Investigating Input Offset Reduction with Timing Manipulation in Low Voltage Sense Amplifiers

by

Matthew Joseph Di Nardo

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2021

© Matthew Joseph Di Nardo 2021

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Static Random Access Memories (SRAMs) are ubiquitous in modern computer systems. They provide a fast and relatively compact method of data storage. SRAM cells are read from and written to using analog differential bitline signals, BL and BLB. To increase operating speed and conserve power during a read cycle, cell access time is limited to a short duration. Since SRAM are often implemented with near-minimum sized devices to maximize memory density, the devices are relatively weak and can only generate a limited differential voltage during this read window, typically between 10mV to 100mV. Standard logic devices cannot read this small signal, so sense amplifiers are used to rapidly amplify it to logic levels.

A key metric for a sense amplifier's performance is its input-referred offset voltage, V_{os}. This dictates the minimum required input voltage to produce a correct decision. A lower V_{os} means that a shorter read window for the SRAM is required, and the overall read cycle can be performed at a higher frequency. Unfortunately, with the trends of technology scaling, the effects of device mismatches from process variation are becoming more significant. In sense amplifiers, this device mismatch will create a statistical spread of V_{os} with a mean and standard deviation of μ_{os} and σ_{os}. To guarantee error-free operation, a lower bound for input differential voltage is set by the worst-case scenario from this spread. Another difficulty introduced with modern trends is low voltage operation. The drive strengths of devices in lower VDD systems are weaker, so any imbalances due to threshold mismatch can become more significant compared to the nominal quantities.

This thesis explores methods of reducing input offset voltage of low voltage SRAM sense amplifiers with a primary goal of reducing σ_{os}. A circuit called the Delayed PMOS VLSA, or DVLSA, is proposed. The DVLSA is based on the common VLSA and uses a timing manipulation technique with its control signals. The circuit design attempts to reduce σ_{os} by reducing the mismatch contribution of the PMOS pull-up pair.

The circuit is tested at 0.4V with the VLSA used as a reference. Statistical simulations show that for the PMOS pull-up pair varying in isolation, the circuit works as intended and σ_{os} is reduced. When all differential devices are varied, the DVLSA has a larger σ_{os}. Investigating the source of the failure using the isolated variation of the other two device pairs shows that the timing manipulation technique has a negative impact on the NMOS pair. It also suggests that the use of the DLVSA architecture introduces additional covariances when all differential devices are varied.

Acknowledgements

I would like to thank my supervisors Dr. Manoj Sachdev and Dr. Derek Wright for their guidance and support throughout my program. To my readers, Dr. John Long and Dr. David Narin, thank you for offering your time. Thanks to Hugo, Qing, and Govind for their mentorship on chip design, layout, and using the Cadence tools.

To Hugo, thanks for being an amazing friend and helping to proofread this thesis.

To my family, I could not have done this without you. Thank you from the bottom of my heart for being there and supporting me all these years.

Table of Contents

AUTHOR'S DECLARATION ii
Abstract iii
Acknowledgements iv
List of Figures vii
List of Tables ix
Chapter 1 Introduction 1
1.1 Motivation 1
1.2 Topics to be Explored 3
1.3 Thesis Organization 3
Chapter 2 Background 4
2.1 SRAM Architecture and Operation 4
2.1.1 Architecture 4
2.1.2 6T SRAM Operation 5
2.2 MOSFET Vt Model 7
2.2.1 Short Channel Effects and Drain-Induced Barrier Lowering 8
2.2.2 Mismatch Model 8
2.3 Cross-Coupled Pairs 9
2.3.1 Introduction and Derivation 9
2.3.2 Effect of Mismatch on The Cross-coupled Pair 10
2.3.3 Cross-coupled Pairs in Digital Circuits 10
2.4 Latch Type Sense Amplifiers 11
2.5 Latch Type Sense Amplifier Vos Yield Model 13
2.6 Review of Offset Reduction Methods 15
2.6.1 Self Calibration / Auto Zeroing 15
2.6.2 Vt Mismatch Cancellation and Compensation 16
2.6.3 Multiple Injection 18
2.6.4 Input Boosting / Multiplication 19
2.6.5 Timing Manipulation 20
2.7 VDD Dependence of Device Mismatch Contribution in VLSA 21

Chapter 3 Methodology 23
3.1 Hypothesis 23
3.2 Proposed Circuit – DVLSA 24
3.2.1 Concept 24
3.2.2 Device Sizing 26
3.3 Reference Circuit – VLSA 27
3.4 Testbench Architecture 27
3.4.1 Sense Amp Selection Cell (SA_SEQ_SWEEP) 28
3.4.2 Sense Amplifier Cores 29
3.4.3 Timing Blocks 31
3.4.4 Output Decision Block 32
3.4.5 Fixed ΔVBL Testbench 33
3.4.6 Successive Approximation of ΔVBL Testbench 34
3.5 Tests 36
3.5.1 Nominal Waveform Comparison 36
3.5.2 Statistical Input Offset – Isolated PMOS Pair 37
3.5.3 Statistical Input Offset – All PMOS and NMOS Symmetric Pairs 38
3.6 Test Conditions 39
Chapter 4 Simulation Results and Analysis 40
4.1 Nominal Waveform Comparison 40
4.1.1 Large input Voltage ($\Delta VBL = 40mV$) 40
4.1.2 Small Input Voltage ($\Delta VBL = 10mV$) 43
4.1.3 Summary 44
4.2 Statistical Input Offset – Isolated PMOS Pair 45
4.3 Statistical Input Offset – All PMOS and NMOS Symmetric Pairs 49
4.3.1 Further Investigation into Source of Failure 54
Chapter 5 Conclusions and Future Research 56
References 57
Appendix A Testbench Verilog-A Code 60

List of Figures

Figure 1.1 Die Photo of an Ultra-Low Power 0.4V SoC for Biomedical Wireless Sensor Nodes [2] ... 1
Figure 1.2 Comparison of Input Offset Voltage of a Sample Sense Amplifier for Multiple Technology Nodes and Different Supply Voltages [3] ... 3
Figure 2.1 Example SRAM Bank Architecture [4] ... 4
Figure 2.2 Conventional 6T SRAM Cell with NMOS Access Transistors and Bitline Capacitances.... 5
Figure 2.3 SRAM Cell Read and Write Cycles ... 6
Figure 2.4 SRAM time for $\Delta VBL = 50mV$ development vs VDD ... 7
Figure 2.5 NMOS Cross-Coupled Pair (a), Ideal Small-Signal Model (b), Differential Mode Small Signal Model (c) ... 9
Figure 2.6 NMOS Cross-coupled Pair Small Signal Model with Threshold Mismatch (a), Differential Mode Small Signal Model (b) ... 10
Figure 2.7 Voltage Latch Sense Amplifier (VLSA) ... 12
Figure 2.8 Current Latch Sense Amplifier (CLSA) ... 13
Figure 2.9 Plot of Required ΔVin Normalized to σos for a Target Yield of a Sense Amplifier ... 14
Figure 2.10 Self Calibrating Dynamic Comparator [13] ... 15
Figure 2.11 Offset Cancelling Sense Amplifier Example [14] ... 16
Figure 2.12 Capacitor Based Compensation for CLSA [15] ... 17
Figure 2.13 HYSA-QZ Schematic [16] ... 18
Figure 2.14 SBLSA [17] ... 19
Figure 2.15 Intrinsic and Extrinsic Offset of VLSA vs SAE Rise Time [18] ... 20
Figure 3.1 DVLSA Schematic (a) and Conceptual Waveforms (b) ... 24
Figure 3.2 NMOS Vt Extraction vs VDS ... 26
Figure 3.3 DVLSA Implementation for SA_SEQ_SWEEP ... 28
Figure 3.4 VLSA Sense Amplifier Core Implementation ... 29
Figure 3.5 DVLSA Sense Amplifier Core Implementation ... 30
Figure 3.6 DVLSA Timing Cell ... 31
Figure 3.7 Output Decision Block ... 32
Figure 3.8 Fixed ΔVBL Testbench ... 33
Figure 3.9 VBL and $VBLB$ Setup vs ΔVBL (VDD=400mV) ... 33
Figure 3.10 Successive Approximation of VOS Testbench ... 34

Figure 3.11 Example Waveforms for SAR Testbench 35
Figure 3.12 VLSA Nominal Waveforms ($\Delta VBL = 40mV$) 36
Figure 3.13 VLSA Input Offset Voltage Histogram – PMOS Pair Only 37
Figure 3.14 VLSA Input Offset Voltage Histogram – PMOS and NMOS Pairs 38
Figure 4.1 Nominal Comparison of VSLA vs DVLSA ($\Delta VBL = 40mV,$ SAEB_delay=2.0ns) 40
Figure 4.2 Nominal Comparison of Currents and Operating Regions ($\Delta VBL = 40mV$, SAEB_delay=2.0ns) 41
Figure 4.3 Nominal Comparison of VSLA vs DVLSA ($\Delta VBL = 10mV$, SAEB_delay=2.0ns) 43
Figure 4.4 Nominal Comparison of Currents and Operating Regions ($\Delta VBL = 10mV$, SAEB_delay=2.0ns) 43
Figure 4.5 DVLSA Input Offset Voltage (SAEB_delay=0.0ns) Histogram – PMOS Pair Only 46
Figure 4.6 DVLSA Input Offset Voltage (SAEB_delay=0.5ns) Histogram – PMOS Pair Only 46
Figure 4.7 DVLSA Input Offset Voltage (SAEB_delay=1.0ns) Histogram – PMOS Pair Only 47
Figure 4.8 DVLSA Input Offset Voltage (SAEB_delay=2.0ns) Histogram – PMOS Pair Only 48
Figure 4.9 DVLSA 0.4V Input Offset Voltage (SAEB_delay=0.0ns) 49
Figure 4.10 DVLSA 0.4V Input Offset Voltage (SAEB_delay=0.5ns) 50
Figure 4.11 DVLSA 0.4V Input Offset Voltage (SAEB_delay=1.0ns) 51
Figure 4.12 DVLSA 0.4V Input Offset Voltage (SAEB_delay=2.0ns) 52

List of Tables

Table 2.1 Offset Variance Contribution of NMOS and PMOS pairs for standard VLSA vs VDD 22
Table 3.1 DVLSA Transistor Sizes 27
Table 3.2 Testbench Default Parameters 39
Table 4.1 Input Offset Voltage Statistics - Isolated PMOS Pair (P1/P2) 48
Table 4.2 DVLSA Mismatch Contribution Summary 53
Table 4.3 Input Offset Voltage Statistics - Isolated PMOS Access Pair (P3/P4) 54
Table 4.4 Input Offset Voltage Statistics - Isolated NMOS Pair (N1/N2) 54

Chapter 1
Introduction

1.1 Motivation

Systems on a Chip (SoC) have been growing in popularity, especially in mobile and IoT applications. They allow us to integrate most of what a system needs to function into a single package on one or multiple closely spaced Integrated Circuit (IC) dice, which shortens the distance for signals to travel. Traditionally, one of the biggest bottlenecks to overall system performance is memory access latency due to the use of off-chip memory. With SoCs, it is becoming very common to embed memory structures directly on-die, and they will often take up more than half of the total die area [1]. As an example, Figure 1.1 shows a die photo of a low power SoC for biomedical applications, where most of the die area is taken up by several blocks of memory that are custom designed for specific tasks.

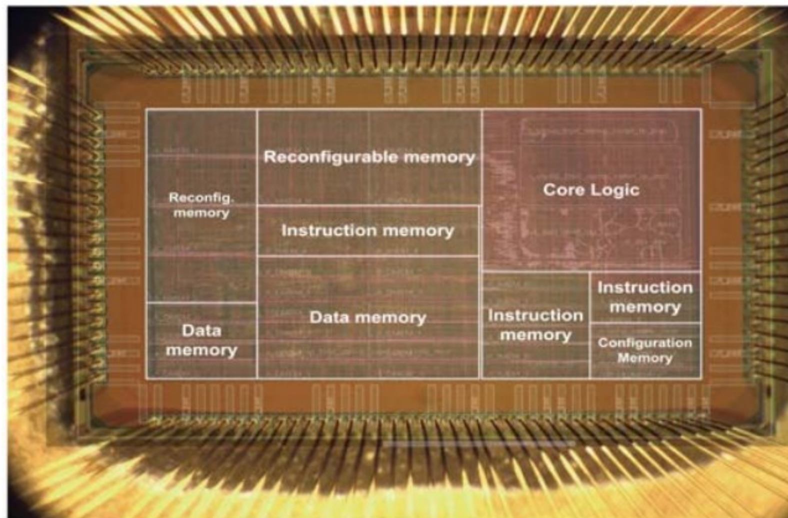

Figure 1.1 Die Photo of an Ultra-Low Power 0.4V SoC for Biomedical Wireless Sensor Nodes [2]

One of the most popular technologies used to implement these memories is Static Random Access Memory (SRAM). In SRAM, the slowest operation to perform is a bit read. This is because SRAM bitcells are usually implemented with transistors near the minimum size allowed by the used technology, and many bitcells are packed tightly together to maximize cell density. During a read

cycle, the bitcells attempt to drive a logic signal onto the data lines (the bitlines), which are large capacitive loads. Due to the small size of the transistors in the bitcells, their drive strength is weak, which results in an unacceptably long access time required for the bitcell to produce logic-level signals on the bitlines. For this reason, sense amplifiers (SA) are used. SAs are an analog or dynamic digital circuit which can rapidly amplify a small signal produced by SRAM bitcells into a logic-level signal. The use of SAs allows for a significantly shorter SRAM bitcell access time and this greatly increases the speed of a read operation.

From the perspective of fast SRAM operation, a key property of a sense amplifier is its input-referred offset voltage V_{os}, which arises from Process-Voltage-Temperature (PVT) conditions and device mismatch. This voltage is statistical and can be described with a mean and standard deviation of μ_{os} and σ_{os}, respectively. A sense amplifier needs at least this much voltage on its inputs to ensure a correct decision. To guarantee the correct operation of an SRAM's sense amplifiers to a certain yield, the minimum required input voltage must be derived from worst-case V_{os} computed as some multiple of σ_{os} away from μ_{os}. The greater σ_{os} is, the bigger this worst-case V_{os} gets. This directly impacts the speed of an SRAM block as it takes more time for bitcells to generate the required signal voltage.

With technology scaling, the minimum size of devices continues to decrease. Short channel effects and device mismatch become more significant with minimum sized devices. Additionally, there is an industry trend to reduce the power consumption of ICs, often via lowering supply voltages. For ultra-low power applications like in IoT, supply voltages are aggressively scaled to 0.4V or lower, which can be at or below the threshold voltage of the transistors used. This places their operating region within or near the subthreshold region, which reduces the drive strength of transistors. These effects culminate as an increased standard deviation of the input offset voltage.

Figure 1.2 from [3] demonstrates this effect by comparing the distribution of a sense amplifier's input offset voltage normalized to VDD across 90nm and 45nm technology nodes, and a VDD of 1V and 0.4V. For the 90nm plots, the 0.4V curve is thinner than the 1V case, indicating that σ_{os} decreased with lower VDD. Relative to the 90nm node the plots for both VDD cases of the 45nm node are wider, which indicates a relative increase in σ_{os} between the technology nodes. For the 45nm plots, the 0.4V trace of the 45nm node is significantly wider than the 1V case, which suggests that σ_{os} at 0.4V is larger than σ_{os} at 1V.

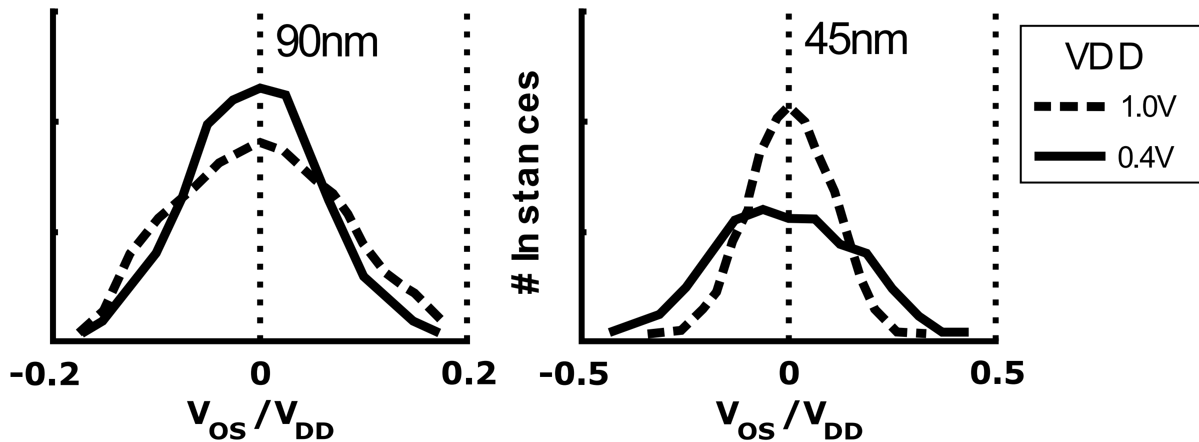

Figure 1.2 Comparison of Input Offset Voltage of a Sample Sense Amplifier for Multiple Technology Nodes and Different Supply Voltages [3]

To continue to scale the performance of SRAM with the industry trends of technology scaling and power reduction, new sense amplifiers need to be designed in ways that reduce or eliminate the input offset voltage.

1.2 Topics to be Explored

This thesis explores SRAM sense amplifiers for use in low voltage environments. The primary goal of this thesis is to find a sense amplifier topology and control scheme that reduces σ_{os}. The circuit tested is a modified VLSA that allows for independent timing of its control signals. The figures of merit used to evaluate sense amplifier performance are the mean and standard deviation, μ_{os} and σ_{os}, of the sense amplifier's input-referred offset voltage V_{os}. The circuits designed are on TSMC's 65nm node due to it being a mature and readily available technology, yet advanced enough to be affected by the modern design challenges of device mismatch and short channel effects.

1.3 Thesis Organization

Chapter 2 covers background information on SRAM organization and SRAM bitcells, threshold voltage mismatch and short channel effects, cross-coupled pairs, latch type sense amplifiers, and some known methods of input offset reduction in sense amplifiers. Chapter 3 details a proposed circuit to test, and the test methodology. Chapter 4 presents and analyzes simulation results. Lastly, Chapter 5 concludes the thesis.

Chapter 2
Background

2.1 SRAM Architecture and Operation

2.1.1 Architecture

A block of SRAM is typically composed of one or more SRAM banks with shared control circuitry. Figure 2.1 shows an example architecture of a bank of SRAM. A large portion of the bank area is taken up by the bitcell array, which is what stores the data. One of the most common bitcell designs used is the 6T SRAM memory cell. Bitcells can be read from or written to using single-ended or differential bitlines, which are exposed to a bitcell's state when the bitcell's wordline signal is activated. The decoder and wordline driver handle activating the wordline signal for a row of bitcells. The precharge array is a set of circuit blocks that charge the bitlines before a read or write operation, typically to VDD.

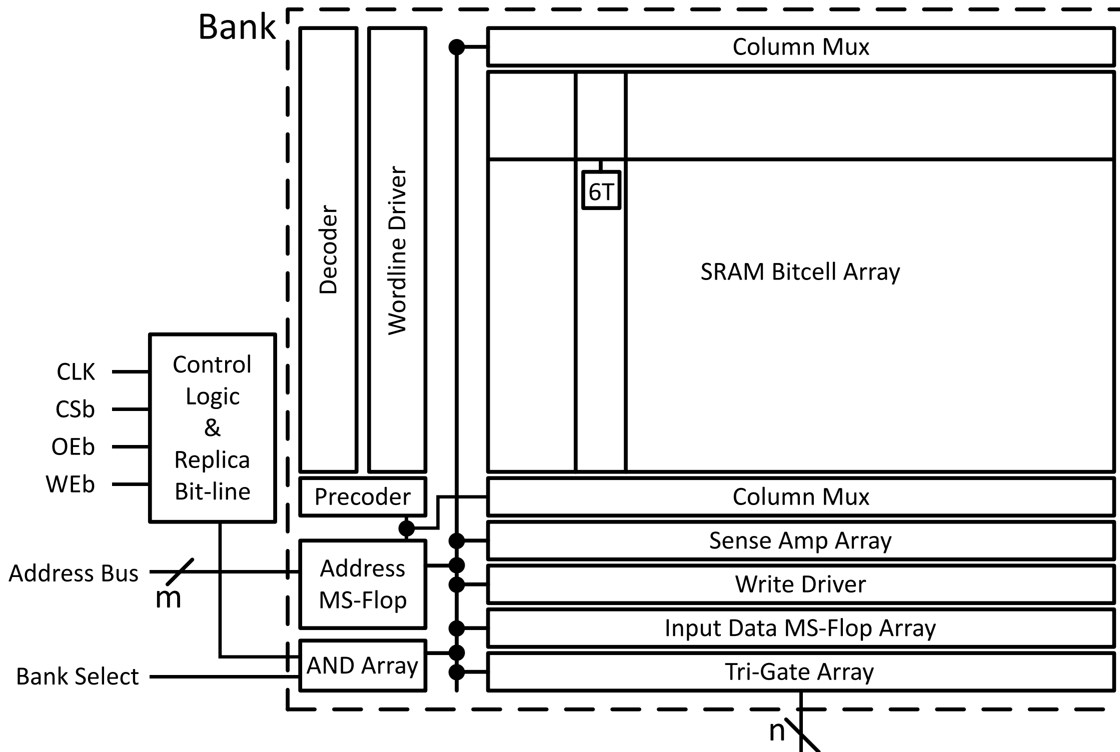

Figure 2.1 Example SRAM Bank Architecture [4]

The column multiplexers are optional switches inserted between the bitlines and some of the read/write circuitry, specifically the sense amplifier and the write driver. Sometimes a sense amplifier implementation is larger than a single column of bitcells, so the column multiplexers allow for two or more columns to share a single sense amplifier. The sense amplifier is tasked with reading a small differential voltage produced on the bitlines during a bitcell read operation and rapidly amplifies it to logic levels. The write driver is responsible for pulling the bitlines to VDD or VSS such that during a write operation, the state of a bitcell is altered.

Lastly, the MS-flop array and Tri-gate array as shown in the diagram represent the latch circuitry required to save the output state of a sense amplifier after a read, and control of writing this latched data to a common data bus only when this bank is selected.

2.1.2 6T SRAM Operation

Figure 2.2 shows the topology of the conventional 6T SRAM cell. Parasitic capacitances from the bitlines are also shown. Device pairs P1/P2 and N1/N2 form back-to-back inverters, which creates a metastable positive feedback loop. Q and QB will latch to VDD or VSS and keep their state unless they are externally driven from the bitlines. For whichever of Q or QB is at VSS, the NMOS transistor in the cross-coupled pair N1/N2 whose drain is connected to it is activated and is driving the node hard to ground. Meanwhile, the other NMOS device of the N1/N2 pair is in the cutoff region and thus is not conducting.

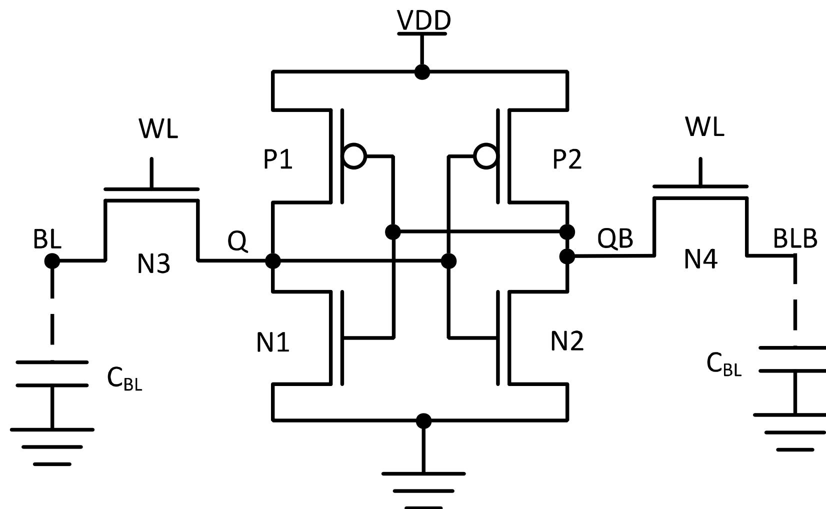

Figure 2.2 Conventional 6T SRAM Cell with NMOS Access Transistors and Bitline Capacitances

To perform a read, the bitlines are first precharged to VDD. After that, the wordline signal is activated (driven to VDD), and the SRAM cell discharges one of the bitlines. For VDD much greater than the threshold voltage, the activated transistor of the NMOS pair N1/N2 is in a state of strong inversion and in the triode region, which makes a strong connection to VSS. This causes the connected pass transistor N3 or N4 to be driven into a state of strong inversion, and in the saturation region. By the MOSFET Square-Law drain current equation, the maximum current that the pass transistors can conduct is proportional to the square of its overdrive voltage, $(V_{GS} - V_t)$ [5]. Since its gate is at VDD, and its source near VSS, we can approximate that its max current is just proportional to VDD^2 . If the bitline capacitance is known, then we can estimate the required time to develop a certain differential bitline voltage ΔV_{BL} with Equation (2.1) below. Note that this equation is merely a rough approximation; simple device models such as the MOSFET square law cannot accurately predict the dynamic latching performance of the bitcell, especially in advanced technology nodes. Additionally, operating the transistors near or within the subthreshold region further degrades the relationship.

$$\tau_{development} \approx C_{BL} \frac{\Delta V_{BL}}{I_{\mathrm{N3,N4\ max}}(VDD)} \propto C_{BL} \frac{\Delta V_{BL}}{VDD^2} \tag{2.1}$$

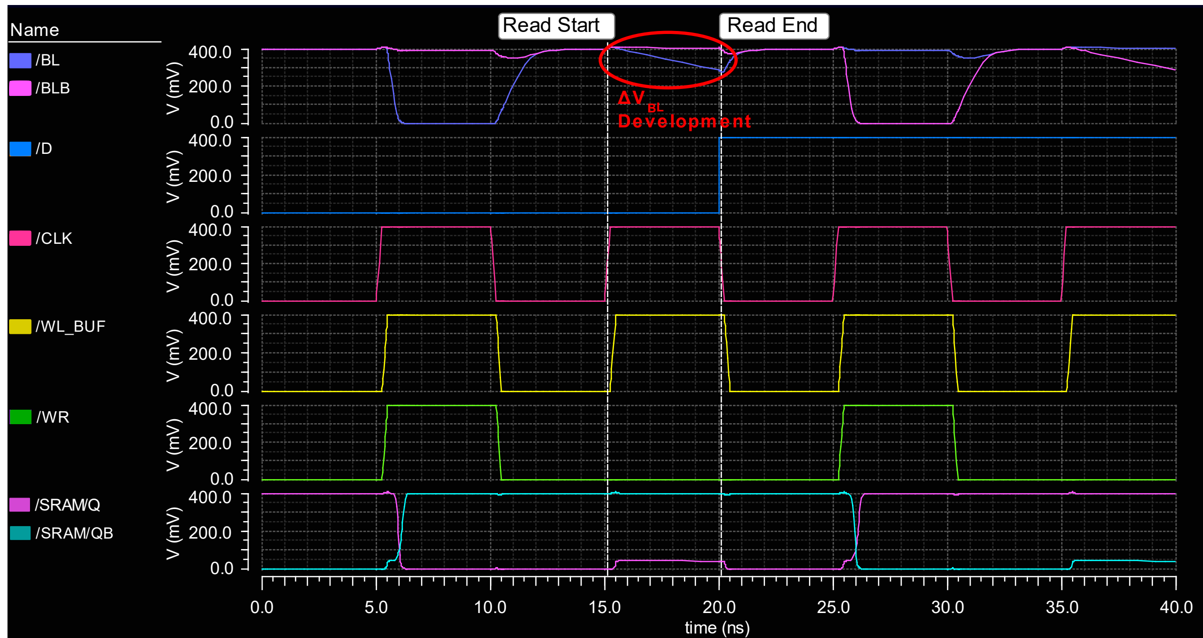

Figure 2.3 SRAM Cell Read and Write Cycles

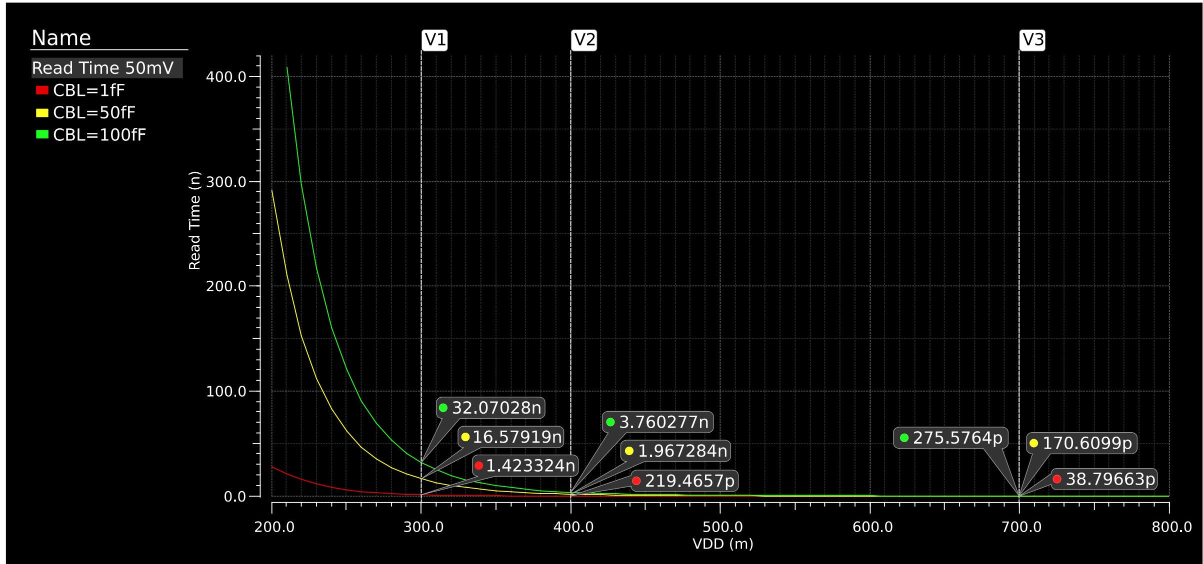

Figure 2.4 SRAM time for $\Delta V_{BL} = 50mV$ development vs VDD

Figure 2.3 shows the waveforms of read and write cycles for one SRAM cell using near minimum size SVT devices in TSMC 65nm. The supply voltage VDD is set to 0.4V, and the bitlines are loaded with 50fF capacitance. The circled region highlights the development of a differential voltage ΔV_{BL} across the bitlines for a read operation of a logical 0 state. Figure 2.4 shows the read time required for the example SRAM cell to develop a ΔV_{BL} of 50mV on the bitlines for a range of VDD, with C_{BL}=1f, 50f, and 100f. These two figures demonstrate that an increase in bitline capacitance and the decrease in the supply voltage VDD both result in an increased read time.

2.2 MOSFET V_t Model

$$V_t = V_{t0} + \gamma\left(\sqrt{\Phi_s - V_{bs}} - \sqrt{\Phi_s}\right), \qquad \gamma = \frac{\sqrt{2q\varepsilon_{si}N_{substrate}}}{C_{oxe}} \tag{2.2}$$

Equation (2.2) from the BSIM4 manual [6] shows the ideal model of MOSFET threshold voltage for a long and wide MOSFET with uniform substrate doping. V_{t0} is the threshold voltage of the device with zero body bias, and γ is the body bias coefficient. For an NMOS device, this equation shows that a positive body voltage relative to the source will produce a negative coefficient with γ and reduce the effective threshold voltage.

2.2.1 Short Channel Effects and Drain-Induced Barrier Lowering

As device dimensions shrink with newer technology nodes, the V_t equation is no longer completely accurate. Short Channel Effects (SCE) and Drain-Induced Barrier Lowering (DIBL) become prevalent and cause deviations in the threshold voltage. The BSIM4 equation that models threshold voltage deviation DV_t due to SCE and DIBL is the following:

$$DV_t(SCE, DIBL) = -\theta_t(L_{eff}) \cdot [2(V_{bi} - \Phi_s) + V_{ds}] \tag{2.3}$$

Where $\theta_t(L_{eff}) = \frac{0.5}{\cosh(L_{eff}/l_t)-1}$ is the short-channel effect coefficient. The first takeaway from this equation is that the MOSFET threshold voltage decreases with an increasing Vds, and the slope is related to the channel length. The second takeaway is that the slope of deviation drastically increases when L_{eff} gets smaller.

The impact of these nonideal phenomena on design is that the transistor performance is bias dependent, and deviations from the bias point can cause significant performance difference for short channel devices. Increasing the channel length can help reduce this threshold voltage deviation.

2.2.2 Mismatch Model

Pelgrom et al in [7] studied the statistical variations of integrated MOSFET devices due to normally distributed process parameters. The key equation from their research was the following:

$$\sigma_{\Delta V_t}^2 = \frac{A_{V_t}^2}{WL} \quad \Rightarrow \quad \sigma_{\Delta V_t} = \frac{A_{V_t}}{\sqrt{WL}} \tag{2.4}$$

Where $\sigma_{\Delta V_t}^2$ is the variance of threshold voltage in $[mV]^2$ between two equally sized, closely spaced devices, $\sigma_{\Delta V_t}$ is the corresponding standard deviation, and A_{V_t} is a technology-dependent scaling factor in $[mV \cdot \mu m]$. This equation shows that the variance of threshold voltage between two devices is inversely proportional to the area of the devices.

With ever-shrinking technology nodes, the minimum device size continues to decrease. There is a preference for using smaller devices because they allow for lower power consumption and higher layout density. Unfortunately, this leads to a larger threshold mismatch between devices.

2.3 Cross-Coupled Pairs

This section covers cross-coupled MOSFET pairs and the effect of device mismatch on their behavior. This circuit topology is important to understand as cross-coupled pairs are often used in sense amplifiers.

2.3.1 Introduction and Derivation

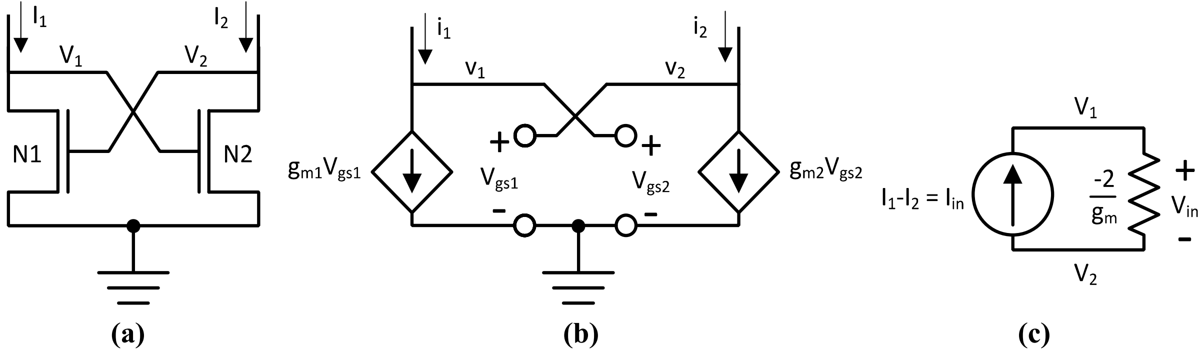

Figure 2.5 NMOS Cross-Coupled Pair (a), Ideal Small-Signal Model (b), Differential Mode Small Signal Model (c)

A cross-coupled pair is two back-to-back NMOS or PMOS transistors, where the gate of each device is connected to the drain of the other. The sources are typically tied together. Figure 2.5a shows the schematic for an NMOS cross-coupled pair with a common source terminal, and Figure 2.5b shows the ideal small-signal model with the output resistance due to channel length modulation omitted.

For a common-mode input, the single-ended impedance is $R_{icm} = \frac{v_{icm}}{g_m v_{icm}} = \frac{1}{g_m}$. Since R_{icm} is positive, any resistive driver attached to the inputs in the common-mode case will be loaded and the voltages limited. For a differential mode input v_{id} the small-signal model can be reduced to the circuit in Figure 2.5c. Assuming devices are matched, the differential input impedance is $R_{id} = \frac{2(v_{id}/2)}{g_m i_{id}} = \frac{v_{id}}{-g_m v_{id}/2} = -\frac{2}{g_m}$ [8]. The metastable point of this cross-coupled pair is when the input differential current I_{in} is zero. Any perturbation from the input current will cause an imbalance in the circuit's voltage terminals. If the circuit is instead driven by a Norton or Thevenin equivalent source with some non-zero resistance, the circuit becomes a positive feedback amplifier. To maintain stability, the loop gain of the amplifier must remain less than unity. However, if the loop gain of this amplifier is greater than unity, then it is unstable and any deviation from the metastable point will

cause the voltage at the cross-coupled pair's terminals to continuously diverge. This is sometimes referred to as regenerative feedback. [9]

2.3.2 Effect of Mismatch on The Cross-coupled Pair

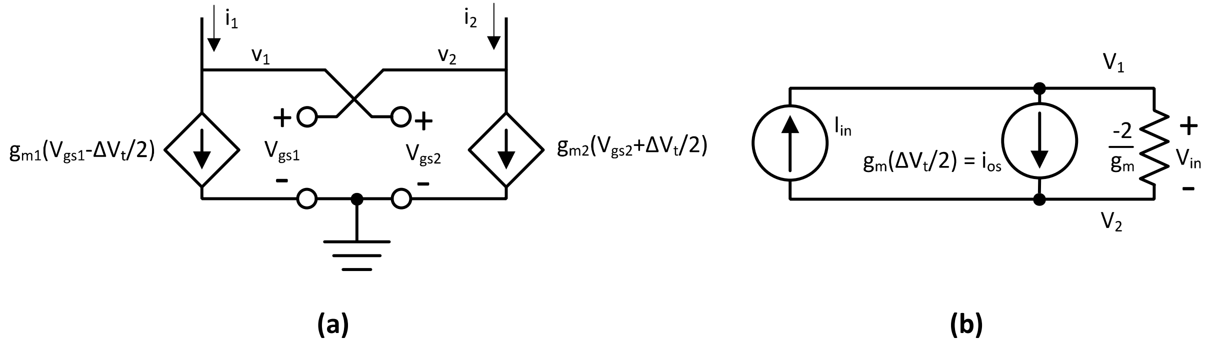

Figure 2.6 NMOS Cross-coupled Pair Small Signal Model with Threshold Mismatch (a), Differential Mode Small Signal Model (b)

The input offset voltage of a differential pair is the differential input voltage required on the gates to balance the drain currents of both transistors in the pair. Ideally, it is zero but due to process variations and other factors, the devices in the pair may not be perfectly matched. Threshold voltage variation is a primary source of mismatch. For a cross-coupled pair, the differential input offset voltage V_{os} is simply equal to the differential input voltage. Figure 2.6a shows the small-signal model of a cross-coupled pair that is modified to incorporate the effects of threshold mismatch. The threshold mismatch is represented by ΔV_t and follows a normal distribution $\mathcal{N}(\mu_{\Delta V_t}, \sigma_{\Delta V_t})$ according to Pelgrom's model. If we assume the mismatch is evenly split between devices (ie: $\mu_{\Delta V_t} = 0$) and approximate that the g_m for each device are equal, then the effect of mismatch can be included with a fixed current source parallel to the dependent one with a magnitude of $g_m \Delta V_t/2$ as seen in Figure 2.6b [10], and $V_{os} = -\left(g_m \frac{\Delta V_t}{2}\right)\left(-\frac{2}{g_m}\right) = \Delta V_t$.

2.3.3 Cross-coupled Pairs in Digital Circuits

In the context of digital circuits, the unstable behavior of the cross-coupled pair is desirable. It allows us to take a small differential voltage or current, and rapidly amplify it to a large differential voltage. Digital circuits are bound by their supply rails VDD and VSS, so the pair's nodes can quickly saturate to these rails. This creates complementary single-ended logic values on the nodes. An additional benefit in the context of digital circuitry is that once saturated to the supply rails, only one

transistor is conducting while the other is in the cutoff region. This reduces static power consumption when compared to, for example, a current mirror.

One caveat of the cross-coupled pair on its own is that there always needs to be something driving one of the inputs to maintain the saturated state. In digital circuits, we usually want to keep devices off as long as possible, and only consume power when a state transition is necessary. A way to make its saturated state stable and reduce power is to incorporate an additional cross-coupled pair of the complementary transistor type.

Figure 2.2 showing a conventional 6T SRAM cell illustrates the connection of the complementary cross-coupled pairs. In this configuration, the complementary pairs form back-to-back inverters. This forces the nodes to be complementary logic states and will remain there until external circuitry tries to drive the nodes.

2.4 Latch Type Sense Amplifiers

Latch type sense amplifiers are similar to a clocked dynamic comparator. The common structure in these types of amplifiers is a cross-coupled transistor pair. There may be one, or a complementary set used. A differential current or voltage is applied to the cross-coupled pair, and on a clock enable signal the cross-coupled pair amplifies the small value to digital levels.

The two traditional classes of latch type sense amplifiers are the voltage latch sense amplifier, and the current latch sense amplifier [11].

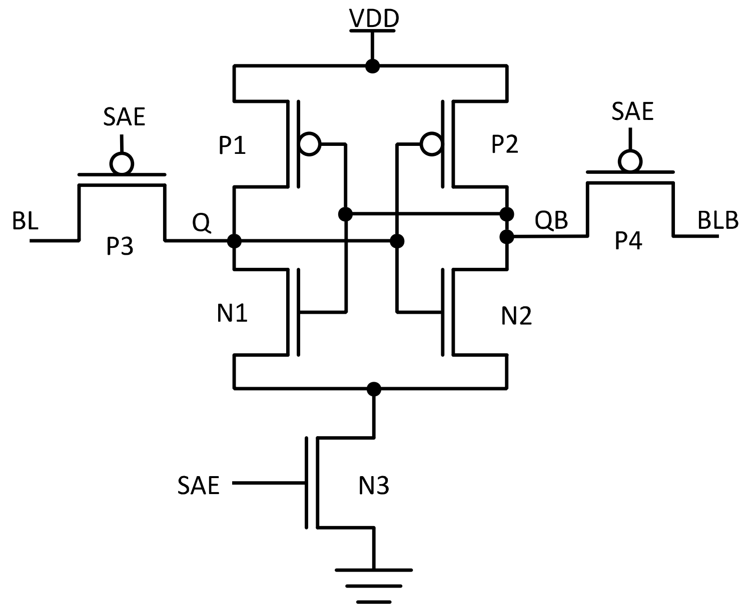

Figure 2.7 Voltage Latch Sense Amplifier (VLSA)

The VLSA as seen in Figure 2.7 above is one of the simplest latch type sense amplifier topologies. Output nodes Q and QB are set to some differential input voltage through P3 and P4 by using the bitlines BL and BLB, which are near VDD. The differential voltage applied means that one of the NMOS devices in the inverter pair has a larger VGS than the other. When the footer NMOS N3 is activated on the clock signal / SAE, the NMOS with the larger VGS will conduct more current than the other NMOS in the pair. This pulls its drain to VSS faster than the other output node. As Q and QB are pulled to VSS, the PMOS devices get closer to turning on. Since Q and QB are at different voltages, one of the PMOS devices turns on earlier than the other one and slows the rate of decrease of its drain voltage. Eventually one of the NMOS devices will turn off, and its drain will be pulled to VDD by the PMOS connected above it. The complementary output node will be pulled to VSS. The read decision produced here will remain as long as SAE is asserted.

Figure 2.8 Current Latch Sense Amplifier (CLSA)

The CLSA as seen in Figure 2.8 above is another popular sense amplifier topology. This class of circuit uses currents to start the regenerative feedback of a cross-coupled inverter pair. Instead of precharging the output nodes to BL and BLB, they are precharged to VDD. The input voltages are applied to the input sensing pair N3/ N4, which converts the signals to the currents. The imbalance in currents due to a differential input voltage causes the source terminals of N1 and N2 to be pulled to VSS at different rates. The side with the higher discharge current will cause the connected device to turn on sooner. The output node that is connected to the drain of this device will be pulled to VSS, while the other side is not driven. This creates an imbalance in the output node voltage, and regenerative feedback begins.

2.5 Latch Type Sense Amplifier Vos Yield Model

To guarantee that a sense amplifier produces the correct output for some input voltage ΔV_{in}, the input voltage must be large enough to overcome the input offset voltage V_{os}. If this condition cannot be met, then the sense amplifier can produce an incorrect result. The probability that a sense

amplifier produces a correct result for some ΔV_{in}, or read yield $Y(\Delta V_{in})$, can be estimated by testing some large N number of randomly selected devices and counting the number of correct results:

$$Y(\Delta V_{in}) = \frac{\text{Number of Correct Results}}{\text{Number of Devices (N)}} * 100\% \tag{2.5}$$

Recall that the input offset voltage of a cross-coupled pair depends on the normally distributed V_t mismatch of its devices. Since cross-coupled pairs are a primary component in sense amplifiers, the read yield of a sense amplifier can be modelled as a probability distribution of the sense amplifier offset voltage [12]:

$$\begin{aligned} Y(\Delta V_{in}) &= P\{\Delta V_{in} \geq V_{os}\} \\ &= \Phi\left(\frac{\Delta V_{in} - \mu_{os}}{\sigma_{os}}\right) \\ &= \frac{1}{2}\left(1 + \text{erf}\left(\frac{\Delta V_{in} - \mu_{os}}{\sigma_{os}\sqrt{2}}\right)\right) \end{aligned} \tag{2.6}$$

Where $V_{os} \sim \mathcal{N}(\mu_{os}, \sigma_{os})$ is a normally distributed random variable, $\Phi(x)$ is the cumulative Gaussian distribution, and $\text{erf}(x)$ is the error function. Equation (2.6) can be rearranged to find the required ΔV_{in} normalized to σ_{os} for a target yield at ΔV_{in}:

$$k = \frac{\Delta V_{in} - \mu_{os}}{\sigma_{os}} = \sqrt{2} * \text{erf}^{-1}(2Y(\Delta V_{in}) - 1) \tag{2.7}$$

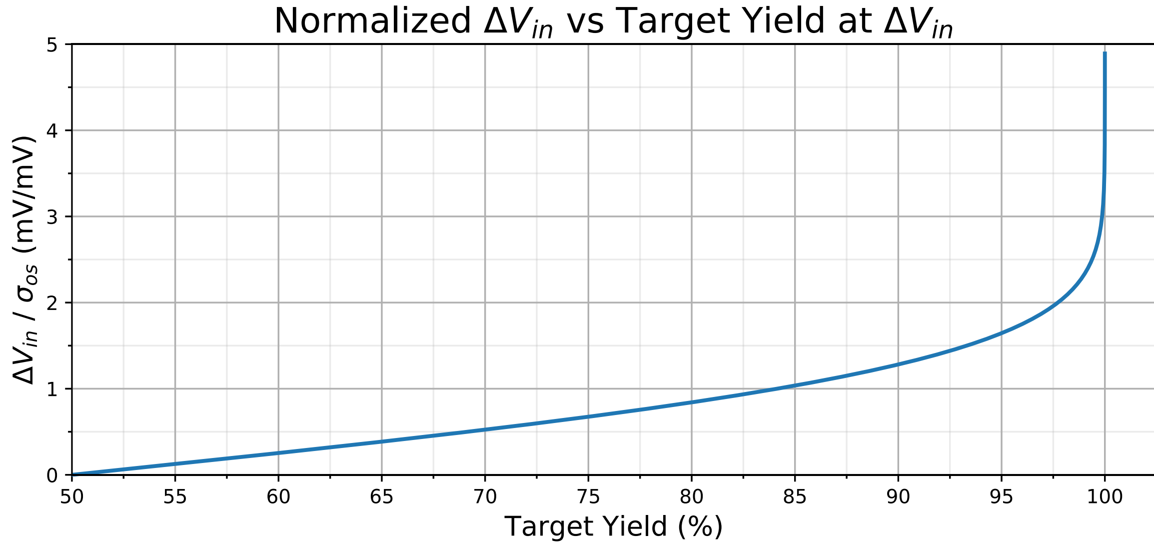

Figure 2.9 Plot of Required ΔV_{in} Normalized to σ_{os} for a Target Yield of a Sense Amplifier

Figure 2.9 plots equation (2.7) with $\mu_{os} = 0$ for a range of target yields. The relationship is roughly linear for yields <= 90%. It starts to curve upward around 95%, and at >= 99% the curve becomes asymptotically vertical. A 99.73% yield (0.23% error rate) is obtained with a $3\sigma_{os}$ offset, while an error rate of 1 part per million (1 ppm) requires a minimum input of $4.75\sigma_{os}$. This shows that there is diminishing returns to reducing error rate by increasing the minimum input offset voltage. A corollary to this is that for high yield designs, the minimum input voltage increases sharply for any marginal improvement to yield. Since ΔV_{BL} development takes time in SRAM circuits, reducing σ_{os} of a sense amplifier will help reduce the time required for the SRAM cells to produce the required ΔV_{in} for a targeted yield and thus reduce the total read cycle time.

2.6 Review of Offset Reduction Methods

This section briefly covers some input offset reduction techniques for VLSA and CLSA style sense amplifiers that other researchers have previously investigated. The following subsections group the techniques by their style of offset reduction.

2.6.1 Self Calibration / Auto Zeroing

Sense amplifiers can be designed with additional structures that allow for active monitoring and zeroing of its input offset voltage. Figure 2.10 below shows an example of this idea.

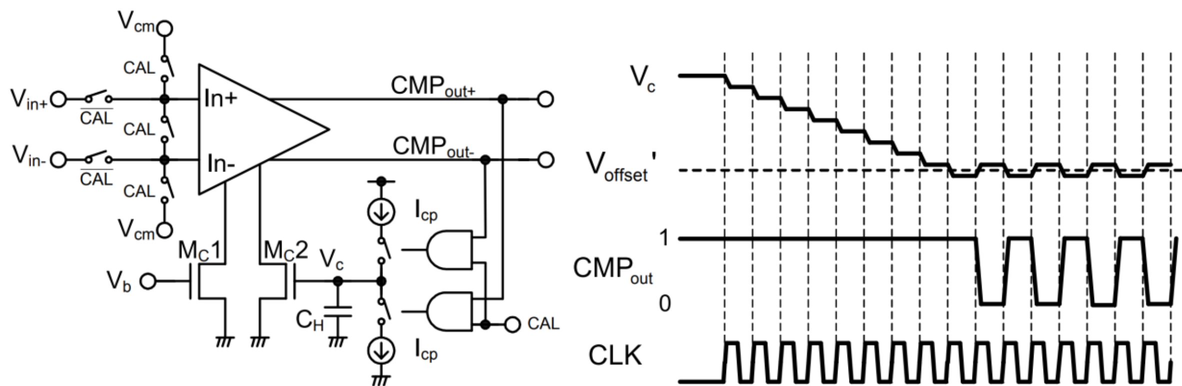

Figure 2.10 Self Calibrating Dynamic Comparator [13]

Mc1 and Mc2 devices are used to trim the input sensing pair of the comparator for zero input offset voltage. By driving the Mc2 gate voltage Vc in a feedback loop with an up/down sensor and a charge pump, spare clock cycles in between memory read operations can be used for a self-calibration routine to zero out the input offset voltage.

2.6.2 V_t Mismatch Cancellation and Compensation

This type of scheme uses a precharge or cancellation phase in its sensing sequence to eliminate the effect of V_t mismatch of devices. This scheme works well in current-mode sense amplifiers, where nodes can be precharged with the threshold voltages of the critical transistors. This balances the current drive strength of the devices when the sensing stage begins.

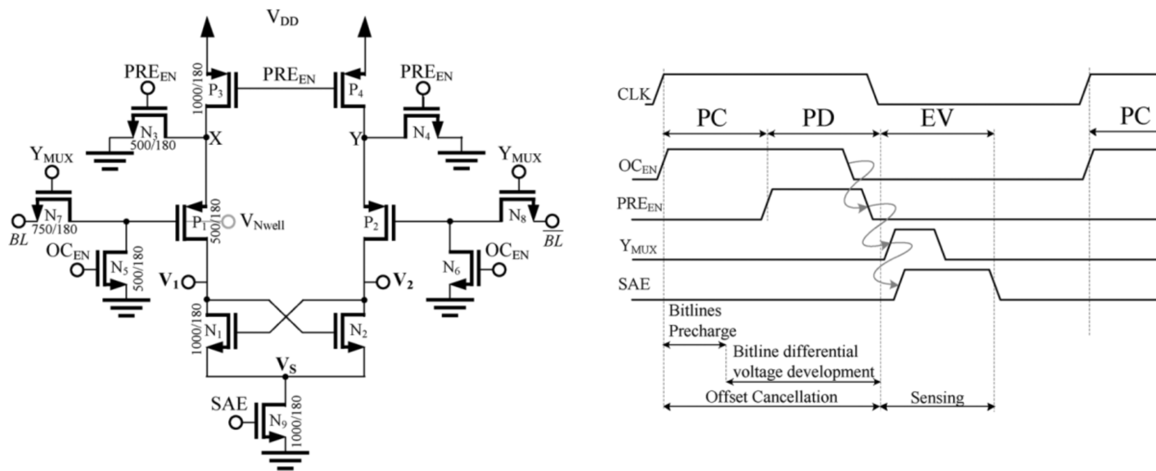

Figure 2.11 Offset Cancelling Sense Amplifier Example [14]

Figure 2.11 from [14] uses a current mode input topology. It has a two-part cancellation phase that pre-loads the threshold voltages of the input pair P1/P2 onto the latching pair N1/N2 without the influence of the bitline voltages. On sense amp enable, the input signal is added to the gates of the input pair to induce a current imbalance going into the N1/N2 pair to trigger the positive feedback loop.

An important feature of this architecture is that the cancellation phase can occur simultaneously with the SRAM bitcells being prepared for access, which is time that the sense amplifier would otherwise not be doing anything.

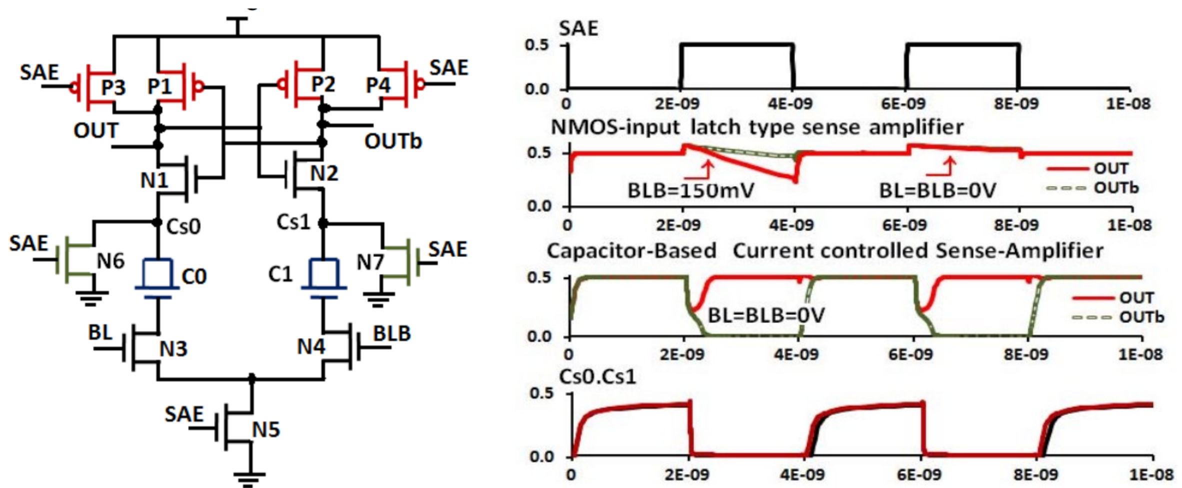

Figure 2.12 Capacitor Based Compensation for CLSA [15]

Figure 2.12 from [15] uses a CLSA structure. MOS caps are inserted in between the source terminals for the N1/N2 pair devices, and the drain terminals of the N3/N4 input sensing pair. These capacitors assist in the precharge phase in setting the node voltages Cs0 and Cs1 to VDD minus the threshold voltages of N1/N2. During the sensing phase, they provide a way to discharge Cs0 and Cs1 to VSS quickly. In combination with the N6/N7 pair, the N3/N4 input pair sets an imbalanced discharge path through the capacitors to VSS that affects the current-mode regeneration of the cross-coupled pair.

Similar to the previously shown sense amplifier, this architecture utilizes dead time for state preparation. Cs0 and Cs1 precharge begins on the falling edge of the sensing clock SAE.

2.6.3 Multiple Injection

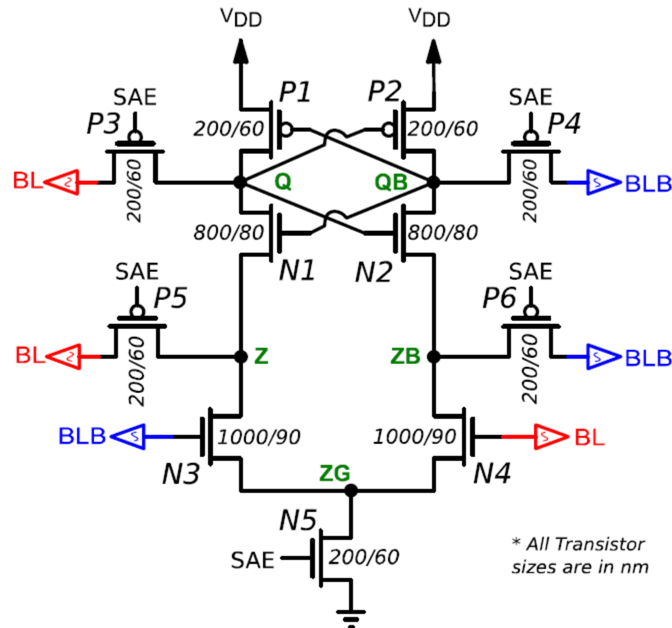

Figure 2.13 HYSA-QZ Schematic [16]

Figure 2.13 from [16] uses a hybrid topology of a CLSA with VLSA-like precharging. Instead of precharging the output nodes Q and QB to VDD and relying only on the input signals at the NMOS pair N3/N4, the pass transistors P3, P4, P5, and P6 are used to inject BL/BLB onto the output nodes Q/QB and the intermediate nodes Z/ZB during a precharge phase. For an input differential voltage $\Delta V_{BL} = BL - BLB$ that is applied to Q/QB and Z/ZB during precharge, then the VGS on N1 and N2 are $-\Delta V_{BL}$ and ΔV_{BL} respectively, which results in a ΔV_{GS} between N1 and N2 of $2\Delta V_{BL}$.

Additionally, since N1 has a negative VGS, it takes some time during the sensing phase for its source node to discharge and produce a large enough positive VGS to form a conducting drain-source channel. Meanwhile, the VGS on N2 is already positive so it will take less time for its VGS to grow enough to start conducting. The key takeaway of this architecture is that multi-point signal injection can be used to shift the effective VGS seen by the transistors in the sense amplifier.

2.6.4 Input Boosting / Multiplication

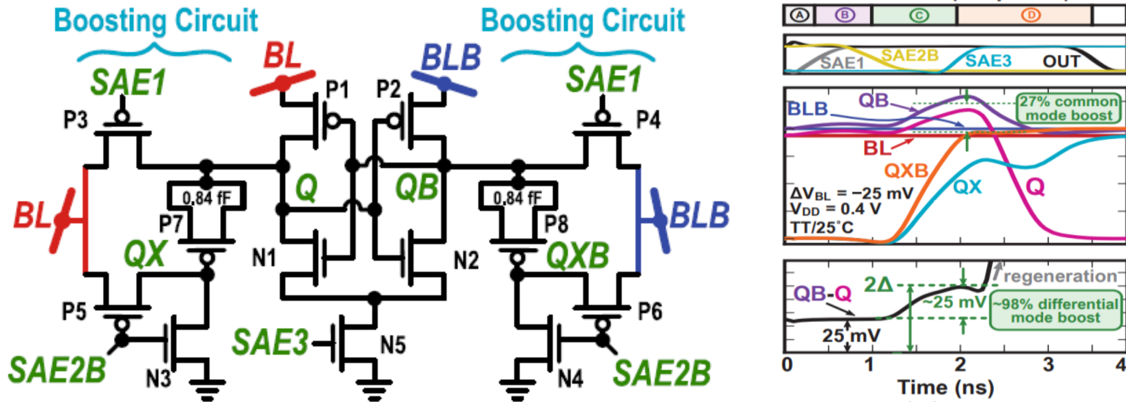

Figure 2.14 SBLSA [17]

Figure 2.14 from [17] uses a VLSA structure targeted for low voltage applications. Boosting circuitry is added to the Q and QB lines. These structures allow for BL and BLB to be loaded onto matched MOS caps while Q and QB are precharged, and during the sensing stage, the MOS cap gates are shifted from VSS to BL and BLB. Since the MOS cap voltage cannot change instantaneously, the voltages that were applied to Q and QB get a two times differential boost. The benefit of this boosting is two-fold: the differential input voltage needed for a given input offset is effectively halved, and the common-mode boost increases the gate overdrive of the NMOS pair which contributes to a significant improvement on sensing delay for subthreshold operation. This circuit also borrows the multiple injection point concept by powering the pull-up PMOS pair P1/P2 with the bitlines.

2.6.5 Timing Manipulation

Timing manipulation refers to the adjustment of the properties of a sense amplifier's control signals. This can be done by inserting timing skews between control signals, or by adjustment of the signal rise and fall times.

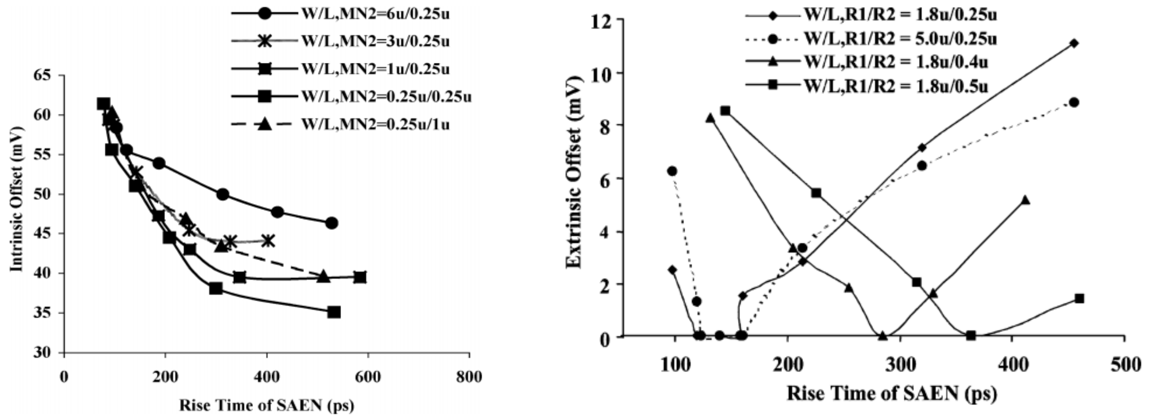

Figure 2.15 Intrinsic and Extrinsic Offset of VLSA vs SAE Rise Time [18]

Singh and Bhat in [18] investigated the effect of the rise time of the SAE signal on the input offset of a VLSA. They used an "input decoupled" VLSA topology, which is the VLSA presented in Figure 2.7 but with additional column multiplexer switch transistors. Offset sources were separated into "intrinsic" and "extrinsic", for the core cross-coupled pairs (N1/N2, and N3/N4) and the pass transistors (P3/P4), respectively. Plots of intrinsic and extrinsic offset sources for a swept SAE rise time as in Figure 2.15 were recorded. The intrinsic offset has a 1/x like response, decreasing as the rise time is increased. For the extrinsic offset, the plots have a V like shape to them, with the center of the V moving closer to 0 seconds as the W/L ratio of the pass transistors is increased. The combined offset results for the whole latch have a 1/x like shape to them that flattens out for larger rise times. The conclusion drawn is that a slower rise time can help reduce the VLSA offset.

Shi in [19] studied both VLSA and CLSA amplifiers. For a VLSA with separate control signals for the NMOS footer transistor (controlled by SAE) and the pass transistors (controlled by PGB), intentionally adding a skew between the control signals can have a positive impact on reducing input offset. Simulation results showed that delaying PGB by a small amount relative to SAE (10ps in their example) produced an approximate 10% reduction in input offset.

2.7 VDD Dependence of Device Mismatch Contribution in VLSA

Pileggi in [20] analyzed the input offset voltage of a VLSA in 65nm technology. He used a basic model to fit the input offset voltage as a linear function of the V_t mismatch of the NMOS and PMOS pairs: $V_{os} = a\Delta V_{tn} + b\Delta V_{tp} + c$. Using this model, the parameters a and b are used to construct a correlation coefficient (CC) that measures the contribution of the PMOS and NMOS mismatch as a ratio to the input offset voltage. He performed a sweep of bitline precharge voltages up to VDD and computed the NMOS and PMOS CC for each point. The plot shows that the closer the bitline precharge voltage is to VDD, the lower the CC of the PMOS pair. The reasoning is that time which the PMOS is off during a read cycle is maximized, during which only the NMOS pair is conducting.

An important detail in Pileggi's analysis is that he performed this at a VDD of 1V, which is the nominal core voltage for 65nm CMOS. This is well above the V_t of the standard V_t devices (300mV to 400mV range for 65nm). At the time of initial discharge, the transistors in the NMOS pair are driven very hard into the saturation region, and the PMOS pair is effectively off. Even accounting for any subthreshold currents in the PMOS devices, the currents in the NMOS pair are orders of magnitude greater.

At lower voltages, however, this is not necessarily the case. The lower the supply voltage, the weaker the drive on the NMOS pair. If the voltage becomes low enough, then the NMOS devices may be operating near or within the subthreshold region. The original assumptions from the 1V case may no longer be valid as the currents in the NMOS devices are now on a much more comparable level to the PMOS devices. This may mean that PMOS mismatch has a more significant impact on the input offset voltage.

To demonstrate this, a standard VLSA circuit is constructed with standard V_t devices in 65nm technology. Input offset is characterized by using a Monte Carlo simulation as described in Section 3.4.6 of this document. The measurement is conducted for a set of supply voltages ranging from 0.2V to 1V. Virtuoso's sensitivity analysis tool is then used to analyze the mismatch contribution of each of the transistors to the variance of the input offset voltage.

Table 2.1 Offset Variance Contribution of NMOS and PMOS pairs for standard VLSA vs VDD

Parameter	Device	Value			
VDD (V)	-	0.20	0.40	0.70	1.00
σ_{os} (mV)	-	11.26	10.79	11.01	9.15
Variance Contribution R^2 total (%)	-	95.34	99.98	100	100
Variance Contribution (%) of σ_{os}^2	N1	39.25	47.10	47.78	47.95
	N2	41.45	51.12	51.84	51.98
	P1	7.08	0.89	0.19	0.04
	P2	7.56	0.87	0.18	0.03
Combined Variance Contribution (%) of σ_{os}^2	NMOS Pair (N1+N2)	80.70	98.23	99.62	99.93
	PMOS Pair (P1+P2)	14.64	1.76	0.37	0.07

Table 2.1 above presents the results of the input offset mismatch contribution analysis. For each VDD point the standard deviation of input offset voltage σ_{os} is reported for reference of scale. The mismatch contribution of each device is reported as a percentage of the offset variance σ_{os}^2, and the NMOS and PMOS contributions are summed to give the mismatch contribution of each cross-coupled pair. Note that VDD near 1V results in a negligible mismatch contribution by the PMOS pair. As VDD lowers to the threshold voltage and below, the PMOS contribution becomes much more significant. This raises a question about the possibility of reducing the PMOS contribution at lower operating voltage.

Chapter 3
Methodology

This chapter presents a sense amplifier circuit topology and timing mechanism with the goal of reducing input offset voltage variation in a low supply voltage environment, specifically at 0.4V. The circuit is based on the standard VLSA circuit and its control mechanism is a type of timing manipulation. Its design attempts to address the mismatch contribution issues of the VLSA as discussed in Section 2.7. Two testbenches and their architectures are described in Section 3.4. Section 3.5 outlines the tests to be performed and Section 3.6 presents the default simulation parameters used.

3.1 Hypothesis

In the standard VLSA, the PMOS pull-ups are initially off due to the Q and QB nodes being precharged to near VDD with some ΔV_{BL}. When the sense amplifier is activated there is some time where it is only the NMOS pair conducting, and Q and QB are discharged to GND. However, once Q and QB have discharged enough that they are $|V_{tp}|$ below VDD, the PMOS devices will begin to turn on and start the regenerative feedback. It is at this point that the mismatch of the PMOS threshold voltage will influence the output voltage development.

From this description of the VLSA's operation, we recognize that there is some time delta between the start of the sensing cycle and the time at which the PMOS devices turn on. Also, as outlined in Section 2.7, there is the possibility of reducing the VLSA's input offset voltage by reducing the PMOS pair's contribution to the VLSA's input offset variation. These two statements raise the following question: would extending this time where the PMOS pair devices are off influence the input offset voltage variation of the sense amplifier? Thus, we hypothesize that introducing this time extension will reduce the PMOS pair's contribution to the input offset voltage of the sense amplifier due to threshold voltage mismatch, and by extension reduce the input offset voltage variation.

3.2 Proposed Circuit – DVLSA

3.2.1 Concept

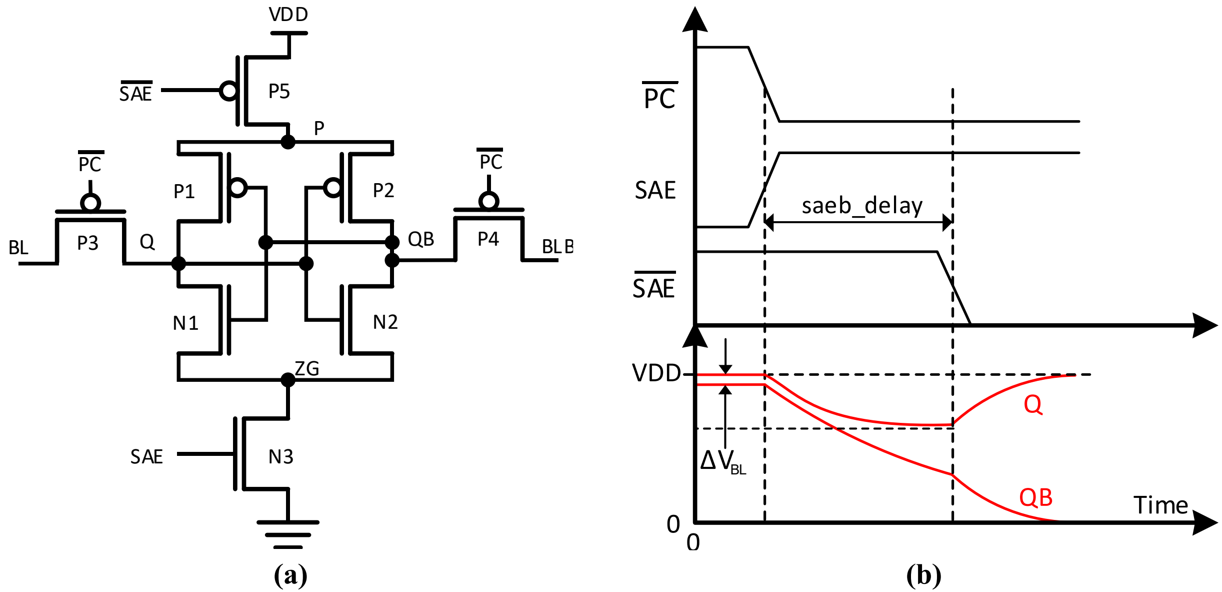

Figure 3.1 DVLSA Schematic (a) and Conceptual Waveforms (b)

The proposed circuit as seen in Figure 3.1 is a VLSA with delayed SAEB signal, or DVLSA. It is a modified version of the conventional VLSA. A header PMOS switch, P5, is inserted between VDD and the PMOS pullup pair P1/P2. The method of control is similar to the standard VLSA, with the difference being the use of P5 (controlled by SAEB) to control the current path from VDD to the P1/P2 pair independently from the other devices.

Observe Figure 3.1b for the circuit's conceptual waveforms to help understand the predicted method of operation. The plots assume ΔV_{BL} is positive (Q > QB), which corresponds to a logic '1'. We expect that this circuit will behave in the following manner:

1. Delaying SAEB extends the period for which the PMOS devices are off, and thus only the NMOS pair is developing a differential signal.
2. With enough SAEB delay, the Q and QB nodes should decay to near the threshold voltages of the NMOS devices. If there is a sufficiently large differential voltage already developed, then the NMOS device with the lower gate voltage will shut off first. Using the conceptual waveforms as a reference this is device N1 and node QB on its gate. The Q node attached to N1's drain should now have no discharge path as both the connected NMOS and PMOS

devices are off, so this node is left floating. Due to the cross-coupled connections the gate of N2 is now a fixed voltage, and thus that NMOS device will continue to discharge QB until it reaches VSS. If this large differential voltage develops correctly, then it should completely compensate for any mismatch in the PMOS devices when they are turned on.

There is some concern relating supply voltage to the effectiveness of this control method. The performance of the circuit relies on P node remaining floating and not discharging through the PMOS pair during the delayed SAEB period. Supply voltages that are at the upper end of the allowed range for the technology will result in voltage swings on Q/QB that are larger than the magnitude of V_t by some multiple. Even if the P node was floating during the beginning part of a read operation, the Q/QB nodes can reach voltages low enough below the voltage of P that the PMOS pair will turn on anyway and discharge the P node. In this event, the PMOS devices will still have some influence on Q/QB voltage development and potentially increase the input offset variation.

If the supply voltage is restricted to the low voltage region that is comparable in magnitude to $|V_t|$, the Q/QB nodes will have to discharge by an amount near the full supply voltage before the PMOS devices gain sufficient V_{GS} to be considered on. Under these conditions, the concern is reduced to possible subthreshold conduction through the PMOS pair. This may be small enough that the performance is not significantly affected. We will target a VDD of 0.4V for testing.

3.2.2 Device Sizing

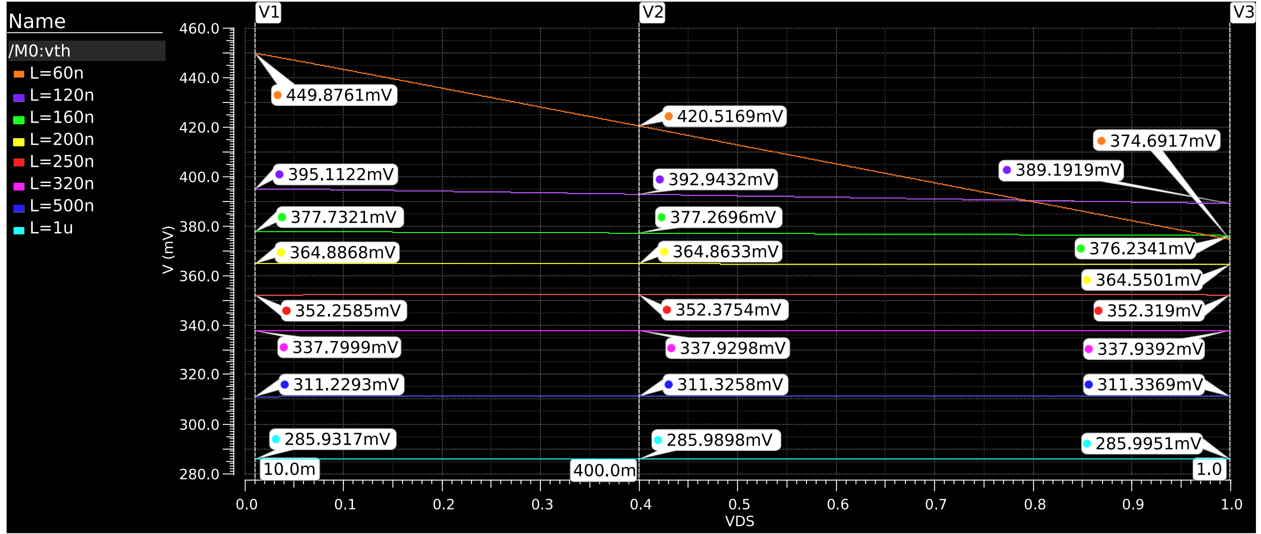

Figure 3.2 NMOS V_t Extraction vs VDS

In sub-micrometer technology nodes such as TSMC 65nm, short channel effects and DIBL can become significant for minimum dimensioned devices. Figure 3.2 is a plot of V_t extracted from a DC sweep of VDS for a set of channel lengths. VGS is held at a fixed value of 0.5V, and the gate area is held at a constant $0.08\mu m^2$. In the plot we can see that the slope of the V_t curve is large for a small L and flattens out as L increases, which agrees with what was discussed in Section 2.1.

From the perspective of input offset voltage reduction, we are concerned with device threshold variation. Depending on how the sense amplifier is precharged and a differential voltage is applied, DIBL can work together with the threshold mismatch to make the input offset voltage even worse. For the critical transistor pairs of a sense amplifier, gate area and W/L ratio must be designed carefully. The area should be selected based on Pelgrom's model to achieve a nominal threshold voltage variation. Designing the W/L ratio is a tradeoff between drive strength and short channel effects; a lower W/L ratio for a given gate area means the channel length is longer which can reduce DIBL but at a cost of a lower device transconductance.

The transistor pair N1/N2 is the primary contributor to the input offset in the VLSA, so their sizing is critical. It is assumed that the N1/N2 pair is also the primary contributor to the DVLSA structure given the similarity to the VLSA. A gate area of $0.08\mu m^2$ is chosen with initial dimensions of 160nm length and 500nm width for an estimated $\sigma_{\Delta V_t} = 8.34mV$.

PMOS pull-ups P1/P2 are given a length of 120nm to significantly reduce DIBL, and the width kept at the device default of 200nm. Bitline access transistors P3,4 are kept at the default width of 200nm, and an increased length of 80nm to reduce leakage. Footer switch NMOS N3 is kept at minimum length and given an increased width of 300nm for increased drive strength. PMOS switch P5 is given a length of 80nm, and a total width of 600nm. The pull-up path to VDD is weak here, so giving a higher W/L ratio can improve drive strength. For layout considerations, this device is split into 2 fingers of 300nm width, which can allow for symmetric placement around P1 and P2. Table 3.1 below provides a summary of the chosen device sizes.

Table 3.1 DVLSA Transistor Sizes

Device	Width (nm)	Length (nm)	Fingers
N1, N2	500	160	1
N3	300	60	1
P1, P2	200	120	1
P3, P4	200	80	1
P5	300	80	2

3.3 Reference Circuit – VLSA

A standard VLSA, as shown in Figure 2.7, will be used as a reference for evaluating the DVLSA's performance. The topologies of the VLSA and DVLSA are very similar, so this reference VLSA implementation is designed with the same device sizes as the DVLSA as in Table 3.1. One exception to this is P5 since it does not exist in VLSA. A benefit of using the same device sizes is that each device should have the same threshold variation when operating individually. This gives higher confidence that any difference in circuit performance is due to the topology and control mechanism rather than just Pelgrom scaling.

3.4 Testbench Architecture

It is often possible to characterize some aspects of analog amplifiers and digital circuitry using only DC simulations. As an example, input offset voltage for an op-amp can be found with a simple DC operating point simulation or a DC sweep. Some circuits however are dynamic and their performance must be evaluated from a time-domain simulation. As mentioned in Section 2.4, latch type sense amplifiers are a form of a dynamic comparator, and thus fall into this latter category of circuits.

This section discusses the testbench architecture and the different testbenches used for evaluating the selected sense amplifier designs. Two testbenches are used. The first testbench is the fixed ΔV_{BL} testbench as seen in Section 3.4.5. It is used to evaluate the nominal operation of the device, and to measure parameters such as sensing delay. The second testbench is the dynamic testbench as seen in Section 3.4.6. It is used to characterize input offset voltage and device mismatch contribution.

Each testbench can be used for both single-point measurements and multi-point statistical runs. The Monte Carlo method is used for the statistical variation of device parameters, and 1000 single points are used per statistical run.

3.4.1 Sense Amp Selection Cell (SA_SEQ_SWEEP)

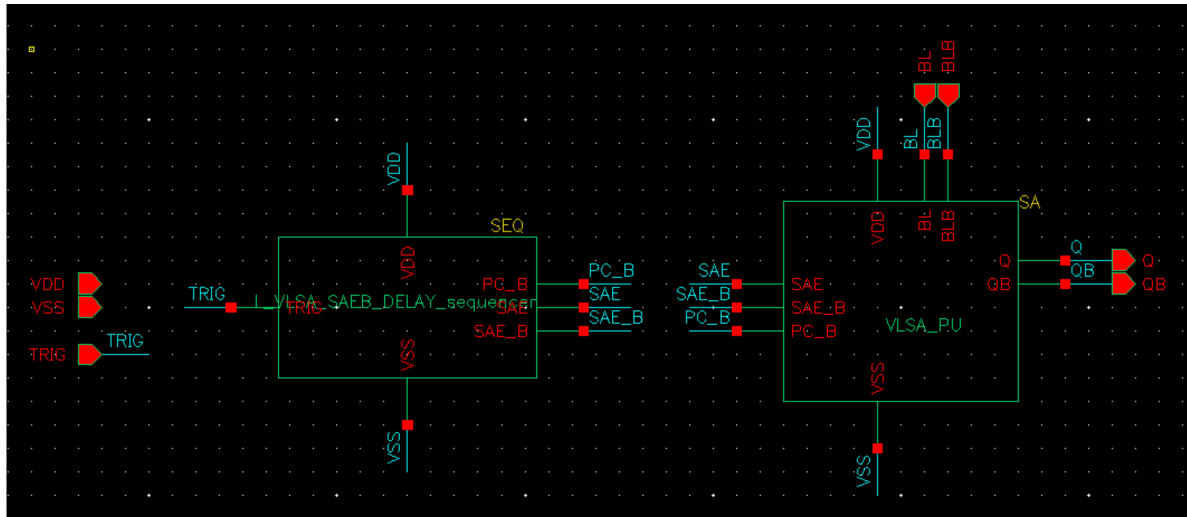

Figure 3.3 DVLSA Implementation for SA_SEQ_SWEEP

For each sense amplifier topology tested, the sense amplifier core and its corresponding timing circuitry are implemented in separate cells. They are then assembled in uniquely implemented schematic views of a cell named SA_SEQ_SWEEP. Figure 3.3 shows the implementation of this cell for the DVLSA. An instance of SA_SEQ_SWEEP is instantiated at the top level of each testbench.

This configuration encapsulates topology-specific signals within the SA_SEQ_SWEEP cell. Only the common signals VDD/VSS, Q/QB, BL/BLB, and TRIG are exposed as ports. By using Virtuoso's config view hierarchies at the top-level schematic of the testbench one can easily swap a SA_SEQ_SWEEP instance's implementation at netlisting time to change the device under test

(DUT). The config views can also be used to swap out variants of timing blocks and core blocks for a given topology. The major advantage of this is the ability to have a single testbench and simulation environment that is shared across every DUT.

3.4.2 Sense Amplifier Cores

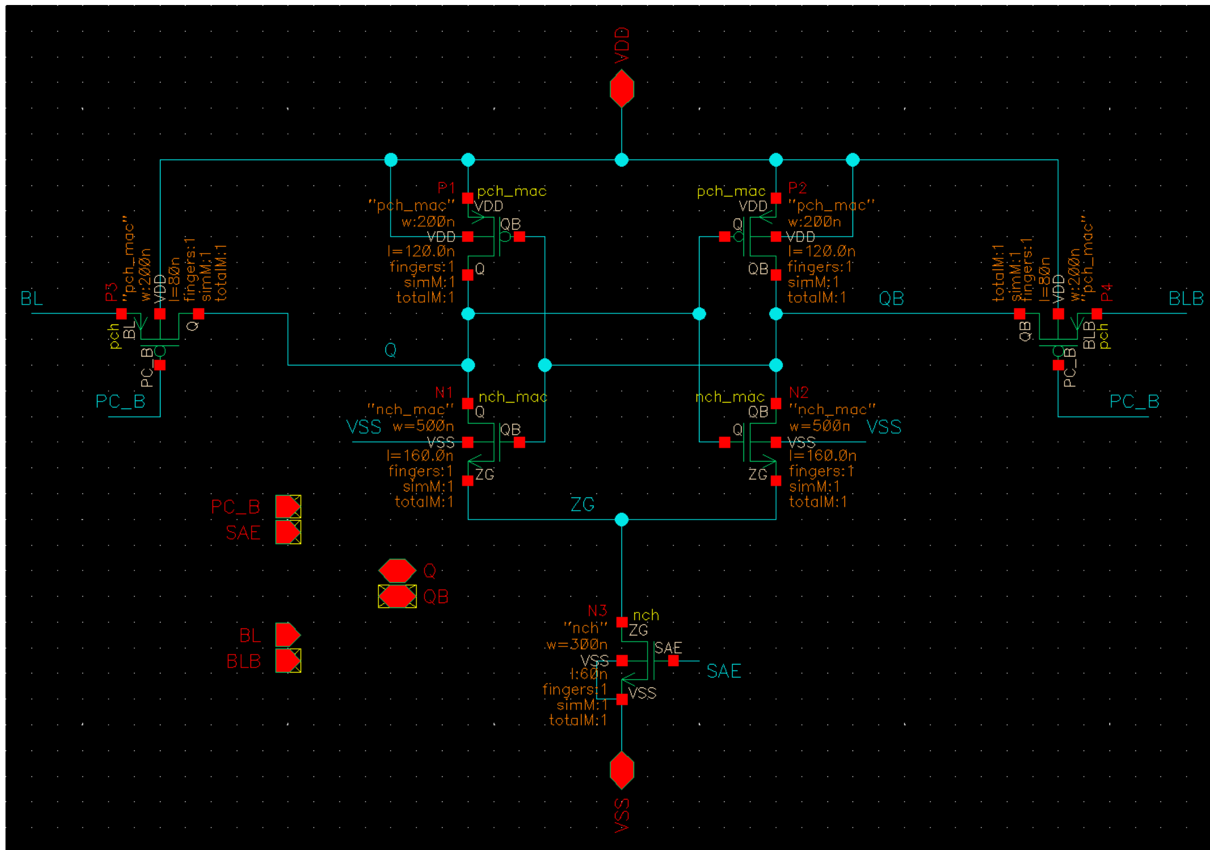

Figure 3.4 VLSA Sense Amplifier Core Implementation

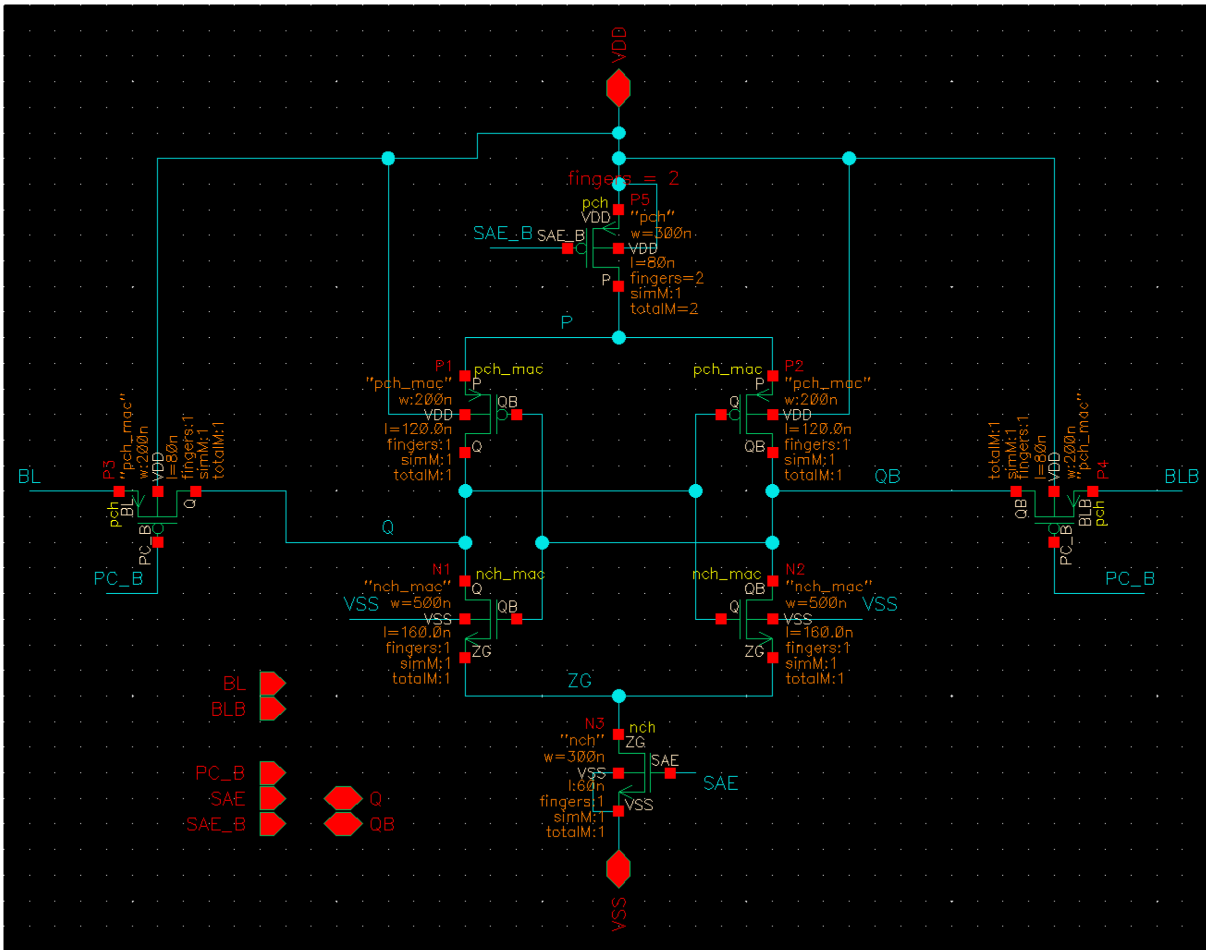

Figure 3.5 DVLSA Sense Amplifier Core Implementation

Figure 3.4 and Figure 3.5 show the implementation of the sense amplifier core cells for the VLSA and DVLSA, respectively. The device sizes in each cell match what is displayed in Table 3.1. The nch and pch instances represent transistors using the standard V_t models of the TSMC 65nm PDK. nch_mac and pch_mac are variants on the standard nch and pch devices; they operate the same under nominal conditions but enable the local variation of that instance's parameters in statistical simulations. By default, the regular nch and pch devices are used for each instance; they are only swapped to the nch_mac and pch_mac variants if a test requires them to be varied.

3.4.3 Timing Blocks

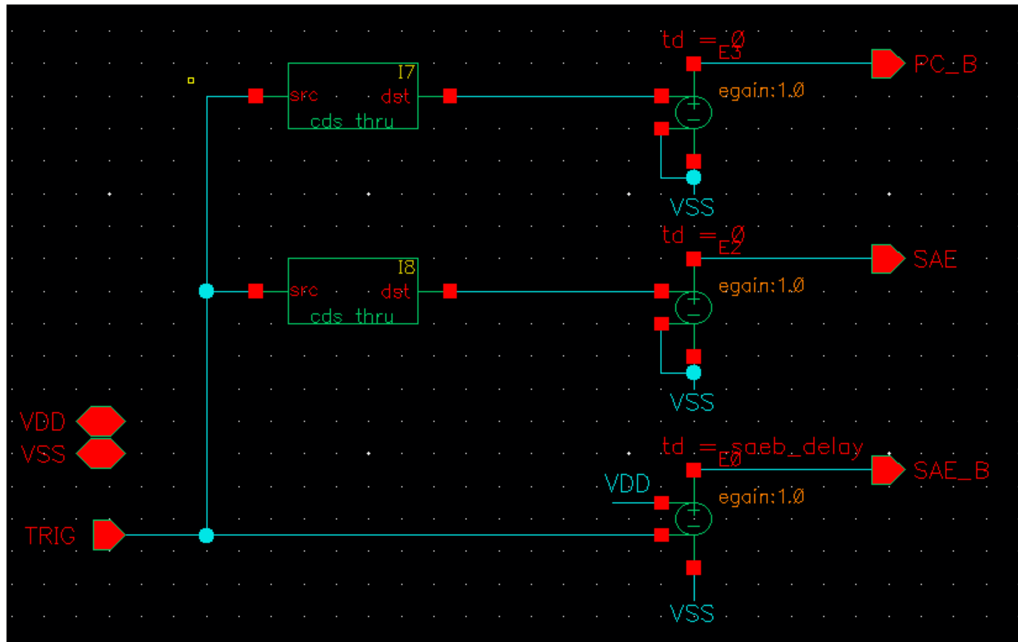

Figure 3.6 DVLSA Timing Cell

Each sense amplifier tested has a custom timing cell. All timing cells have a common TRIG input from which the design-specific output signals are derived. For consistency in simulation, each implementation is composed of a set of voltage-controlled voltage sources (VCVS). This allows for consistent rise times to be applied to every output signal and is dependent only on the input trigger signal. Each VCVS can also have a fixed propagation delay specified as an argument set by the testbench environment. Figure 3.6 shows the timing block implementation for the DVLSA following this methodology. The implementation for the reference VLSA's timer is similar, but all VCVS delays are set to 0.

3.4.4 Output Decision Block

Figure 3.7 Output Decision Block

Figure 3.7 shows the output decision block, tb_SA_SEQ_decision, which is implemented in Verilog-A (refer to Appendix A for the code). This block continuously reads the analog signals of the sense amplifier's Q and QB output nodes and converts it into a binary state for a testbench to sample on its clock edges. SGN determines if (VQ-VQB) is positive or negative, and DECISION tests if VQ>VQB. These are simple calculations that could output a favourable decision even if the sense amplifier's output delta is quite small. An actual circuit sampling the sense amplifier output would most certainly be unable to correctly determine the output state in this case. Therefore CRIT_SGN and CRIT_DECISION are provided to decide on the sense amplifier's output with additional conditions. CRIT_SGN is programmed to output the sign of the input delta only if the magnitude is above a specified threshold (default to 90% of VDD). Otherwise, it remains at zero. CRIT_DECISION is programmed to output a binary test of if CRIT_SGN equals 1.

Using CRIT_DECISION as the signal to sample allows a testbench to determine if the sense amplifier can read a logic '1' from its inputs and produce a large enough output delta for other circuits in the system to correctly interpret the result.

3.4.5 Fixed ΔV_{BL} Testbench

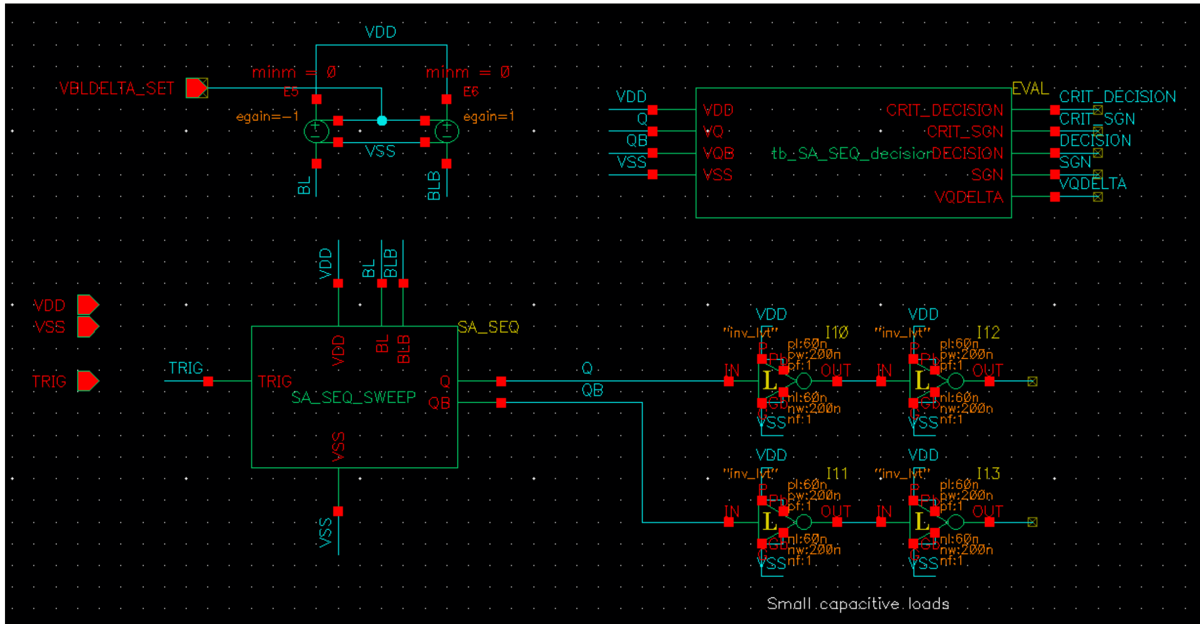

Figure 3.8 Fixed ΔV_{BL} Testbench

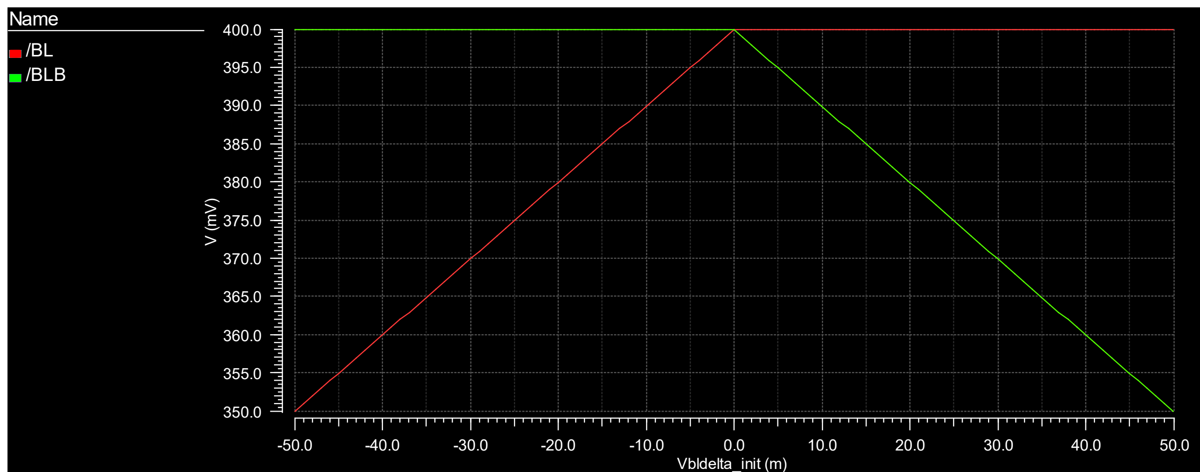

Figure 3.9 V_{BL} and V_{BLB} Setup vs ΔV_{BL} (VDD=400mV)

Figure 3.8 shows a testbench that allows for sense amp reads with a fixed ΔV_{BL} on the bitlines. Two voltage sources are used to set the voltages on BL and BLB using a single differential control voltage VBLDELTA_SET. Figure 3.9 shows how BL and BLB are set up for a given ΔV_{BL}.

This configuration limits the max voltage on either of the lines to be VDD, regardless if ΔV_{BL} is positive or negative.

The testbench can run a single read cycle (default) or multiple read cycles. This testbench allows for inspection of the DUT's operation at nominal conditions, and statistical characterization of directly measurable parameters such as read time (t_{sense}). To model a realistic capacitive load on the sense amplifier output that would be expected in a physical implementation, a small chain of near-min-sized inverters is added to the outputs of Q and QB.

3.4.6 Successive Approximation of ΔV_{BL} Testbench

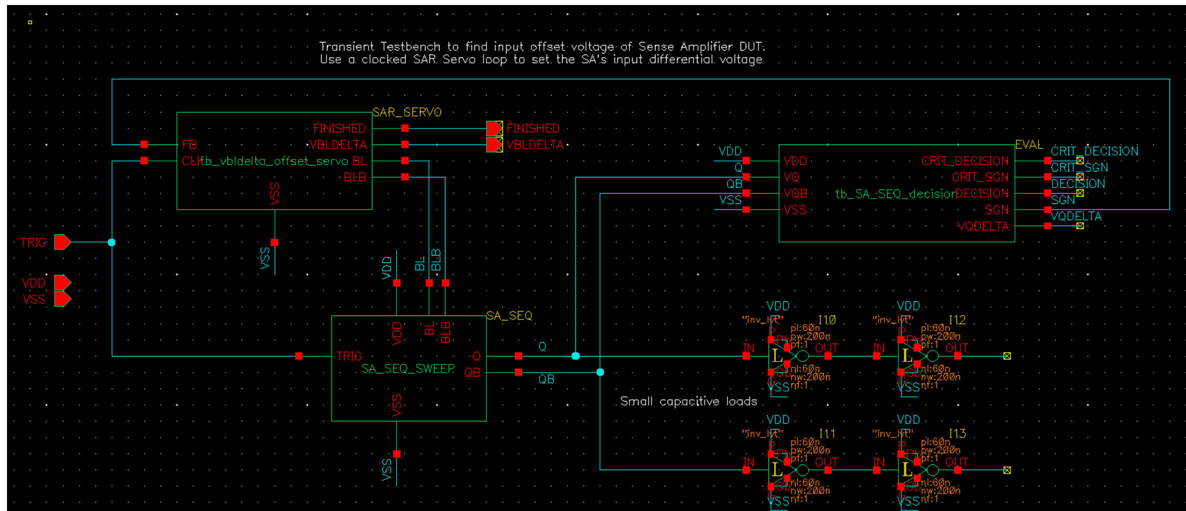

Figure 3.10 Successive Approximation of V_{OS} Testbench

Figure 3.10 shows the testbench for input offset characterization using a Successive Approximation (SAR) algorithm in a clocked servo loop. The benefit of a SAR algorithm is its logarithmic time complexity, which offers a significant reduction in simulation time for high-resolution tests [21]. If the controller starts the search at 0 and the first step is VDD, the resolution of the search will be equal to $V_{LSB} = \frac{VDD}{2^{N-1}}$, where N is the number of cycles.

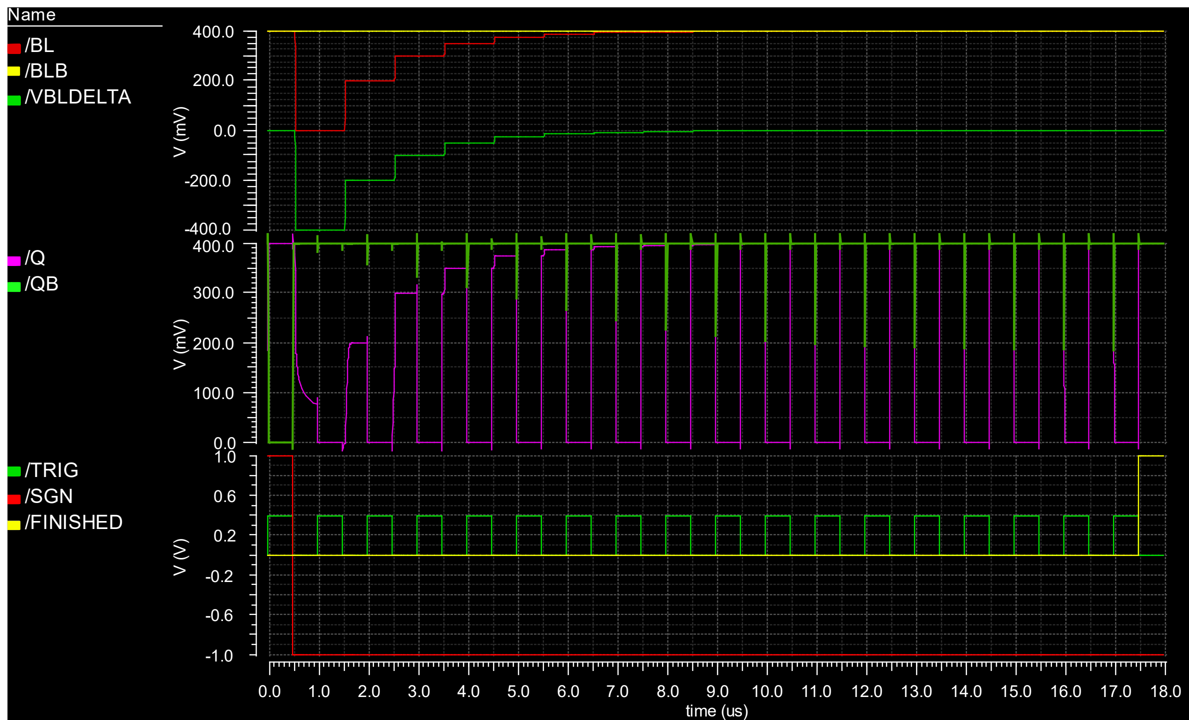

Figure 3.11 Example Waveforms for SAR Testbench

Figure 3.11 shows an example set of waveforms for this testbench. The SAR controller sets BL and BLB voltages with a differential voltage of BL_DELTA and follows the same function seen in Figure 3.9 in the fixed ΔV_{BL} testbench. On the rising edge of the testbench clock, the sense amplifier DUT will perform its sensing operation. The evaluator block will read the differential output voltage of the sense amplifier and apply some function to the output. The result is fed back to the SAR controller so that it can decide to step up or down for the next read cycle. Once enough cycles are performed to reach the desired accuracy level, the FINISH signal is asserted immediately so that the testbench measurements can read the last used BL_DELTA quantity. This becomes the resulting input offset voltage for the simulation.

Refer to Appendix A for the Verilog-A code of the SAR controller.

3.5 Tests

3.5.1 Nominal Waveform Comparison

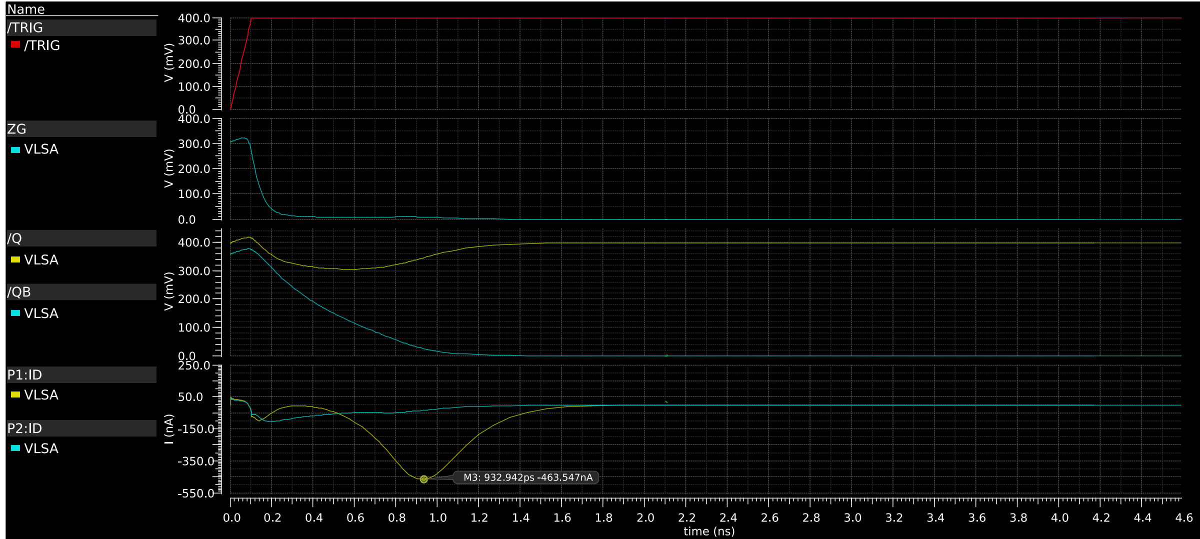

Figure 3.12 VLSA Nominal Waveforms ($\Delta V_{BL} = 40mV$)

This test uses the basic testbench from Section 3.4.5 to perform a sensing operation for the VLSA and DVLSA under nominal conditions with the same selected parameters. By the circuit behavior predicted in Section 3.2.1, there should be some duration where the P1 and P2 transistors are off and thus do not influence the discharging of the Q/QB nodes. To evaluate this, we will examine the waveforms for the drain currents of P1 and P2. The waveform for the ZG node will also be used to help determine if there is any subthreshold leakage through the PMOS devices. Figure 3.12 shows an example set of these waveforms collected for the reference VLSA implementation at $\Delta V_{BL} = 40mV$.

3.5.2 Statistical Input Offset – Isolated PMOS Pair

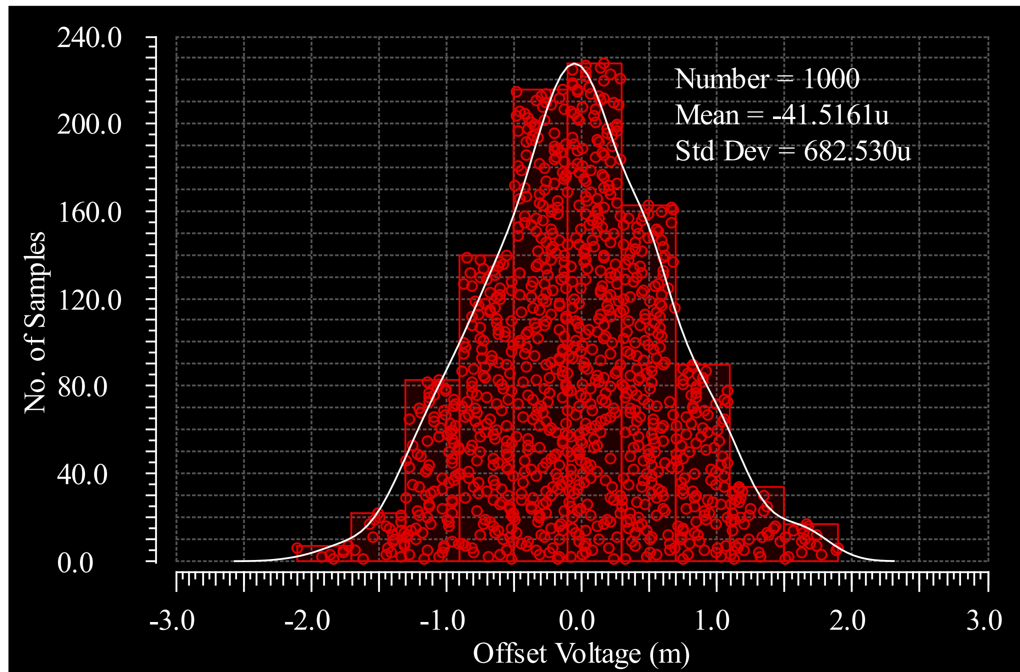

Figure 3.13 VLSA Input Offset Voltage Histogram – PMOS Pair Only

This test will run a Monte Carlo simulation on SAR testbench from Section 3.4.6 with only the PMOS pair devices P1 and P2 varying. The goal of this test is to determine if the DVLSA topology can achieve a smaller standard deviation of V_{os} relative to the VLSA when the PMOS pair devices P1 and P2 are varied in isolation from other devices. Since only two devices from a single pair are being varied there is no need to perform a mismatch contribution analysis. Figure 3.13 above shows a set of this data collected for the reference VLSA implementation.

3.5.3 Statistical Input Offset – All PMOS and NMOS Symmetric Pairs

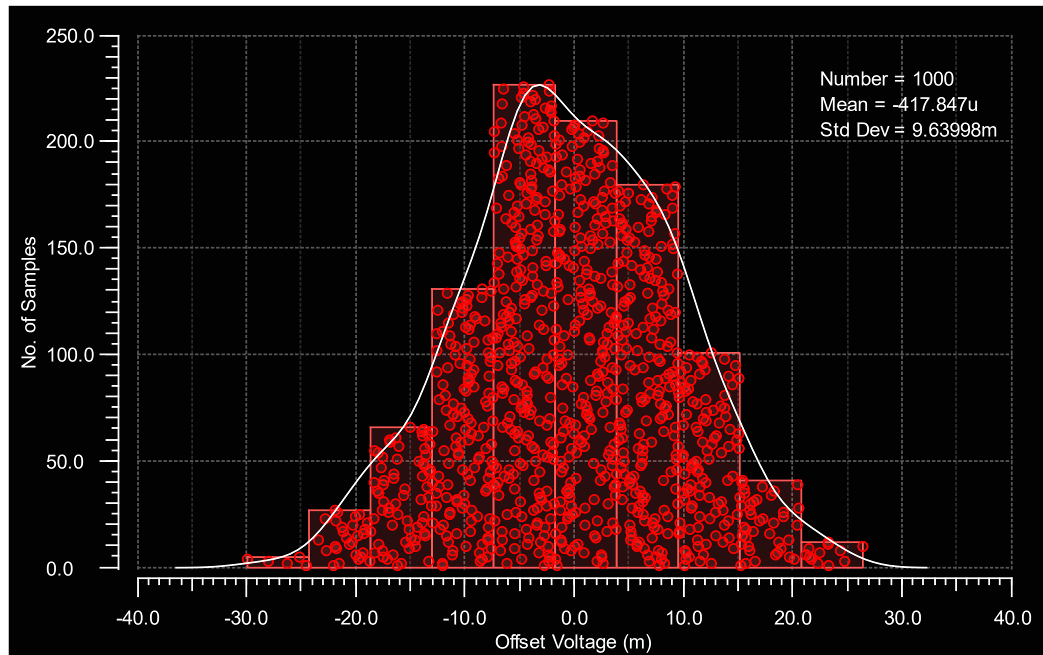

Figure 3.14 VLSA Input Offset Voltage Histogram – PMOS and NMOS Pairs

This test will run a Monte Carlo simulation on SAR testbench from Section 3.4.6 where all differential devices are being varied. This set of devices includes the NMOS cross pair N1/N2, the PMOS cross pair P1/P2, and the PMOS access pair P3/P4. Histogram plots and input offset voltage statistics are to be collected. Figure 3.14 above shows a set of this data for the reference VLSA implementation. The goal of this test is to determine if the DVLSA topology can achieve a smaller standard deviation of V_{os} relative to the VLSA when all the devices in the NMOS and PMOS pairs are varied. In a physical implementation of this circuit, all devices will have some variation. Therefore, testing with most of the devices varying will provide a more realistic prediction of the circuit's performance than the previous isolated PMOS test.

Note that the TSMC 65nm PDK uses two standard gaussian parameters per transistor to determine the variation of its properties. For NMOS they are parn1 and parn2, and for PMOS they are parp1 and parp2. In the test from the previous section, there is only a single device pair varying. As such, all variation to input offset voltage is directly attributable to that one device pair, hence these varied input parameters are not as important. In this test, however, there are several devices across multiple device pairs that are being varied, and thus we cannot directly determine how much each

device pair affects the input offset voltage. To obtain that level of information it becomes necessary to perform a mismatch contribution analysis.

The mismatch contribution analysis tool in Virtuoso performs a multivariate regression on a measured result to correlate its variation with one or more varied input parameters. The tool uses this information to compute how much each varied input parameter contributed to the output parameter's variation and reports these quantities as a percentage of the total variance of the output parameter. In this test, the target result to measure is the input offset voltage, and the varied input parameters are the parn and parp variables of each varying device. By summing the variance contribution of parn1, parn2, parp1, and parp2 for each device we can obtain an estimate of how much that device contributed to the input offset voltage variation. This step can be repeated across devices to determine how much a pair or group of devices contributed to the input offset voltage variation.

3.6 Test Conditions

Table 3.2 Testbench Default Parameters

Technology	TSMC 65nm GP
Transistor V_t Type	Standard V_t
Temperature	27 $°C$
V_{DD}	0.4V
f_{clk}	1MHz
t_{rise}	250ps
SAEB_delay	2ns
SAR Testbench Cycles	18
SAR Testbench V_{LSB}	3.05uV

All tests are conducted with the conditions in Table 3.2 unless otherwise specified. Due to how the SAR controller is programmed, the sense amplifier experiences differential voltage steps as large as VDD in the early cycles of the offset testbench. At the low voltage of VDD=0.4V, the transistors are weaker so the precharge response to voltage steps has a large time constant. A frequency of 1MHz ensured that there is more than enough time for the internal nodes to settle for most of the input range. The SAR testbench uses 18 clock cycles to locate the offset voltage. Using the equation in Section 3.4.6 for SAR resolution, the input offset measurement should be accurate to: $V_{LSB} = 0.4V * 2^{-(18-1)} = 3.05$uV.

Chapter 4

Simulation Results and Analysis

4.1 Nominal Waveform Comparison

For the results presented below, a SAEB_delay of 2ns is used. The nominal waveforms are collected for a ΔV_{BL} of 10mV and 40mV to compare the performance when using a small input voltage versus a large input voltage. For each SAEB configuration of the DVLSA tested, the plots of the collected waveforms have the waveforms of the reference VLSA superimposed to simplify the comparison between the circuits.

Note that some of the plots below report the region of operation of certain devices. These waveforms are integer enumerations of the operating region as described in the BSIM manual: 0 = cutoff, 1 = triode, 2 = saturation, and 3 = subthreshold.

4.1.1 Large input Voltage ($\Delta V_{BL} = 40mV$)

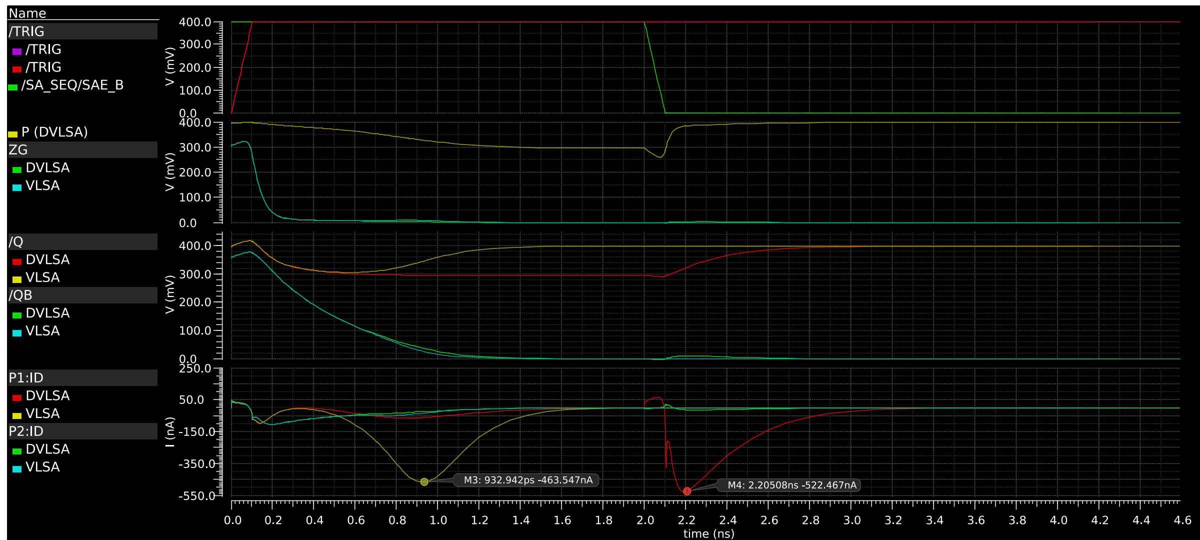

Figure 4.1 Nominal Comparison of VSLA vs DVLSA ($\Delta V_{BL} = 40mV$, SAEB_delay=2.0ns)

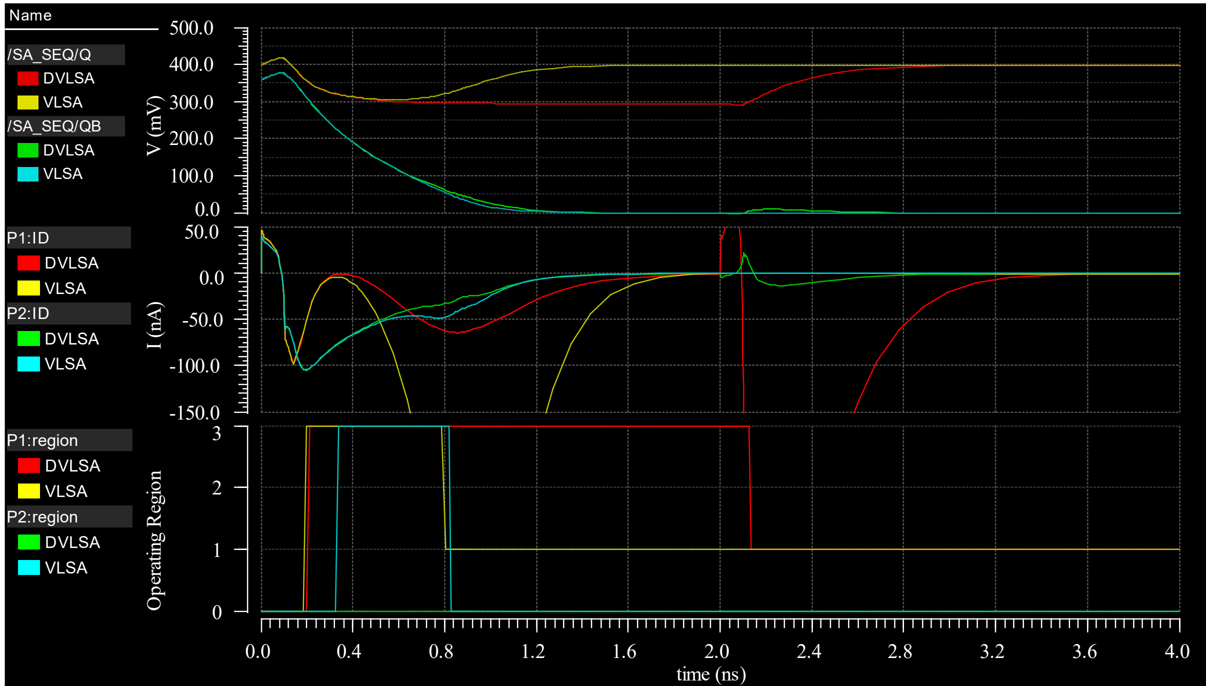

Figure 4.2 Nominal Comparison of Currents and Operating Regions ($\Delta V_{BL} = 40mV$, SAEB_delay=2.0ns)

Figure 4.1 shows the nominal waveforms of the DVLSA with a SAEB_delay of 0ns VLSA when $\Delta V_{BL} = 40mV$. Initial observations of the Q and QB waveforms at the rising edge of TRIG that the DVLSA appears to perform as expected in the concept description, with the exception being the transient spike at the rising edge of TRIG. We see as the ZG node gets discharged to VSS through N3, both Q and QB discharge but at different rates. At about 0.5ns we see that the Q waveform of both the VLSA and DVLSA levels off to approximately 300mV, while QB continues to discharge. In the VLSA, Q reverses direction and charges back to VDD, while in the DVLSA the Q node remains nearly flat at this level. This suggests that the PMOS pair of the VLSA turned on and the circuit's regenerative feedback started, while in the DVLSA the PMOS pair remained off.

To confirm these details, we analyze Figure 4.2. This figure plots the regions of operation for the PMOS pair P1 and P2, in addition to a zoomed plot of these devices' drain currents. The Q and QB waveforms are included as a common reference with the previous figure. At the rising edge of TRIG, we can see the transient spikes in the current waveforms. From 0.0 to 0.2ns the region of operation reported for P1 and P2 across both the VLSA and DVLSA are reported to be in the cutoff

region. In this region of operation, these devices should not conduct any current, so these current transients are most likely due to capacitive coupling with the control signals.

After 0.2ns we see P1 in both the VLSA and DVLSA entering the subthreshold conduction region. At 0.3ns we see P2 for the VLSA entering the subthreshold conduction region, but P2 for the DVLSA remains in the cutoff region. P2 for the DVLSA remaining in cutoff matches our expectations for the circuit's operation, but P1 going into subthreshold differs. To explain this discrepancy, we note that at 0.5ns Q has discharged to 355mV, QB to 310mV, and the P node to 391mV. This gives a $|V_{GS}|$ on P1 of 81mV while $|V_{GS}|$ on P2 is less than half that at 36mV. Both Q and P continue to discharge with Q being much more rapid than P. This grows the $|V_{GS}|$ of P1, pushing it closer to the saturation region. QB however slows its rate of discharge. With the P node discharging through P1, there is a point where the $|V_{GS}|$ of P2 starts to decrease. Near 0.5ns this $|V_{GS}|$ reaches a peak of 66mV, which is less than the 81mV that P1 experienced when it transitioned to subthreshold conduction.

Another interesting feature of the current waveforms is the P2:ID current behavior that occurs just after the transient spike. We note that it has a negative peak of approximately -100nA and it takes more than 1ns to decay to near zero. As stated previously, P1 for the VLSA and DVLSA transition to the subthreshold region around 0.2ns. At this time, P1:ID is -60nA to -45nA between the cases, while P2:ID is -100nA. Even though the P2:ID current is double that of P1:ID when P1 is considered to be in the subthreshold region, P2 is still considered to be in the cutoff region. Combined with the earlier observation that the $|V_{GS}|$ of P2 around this time is about half that of P1, we would expect if anything that the P2 current is lower than the current of P1. If the current is not due to P2 conducting, then we suspect it is due to the rapidly declining QB voltage that induces some capacitive coupling between P2's drain node and other internal nodes.

4.1.2 Small Input Voltage ($\Delta V_{BL} = 10mV$)

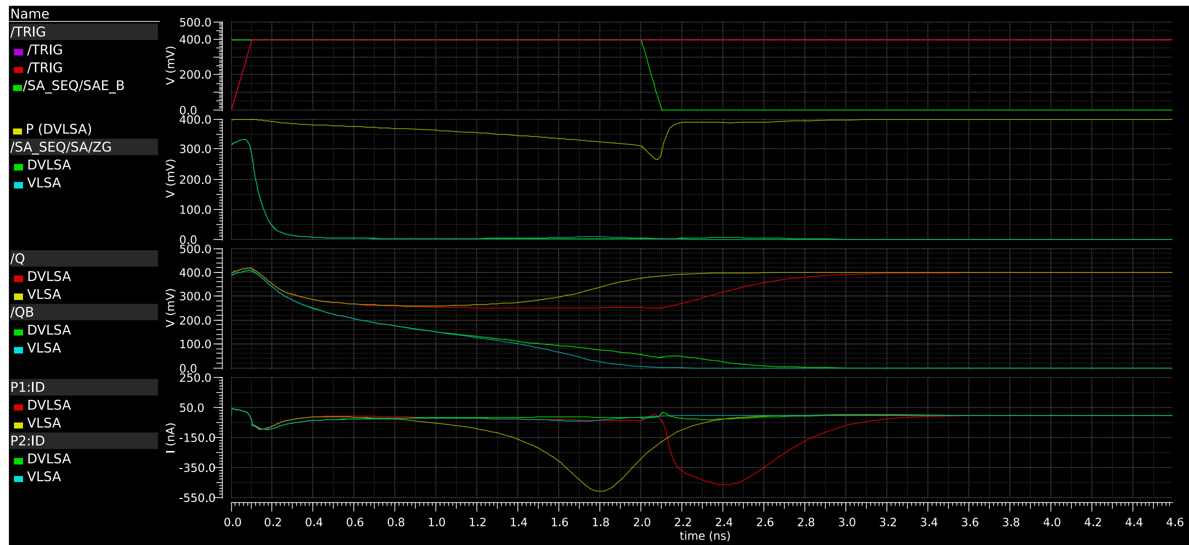

Figure 4.3 Nominal Comparison of VSLA vs DVLSA ($\Delta V_{BL} = 10mV$, SAEB_delay=2.0ns)

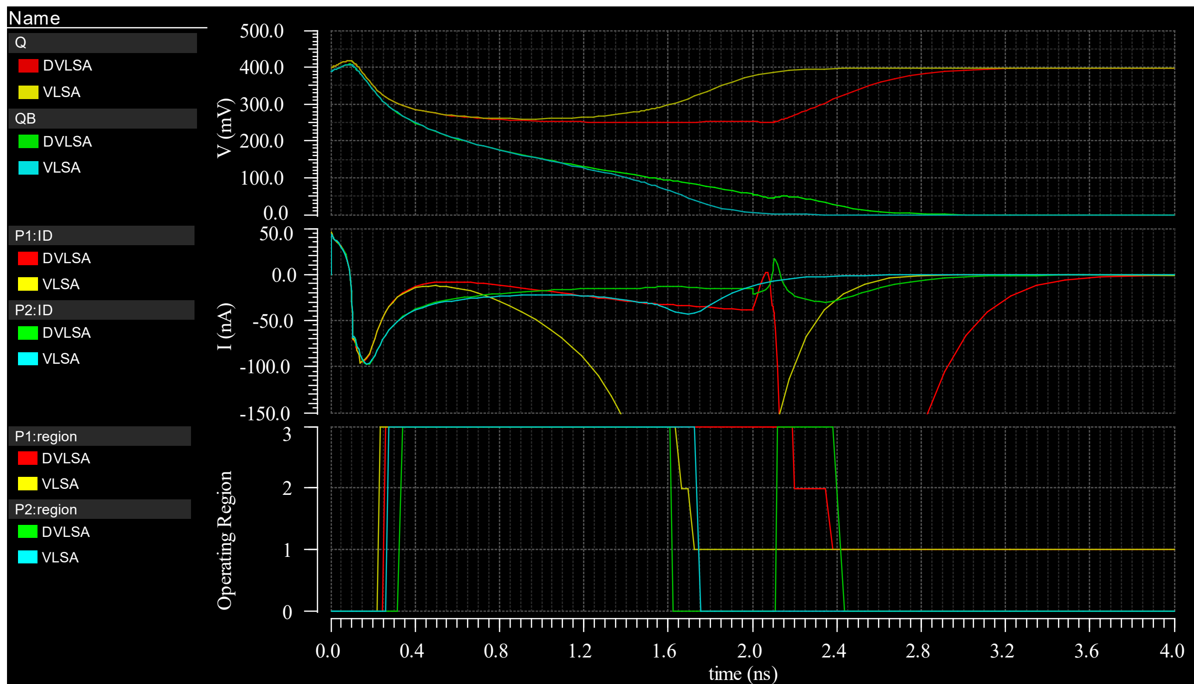

Figure 4.4 Nominal Comparison of Currents and Operating Regions ($\Delta V_{BL} = 10mV$, SAEB_delay=2.0ns)

Figure 4.3 shows the nominal waveforms of the DVLSA with a SAEB_delay of 0ns VLSA when $\Delta V_{BL} = 10mV$, and Figure 4.4 is a plot of P1 and P2's drain currents and their operating regions. Initial observations of the Q and QB waveforms indicate that they follow the conceptual waveforms. After the initial transient spike, both Q and QB begin to decay at different rates. The differential voltage develops at a slower rate than the previous test case due to the smaller initial input voltage delta. The consequence of this is that both Q and QB decay further before N1 shuts off and Q levels off. The voltage that Q settles at is approximately 250mV, which is 50mV lower than the 300mV in the previous test case.

A consequence of the lower Q and QB voltages combined with the slower discharge of the P node is that both P1 and P2 enter the subthreshold region and conduct a small amount of current. After both turn on, we note that both currents remain less than 50nA in magnitude for the remainder of the time until SAEB is activated.

We once again direct our attention to the P2:ID current behavior just after the transient spikes. Notice that it decays towards zero much more rapidly than the other test case. It experiences a sharper kink in the 0.3ns region. At this time the P2 device goes into the subthreshold region, and the current waveform decays to zero at a noticeably slower rate. Correlating this initial rapid decay of the current with the discharging of QB that is noticeably slower than the previous test case, this further supports the suggestion that this current is due to capacitive coupling through P2's drain and not channel conduction.

4.1.3 Summary

The 10mV and 40mV cases tested have Q and QB waveforms that perform as expected in the DVLSA's concept description. However, the details of the PMOS pair devices' operating regions are not entirely as expected. We predicted that the DVLSA would keep the PMOS pair off until SAEB is activated, but the simulation results showed that this is not entirely true. While the devices stayed out of the saturation and triode regions during the SAEB_delay period, they did manage to enter the subthreshold conduction region. For the large input voltage case, P2 remained in the cutoff region as predicted but P1 entered the subthreshold conduction region. For the small input voltage case, both P1 and P2 entered the subthreshold conduction region between 0.2ns to 0.3ns.

Note that for much of the SAEB_delay time, the drain currents of P1 and P2 remained under 50nA. This is roughly a factor of 10 lower than the peak currents achieved once SAEB is activated, or

the peaks in the VLSA when the regenerative feedback starts. Although the PMOS pair devices may not truly be off, the currents that pass are small enough that we believe they should not cause any issues once statistical variation is incorporated.

A general observation is that the waveforms in the $\Delta V_{BL} = 10mV$ test case appear to be stretched out in comparison to the $\Delta V_{BL} = 10mV$ test case. Due to the smaller initial input voltage and the slower rate of signal development the Q and QB nodes of the DVLSA and VLSA are similar for a longer period before the VLSA's regenerative feedback activates. For smaller input voltages or voltages that are close to the sense amplifier's input offset voltage, this suggests that short SAEB_delay quantities may not have much of an effect and that the performance improvements will be expected with the longer SAEB_delay quantities.

We also noticed some odd behavior in the P2:ID waveform just after the initial transient spike that is likely due to capacitive coupling. With the data collected up until this point, we cannot make a definitive claim on whether this has some effect on the statistical performance of the circuit. This feature may warrant further investigation in future research.

4.2 Statistical Input Offset – Isolated PMOS Pair

Monte Carlo simulations are performed with only the PMOS pair devices P1 and P2 varying. The statistics and histogram plot of the sense amplifier's input offset voltage is recorded for each statistical run. For the DVLSA the process is repeated for a SAEB_delay of 0.0ns, 0.5ns, 1.0ns, and 2.0ns. The statistics and histogram for the reference VLSA can be found in Figure 3.13 under Section 3.5.2 where this test is initially described.

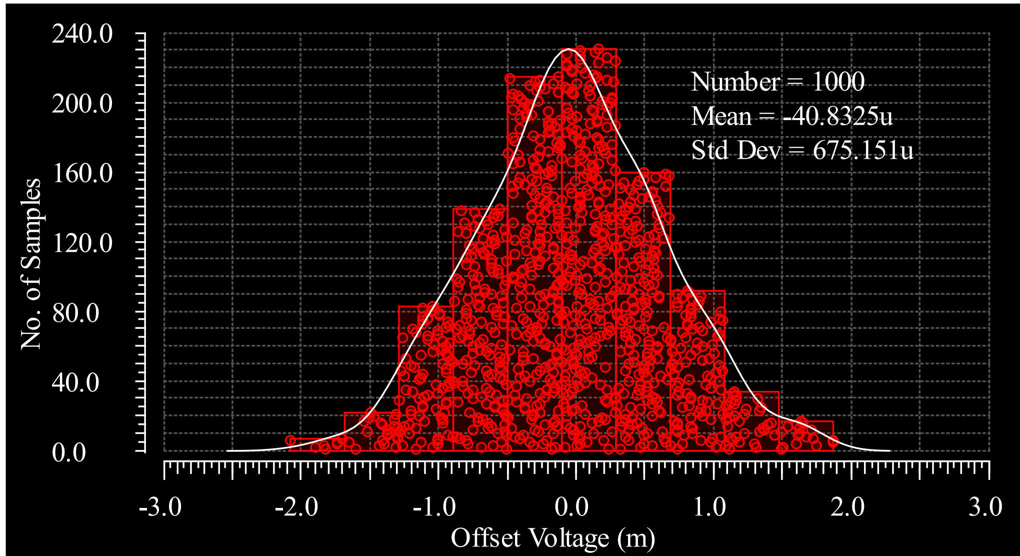

Figure 4.5 DVLSA Input Offset Voltage (SAEB_delay=0.0ns) Histogram – PMOS Pair Only

Figure 4.5 above presents the histogram of the DVLSA's input offset voltage when SAEB_delay is set to 0.0ns. The mean input offset μ_{os} is 40.83uV and the standard deviation σ_{os} is 675.2uV. Compared to (μ_{os}, σ_{os}) = (-41.52uV, 682.5uV) for the reference VLSA, this configuration of the DVLSA exhibits an improvement in both parameters. The nominal waveform analysis in the previous section showed that for the first couple hundred picoseconds the waveforms between the VLSA and DVLSA are near identical. For this reason, we did not expect this DVLSA configuration to show much of an improvement, if at all.

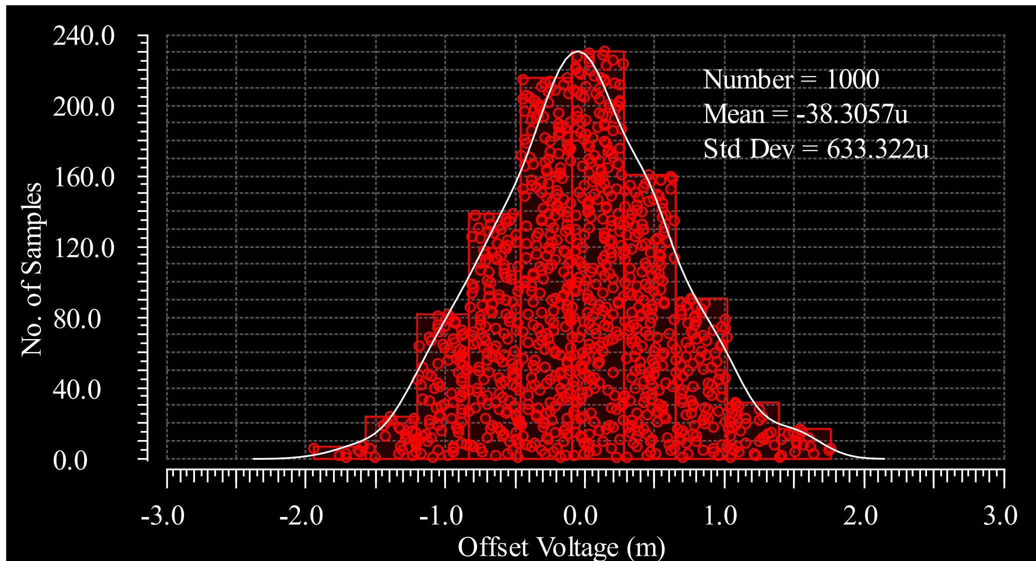

Figure 4.6 DVLSA Input Offset Voltage (SAEB_delay=0.5ns) Histogram – PMOS Pair Only

Figure 4.6 above presents the histogram of the DVLSA's input offset voltage when SAEB_delay is set to 0.5ns. The mean input offset μ_{os} is 38.31uV and the standard deviation σ_{os} is 633.3uV. Compared to (μ_{os}, σ_{os}) = (-41.52uV, 682.5uV) for the reference VLSA, this configuration of the DVLSA exhibits an improvement in both parameters.

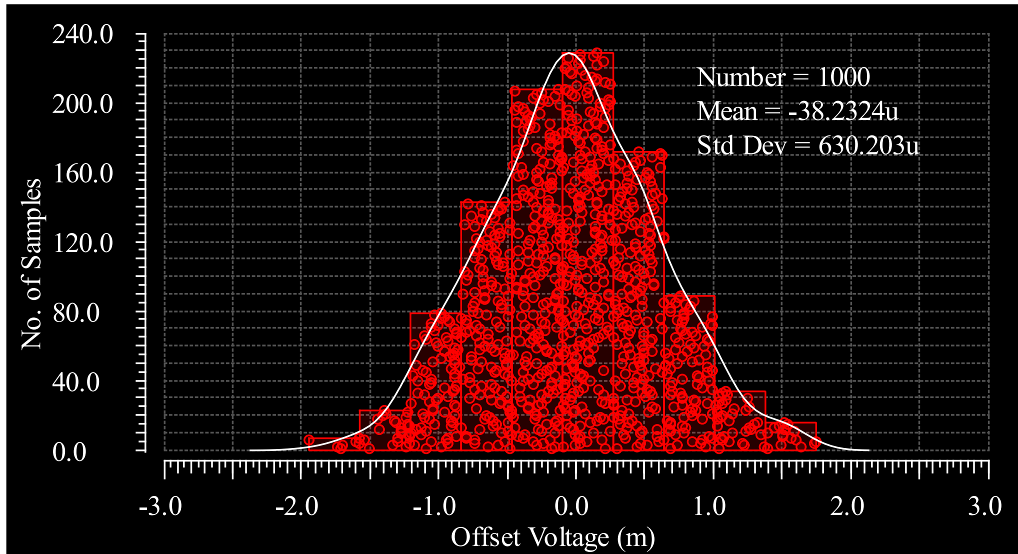

Figure 4.7 DVLSA Input Offset Voltage (SAEB_delay=1.0ns) Histogram – PMOS Pair Only

Figure 4.7 above presents the histogram of the DVLSA's input offset voltage when SAEB_delay is set to 01.0ns. The mean input offset μ_{os} is -38.23uV and the standard deviation σ_{os} is 630.2uV. Compared to (μ_{os}, σ_{os}) = (-41.52uV, 682.5uV) for the reference VLSA, this configuration of the DVLSA exhibits an improvement in both parameters.

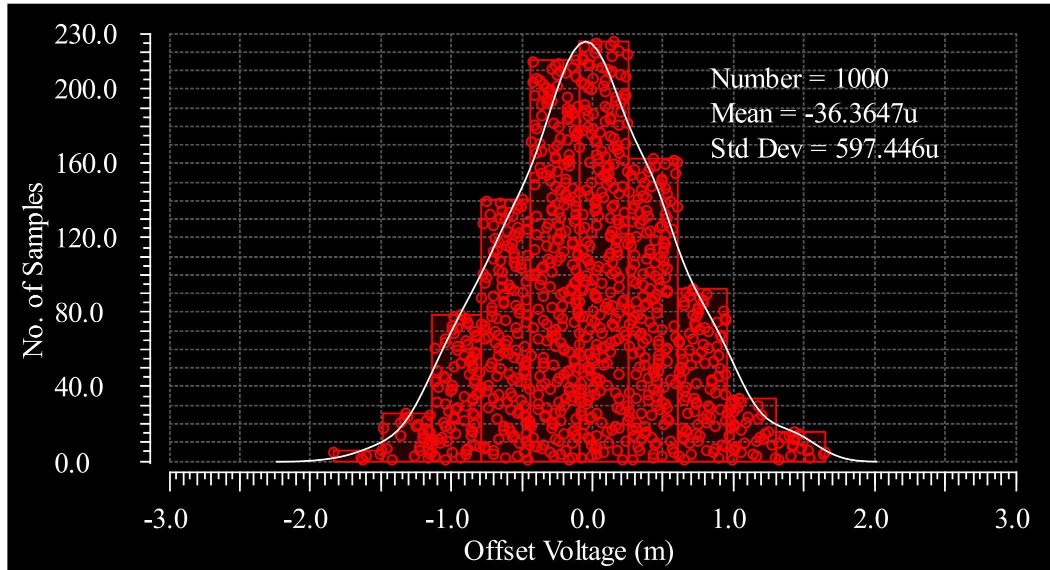

Figure 4.8 DVLSA Input Offset Voltage (SAEB_delay=2.0ns) Histogram – PMOS Pair Only

Figure 4.8 above presents the histogram of the DVLSA's input offset voltage when SAEB_delay is set to 2.0ns. The mean input offset μ_{os} is -36.36uV and the standard deviation σ_{os} is 597.4uV. Compared to (μ_{os}, σ_{os}) = (-41.52uV, 682.5uV) for the reference VLSA, this configuration of the DVLSA exhibits an improvement in both parameters.

Table 4.1 Input Offset Voltage Statistics - Isolated PMOS Pair (P1/P2)

	Value (VLSA)	**Value (DVLSA)**			
SAEB Delay (ns)	-	0.00	0.50	1.00	2.00
$\boldsymbol{\mu_{os}}$ **(uV)**	-41.52	-40.83	-38.31	-38.23	-36.36
$\boldsymbol{\sigma_{os}}$ **(uV)**	682.53	675.15	633.32	630.20	597.45

Table 4.1 shows a summary of the input offset voltage statistics of the DVLSA across SAEB sweeps and the reference VLSA. We can see that across all SAEB delay choices of the DVLSA that both the mean and standard deviation of the input offset voltage are lower than that of the reference VLSA. Additionally, we see a clear trend in both the parameters decreasing as the delay is increased. On a percentage basis, the improvement of μ_{os} ranges from 1.66% to 12.42% and σ_{os} ranges from 1.07% to 12.47%.

These results support the hypothesis of this thesis. However, they cannot be used to make a definitive claim that this architecture does reduce input offset. In a fabricated chip all devices will have some variation of their parameters, so to increase our confidence in this result we must perform additional tests with more devices being varied.

4.3 Statistical Input Offset – All PMOS and NMOS Symmetric Pairs

Monte Carlo simulations are performed with P1, P2, P3, P4, N1, and N2 varying. The statistics and histogram plot of the sense amplifier's input offset voltage is recorded for each statistical run. For the DVLSA the process is repeated for a SAEB_delay of 0.0ns, 0.5ns, 1.0ns, and 2.0ns. The statistics and histogram for the reference VLSA can be found in Figure 3.14 under Section 3.5.3 where this test is originally described.

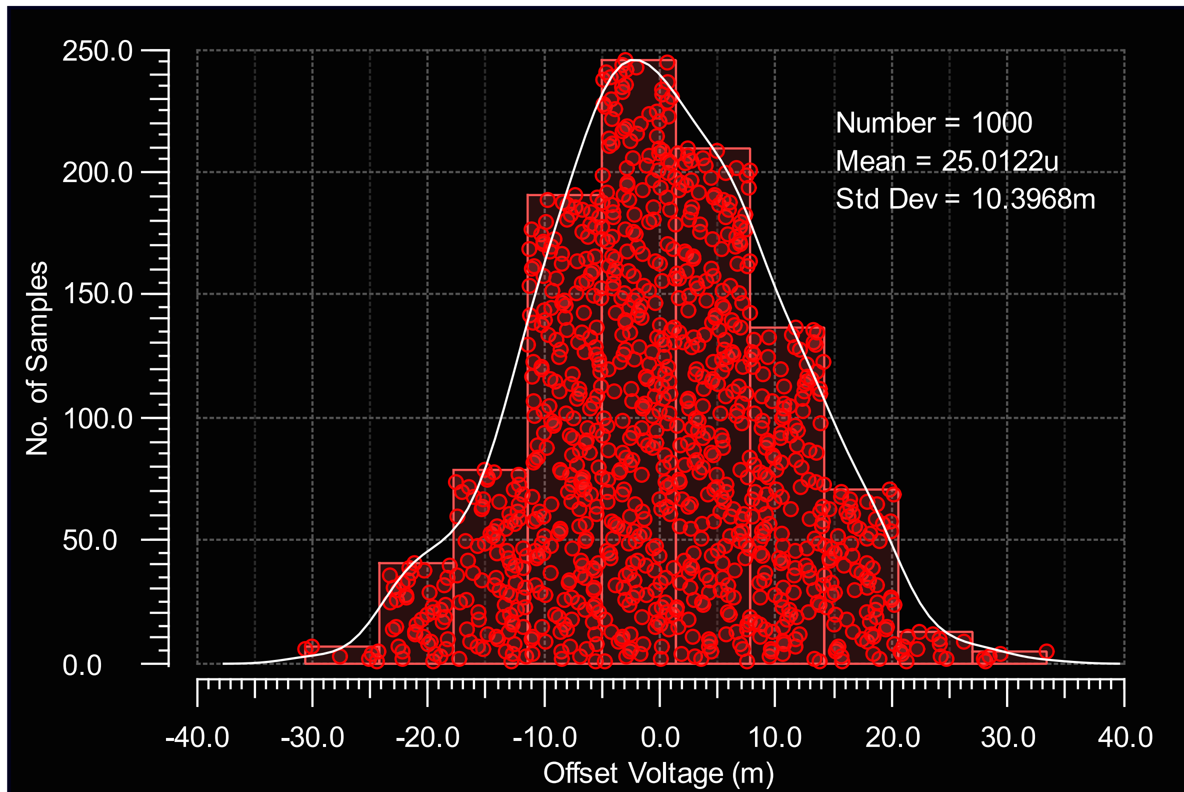

Figure 4.9 DVLSA 0.4V Input Offset Voltage (SAEB_delay=0.0ns)

Figure 4.9 shows the input offset voltage histogram of the DVLSA at 0.4V with a SAEB_delay of 0ns. It is centered at 42.35uV with a standard deviation of 10.40mV. Compared to

(μ_{os}, σ_{os}) = (-417.8uV, 9.64mV) for the reference VLSA, this configuration of the DVLSA exhibits an improvement in μ_{os} by 94.00% and a degradation in σ_{os} by 7.88%.

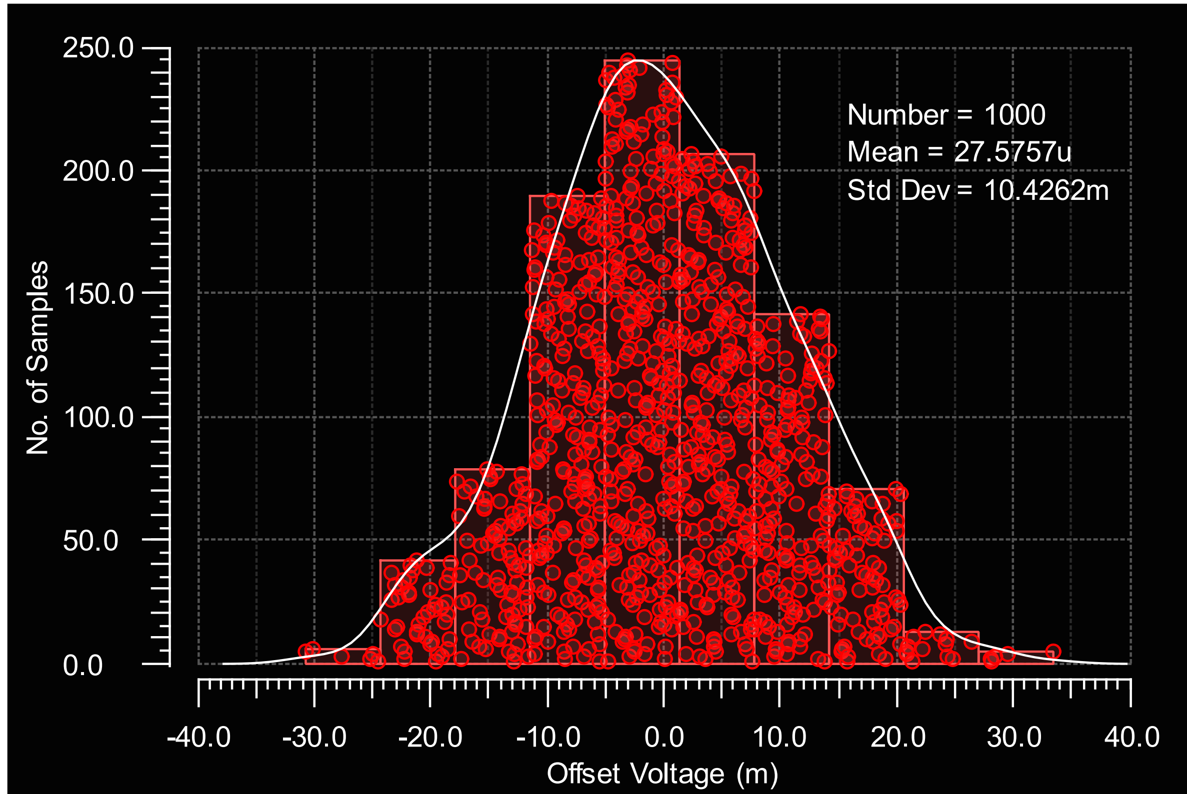

Figure 4.10 DVLSA 0.4V Input Offset Voltage (SAEB_delay=0.5ns)

Figure 4.10 shows the input offset voltage histogram of the DVLSA at 0.4V with a SAEB_delay of 0.5ns. It is centered at 27.58uV with a standard deviation of 10.43mV. Compared to (μ_{os}, σ_{os}) = (-417.8uV, 9.64mV) for the reference VLSA, this configuration of the DVLSA exhibits an improvement in μ_{os} by 93.38% and a degradation in σ_{os} by 8.20%.

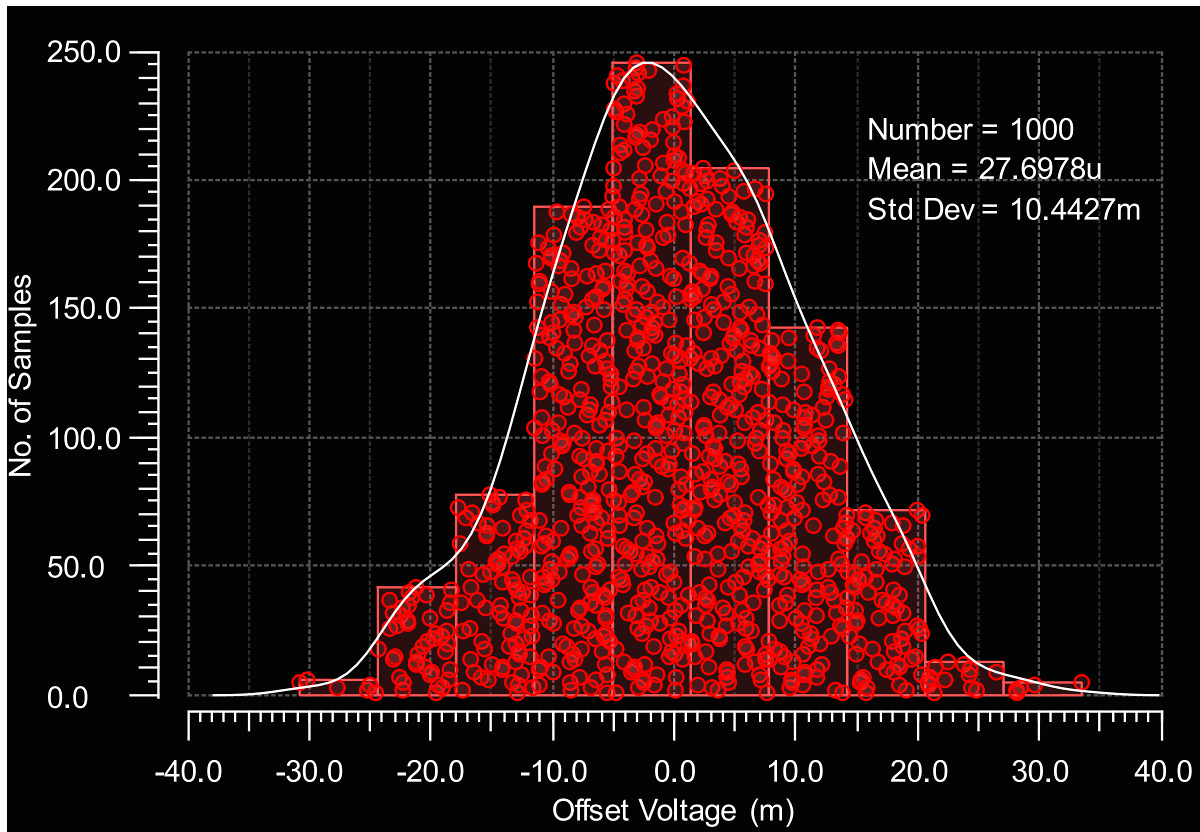

Figure 4.11 DVLSA 0.4V Input Offset Voltage (SAEB_delay=1.0ns)

Figure 4.11 shows the input offset voltage histogram of the DVLSA at 0.4V with a SAEB_delay of 1ns. It is centered at 27.70uV with a standard deviation of 10.44mV. Compared to (μ_{os}, σ_{os}) = (-417.8uV, 9.64mV) for the reference VLSA, this configuration of the DVLSA exhibits an improvement in μ_{os} by 93.36% and a degradation in σ_{os} by 8.30%.

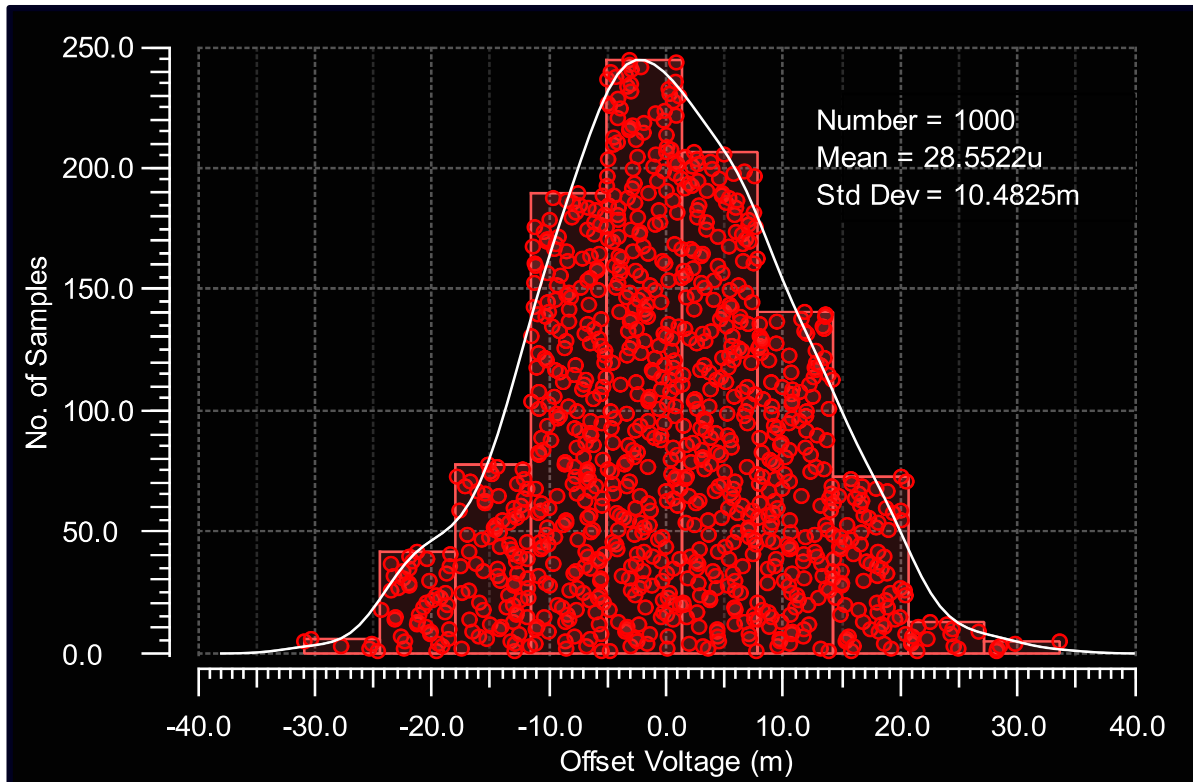

Figure 4.12 DVLSA 0.4V Input Offset Voltage (SAEB_delay=2.0ns)

Figure 4.12 shows the input offset voltage histogram of the DVLSA at 0.4V with a SAEB_delay of 2ns. It is centered at 28.55uV with a standard deviation of 10.48mV. Compared to (μ_{os}, σ_{os}) = (-417.8uV, 9.64mV) for the reference VLSA, this configuration of the DVLSA exhibits an improvement in μ_{os} by 93.16% and a degradation in σ_{os} by 8.71%.

Table 4.2 DVLSA Mismatch Contribution Summary

	Device	Value (VLSA)	Value (DVLSA)			
SAEB_delay (ns)	-	-	0.00	0.50	1.00	2.00
σ_{os} (mV)	-	9.64	10.40	10.43	10.44	10.48
μ_{os} (uV)	-	-422.85	25.01	27.58	27.70	28.55
Variance Contribution (%) to σ_{os}^2	N1	49.74	50.60	50.64	50.65	50.67
	N2	49.51	48.70	48.74	48.74	48.76
	P1	0.28	0.25	0.21	0.21	0.19
	P2	0.25	0.23	0.20	0.19	0.17
	P3	0.11	0.11	0.10	0.10	0.10
	P4	0.11	0.11	0.10	0.10	0.10
Combined Variance Contribution (%) of σ_{os}^2	NMOS Pair (N1+N2)	99.25	99.30	99.38	99.39	99.43
	PMOS Pair (P1+P2)	0.54	0.48	0.41	0.41	0.37
	Other (P3+P4)	0.22	0.22	0.20	0.20	0.20

Table 4.2 compares the mean and standard deviation of the input offset voltage of the DVLSA across SAEB_delay sweeps to the reference VLSA. It also shows the contribution of device mismatch to the input offset voltage variance, presented as a percentage of the total variance. The mismatch contributions from the access transistors P3 and P4 remain approximately constant between the VLSA and all configurations of the DVLSA; the difference between 0.11% and 0.10% across the cases is due to rounding errors. For all DVLSA configurations, we see that the PMOS pair mismatch contribution is lower than that in the VLSA. There is a clear trend of the PMOS mismatch contribution decreasing as SAEB_delay increases. This confirms the first half of our hypothesis.

For all cases of the DVLSA, the measured μ_{os} is reduced by 93-94% relative to that of the VLSA, but unfortunately σ_{os} is higher by 7.9-8.2%. There is also a trend of both μ_{os} and σ_{os} increasing as SAEB_delay is increased. This result is the opposite of what is observed in the isolated PMOS pair mismatch test where the DVLSA showed some improvement. To identify the cause for this discrepancy in the results, further investigation is needed.

4.3.1 Further Investigation into Source of Failure

The data suggests the source of the raised σ_{os} could be due to the mismatch performance of one of the other device pairs, and possible covariances between device pairs that only arise when all devices are varied. To test this, the isolated mismatch test is performed again for the NMOS pair varying in isolation, and for the PMOS access pair varying in isolation.

Table 4.3 Input Offset Voltage Statistics - Isolated PMOS Access Pair (P3/P4)

	Value (VLSA)	Value (DVLSA)			
SAEB_delay (ns)	-	0.00	0.50	1.00	2.00
σ_{os} (uV)	480.2	500.9	480.6	481.3	481.9
μ_{os} (uV)	30.70	32.09	30.99	31.05	30.87

Table 4.3 shows the input offset results for the PMOS access pair P3/P4 varying in isolation. The results of the DVLSA are comparable to the reference VLSA. μ_{os} for all cases differs by up to 1.3uV, which is less than the minimum resolution of the testbench. σ_{os} for the 0.0ns delay case is higher by 20uV than the reference VLSA, but the remaining three configurations differ by less than 2uV. This consistency in σ_{os} correlates well with the consistency in mismatch contribution seen in the comprehensive simulation. Therefore, due to the small magnitude of σ_{os} being less than 1mV and the relatively small differences across DVLSA configurations we do not believe this pair is the cause of the negative result.

Table 4.4 Input Offset Voltage Statistics - Isolated NMOS Pair (N1/N2)

	Value (VLSA)	Value (DVLSA)			
SAEB_delay (ns)	-	0.00	0.50	1.00	2.00
σ_{os} (mV)	9.907	9.905	9.936	9.953	9.992
μ_{os} (uV)	504.6	504.7	506.3	506.9	508.9

Table 4.4 shows the input offset results for the NMOS cross pair N1/N2 varying in isolation. At a delay of 0.0ns, we see the performance of the DVLSA is near identical to the VLSA. As SAEB_delay increases however we see that both parameters increase. μ_{os} increases by 4.2uV over the sweep range. σ_{os} increases by 98.7uV over the SAEB_delay sweep range. As per Table 4.2 for the comprehensive mismatch result, σ_{os} ranges from 10.4mV to 10.48mV, which has a delta of 80uV over the sweep range. This delta is comparable to the 98.7uV seen in Table 4.4. Given the similarity

in the increase of σ_{os} from the isolated NMOS results to the comprehensive mismatch simulation, we conclude this NMOS pair is the reason for the offset degradation.

With the data collected for all three device pairs varied in isolation, we can as a sanity check estimate what the expected σ_{os} should be, assuming all mismatch sources are uncorrelated. For the 0.0ns delay case, $\sigma_{os,expected}\ (\text{mV}) = \sqrt{\sigma^2_{os,P1+P2} + \sigma^2_{os,P3+P4} + \sigma^2_{os,N1+N2}} = \sqrt{9.905^2 + 0.501^2 + 0.675^2} = 9.941\text{mV}$. This is 459uV or 4.4% lower than the simulation measurements.

The discrepancy could be due to several factors. One source could be is variations in device properties not considered in this thesis. Another source could be is covariances in mismatch parameters between devices that can only arise when all devices are varied. A third source could be the odd P2:ID capacitive current seen in Section 4.1; if the variation of P1 and P2 parameters caused a difference in capacitance on the Q and QB nodes, then this could affect the varying NMOS pair and cause additional offset.

Chapter 5
Conclusions and Future Research

A VLSA circuit modified with an additional PMOS switch to VDD was tested. A fixed delay between the sense amp enable signal (SAE) and the turn-on of this additional PMOS switch (controlled by SAEB) was added in an attempt to reduce the mismatch contribution of the PMOS pull-up cross pair to the input offset voltage. Nominal simulations were run to observe the difference of behavior between this topology and the reference VLSA implementation. Statistics on input offset were also collected for isolated variation of the PMOS pull-up pair, and all differential devices varied. Isolated PMOS mismatch results showed that both the mean and standard deviation were reduced, with further reduction for increasing SAEB delay.

The statistics for the more comprehensive statistical run showed that μ_{os} was reduced in all cases, but σ_{os} was higher in all cases. The data had a trend that opposed the isolated PMOS mismatch results, where both μ_{os} and σ_{os} increased with increasing SAEB delay. Further investigation showed in an isolated NMOS pair simulation that μ_{os} and σ_{os} experienced a similar positive correlation with increased SAEB delay. In conjunction with the large magnitude of σ_{os} relative to that from the other isolated mismatch simulations, the data suggests that the NMOS pair is the root cause of the negative result.

Given the results collected, the conclusion drawn is that the DVLA topology and method of control as presented in this thesis is inferior to the reference VLSA for the design goal of reducing σ_{os} of a sense amplifier. As a result, the recommendation would be to use a different combination of topology and control mechanism to achieve a reduced σ_{os}.

Further research could investigate in more detail the effect of the negative current out of device P2 after the initial transient spike. This current appears to be due to capacitive coupling and it may be a source of offset. Within this context, the DVLSA idea can be revisited with a modification that aims to reduce the capacitance on the Q/QB nodes as seen by the N1/N2 device pair. Another research path could be combining the DVLSA topology with other offset-reducing methods, such as input multiplication.

References

[1] D. J. Rennie, T. Shakir and M. Sachdev, "Design Challenges in Nanometric Embedded Memories," in *2009 3rd International Conference on Signals, Circuits and Systems (SCS)*, Medenine, 2009.

[2] M. Konijnenburg, Y. Cho, M. Ashouei, T. Gemmeke, C. Kim, J. Hulzink, J. Stuyt, M. Jung, J. Huisken, S. Ryu, J. Kim and H. d. Groot, "Reliable and energy-efficient 1MHz 0.4V dynamically reconfigurable SoC for ExG applications in 40nm LP CMOS," in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, San Francisco, CA, 2013.

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

[4] M. R. Guthaus, J. E. Stine, S. Ataei, B. Chen, B. Wu and M. Sarwar, "OpenRAM: An Open-Source Memory Compiler," in *Proceedings of the 35th International Conference on Computer-Aided Design (ICCAD '16)*, New York, NY, USA, 2016.

[5] A. Sedra and K. C. Smith, Microelectronic Circuits, 7th ed., Oxford University Press, 2014.

[6] W. Liu, X. Jin, K. M. Cao and C. Hu, "BSIM 4.1.0 MOSFET Model-User's Manual," Oct. 2000.

[7] M. J. M. Pelgrom, C. J. Duinmaijer and A. P. G. Welbers, "Matching Properties of MOS Transistors," *IEEE Journal Solid-State Circuits,* vol. 24, no. 5, pp. 1433-1440, Oct. 1989.

[8] B. Razavi, "The Cross-Coupled Pair - Part II [A Circuit for All Seasons]," *IEEE Solid-State Circuits Magazine,* vol. 6, no. 4, pp. 9-12, 2014.

[9] B. Razavi, "The Cross-Coupled Pair?Part III [A Circuit for All Seasons]," *IEEE Solid-State Circuits Magazine,* vol. 7, no. 1, pp. 10-13, 2015.

[10] A. Abidi and H. Xu, "Understanding The Regenerative Comparator Circuit," in *Proceedings of the IEEE 2014 Custom Integrated Circuits Conference*, San Jose, CA, 2014.

[11] T. Na, S.-H. Woo, J. Kim, H. Jeong and S.-O. Jung, "Comparative Study of Various Latch-Type Sense Amplifiers," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* vol. 22, no. 2, pp. 425-429, Feb 2014.

[12] B. Wicht, T. Nirschl and D. Schmitt-Landsiedel, "Yield and Speed Optimization of a Latch-Type Voltage Sense Amplifier," *IEEE Journal of Solid-State Circuits,* vol. 39, no. 7, pp. 1148-1158, 2004.

[13] M. Miyahara, Y. Asada, D. Paik and A. Matsuzawa, "A low-noise self-calibrating dynamic comparator for high-speed ADCs," in *2008 IEEE Asian Solid-State Circuits Conference*, Fukuoka, 2008.

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

[15] S. V. Yamani, U. R. Nsai and R. Vaddi, "A Low Voltage Capacitor Based Current Controlled Sense Amplifier for Input Offset Compensation," in *International SoC Design Conference (ISOCC)*, Seoul, 2017.

[16] D. Patel, A. Neale, D. Wright and M. Sachdev, "Hybrid Latch-Type Offset Tolerant Sense Amplifier for Low-Voltage SRAMs," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 66, no. 7, pp. 2519-2532, 2019.

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

[18] R. Singh and N. Bhat, "An Offset Compensation Technique for Latch Type Sense Amplifiers in High Speed Low Power SRAMs," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* vol. 12, no. 6, pp. 652-657, 2004.

[19] K. Shi, "A Comparative Analysis of SRAM Sense Amplifiers," University of California Los Angeles, 2017.

[20] L. Pileggi, G. Keskin, X. Li, K. Mai and J. Proesel, "Mismatch analysis and statistical design at 65 nm and below," in *IEEE Custom Integrated Circuits Conference*, San Jose, CA, 2008.

[21] A. J. Ginés, E. Peralías, G. Leger and A. Rueda, "Closed-loop simulation method for evaluation of static offset in discrete-time comparators," in *2014 21st IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, Marseille, 2014.

[22] A. Hajimiri and R. Heald, "Design issues in cross-coupled inverter sense amplifier," in *ISCAS '98. Proceedings of the 1998 IEEE International Symposium on Circuits and Systems*, Monterey, CA, 1998.

[23] B. Ullmann, K. Puschkarsky, M. Waltl, H. Reisinger and T. Grasser, "Evaluation of Advanced MOSFET Threshold Voltage Drift Measurement Techniques," *IEEE Transactions on Device and Materials Reliability,* vol. 19, no. 2, pp. 358-362, 2019.

[24] T. Singh, S. Rangarajan, D. John, R. Schreiber, S. Oliver, R. Seahra and A. Schaefer, "2.1 Zen 2: The AMD 7nm Energy-Efficient High-Performance x86-64 Microprocessor Core," in *IEEE International Solid- State Circuits Conference - (ISSCC)*, San Francisco, CA, USA, 2020.

[25] M. Sinha, S. Hsu, A. Alvandpour, W. Burleson, R. Krishnamurthy and S. Borkar, "High-Performance and Low-Voltage Sense-Amplifier Techniques for sub-90nm SRAM," in *IEEE International [Systems-on-Chip] SOC Conference, 2003. Proceedings.*, Portland, OR, USA, 2003.

[26] V. Nautiyal, G. Singla, S. Singh, F. a. Bohra, J. Dasani, L. Gupta and S. Dwivedi, "Charge recycled low power SRAM with integrated write and read assist, for wearable electronics, designed in 7nm FinFET," in *IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*, Taipei, 2017.

Appendix A

Testbench Verilog-A Code

A.1 Evaluator

```
// Verilog-A for mtw_sense_amp_2, tb_SA_SEQ_decision, Verilog-A

`include "constants.vams"
`include "disciplines.vams"

// Analog output voltage sampler for SA SEQ testbench
// Monitors output voltage delta between Q and QB.
// outputs digital sgn(x) of delta voltage; -1, 0, 1 - also makes a binary
decision, if (sgn(x) == 1)
// also outputs a variant that adds an absolute ammplitude threshold
condition as a % of VDD
// default:  crit_sgn(x) = sgn(x) if abs(x) >= VDD*threshold_ratio else 0
module tb_SA_SEQ_decision (VDD, VSS, VQ, VQB, VQDELTA, SGN, CRIT_SGN,
DECISION, CRIT_DECISION);
electrical VDD, VSS, VQ,VQB, VQDELTA, SGN, CRIT_SGN, DECISION,
CRIT_DECISION;
input VDD, VSS, VQ, VQB;
output VQDELTA, SGN, CRIT_SGN, DECISION, CRIT_DECISION;

real vqdelta;
real crit_threshold_ratio = 0.9;
real crit_threshold;
integer sgn;
integer crit_sgn;

analog begin
        vqdelta = V(VQ,VSS)-V(VQB,VSS);
        crit_threshold = V(VDD,VSS) * crit_threshold_ratio;

        sgn = (vqdelta > 0) ? 1 : (vqdelta < 0) ? -1 : 0;
        crit_sgn = (abs(vqdelta) >= crit_threshold) ? sgn : 0;

        V(VQDELTA, VSS) <+ vqdelta;
        V(SGN, VSS) <+ sgn;
        V(DECISION, VSS) <+ (sgn == 1);
        V(CRIT_SGN, VSS) <+ crit_sgn;
        V(CRIT_DECISION, VSS) <+ (crit_sgn == 1);
end
endmodule
```

A.2 SAR Controller

```
// Verilog-A for mtw_sense_amp_2, tb_vqdelta_offset_servo, Verilog-A

`include "constants.vams"
`include "disciplines.vams"

module tb_vqdelta_offset_servo(VSS, CLK, FB, BL, BLB, VBLDELTA, FINISHED);
electrical VSS, CLK, FB, BL, BLB, VBLDELTA, FINISHED;
input VSS;
input CLK, FB;
output BL, BLB, VBLDELTA, FINISHED;

parameter real VDD = 0.4;
parameter real fclk = 1e6;
real period = 1/fclk;
parameter real ncycles = 18;

real tr = 250.0e-12;
real bl_update_delay = 0.05*period;
parameter real t_delay = 0;


real bl_out = 0;
real blb_out = 0;
real delta_out = 0;
real delta_step = 0;
real sgn_sample = 0;
integer finished = 0;
integer cycle_count = 0;

analog begin
        @(initial_step) begin
                delta_out = 0;
                delta_step = VDD;
                finished=0;
                cycle_count = 0;
        end


        @(cross(V(CLK,VSS) - VDD/2, -1)) begin
                sgn_sample = V(FB, VSS);
                if(sgn_sample < 0) delta_out = delta_out + delta_step;
                if(sgn_sample > 0) delta_out = delta_out - delta_step;
                delta_out = max(-VDD, min(VDD, delta_out)); // bound to {-
VDD, VDD}
                delta_step = delta_step/2;
                cycle_count = cycle_count + 1;
                if(cycle_count >= ncycles) finished = 1;
        end
```

```
        V(FINISHED, VSS) <+ finished; // instantaneous transition to latch
VBLDELTA before transition
        // slightly delayed transition to update other values
        V(VBLDELTA, VSS) <+ transition(delta_out, 0.05*period, tr);
        V(BL,  VSS) <+ transition(VDD + (delta_out < 0 ? delta_out : 0),
bl_update_delay, tr);
        V(BLB, VSS) <+ transition(VDD - (delta_out > 0 ? delta_out : 0),
bl_update_delay, tr);

end

endmodule
```