flavor transistors in proposed HYSA-QZ topology ultimately provides

(a)

(b)

(c)

Fig. 22. (a) Predicted $P_{dy-read}$ improvement with proposed R/L-HYSA-QZ at $V_{DD} = 0.4$ V. (b) Minimum bitline capacitance (C_{BL}) crossover point for $P_{dy-read}$ improvement greater than 0% across different values of ζ factor. (c) Predicted $P_{dy-read}$ improvement with proposed R/L-HYSA-QZ at $V_{DD} = 1$ V.

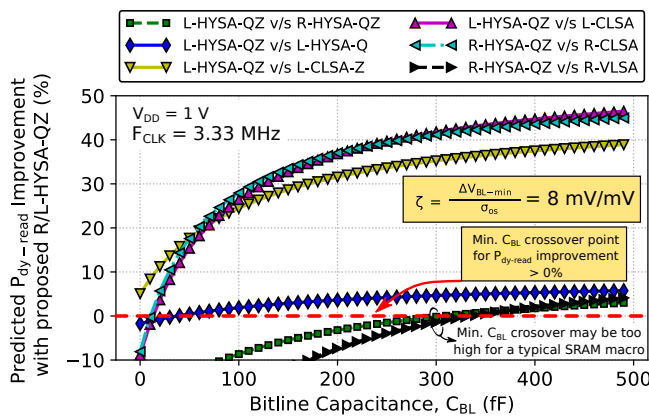$P_{dy-read}$ savings compared to using R-V_T transistors around threshold region.

To further see the impact of the proposed R/L-HYSA-QZ with nominal supply in superthreshold region of operation, the analysis from Fig. 22 (a) was repeated with $V_{DD} = 1$ V and $\zeta = 8$ mV/mV, and is shown in Fig. 22 (c). From Fig. 22 (c) it was evident that the R/L-HYSA-QZ can still provide savings in $P_{dy-read}$ compared to other topologies within typical C_{BL} range for an SRAM macro except R-VLSA (black line). R-HYSA-QZ can only provide $P_{dy-read}$ improvement over R-VLSA at 1 V and $\zeta = 8$ mV/mV if C_{BL} is > 350 fF which is relatively high for a typical SRAM macro. Hence, in such situations, R-VLSA is preferred over the proposed R-HYSA-QZ. Similarly, the green line highlights that in such situations, R-HYSA-QZ is preferred over L-HYSA-QZ. In general, with superthreshold supply voltages, the choice of an SA remains debatable between R-VLSA, R-HYSA-QZ, and L-HYSA-QZ depending on the technology and C_{BL}.

$$P_{dy-read} = P_{dy-BL/BLB} + P_{dy-SA} \qquad (1)$$

$$P_{dy-BL/BLB} = C_{BL} \cdot V_{DD} \cdot \Delta V_{BL-min} \cdot F_{CLK} \qquad (2)$$

$$P_{dy-SA} = P_{dy-SA-V_{DD}} + P_{dy-SA-BL/BLB} + P_{dy-SA-SAE} \qquad (3)$$

$$\Delta V_{BL-min} = \zeta \cdot \sigma_{OS} \qquad (4)$$

$$\zeta = \frac{\Delta V_{BL-min}}{\sigma_{OS}} \left(\frac{mV}{mV}\right) \qquad (5)$$

The state-of-the-art on-chip fabricated SA topologies are compared with the proposed R/L-HYSA-QZs in Table II. The proposed SA in [19] relies on small-signal pre-amplification, offering 22% iso-layout-area improvement in ΔV_{BL-min} compared to VLSA and V_{DD-min} of 450 mV. The work from [22] deployed with small-signal preamplification circuit and fine-tuned MOM capacitor-based offset mitigation lowers the bit-error-rate (BER) from 12% to 0% with two MOM capacitors area penalty in 2 metal layers compared to VLSA. The offset compensation technique in [23] relies on MOS capacitor-based threshold matching whereas [24] relies on body biasing-based calibration on power-up. SAs in both [23] and [24] achieve ~50% offset improvement with over ~3% overall SRAM layout area penalty compared to VLSA and CLSA, respectively. Our recent work in [25] employed with large signal differential and common mode boosting also implemented in 65 nm-GP CMOS achieves 23% offset improvement compared to R-VLSA at 0.3 V with 12% layout area overhead. The offset improvement in [25] is sensitive to the accuracy of its multi-phase timing and the supply voltage where the boosting benefits diminishes as the supply voltage is increased; as a result, optimum offset improvement was only within narrow supply range (0.3 V – 0.6 V). The advantages of simply precharging internal nodes with bitlines in HYSA-QZ are simpler timing and relatively stable offset improvement over wider supply range (0.4 V – 1 V) as evident in Fig. 18 (c). The on-chip work in [16], similar to L-HYSA-Q, brings down BER from 2.8% to 1% with the layout area penalty of 16.8% compared to CLSA while offering record low subthreshold operation at 140 mV. The SA in [16] was implemented in low-leakage and Low-Power (LP) technology whereas this work was implemented in regular General-Purpose (GP) technology where the leakage is relatively higher due to significant degradation in I_{ON}/I_{OFF} ratio, ultimately constraining the data reproducibility in this work.

TABLE II.   COMPARISON WITH THE STATE-OF-THE-ART OFFSET TOLERANT SENSE AMPLIFIERS

Designs in CMOS	[19] A-SSCC 2016	[22] ISSCC 2014	[23] JSSC 2016	[24] JSSC 2014	[25] LSSC 2018	[16] JSSC 2017	This work R-HYSA-QZ		This work L-HYSA-QZ	
Technology	28 nm-HPM	28 nm	28 nm-HP	65 nm-LP	65 nm-GP	65 nm-LP	65 nm-GP			
Measured Samples	432 (6 ICs)	352 (22 ICs)	2080 (1 IC)	512 (1 IC)	8192 (16 ICs)	16 (1 IC)	5120 (10 ICs)			
Latch Type Sensing Topology	Self-timed small-signal pre-amp. and latching	Multi-phase small-signal pre-amp., & MOM Cap. based offset mitigation	Multi-phase MOS cap. based threshold matching	CLSA with Body-Bias offset calibration on power-up	Large signal differential and comnon mode boost, and using bitlines as SA supply	Precharging SA outputs with bitlines	Precharging multiple internal nodes with bitlines			
# of Devices	15T	10T + 2 MOM Caps	11T+ 2 MOS Caps + 5 INV	15T + 2 NOR + 2 NAND + 3 INV + 1 Latch	11T + 2 MOS cap	9T	11T			
Design Effort Overhead	None	MOM Cap. design + Complex Timing	None	Body Bias + Calibration	Complex Timing	None	None			
SA Area Overhead [a.]	0% (w.r.t VLSA)	0% + 2 MOM Caps (w.r.t VLSA) [b.]	3.2% overhead in overall 128 kb SRAM (w.r.t VLSA)	3.5% overhead in overall 128 kb SRAM (w.r.t CLSA)	12% (w.r.t VLSA)	16.8% (w.r.t CLSA)	6.5% [c.] 4.5% [d.] w.r.t R-CLSA	30.7% [c.] 18.8% [d.] w.r.t R-VLSA	6.5% [c.] 4.5% [d.] w.r.t R-CLSA	30.7% [c.] 18.8% [d.] w.r.t R-VLSA
Offset Improvement	22% [a.] (in simulated $\Delta V_{BL\text{-}min}$) @ 0.45 V, 25 °C	BER improvement from 12% to 0% [a.] @ 1 V, 27 °C	49% [a.] @ 0.5 V, 85 °C	50% [a., e.] @ 1.2 V	23.3% in Std. of offset @ 0.3 V, 25 °C	BER improvement from 2.8% to 1.1% [a.] @ 0.14 V, 25 °C	50% [e.] 46.6% [f.]	22.8% [e.] 7.6% [f.]	56.5% [e.] 53.3% [f.]	32.9% [e.] 18.1% [f.]
							@ 0.4 V, 25 °C		@ 0.4 V, 25 °C	
$V_{DD\text{-}min}$ @ 25 °C	450 mV	Only tested @ 1V	500 mV	370 mV	230 mV	140 mV	260 mV		260 mV	

a. Compared to conventional topology implemented in respective work
b. Area overhead in 2 Metal layers only (not in active region)
c. Total gate area overhead
d. Layout area overhead
e. Improvement in measured $\Delta V_{BL\text{-}min}$
f. Iso-gate area improvement in measured σ_{OS}

The proposed R/L-HYSA-QZ topology is simple to implement without any timing or layout complexity and can easily be scaled with technology. Also, it does not depend on the body-biasing effect or capacitor non-linearity. Thus, conventional SAs, VLSA and CLSA can simply be replaced by the R/L-HYSA-QZ while offering reliable and higher offset tolerance around threshold region.

VII. CONCLUSION

The offset tolerance in proposed HYbrid Latch-Type Sense Amplifier, HYSA-QZ for low-voltage SRAMs is significantly increased by pre-charging multiple internal nodes with bitline signals. Along with HYSA-QZ, two intermediate formulations of HYSA-QZ (CLSA-Z and HYSA-Q), and conventional CLSA and VLSA topologies were also analyzed and implemented in 65nm-GP CMOS test chips. Offset measurements were performed across 5120 SAs from 10 ICs. It showed that HYSA-QZ achieves 50.0% and 22.8% lower $\Delta V_{BL\text{-}min}$ at 0.4 V with 6.5% (or 4.5%) and 30.7% (or 18.8%) total gate (or layout) area overhead compared to R-CLSA and R-VLSA implemented with R-V_T transistors, respectively. Reported offset improvement with respective gate area penalty is justified by Pelgrom's model where R-HYSA-QZ achieves 46.6% and 7.7% iso-gate-area improvement in σ_{OS} compared to R-CLSA and R-VLSA, respectively. Moreover, measured shmoo plots showed that R-HYSA-QZ, R-CLSA and R-VLSA required $V_{DD\text{-}min}$ of 0.26 V, 0.30 V and 0.26 V at 25°C, respectively. Utilizing L-V_T transistors in L-HYSA-QZ further lowered its $\Delta V_{BL\text{-}min}$ by 13.0% and improved its operational coverage across temperature and frequency at 0.4 V.

Finally, the overall SRAM $P_{dy\text{-}read}$ benefit was analyzed with measured V_{OS} statistics and resulting $P_{dy\text{-}SA}$ of all SAs. It suggested that offset improvement in proposed HYSA-QZ ultimately leads to an overall SRAM $P_{dy\text{-}read}$ improvement compared to other analyzed SAs around threshold region. For the nominal supply operations, the choice between R-VLSA, R-HYSA-QZ and L-HYSA-QZ is subject to targeted C_{BL} of an SRAM macro and technology.

REFERENCES

[1] K. Zhang, K. Hose, V. De and B. Senyk, "The scaling of data sensing schemes for high speed cache design in sub-0.18 /spl mu/m technologies," *2000 Symposium on VLSI Circuits. Digest of Technical Papers (Cat. No.00CH37103)*, Honolulu, HI, USA, 2000, pp. 226-227.

[2] R. Houle, "Simple Statistical Analysis Techniques to Determine Minimum Sense Amp Set Times," *2007 IEEE Custom Integrated Circuits Conference*, San Jose, CA, 2007, pp. 37-40.

[3] B. Mohammad, P. Dadabhoy, K. Lin and P. Bassett, "Comparative study of current mode and voltage mode sense amplifier used for 28nm SRAM," *2012 24th International Conference on Microelectronics (ICM)*, Algiers, Algeria, 2012, pp. 1-6.

[4] B. Wicht, T. Nirschl and D. Schmitt-Landsiedel, "Yield and speed optimization of a latch-type voltage sense amplifier," in *IEEE Journal of Solid-State Circuits*, vol. 39, no. 7, pp. 1148-1158, July 2004.

[5] A. Bhavnagarwala *et al.*, "Fluctuation limits & scaling opportunities for CMOS SRAM cells," *IEEE InternationalElectron Devices Meeting, 2005. IEDM Technical Digest.*, Washington, DC, 2005, pp. 659-662.

[6] M. H. Abu-Rahma *et al.*, "Characterization of SRAM sense amplifier input offset for yield prediction in 28nm CMOS," *2011 IEEE Custom Integrated Circuits Conference (CICC)*, San Jose, CA, 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," in *IEEE Journal of Solid-State Circuits*, vol. 35, no. 8, pp. 1200-1204, Aug. 2000.

[8] D. G. Laurent, "Sense amplifier signal margins and process sensitivities [DRAM]," in *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 3, pp. 269-275, Mar 2002.

[9] J. S. Shah, D. Nairn and M. Sachdev, "An Energy-Efficient Offset-Cancelling Sense Amplifier," in *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 60, no. 8, pp. 477-481, Aug. 2013.

[10] L. Pileggi, G. Keskin, X. Li, Ken Mai and J. Proesel, "Mismatch analysis and statistical design at 65 nm and below," *2008 IEEE Custom Integrated Circuits Conference*, San Jose, CA, 2008, pp. 9-12.

[11] J. F. Ryan and B. H. Calhoun, "Minimizing Offset for Latching Voltage-Mode Sense Amplifiers for Sub-Threshold Operation," *9th International Symposium on Quality Electronic Design (isqed 2008)*, San Jose, CA, 2008, pp. 127-132.

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

[13] S. H. Dhong *et al*., "A 4.8GHz fully pipelined embedded SRAM in the streaming processor of a CELL processor," *ISSCC. 2005 IEEE International Digest of Technical Papers. Solid-State Circuits Conference, 2005.*, San Francisco, CA, 2005, pp. 486-612 Vol. 1.

[14] Chen Nan, and Ritu Chaba. "Dual Sensing Current Latched Sense Amplifier." U.S. Patent 12 731 623, Sept. 29, 2011.

[15] M. Tsai, J. Tsai, M. Fan, P. Su and C. Chuang, "Variation tolerant CLSAs for nanoscale Bulk-CMOS and FinFET SRAM," *2012 IEEE Asia Pacific Conference on Circuits and Systems*, Kaohsiung, 2012, pp. 471-474.

[16] K. Sarfraz, J. He and M. Chan, "A 140-mV Variation-Tolerant Deep Sub-Threshold SRAM in 65-nm CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 52, no. 8, pp. 2215-2220, Aug. 2017.

[17] B. Razavi, "The StrongARM Latch [A Circuit for All Seasons]," in *IEEE Solid-State Circuits Magazine*, vol. 7, no. 2, pp. 12-17, Spring 2015.

[18] A. Abidi and H. Xu, "Understanding the regenerative comparator circuit," *Proceedings of the IEEE 2014 Custom Integrated Circuits Conference*, San Jose, CA, 2014, pp. 1-8.

[19] P. F. Chiu, B. Zimmer and B. Nikolić, "A double-tail sense amplifier for low-voltage SRAM in 28nm technology," *2016 IEEE Asian Solid-State Circuits Conference (A-SSCC)*, Toyama, 2016, pp. 181-184.

[20] M. E. Sinangil and A. P. Chandrakasan, "Application-Specific SRAM Design Using Output Prediction to Reduce Bit-Line Switching Activity and Statistically Gated Sense Amplifiers for Up to 1.9x Lower Energy/Access," in *IEEE Journal of Solid-State Circuits*, vol. 49, no. 1, pp. 107-117, Jan. 2014.

[21] Xiaoyao Liang, Kerem Turgay and D. Brooks, "Architectural power models for sram and cam structures based on hybrid analytical/empirical techniques," *2007 IEEE/ACM International Conference on Computer-Aided Design*, San Jose, CA, 2007, pp. 824-830.

[22] 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," *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, San Francisco, CA, 2014, pp. 242-243.

[23] M. E. Sinangil *et al*., "A 28 nm 2 Mbit 6 T SRAM With Highly Configurable Low-Voltage Write-Ability Assist Implementation and Capacitor-Based Sense-Amplifier Input Offset Compensation," in *IEEE Journal of Solid-State Circuits*, vol. 51, no. 2, pp. 557-567, Feb. 2016.

[24] Y. Sinangil and A. P. Chandrakasan, "A 128 Kbit SRAM With an Embedded Energy Monitoring Circuit and Sense-Amplifier Offset Compensation Using Body Biasing," in *IEEE Journal of Solid-State Circuits*, vol. 49, no. 11, pp. 2730-2739, Nov. 2014.

[25] D. Patel and M. Sachdev, "0.23-V Sample-Boost-Latch-Based Offset Tolerant Sense Amplifier," in *IEEE Solid-State Circuits Letters*, vol. 1, no. 1, pp. 6-9, Jan. 2018.

Dhruv Patel (M'12) received BASc. degree (Co-op, Hons.) in Electrical Engineering from the University of Waterloo in 2016 and is currently pursuing MASc. degree at the University of Toronto in the area of mixed-signal integrated circuits for Optical links. He previously held internships at Apple, Arista, BlackBerry and Christie Digital. During his undergraduate degree, he was involved with variation tolerant sub-threshold SRAM circuits research and held two NSERC undergraduate student research assistantship awards. He received Queen Elizabeth II Graduate Scholarship and Edward Rogers Sr. Graduate Scholarship during 2016 – 2018 for his MASc. degree at the University of Toronto.

Adam Neale (M'08) received the BASc., MASc. and Ph.D. degrees in Electrical and Computer Engineering from the University of Waterloo in 2008, 2010 and 2015 respectively. He is currently working as a Quality and Reliability Engineer with Intel Corporation in Hillsboro, Oregon where he is responsible for developing radiation-induced upset reliability models. Dr. Neale was the recipient of the Best Poster Paper Award at the 2014 IEEE Custom Integrated Circuits Conference (CICC) for his work on multiple adjacent-bit upset error correcting codes for SRAMs.

Derek Wright (M'09) is a faculty member in the Electrical and Computer Engineering department at the University of Waterloo. His research interests include low-power digital circuit design, multidomain modeling and simulation, and biomedical devices, circuits and systems. He received the BASc. and MASc. degrees in Electrical and Computer Engineering from the University of Waterloo in 2003 and 2005, and the PhD degree in collaborative Electrical and Biomedical Engineering from the University of Toronto in 2010.

Manoj Sachdev (M'85-SM'97-F'12) received his B.E. degree (with Honors) in Electronics and Communication Engineering from IIT Roorkee (India), and Ph.D. from Brunel University (UK). He was with Semiconductor Complex Limited, Chandigarh (India) from 1984 till 1989 where he designed CMOS Integrated Circuits. From 1989 till 1992, he worked in the ASIC division of SGS-Thomson at Agrate (Italy). In 1992, he joined Philips Research Laboratories, Eindhoven (The Netherlands), where he researched on various aspects of VLSI testing and manufacturing.

Dr. Sachdev is a professor in the Electrical and Computer Engineering department, at the University of Waterloo, Canada. His research interests include low-power and high-performance digital circuit design, mixed-signal circuit design, test and manufacturing issues of integrated circuits. He has contributed to five books, two book chapters, and has co-authored over 200 technical articles in conferences and journals. He holds more than 30 granted and several pending US patents on various aspects of VLSI circuit design, reliability, and test. He, his students, and his colleagues have received several international awards. In 1997, at the IEEE European Design and Test Conference, he received the best paper award. In 1998, he was a co-recipient of the honorable mentioned award in the IEEE International Test Conference. He received the best panel award in 2004 IEEE VLSI Test Symposium. In 2011, he was a co-recipient of the best paper award in IEEE International Symposium on Quality Electronics Design. In 2015, he was a co-recipient of the best poster award in IEEE Custom Integrated Circuits Conference.

Professor Sachdev is an IEEE Fellow, Fellow of Engineering Institute of Canada, Fellow of Canadian Academy of Engineering; and has served on several conference committees. In 1999, he was the Technical Program Co-Chair for IEEE Mixed-signal Test Workshop. He also served as the Technical Program Chair from IEEE IDDQ Test Workshop in 1999, and 2000. In 2007-08, he was an Associate Editor for IEEE Transactions on Vehicular Technology. In 2008 and 2009, he was the Program Chair for Microsystems and Nanoelectronics Research Conference, Ottawa, Canada. He serves on the editorial board of the Journal of Electronic Testing: Theory and Applications. He was a member of technical program committee of IEEE Custom Integrated Circuits Conference from 2008 till 2014, and since 2016 he is a program committee member of IEEE Design and Test in Europe (DATE) conference.