

SELF-CALIBRATING
ON-CHIP INTERCONNECTS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL
ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

James Chen

March 2012

© Copyright by James Chen 2012
All Rights Reserved

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(William J. Dally) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Mark A. Horowitz)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Bruce A. Wooley)

Approved for the University Committee on Graduate Studies.

Abstract

In an era where computing systems are becoming increasingly power limited, there is a growing need for energy-efficient on-chip interconnects. Low-swing interconnects (LSIs) use specialized transceivers to restrict the signal swing to a small fraction of the supply voltage. They have the potential to substantially reduce wire energy over the conventional, full-swing interconnects. The capacitively-driven interconnect (CDI), which does not require a separate supply and benefits from wire equalization, is a particularly attractive candidate. Unfortunately, the deteriorating device mismatch in modern CMOS processes, which sets a lower bound in signal swing, combined with the decreasing supply voltage, have made it progressively more difficult to build an energy-efficient CDI with sufficient reliability.

In this dissertation we propose the self-calibrating interconnect (SCI), a new type of CDI that use feedback and charge-pumps to automatically neutralize receiver offsets caused by device mismatch. It can operate with extremely low signal swings, which minimizes energy, without compromising reliability. A 2 mm prototype built in a 90 nm low-power CMOS process consumes 77 fJ/bit of energy at 1.5 GHz, and sustains a bit-error rate (BER) below 10^{-30} with just 32 mV of swing.

As signal swing is reduced, the effect of thermal noise becomes more pronounced. To this end, we present a new way to reduce the noise level of comparators by adjusting the size of the precharge devices. Specifically, for a given energy budget, we demonstrate that by upsizing the precharge devices we can reduce the input-referred thermal noise of the popular StrongARM latch by up to 30% over previously known sizing techniques.

Acknowledgements

This dissertation, and the work leading up to it, would not have been possible without the assistance of numerous people.

First of all, I would like to thank my advisor Bill Dally. Bill's raw brain power and the extent of his knowledge, coupled with his infectious sense of optimism and adventurous spirit, makes him one of the most interesting people that I have ever had the privilege of working with. Bill pushes his students to excel, but he also gives them enough room to develop as individuals. I appreciate all the intellectually stimulating discussions Bill had with me, and the ways he strives to befriend his students at the annual ski trip or the summer picnic.

I want to express my utmost respect for my co-advisor Mark Horowitz. It was during Mark's fun and insightful classes that I first developed the desire to do research in the field of integrated circuits. Mark's passion for teaching is commendable, and his wise and intuitive way of problem solving has been a source of inspiration for me. I can always depend on Mark to give me great advices.

I must salute the entire Electrical Engineering and Computer Science faculty at Stanford University. Every class I have taken is well-thought-out, practically relevant, and meticulously executed. It is an honor to be in an academic environment where excellence is so pervasive that it is the norm. I want to personally thank Bruce Wooley and John Ousterhout for taking the time out of their busy schedules to be on my PhD orals committee.

Apart from the faculty, I am indebted to my outstanding colleagues and the helpful administrators. I want to thank Patrick Chiang, Zain Asgar, Byongchan Lim, Ofer Shacham and Megan Wachs for their assistance during the development and

fabrication of my prototype. I also want to thank Daniel Becker for his dependable IT support on our computer cluster. My heart goes out to Jane Klickman, Uma Mulukutla and Sue George for working behind the scenes to make my life as a graduate student that much easier.

I want to make a special shout-out to John Poulton, Trey Greer, Steve Tell and Robert Palmer in the NVIDIA Circuits Research Group. I thoroughly enjoyed the opportunity to collaborate with them over the summer of 2010, and the great discussions that we had led to the development of brand new insights that made this dissertation successful. I will never forget the fun, July-4th party that we had at the Poultons' ranch. Trey's BBQ was legendary, and the homemade trebuchet and potato cannon were spectacular.

My research could not have taken place without the backing of individuals and institutions who are committed to education and knowledge creation. I am deeply touched by the generosity of Sequoia Capital (SGF Fellowship), NSA (H98230-08-C-0272), NSF (CCF-0702341) and SRC (2009-HJ-1976) for providing financial support throughout my PhD.

I can never express enough gratitude for the love and support that my Dad (Yu-Lin), my Mum (Ling-Yu) and my big brother (David) has shown me. Throughout every stage of my life, in every country we have lived in (Taiwan, New Zealand and the United States), whether in good or bad times, they have always been right there with me. Their love gives me the confidence to be myself, the courage to face new challenges, and a reason to smile everyday.

Last but not least, I want to thank God and the spiritual family that He has blessed me with in the Bay Area Christian Church. The friendships that I have there sharpens me and keeps me rooted in the community. They prevent me from becoming conceited and gives me a sense of hope and purpose for the future.

"For I know the plans I have for you, declares the LORD, plans to prosper you and not to harm you, plans to give you hope and a future."

Jeremiah 29:11

Contents

Abstract	v
Acknowledgements	vi
1 Introduction	1
1.1 Contributions	5
1.2 Dissertation Outline	5
2 Low-Swing Interconnects	7
2.1 Multi-Supply Interconnect	8
2.1.1 Wire Energy	9
2.1.2 Low-Swing Supply	9
2.1.3 Bandwidth Reduction	10
2.1.4 Supply Noise Sensitivity	11
2.2 Capacitively-Driven Interconnect	12
2.2.1 Supply Simplicity and Flexibility	12
2.2.2 Equalized Signaling	13
2.2.3 Supply Noise Filtering	15
2.2.4 Wire Energy	15
2.2.5 DC Bias	16
2.3 Technology Trends	18
2.4 Offset Compensation	21

3	Self-Calibrating Interconnect	25
3.1	Bias Detection	26
3.2	Calibration Process	28
3.2.1	Interconnect Availability	29
3.2.2	Incremental Calibration	29
3.3	Voltage Adjustment Circuits	30
3.3.1	Decrementer	30
3.3.2	Incrementer	32
3.4	Leakage	35
3.4.1	Receiver Inputs	36
3.4.2	Coupling Capacitors	36
3.4.3	Direct Channel Connections	37
3.5	Complete Design	38
3.6	Control Signals	39
3.6.1	Timing	39
3.6.2	Generation	40
3.6.3	Energy Overhead	40
3.7	Operation	42
3.7.1	Reset Phase	43
3.7.2	Warm-Up Phase	43
3.7.3	Maintenance Phase	43
3.8	Performance	43
4	Noise	45
4.1	Theory	46
4.1.1	Noise Analysis in LTV Systems	48
4.1.2	Input-Referred Noise	51
4.1.3	Bit Error Rate	52
4.2	Simulation	54
4.3	Noise Reduction	58
4.3.1	Upsize	59

4.3.2	Increase Aperture	59
4.3.3	Preamplifier	61
4.3.4	Integrating Comparator	62
4.4	Impact of Precharge Device Sizing	63
5	Prototype Evaluation	69
5.1	Chip Design	70
5.1.1	SCI	71
5.1.2	Reference Interconnect	73
5.1.3	Zero Phase Detector	73
5.1.4	Traffic Generator	74
5.1.5	Calibration Controller	74
5.1.6	Error Counter	75
5.1.7	Analog Probe	75
5.1.8	StrongARM Array	75
5.1.9	JTAG Controller	75
5.2	Test Board and Experimental Setup	76
5.3	SCI Experiments	77
5.3.1	Calibration Effectiveness	77
5.3.2	Eye Opening	79
5.3.3	Leakage	80
5.3.4	Energy and Area	81
5.4	Comparator Noise Experiment	83
6	Conclusion	86
A	Transfer Function Derivation for Capacitor-Coupled Wire Segment	89
B	Registers Accessible via JTAG	91
	Bibliography	95

List of Tables

5.1	Comparison of the SCI with other energy-efficient on-chip interconnects.	82
B.1	Control and status registers accessible via the JTAG interface on the prototype chip.	91

List of Figures

1.1	Comparison of computation energy to on-chip communication energy in a 40 nm process.	2
1.2	ITRS projected wire to device capacitance ratio.	3
1.3	Conventional interconnect with repeated inverter-driven wire segments.	4
2.1	Multi-supply interconnect architecture.	9
2.2	Energy and frequency tradeoff for different interconnect architectures over a distance of 4 mm.	11
2.3	Capacitively-driven interconnect architecture.	12
2.4	RC models for plain and equalized wire segments.	14
	(a) Plain	14
	(b) Equalized	14
2.5	DC bias through dynamic refresh.	16
2.6	DC bias with leaky PFET.	17
2.7	DC bias with transconductance and load resistance.	18
2.8	Changes in supply voltage over different CMOS process generations.	19
2.9	The StrongARM latch and its equivalent circuit model in the presence of mismatch.	20
	(a) Transistor level schematic	20
	(b) Model with input offset voltage	20
2.10	Threshold mismatch for minimum-sized NFETs over different CMOS process generations.	21
2.11	Offset compensation with digitally controlled current sources.	22
2.12	Offset compensation with digitally adjustable capacitive loads.	24

3.1	Combined DC bias and offset compensation in the SCI.	26
3.2	Bias detection using the reset signal.	27
3.3	Offset compensation through iterative refinement.	28
3.4	Trade-off between convergence rate and calibration resolution.	29
3.5	Charge pump for decrementing line voltage.	31
3.6	Decrementer schematics.	32
3.7	Charge pump for incrementing line voltage.	33
3.8	Bootstrapped charge pump for incrementing line voltage.	34
3.9	Bootstrapped incrementer schematics.	35
3.10	Effect of differential and common-mode leakage on line voltages.	36
3.11	Leakage optimized voltage decrementer.	37
3.12	Full SCI Schematics.	38
3.13	Timing relationship between the calibration control signals.	39
3.14	Generating control signals from one global calibration signal.	41
3.15	SCI operating under 70 mV of input offset with a 20 mV signal swing.	42
3.16	Comparison of energy and maximum operating frequency of the SCI against other interconnect architectures.	44
4.1	The StrongARM latch.	46
4.2	The five operating phases of the StrongARM latch.	47
4.3	Duality between the impulse response and the ISF.	49
4.4	Modeling the effect of noise with input-referred noise voltage.	51
4.5	Effect of increasing VSNR on BER.	52
4.6	Representative 64-node network-on-chip.	54
4.7	Noise simulation setup based on LTV circuits theory.	56
4.8	Representative outputs from a noise simulation run.	57
4.9	Apertures of comparators.	60
4.10	Improving comparator SNR with a preamplifier.	61
4.11	Improving comparator SNR with integrating stage.	62
4.12	Variation of input-referred noise voltage and input offset with precharge device size.	63

4.13	Impact of increasing precharge device width on noise contributions. . .	65
4.14	Effect of sizing on small signal parameters of the sensing device. . . .	66
4.15	Energy and noise tradeoff of the StrongARM latch under different up- sizing techniques.	67
4.16	Variation of StrongARM latch decision latency with precharge device size.	68
5.1	Die photograph of prototype chip in the IBM 90 nm low-power process.	69
5.2	Prototype Chip Block Diagram.	70
5.3	Modified SCI for testing.	72
5.4	Zero phase detector built from cross-coupled flip-flops.	73
5.5	LFSR based traffic generator.	74
5.6	Board and lab setup for prototype testing.	76
5.7	Dependence of BER on signal swing with and without calibration. . .	78
5.8	Variation of BER with the phase of the receiver clock.	79
5.9	Signal swing required to sustain a BER of 10^{-9} for different calibration intervals.	80
5.10	Measuring input-referred noise through curve-fitting.	84
5.11	Comparison of measured and simulated input-referred noise for Stron- gARM latches with different precharge device sizes.	84
5.12	Comparison of measured and simulated input offset for StrongARM latches with different precharge device sizes.	85

Chapter 1

Introduction

As computing systems advance, energy-efficiency has emerged as one of the most important design metrics. Smart-phones, tablets and many other mobile devices that are an integral part of our daily activities have limited battery lives. In order to make increasingly sophisticated applications available to the end users, engineers must find ways to boost performance while reducing the per unit power consumption. In commercial data centers, the cost of electricity and cooling for the servers is becoming a large fraction of the total operating cost, creating a strong financial incentive to reduce computation power. A recent study shows that the cost of electricity can exceed the initial hardware purchase cost in just 3 years [33]. Excessive power can also directly limit the maximum throughput that we can realistically achieve for any single computer. Indeed, experts agree that power is the most pervasive design challenge to building an exascale computer by 2015 [29].

Taking energy-efficiency of computers to the next level requires a deep, vertically-integrated collaboration between engineers at every level of the design hierarchy, from software development down to device and process engineering. At a chip level, the power of on-chip interconnects is a critical concern for circuit designers. Figure 1.1 shows that it takes approximately 20 pJ of energy to perform a 64-bit floating-point operation in a state-of-art 40 nm process [26]. In contrast, the energy required to move two 64-bit operands across an on-chip distance of 10 mm is over 120 pJ, a factor of 6 higher. Traditionally, computer architectures exploit locality to amortize the cost

of communication over multiple computations. However, as the number of cores increases and on-chip networks become more pervasive, the systems are increasingly dominated by communication. More than ever before, there is a pressing need for energy-efficient on-chip interconnects.

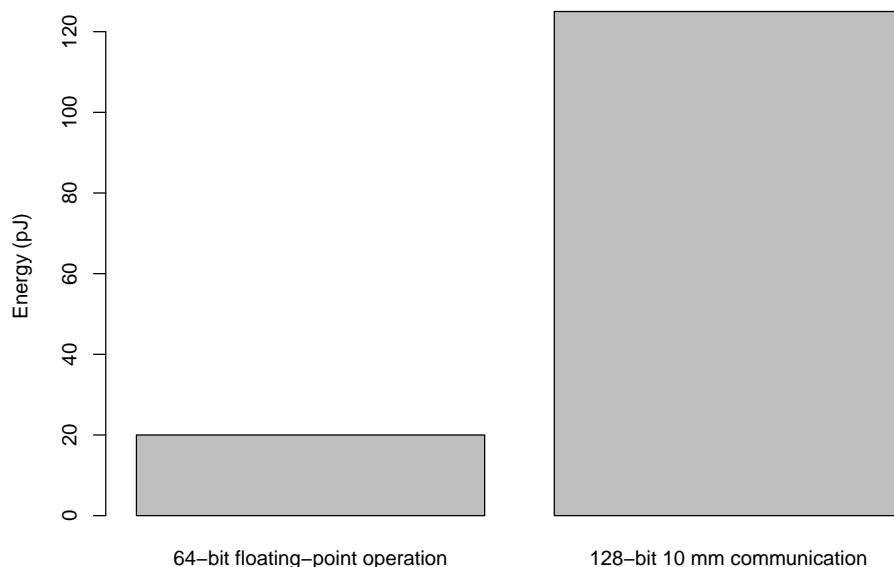


Figure 1.1: Comparison of computation energy to on-chip communication energy in a 40 nm process.

In CMOS technology, wire dimensions and hence capacitances continue to scale down at a slower rate than the devices. Since energy is proportional to capacitance, the energy gap between wires and devices widens with every new process generation. Figure 1.2 shows how the ratio of wire capacitance to device capacitance is predicted to change by the International Technology Roadmap for Semiconductors (ITRS) [1, 2]. Over the next 12 years, this ratio is expected to increase by approximately another 40%. Without a breakthrough in the design, large fractions of power will be wasted moving data.

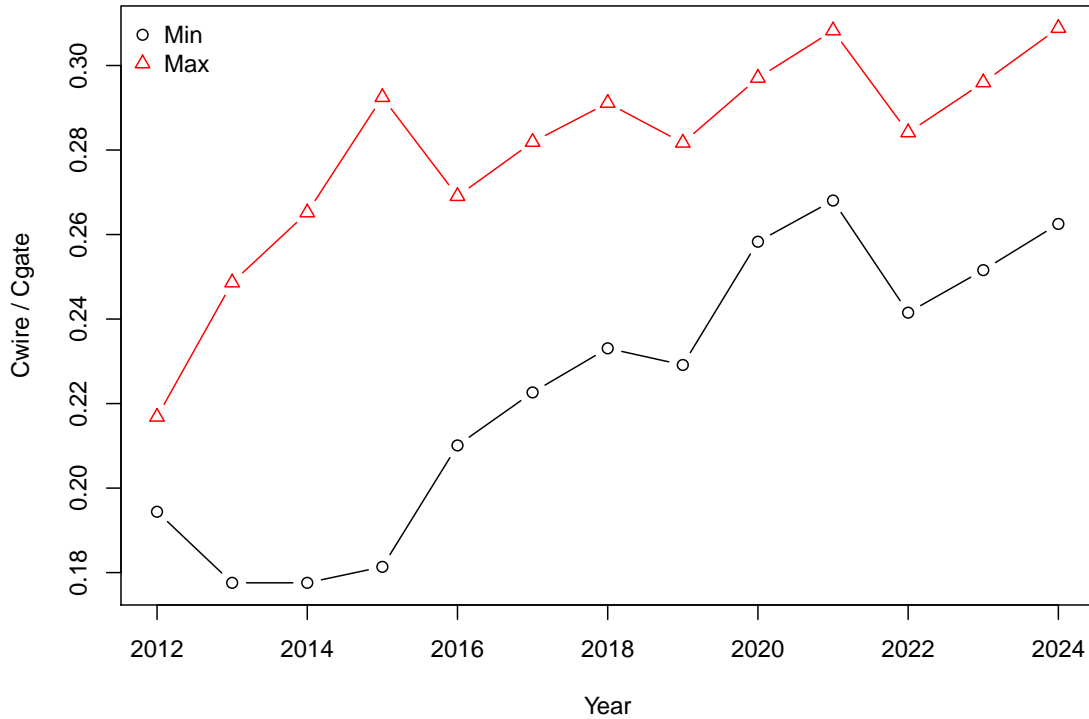


Figure 1.2: ITRS projected wire to device capacitance ratio.

Wire resistance and wire capacitance both increase linearly with wire length. Propagation delay, which is proportional to the product of resistance and capacitance, increases quadratically with wire length. In a conventional interconnect, henceforth referred to as the full-swing interconnect (FSI), repeaters (inverters) are placed at regular intervals along a long wire to minimize delay as shown in figure 1.3. The inverters divide the wire into shorter segments, providing both isolation and buffering at the breakpoints. With each segment decoupled from its neighbors, total delay now depends on the number of segments, which only increases linearly with wire length. While effective at reducing delay, the FSI drives the entire wire capacitance to the extremes of the power supply at every data transition, dissipating a large amount of energy.

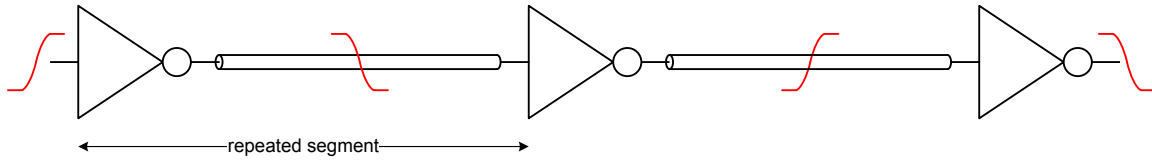


Figure 1.3: Conventional interconnect with repeated inverter-driven wire segments.

Over the years, a number of low-swing interconnects (LSIs) have been developed to improve the energy-efficiency of on-chip communication [50, 23, 36, 24, 41, 30]. LSIs use more advanced transmitters to restrict the signal swing on the wires to a small fraction of the power supply. At the receivers, clocked comparators (also called sense amplifiers) are used to recover the original full-swing signal. Since energy is proportional to the voltage swing, in longer interconnects where wire capacitance dominate, LSIs have the potential to realize large energy savings over the FSI. Depending on the design and the process used, energy savings from 3 to 10 times have been reported.

To improve robustness, most LSIs use differential signaling to neutralize common-mode coupling from nearby wires and transient supply variations. This doubles the number of wires and thus requires a lower signal swing to achieve the same amount of energy saving. As supply voltage decreased over time, there is pressure to use lower signal swings in order to justify the energy savings of the LSIs. Unfortunately, deteriorating device mismatch in modern CMOS processes is making it increasingly difficult to build comparators with a low input offset voltage. This in turn makes it difficult to detect low voltages with sufficient reliability. As a result, it is becoming more and more challenging to build good LSIs.

This dissertation presents a new kind of LSI, the self-calibrating interconnect (SCI), to simultaneously provide energy-efficiency and reliability for on-chip communication, even in the presence of large device mismatch. Using minimal overheads, the SCI periodically calibrate itself against the intrinsic input offset voltage at its receiver, neutralizing the effect of mismatch. In doing so, the SCI can sustain reliable data transmission with a very low signal swing, which minimizes wire energy. Moreover, unlike its predecessors where the comparators need to be preemptively upsized to

minimize worst case mismatch, the SCI allows the use of much smaller comparators, substantially reducing the receiver energy.

LSIs also need to exhibit low noise characteristics to maintain an acceptable bit error rate (BER). For most designs, the largest source of random noise is the thermal noise internal to the comparators. Traditionally, thermal noise is reduced by increasing the widths of the critical devices. However, up to now, little attention has been paid to how the precharge devices affect the noise performance of a comparator. Most of the time, the smallest precharge devices that meet timing are chosen to minimize the clock load. We have discovered a previously unnoticed interaction between the precharge devices and the input-referred noise of a comparator. In this dissertation, we present our key findings and show that precharge device sizing can be used to achieve better noise performances than what is attainable using previously known sizing techniques.

1.1 Contributions

This dissertation makes two key contributions:

1. We invent the self-calibrating interconnect for energy-efficient and low-latency on-chip communication.
2. We present a novel noise reduction technique for clocked comparators using precharge device sizing.

1.2 Dissertation Outline

The remainder of this dissertation is organized as follows:

Chapter 2 gives a comprehensive review on LSIs. We examine different architectures, describe their operations, and discuss the advantages and disadvantages of each design. We take an in-depth look at various technology trends and identify key challenges that need to be overcome. Along the way, we present a survey of existing

techniques that have been proposed to address some of these problems and discuss their strengths and weaknesses.

Chapter 3 builds on the foundation laid in chapter 2 and formally introduces the SCI. We begin with the high-level concept and then dive into the implementation details of the different building blocks. Where applicable, challenges unique to our approach are discussed along with our proposed solutions. Simulations are presented to demonstrate the SCI operating in the presence of device mismatch, and to compare its performance to other LSIs.

Chapter 4 studies the noise performance of clocked comparators. We present a simulation technique that exposes the noise contributions of individual devices in the comparator, and use it to show how precharge device sizing can be used to reduce the input-referred noise. We compare precharge device sizing to previously known sizing techniques and demonstrate that, for a given energy constraint, lower noise can be obtained by using precharge devices larger than what is typically considered necessary for timing.

Chapter 5 covers the evaluation of our prototype chip in a commercial 90 nm low-power CMOS process. We start by describing the implementation details of our chip and its associated test board. From there, we discuss the results from a series of key experiments performed on the SCI. Finally, we present noise and mismatch measurements from an array of comparators and compare them to the predictions from our simulations.

Chapter 6 concludes this dissertation and identifies specific areas that can be studied further in future work.

Chapter 2

Low-Swing Interconnects

Conceptually, all on-chip interconnects consist of three parts: the transmitter (TX) where the data originates, the channel (CH) which are the wires or repeaters through which the data propagates, and the receiver (RX) where the data is used. The total energy per bit of communication, E_{bit} , is given by

$$E_{bit} = E_{TX} + E_{CH} + E_{RX} \quad (2.1)$$

where E_{TX} , E_{CH} , and E_{RX} are the transmitter, channel and receiver energies. Due to the large physical dimensions of the wire, the channel energy usually dominates prior to any optimization. The channel energy can be further expressed as

$$E_{CH} = E_{static} + E_{dyn}, \quad (2.2)$$

where E_{static} is the static energy dissipated by DC currents, and

$$E_{dyn} = \alpha C_{CH} V_{driver} V_{swing} \quad (2.3)$$

is the dynamic energy due to the charging and discharging of capacitances. Here α is the data activity factor, C_{CH} is the effective channel capacitanceⁱ, V_{driver} is the driver

ⁱThis includes the capacitance of the wires, parasitic capacitances from any repeaters, and a scale factor for crowbar currents, where applicable.

voltage, and V_{swing} is the signal swing.

In most interconnects, E_{static} can be kept low provided that we avoid circuits that require a static bias currentⁱⁱ. Once a target process is chosen, wire capacitance is mostly fixed by the floor plan and other high level design decisionsⁱⁱⁱ. Similarly, the activity factor is purely a function of the application. From a design perspective, this leaves voltage as the only parameter that can be used to reduce energy. For this reason, virtually all energy-efficient interconnects involve some way of reducing either the driver voltage or the signal swing. In the FSI,

$$V_{driver} = V_{swing} = V_{dd}, \quad (2.4)$$

where V_{dd} is the supply voltage. Since there are no voltage reductions, the wire energy is high. In the following sections, we examine the two most ubiquitous low-swing interconnects (LSIs) found in the literatures.

2.1 Multi-Supply Interconnect

The multi-supply interconnect (MSI), shown in figure 2.1, uses a dedicated second supply to control the signal swing on the wires [12, 50, 23, 30]. At the transmitter, the data is first decomposed into its true and complement values. A pair of NFET drivers, powered by the low-voltage supply V_{LS} , is then used to drive the signal differentially onto the wires. NFET pull-ups are used because for low values of V_{LS} , the gate of a PFET cannot be driven low enough for it to conduct. At the receiver, a clocked comparator is used to regenerate the low-swing signals back into full-swing values at the rising clock edge. During the negative half of the clock cycle, the RS-latch at the output holds the previously regenerated value while the comparator precharges and prepares for the detection of the next bit.

ⁱⁱFor most interconnects leakage power is negligible because the fraction of area occupied by active devices is small. For structures like memories, where the great majority of area consist of idling devices, leakage becomes more significant.

ⁱⁱⁱIn a best case scenario where congestion is not a problem, wire capacitance can be reduced by approximately a factor of 2 by increasing the pitch of the wires.

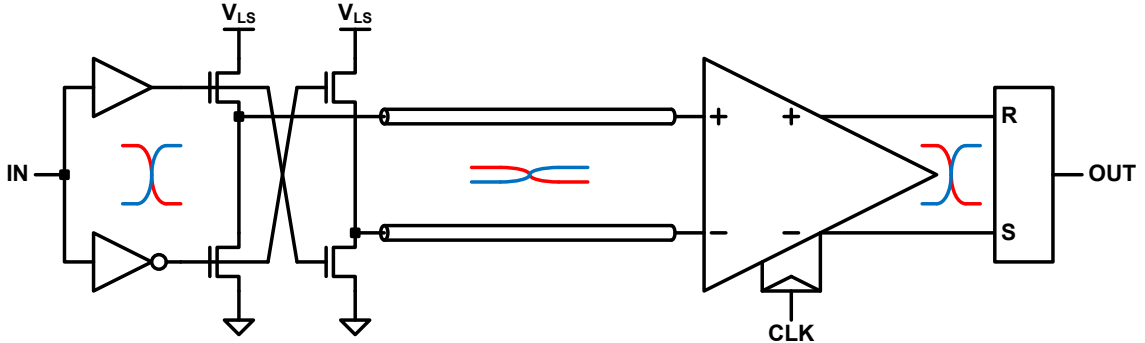


Figure 2.1: Multi-supply interconnect architecture.

2.1.1 Wire Energy

For the MSI, $V_{driver} = V_{swing} = V_{LS}$. Substituting this into equation 2.3, and taking into account of the doubling in wire capacitance due to differential signaling^{iv}, the ratio of the FSI wire energy to the MSI wire energy is given by

$$\frac{E_{dyn(FSI)}}{E_{dyn(MSI)}} = \frac{1}{2} \left(\frac{V_{LS}}{V_{dd}} \right)^2. \quad (2.5)$$

In other words, the energy saving is proportional to the square of the ratio of the low-swing supply to the core supply. Because of this quadratic dependency, even moderate reductions in voltage can result in large improvements. For example, if V_{LS} is one-fifth of V_{DD} then we can expect a 12.5 times reduction in wire energy. In practice, the energy-efficiency of the MSI is limited by its transmitter and receiver, which are more complex than their FSI counterparts.

2.1.2 Low-Swing Supply

A drawback of the MSI is the need for a dedicated second supply. Modern processors can draw a large current. High-end CPUs from Intel and GPUs from NVidia routinely

^{iv}In this analysis we assume that the channel capacitance is well approximated by the wire capacitances. In reality the channel capacitance of the FSI will be slightly higher due to the presence of the repeaters.

have peak power ratings in excess of 100 W [4, 3, 6]. Since these processors operate at a core supply around 1 V, this translates to about 100 A of peak supply current. Distributing such large currents while maintaining supply integrity is one of the most difficult challenges faced by chip designers. Introducing yet another supply takes away the already scarce wiring resources away from the core supply, making it even more difficult to build a reliable power distribution network.

Apart from distribution, the generation of the low-swing supply itself can also pose challenges. Strictly speaking, the quadratic energy saving can only be achieved if the low-swing supply can be generated at zero loss. In practice, all DC-DC converters have limited efficiency so some loss is inevitable. To achieve a high conversion efficiency a switching regulator is typically needed. This usually require large passive components which are difficult to build on-chip. Alternatively, the low-swing supply can be generated off-chip. However, this requires the supply to be brought on-chip which puts more pressure on the limited number of package pins.

2.1.3 Bandwidth Reduction

In the FSI, the inverters used for buffering are independent from the flip-flops used for sequencing. As wire length grows, more inverters can be added to keep the increase in delay linear. In the MSI, the need for a clocked comparator means that buffering can only be added at clock boundaries. Unless clocks with different phases can be used, which can be costly to generate, the inability to add additional buffering result in a wire delay that increases quadratically with wire length. In longer interconnects, MSIs can often only achieve a fraction of the bandwidth that can be realized by the FSI.

Figure 2.2 shows the simulated energy and frequency tradeoff for different interconnects in a 90 nm low-power process over a distance of 4 mm. Typical operating conditions are assumed and a data activity factor of 0.25 is used in all energy calculations. We simulate the FSI with different number of inverter stages, and the number of stages with the best energy at a given frequency is plotted. For the MSI, two different values of low-swing supply are simulated: 200 mV and 400 mV respectively. We

also assume ideal supplies and that a minimum signal swing of 100 mV is required for reliable detection at the receivers. Compared to the FSI, the MSI can achieve about a factor of 9 reduction in energy at low-frequencies. However, its maximum frequency is limited to only about 50-60% of the FSI. Since the MSI already use twice the number of wires, the frequency limitation further degrades the maximum bandwidth that can be realized for a given cross-section^v.

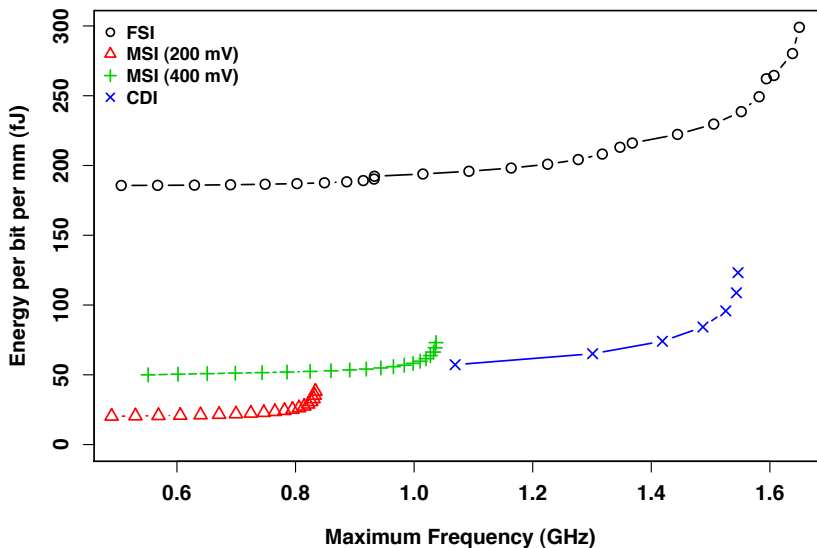


Figure 2.2: Energy and frequency tradeoff for different interconnect architectures over a distance of 4 mm.

2.1.4 Supply Noise Sensitivity

In the MSI the lines are always directly connected to either the power or the ground rail. This means that any noise on the supplies will also be seen on the lines, and hence at the inputs of the receiver, during normal operation. Studies have shown that it is common for the supply rails to exhibit 50 mV or more peak-to-peak ripple around their nominal DC values [8, 31]. This makes it difficult to have robust communication at a very low signal swing.

^vFrequency limitations can be mitigated by pipelining the interconnect, but this increases communication latency.

2.2 Capacitively-Driven Interconnect

The capacitively-driven interconnect (CDI), shown in figure 2.3, generates a low-swing signal without a second supply by using a coupling capacitor in series with the wire [36, 24, 41]. The coupling capacitor forms a voltage divider with the capacitance of the wire, creating a signal swing given by

$$V_{swing} = \frac{C_c}{C_c + C_{wire}} V_{dd} \quad (2.6)$$

where C_c is the size of the coupling capacitor. Apart from the way the signals are generated, the operation of the CDI is similar to the MSI. The signals are transmitted differentially, and a clocked comparator is used to recover the full-swing signal at the receiver.

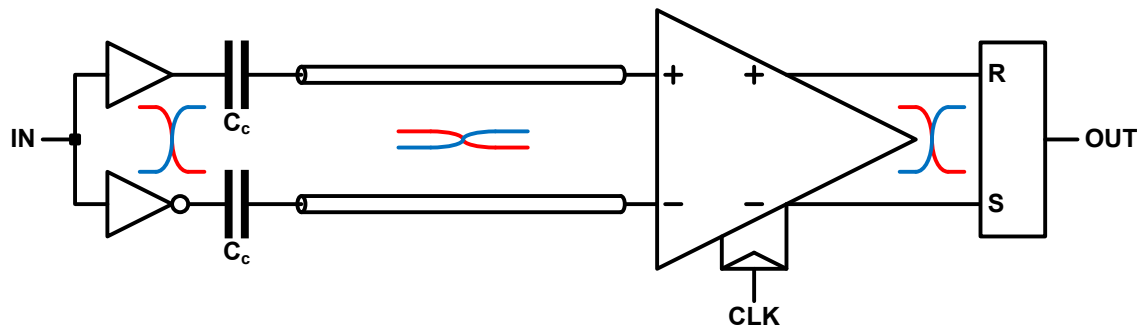


Figure 2.3: Capacitively-driven interconnect architecture.

2.2.1 Supply Simplicity and Flexibility

Since the CDI does not use a second supply it does not disrupt the core supply network. In addition, by adjusting the ratio of the coupling capacitor to the wire capacitance, designers can vary the signal swing as needed for different interconnects on the same chip. This is useful, for example, to fine tune the energy performance of interconnects of different lengths. For a shorter interconnect, a higher swing can be selected so that a smaller sense amplifier can be used, minimizing clock power. For

a longer interconnect, where the wire is more dominant, a lower swing can be used to minimize wire energy. MSI cannot provide this kind of flexibility because it would require multiple additional supplies.

2.2.2 Equalized Signaling

Apart from generating the low-swing signal, the coupling capacitor also improves the frequency response of the wire, allowing the CDI to achieve a higher bandwidth than the MSI. Consider the first order RC-model of a wire segment as shown in figure 2.4a. The transfer function for this system can be readily derived as

$$H(s) = \frac{V_o}{V_i}(s) = \frac{1}{1 + s\tau} \quad (2.7)$$

where we have defined $\tau \triangleq \frac{R_w C_w}{2}$ for convenience. This is a first order system with a cutoff frequency given by

$$\omega_{-3dB,plain} = \frac{1}{\tau}. \quad (2.8)$$

Theoretically, if we can preprocess the signal to be transmitted through a another filter whose frequency response is the inverse of that of the wire segment,

$$H_{prep}(s) = H^{-1}(s) = 1 + s\tau, \quad (2.9)$$

then the combined transfer function of the system will be

$$H_{combined}(s) = H^{-1}(s)H(s) = 1. \quad (2.10)$$

This creates an ideal communication channel where the signal is instantaneously transmitted to the receiver with zero delay. In practice, the ideal inverse filter cannot be realized as it would require infinite power at high frequencies. However, by pre-shaping the signal appropriately, it is possible to significantly extend the bandwidth of the wire segment. This technique, known as equalized signaling, was used earlier by Dally et al. for high-speed I/O circuits [14].

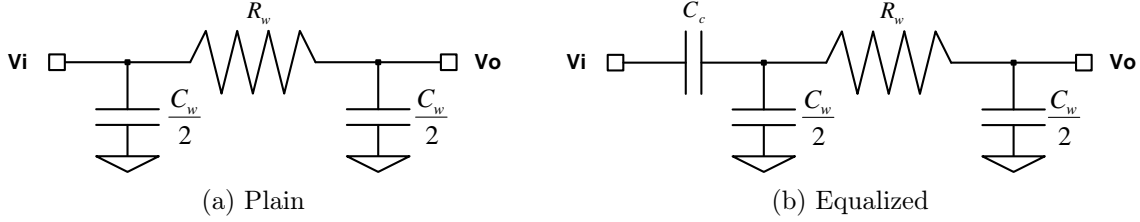


Figure 2.4: RC models for plain and equalized wire segments.

To see how the CDI benefits from equalized signaling, consider the first-order RC model for the capacitor-coupled wire segment in figure 2.4b. With a bit of algebra, the transfer function for this system can be derived as

$$H_{equalized}(s) = \frac{1}{a_1 s + a_0} \quad (2.11)$$

where

$$a_0 = \rho + 1, \quad a_1 = (0.5\rho + 1)\tau, \quad (2.12)$$

and $\rho = \frac{C_w}{C_c}$ is the ratio of the wire capacitance to the coupling capacitance^{vi}. The cutoff frequency for this system is given by

$$\omega_{-3dB,equalized} = \frac{a_0}{a_1} = \frac{\rho + 1}{(0.5\rho + 1)\tau}. \quad (2.13)$$

Dividing equation 2.13 by equation 2.8, the bandwidth ratio of the equalized wire segment to the plain wire segment can be expressed as

$$\frac{\omega_{-3dB,equalized}}{\omega_{-3dB,plain}} = \frac{\rho + 1}{0.5\rho + 1} \quad (2.14)$$

which is purely a function of ρ . For most designs, the signal swing is kept low which

^{vi}See appendix A for details of the derivation.

implies that $\rho \gg 1$, and equation 2.14 can be further simplified to

$$\frac{\omega_{-3dB,equalized}}{\omega_{-3dB,plain}} \approx \frac{\rho}{0.5\rho} = 2. \quad (2.15)$$

Equation 2.15 essentially states that under most circumstances the CDI will have approximately twice the wire bandwidth of an equivalent MSI, which does not use equalized signaling. This effect can be clearly seen in figure 2.2 where the energy-frequency tradeoffs for the CDI are plotted alongside the FSI and the MSI. In this example, the CDI peaks at over 90% of the maximum FSI frequency, a more than 50% improvement over the fastest MSI.

2.2.3 Supply Noise Filtering

In the CDI, supply noise is not directly passed onto the lines at the transmitter. Instead, the noise is attenuated by the voltage-divider formed by the coupling-capacitor and the wire in the same way that the low-swing signal is generated. As the signal swing is reduced, supply noise on the lines are also reduced proportionally. This makes signaling at very low swings more attainable.

2.2.4 Wire Energy

Substituting equation 2.6 into equation 2.3, the ratio of the FSI wire energy to the CDI wire energy can be expressed as

$$\frac{E_{dyn(FSI)}}{E_{dyn(CDI)}} = \frac{1}{2} \frac{C_c}{C_c + C_{wire}} \approx \frac{1}{2} \rho. \quad (2.16)$$

Unlike in the MSI, as we reduce the signal swing in the CDI (increase ρ), wire energy only improves linearly. In order for the CDI to achieve the same wire energy saving as the MSI, a lower swing is needed.

2.2.5 DC Bias

In a CDI, the lines are DC-isolated from the supplies by the coupling capacitors. For proper operation, the line voltages must be initially set and then maintained at the required levels in spite of any leakage in the system. The most straightforward approach, which we call dynamic refresh, is illustrated in figure 2.5. For clarity, only one differential branch is shown. Here the transmitter outputs are periodically forced high by asserting PC while both lines are precharged to the supply voltage. Provided that this occurs frequently enough, the line voltages cannot drift far enough to cause problems^{vii}. A side-effect of using dynamic refresh is that data cannot be transmitted while the interconnect is being refreshed. Depending on the application, this may or may not be a problem. If necessary, redundant interconnects can be introduced so that when an interconnect is being refreshed, another can be switched in its place to ensure that there are no interruptions in the data transmission.

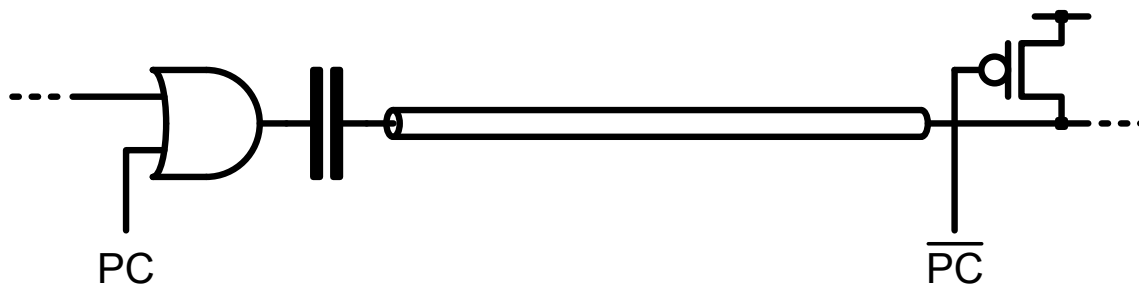


Figure 2.5: DC bias through dynamic refresh.

A different bias scheme, shown in figure 2.6, connects each line to the supply through a leaky PFET [24]. During initialization, the STARTUP signal is held high. This forces the output of the driver to approximately half the supply voltage, and leakage through the PFET charges the line to the full supply voltage. During normal operation, STARTUP is low and the output of the driver swings full-rail. This in turn cause the line voltage to swing both above and below the nominal supply voltage. If

^{vii}If the frequency of refresh can be kept at 2 to 3 orders of magnitude below the operating frequency, the energy overhead of this approach is negligible.

the data is DC balanced over the the time constant of the leaky PFET, the common mode line voltage will not drift too far from their initial value.

While this approach overcomes the need for periodic precharge, the need for a DC balanced data limits its usability. In general, there is no guarantee that there will be an equal number of ones and zeros over any period of time. For example, if an interconnect is used to carry the most significant bit of a long integer then the probability of observing a zero will be much higher than the probability of observing a one. Moreover, due to the burstiness of the traffic, an interconnect can go through periods of inactivity, where it stays at one logical value for a prolonged period of time. When this happens, the leaky PFET slowly pulls the line back towards the supply voltage, eroding the magnitude of the signal swing. Another concern of using this scheme is that the line can swing above the supply voltage during normal operations. In addition to increasing the leakage through the PFET, this can also put excessive stress on the gate oxides at the receiver.

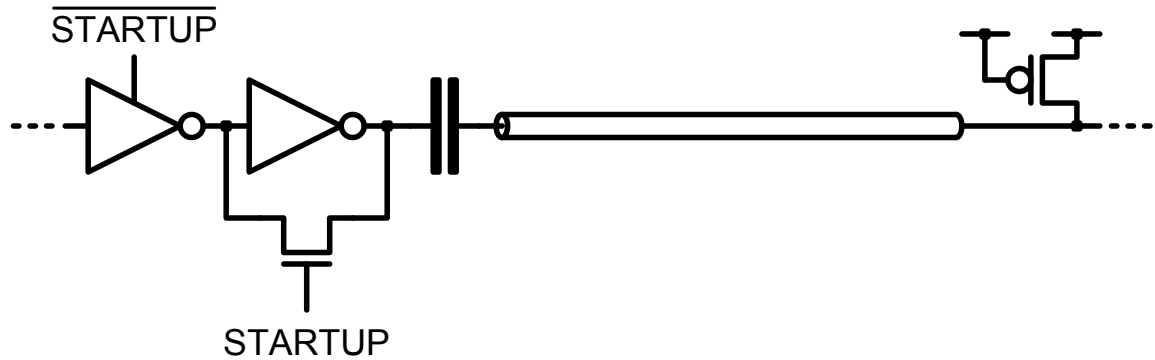


Figure 2.6: DC bias with leaky PFET.

The DC bias can be established directly by adding a load resistance R_L , implemented with a PFET, and a transconductance G_m , implemented with an NFET controlled by the input voltage, on each of the capacitor-coupled lines [36]. Figure 2.7 illustrates this technique. When the output is high, the NFET is off and the the PFET keep the line at the supply voltage. When the output is low, the NFET turns on and a new DC equilibrium is established at $(1 - G_m R_L)V_{DD}$. By adjusting the value of

G_m and R_L , this value can be set to equal the ideal final line voltage determined by the capacitor ratio. This way, no matter what logical state the interconnect is in, the DC voltage on the lines will not drift. In practice, process, voltage and temperature (PVT) variations will cause the DC equilibrium voltage to drift from its ideal value, creating an overhead that must be compensated for by increasing the signal swing. Furthermore, biasing the lines this way leads to static energy dissipation. This can be a problem at low data activity factors or in shorter interconnects where there are less room for amortization.

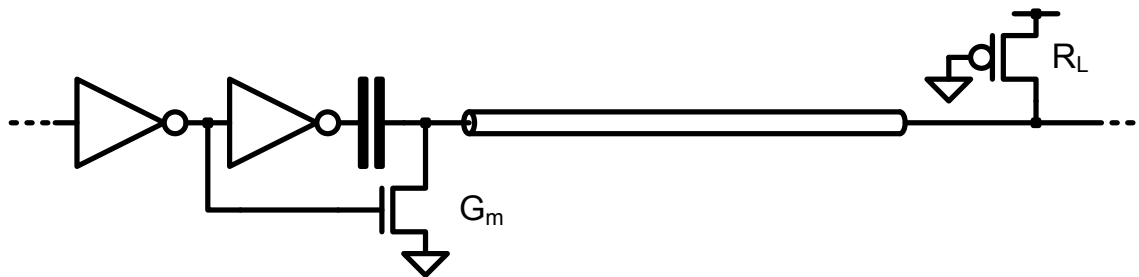


Figure 2.7: DC bias with transconductance and load resistance.

2.3 Technology Trends

The wire energy of the FSI is proportional to the square of the supply voltage. As device feature size scales down supply voltage also decreases. Figure 2.8 shows that between the 250 nm and the the 32 nm node supply voltage reduced by a factor of about 2.5^{viii} [7]. Everything else equal, this corresponds to an over a six-fold reduction in wire energy^{ix}.

In contrast, the wire energy of an LSI is proportional to the signal swing. The minimum swing that can be used is the smallest voltage that can be reliably detected at the receiver. For modern processes, the largest limitation on signal swing is device mismatch, where two identically-drawn devices exhibit different voltage-current

^{viii}Supply voltage scaling has slowed down significantly recently due to leakage considerations.

^{ix}It should be noted that while wire energy is decreasing total power is still rising due to increasing density and higher operating frequency.

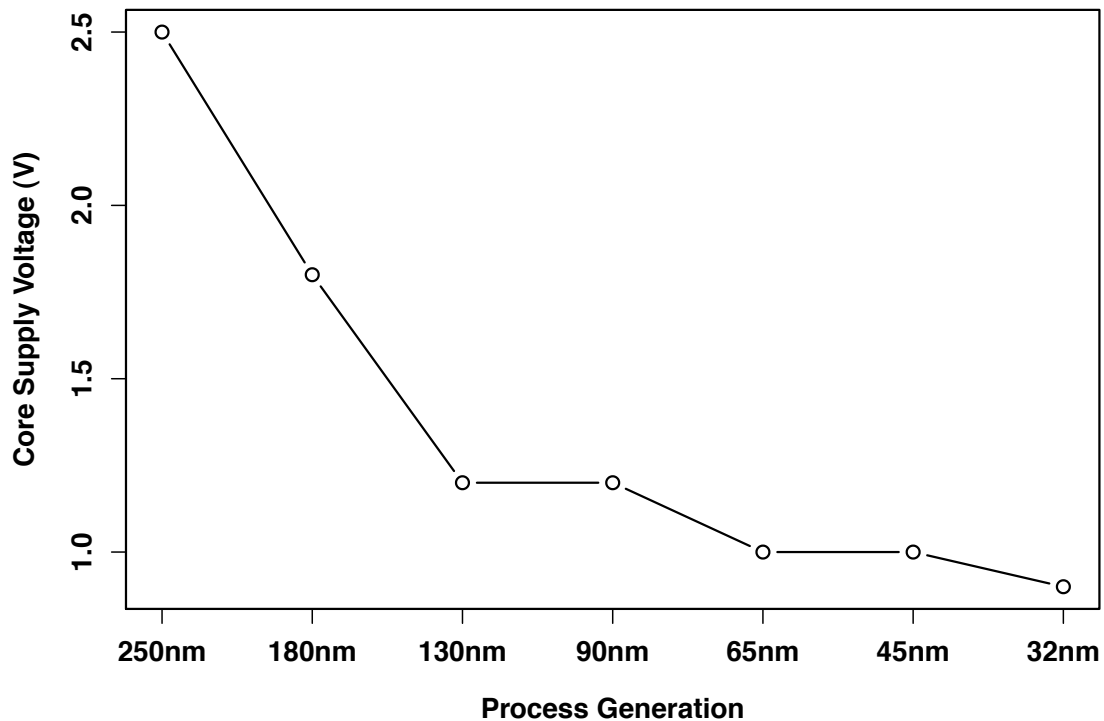


Figure 2.8: Changes in supply voltage over different CMOS process generations.

relationships after fabrication. Device mismatch is caused by imperfections during fabrication such as random dopant fluctuation, line edge roughness, oxide thickness variations, and proximity effects [35, 18, 10, 13, 34, 9, 37, 44].

Figure 2.9a shows the StrongARM latch [38], a popular comparator used in LSI receivers. Ideally, at a rising clock edge, the small differential voltage on the inputs INP and INN is converted into a differential current through M1A and M1B. The differential current is then amplified through positive feedback by the cross-coupled inverters formed by M2A, M2B, M3A and M3B, generating a full-swing differential output across OUTP and OUTN. In the presence of mismatch, a differential current can exist between M1A and M1B even when there is no differential voltage across the inputs. Moreover, mismatch in the cross-coupled inverters can lead to asymmetry in the positive feedback mechanism. The combined effect is that the comparator is predisposed to regenerate to one logical value over the other. As shown in figure 2.9b,

this is typically modeled as a DC voltage source that is connected in series with the input of an ideal comparator. The magnitude of the source, known as the input-offset voltage, sets a lower-bound on the signal swing that can be reliably used.

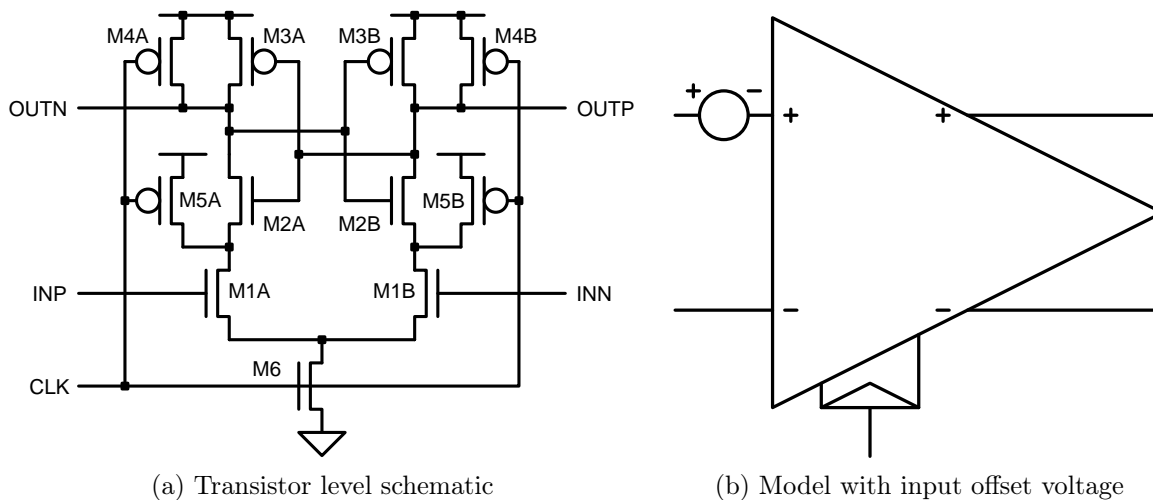


Figure 2.9: The StrongARM latch and its equivalent circuit model in the presence of mismatch.

Figure 2.10 shows that device mismatch is becoming worse with technology scaling. Between the 130 nm and the 45 nm node, threshold mismatch increased by approximately 67%. For LSIs, this means that the input offset voltage, and hence the minimum signal swing, is increasing. Compounded by the decreasing supply, it is becoming more difficult to save wire energy with LSIs. This is particularly a problem for the CDI, where wire energy is only linearly dependent on swing.

Device mismatch can be countered by either increasing the signal swing or increasing the size of the comparator. Studies have shown that the standard deviation of the threshold and current factor mismatch can be expressed as

$$\sigma_{V_t} = \frac{A_{V_t}}{\sqrt{WL}}, \quad \sigma_{\beta} = \frac{A_{\beta}}{\sqrt{WL}} \quad (2.17)$$

where A_{V_t} and A_{β} are process dependent constants, and W and L are the width and channel length of the matching devices respectively [40, 37, 44]. Increasing device sizes, either W or L , reduces the amount of mismatch and allows a lower signaling

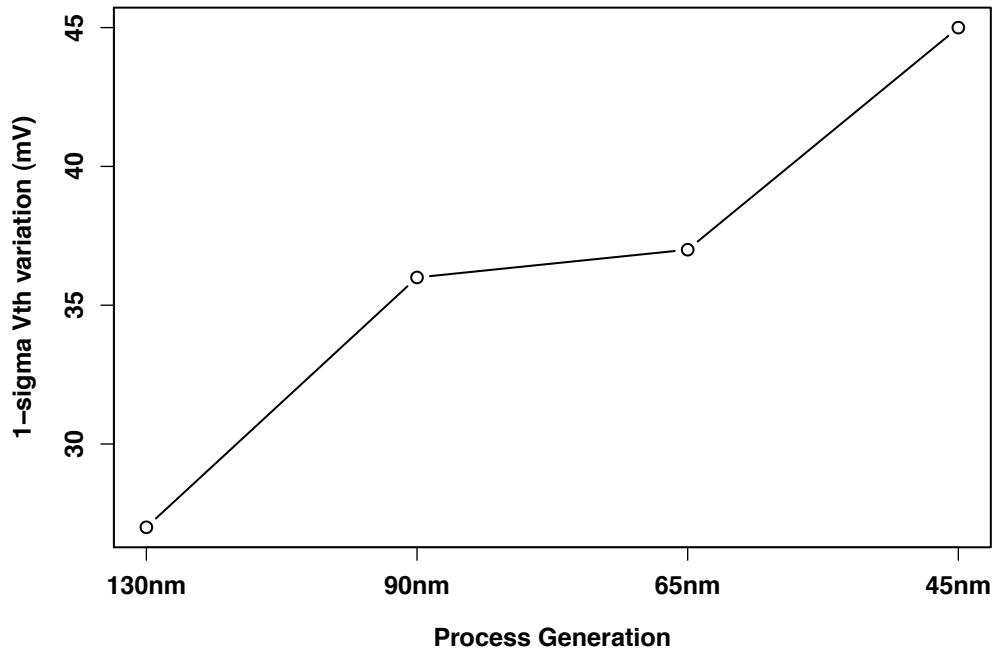


Figure 2.10: Threshold mismatch for minimum-sized NFETs over different CMOS process generations.

voltage to be used. Unfortunately, upsizing devices also increases energy, and this is particularly problematic for comparators since they have high activity factors. Moreover, the square root dependency in equation 2.17 means that large increases in device dimensions are needed for a small improvement in mismatch.

2.4 Offset Compensation

Offset compensation is the collection of design techniques used to minimize the impact of device mismatch. They were first applied to comparators in memory arrays to improve bitline energy and read access time [45, 19]. These earlier techniques adjust the bitline voltages by precharging them through the matching devices. They are ineffective against current factor mismatch and require a long precharge period, which degrades bandwidth. Precharging also increases the wire activity factor, and hence

the energy. In other studies, offset compensation has been applied to current sense amplifiers and designs based on continuous-time amplifiers [42, 47]. These techniques can improve latency but require sizable static bias currents, making them energy-inefficient.

A digital offset compensation scheme, first adopted by Ellersick et al. as part of an equalized multi-level link, is illustrated in figure 2.11 [17, 11, 48]. In this approach, digitally controlled current sources (DCCS) are added in parallel with the input sensing devices. By programming the appropriate binary values into the control registers, current can be added to either branch of the comparator to neutralize intrinsic offsets. This technique is an improvement over the previous methods because it is effective against both threshold and current factor mismatch, requires no bias current, and does not need any precharge periods.

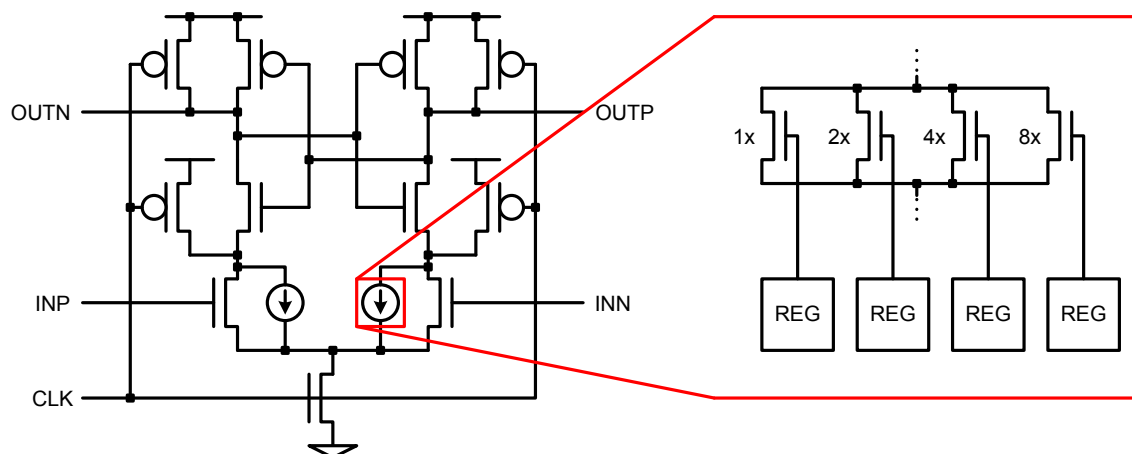


Figure 2.11: Offset compensation with digitally controlled current sources.

The primary cost of the DCCS scheme is the area overhead of the registers and the control logic needed to read and write those registers. During startup, the controller must determine the appropriate binary value to store into the control register for every compensated comparator. To do this, the inputs of the sense amplifier are tied to a common voltage, typically the supply, and the output is monitored to determine which branch of the amplifier needs extra current and the magnitude of that current. A counter is needed to increment or decrement the binary value until the output

oscillates. This value is then written into the control register and fixed for the duration of the operation. When multiple comparators are present, each comparator must be individually addressed and monitored. This requires a decoder to be built for the control registers and a multiplexer for the comparator outputs. Fortunately, most of these circuits can be disabled during normal operation, so they will not consume excessive additional power.

DCCS based offset compensation works well for I/O circuits where the comparators are large. For on-chip interconnects, where the comparators are smaller due to energy constraints, additional challenges must be considered. A smaller comparator exhibits more mismatch, so there is inherently a larger tradeoff between compensation coverage and resolution. For example, in a 90 nm process, a reasonably sized LSI comparator can have a five standard deviation input offset voltage exceeding 200 mV. In order to ensure that this worst case scenario can be compensated, and assuming a 4-bit control register, the minimum compensation step size is given by

$$V_{step} = \frac{200 \text{ mV}}{2^4} = 12.5 \text{ mV}. \quad (2.18)$$

This means that any offset less than 12.5 mV will not be able to be corrected. The only way to increase compensation resolution is to increase the number of bits in the control register, but this cannot be extended indefinitely because every new current branch adds more parasitic capacitance to the sense amplifier. Adding more branches also further complicates the layout, making it more difficult to maintain the symmetry needed for good matching.

Even if the number of bits is not an issue, minimum device dimensions can still limit the resolution of the compensation. To see this, consider the current ratio required to compensate for a 10 mV offset. In the 90 nm process, the supply voltage is 1.2 V, and the nominal threshold voltage of the devices is around 300 mV. This leaves 900 mV of overdrive under normal operations. A 10 mV mismatch corresponds to a one-ninetieth change in overdrive, which to first order translate to approximately a one-ninetieth change in device current. In order to compensate for the 10 mV mismatch, a branch with only one-ninetieth the current of the input device is required.

In LSI comparators, the input device is only around ten times the size of the minimum device. This leaves a factor of nine that cannot be accounted for. One way to generate low-current branches is to use a reference current, but this creates static power consumption and complicates the layout. Furthermore, mismatch between devices in the branches and the reference makes it difficult to precisely adjust the size of the currents.

Another compensation scheme closely related to the DCCS approach is illustrated in figure 2.12 [32]. Instead using current branches, this technique neutralizes offsets by using digitally controlled capacitors (DCCAP) to add extra loading on the stronger branch of the comparator. The DCCAP approach is less limited by minimum device dimensions because, to first order, the compensation resolution is determined by the ratio of the minimum extra capacitance to the total capacitance of that branch, not just the capacitance of the input device. The downside is that while the DCCS works by strengthening the weaker branch, the DCCAP works by weakening the stronger branch. This increases the regeneration latency. Moreover, since difference in current is compensated with capacitance, the DCCAP approach tends to have inferior tracking over temperature and supply variations.

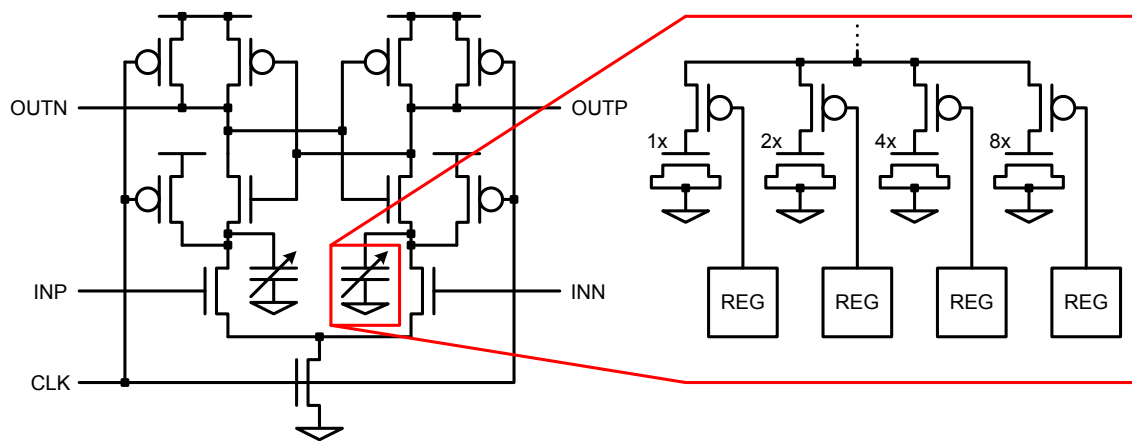


Figure 2.12: Offset compensation with digitally adjustable capacitive loads.

Chapter 3

Self-Calibrating Interconnect

In chapter 2 we showed that the CDI has the potential to achieve low-latency, energy-efficient on-chip communication without a second supply. However, in order for the CDI to realize its full potential, we must establish the DC line voltages and compensate for the increasing input offset voltage caused by device mismatch. Up to now, DC biasing and offset compensation have been addressed as two separate problems. This creates two sets of overheads. In the self-calibrating interconnect (SCI), we take a unified approach where we periodically set, or "calibrate", the DC line voltages to cancel the receiver offset, establishing DC bias and offset compensation in a single mechanism. In essence, the SCI is a CDI augmented with calibration circuits to perform more elaborate line voltage adjustments.

The basic concept of the SCI is illustrated in figure 3.1. Like before, we model the receiver offset as a DC voltage source in series with the input to an ideal comparator. The line that is favored by device mismatch is referred to as the stronger side, and the other line is referred to as the weaker side. In this particular example, the positive line is the stronger side. As in dynamic refresh, data transmission is periodically paused and the outputs of the transmitter are forced high. In this case, however, only the weaker line is precharged to the supply voltage, while the stronger line is precharged to one input offset voltage below the supply. The reduced DC voltage on the positive line neutralizes the intrinsic offset, leaving the lines perfectly matched at the input of the ideal comparator.

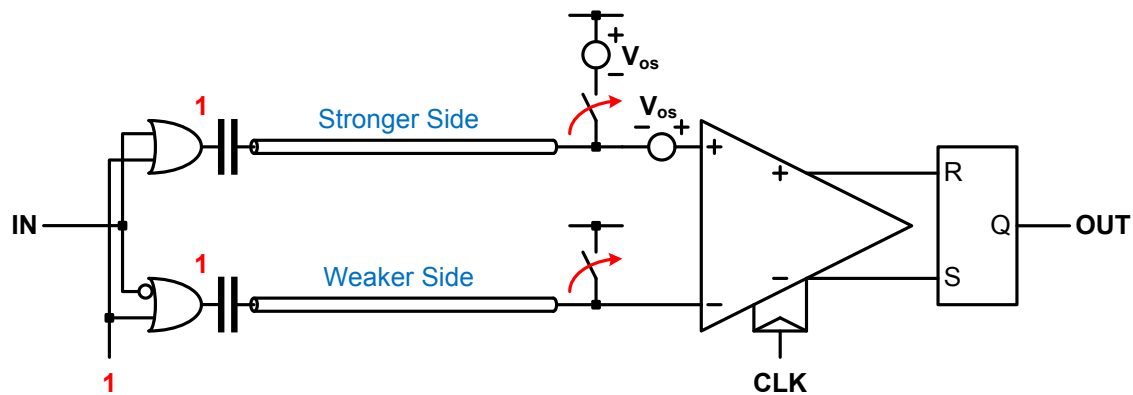


Figure 3.1: Combined DC bias and offset compensation in the SCI.

In practice, since the voltage adjustment is different for the two sides, a mechanism is required to differentiate the stronger side from the weaker side prior to calibration. Moreover, because the input offset is not known at design time, the calibration circuits must be able to dynamically adapt to any post-fabrication variations. In the following sections, we explain how these goals can be accomplished with the smallest amount of circuit overhead.

3.1 Bias Detection

The goal of bias detection is to differentiate the stronger side of the comparator from its weaker side, and making this information available to the voltage adjustment circuits during calibration. As illustrated in figure 3.2, the SCI uses the global reset signal (RES) to perform bias detection. When RES is asserted, the outputs of the transmitter are forced high and both lines are held at the supply voltage. Since the comparator inputs are tied to the same potential, its outputs reflect the intrinsic offset. If the comparator is "zero-biased", that is, the positive line (LINEP) is stronger, then the positive output will be zero and the negative output will be one. Conversely, if the comparator is "one-biased", then the positive output will be one and the negative output will be zero. An RS-latch gated by RES is used to monitor the outputs of the comparator during reset. When RES is deasserted, the latch records the polarity of

the offset in its outputs, SP and SN, for the duration of the operation. By examining SP and SN, the calibration circuits can distinguish the stronger side from the weaker side and adjust the voltages appropriately.

Rarely, random noise can overpower the receiver offset during bias detection and cause the wrong polarity to be recorded. The result is that instead of reducing the line voltage of the stronger side, the calibration circuits will attempt to increase the line voltage of the weaker side to compensate for the offset. Fortunately, since the noise standard deviation is typically only on the order of 3 mV, it is extremely unlikely that a comparator with any sizable offset will be recorded with the wrong polarity. Accordingly, the probability that a line voltage will be raised high enough above the supply to cause problems is vanishingly small. If desired, hysteresis can be built into the RS-latch to further improve the probability of recording the right offset polarity. This can be accomplished, for example, by reducing the drive strength of the devices within the RS-latch so that two consecutive decisions of the same polarity are needed to toggle the state of the latch¹.

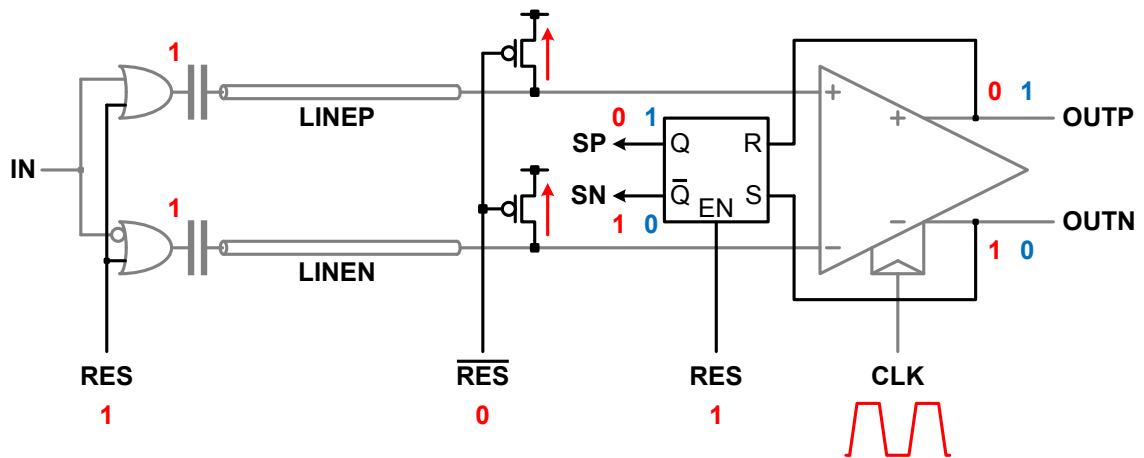


Figure 3.2: Bias detection using the reset signal.

¹Alternatively, a scheme that picks the majority decision over a number of tests can be implemented, although a shared controller is needed in this case due to the higher overhead.

3.2 Calibration Process

The SCI holds one line at the supply voltage and adjusts the voltage of the other line to compensate for the offset. Since the offset is unknown at design time, we use a feedback-driven, iterative approach for voltage adjustments as shown in figure 3.3. Here we assume that bias detection is already complete and that the comparator is identified as being zero-biased. The positive line (LINEP) is held at the supply and voltage adjustments are performed on the negative line (LINEN). At each rising clock edge, if a "1" is observed at the negative output ("0" at the positive output) then the voltage on LINEN is reduced by a fixed amount, V_{step} . Similarly, if a "0" is observed at the negative output ("1" at the positive output), then the voltage on LINEN is increased by the same amount. In this example, both lines start at the supply, so the voltage on LINEN is initially stepped down every cycle. Eventually, the voltage difference between the two lines exceeds the offset voltage, V_{os} , and the comparator outputs begins to toggle. In response, the voltage on LINEN plateaus and oscillates around one offset voltage below LINEP. At this point, the offset is neutralized to within ϵ , the residual offset.

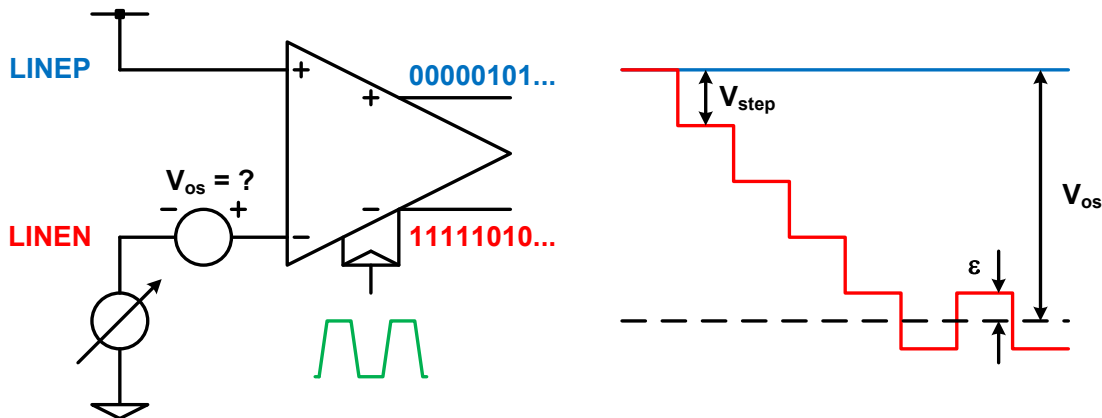


Figure 3.3: Offset compensation through iterative refinement.

3.2.1 Interconnect Availability

Since data transmission is paused during calibrations, it is desirable to minimize both the frequency and duration of the calibrations. In our design, calibration interval refers to the number of cycles between successive calibrations, while calibration duration is the number of cycles devoted to each calibration. The availability of the interconnect, or the fraction of time that it can be used for data transmission, is given by

$$\textit{Availability} = 1 - \frac{\textit{Duration}}{\textit{Interval}}. \quad (3.1)$$

To achieve a high availability, we want a large calibration interval and a small calibration duration.

3.2.2 Incremental Calibration

Figure 3.4 shows that there is a trade-off between the convergence rate and the maximum achievable resolution of the calibration. A larger step size (left) converges to the final equilibrium value in fewer cycles (N_1), but leaves a larger residual offset upon completion (ϵ_1). In contrast, a smaller step size (right) takes more cycles to converge (N_2), but leaves a smaller residual offset (ϵ_2).

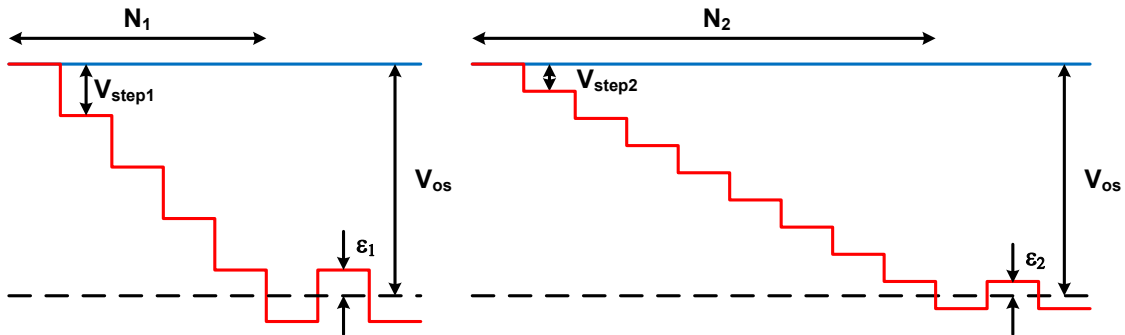


Figure 3.4: Trade-off between convergence rate and calibration resolution.

As mentioned earlier, comparators for on-chip interconnects can have very high worst case offsets. This motivates the use of larger step sizes in order to keep the calibration duration at a moderate value. On the other hand, we would also like to keep the residual offset low because it directly affects the minimum signal swing that can be reliably used. That motivates the use of smaller step sizes. To avoid this conflict of interest, the SCI by construction uses incremental calibration, where each calibration starts at where the previous left off. Under this scheme, once the interconnect is warmed up, only drifts caused by leakage in-between calibrations need to be corrected for instead of the entire worst case offset. This allows the use of very small step sizes, typically around 1 mV, while still maintaining a fast convergence rate, usually less than 20 cycles.

An alternative to using incremental calibration is to use variable step sizes. The idea here is to use a larger step size initially to speed up convergence, but reduce the step size towards the end to minimize the residual offset. While very appealing in theory, the overhead required to implement this technique for each bit of the interconnect can be quite significant.

3.3 Voltage Adjustment Circuits

The SCI use switched-capacitor charge pumps to perform line voltage adjustments. Switched-capacitors are attractive because capacitor ratios, unlike device current, can be more precisely controlled in the presence of process variations. Moreover, their switched operations fit very nicely into the discrete-time feedback loop of the comparator.

3.3.1 Decrementer

The operation of the charge pump used to reduce the line voltage is shown in figure 3.5. Here we model the wire of the interconnect as a lumped capacitance, C_{wire} , to ground. In the first half cycle, a small adjustment capacitor, C_{adj} , is discharged to ground. In the second half cycle, the adjustment capacitor is connected in parallel with the

wire capacitance. Using the conservation of charge, we can derive the change in line voltage, V_{wire} , between two successive time steps, n and $n + 1$, as

$$\Delta V_{wire} = V_{wire}[n + 1] - V_{wire}[n] \quad (3.2)$$

$$= \frac{C_{wire} V_{wire}[n]}{C_{wire} + C_{adj}} - \frac{(C_{wire} + C_{adj}) V_{wire}[n]}{C_{wire} + C_{adj}} \quad (3.3)$$

$$= -\frac{C_{adj}}{C_{wire} + C_{adj}} V_{wire}[n]. \quad (3.4)$$

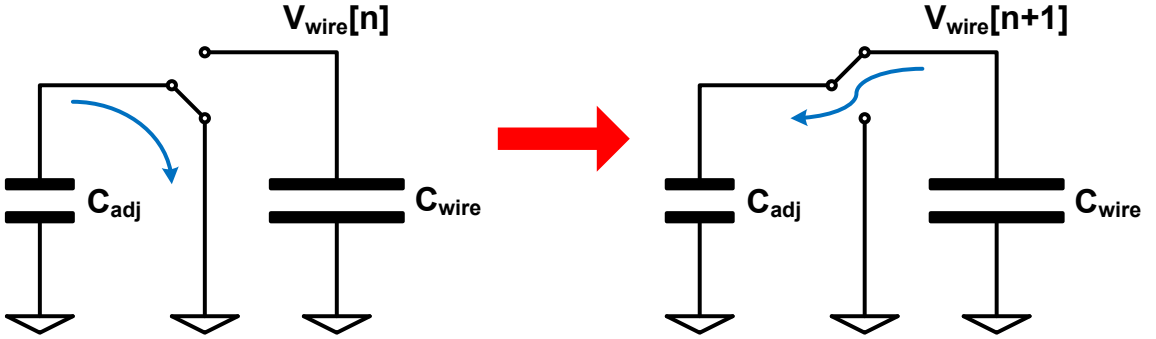


Figure 3.5: Charge pump for decrementing line voltage.

Like most CDIs, the SCI use supply referenced low-swing signaling where the line voltages are kept near the supply voltage during normal operations ($V_{wire} \approx V_{dd}$). In addition, for the step sizes of interest, the adjustment capacitance is usually 2 to 3 orders of magnitude below the wire capacitance ($C_{adj} \ll C_{wire}$). Under these assumptions, equation 3.4 simplifies to

$$\Delta V_{wire} \approx -\frac{C_{adj}}{C_{wire}} V_{dd}, \quad (3.5)$$

where the step size is proportional to the ratio of the capacitances.

Figure 3.6 shows the transistor level implementation of the voltage decremter, where the switching operation is controlled by a qualified output of the comparator. When the comparator is precharged, the adjustment capacitor is discharged through the NFET. Once a decision of the appropriate polarity is made, the adjustment capacitor is connected in parallel with the wire capacitance through the PFET. Even though this is a pull-down operation, PFET is used for the access device because the final voltage across both capacitors is close to the supply.

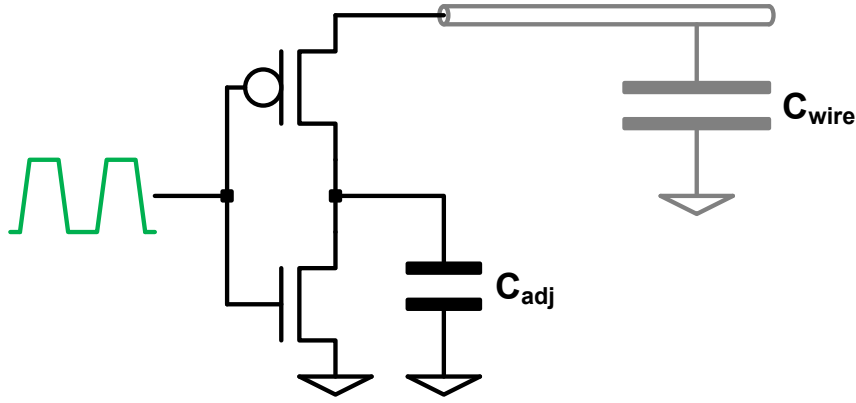


Figure 3.6: Decremter schematics.

3.3.2 Incrementer

As shown in figure 3.7, a similar construction can be used to increment the line voltage. Here the adjustment capacitor is first charged to the supply voltage in the first half cycle and then connected in parallel with the wire capacitance in the second half cycle. The voltage step size can be derived as

$$\Delta V_{wire} = V_{wire}[n+1] - V_{wire}[n] \quad (3.6)$$

$$= \frac{C_{wire}V_{wire}[n] + C_{adj}V_{dd}}{C_{wire} + C_{adj}} - \frac{(C_{wire} + C_{adj})V_{wire}[n]}{C_{wire} + C_{adj}} \quad (3.7)$$

$$= \frac{C_{adj}}{C_{wire} + C_{adj}} (V_{dd} - V_{wire}[n]). \quad (3.8)$$

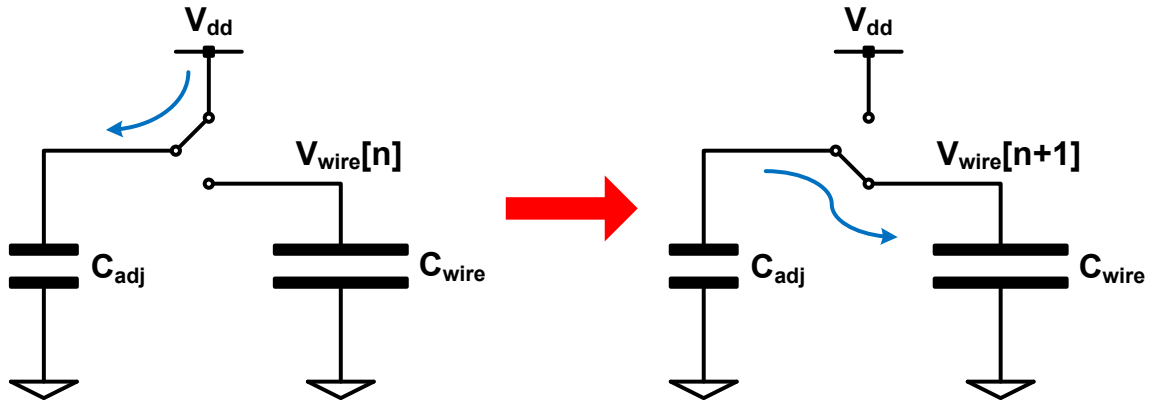


Figure 3.7: Charge pump for incrementing line voltage.

This time, however, when we make the substitutions $V_{wire} \approx V_{dd}$ and $C_{adj} \ll C_{wire}$ equation 3.8 becomes

$$\Delta V_{wire} \approx 0. \quad (3.9)$$

The problem here is that the wire is already at a voltage close to the supply, so connecting another small capacitance charged to the supply does not lead to much additional charge injection. We can get a larger step size by increasing the size of the adjustment capacitor. Unfortunately, as the line voltage gets closer and closer to the supply the size of the adjustment capacitor required becomes exponentially larger, rendering this technique impractical.

To avoid the problem of asymmetric step size the SCI uses a bootstrapped voltage incremter as shown in figure 3.8. Here, as before, the adjustment capacitor is charged to the supply during the first half cycle. However, in the second half cycle, instead of just connecting the capacitors in parallel, the tail end of the adjustment capacitor is simultaneously pulled-up to the supply. This induces a larger voltage change across adjustment capacitor, forcing more charge to be injected into the wire.

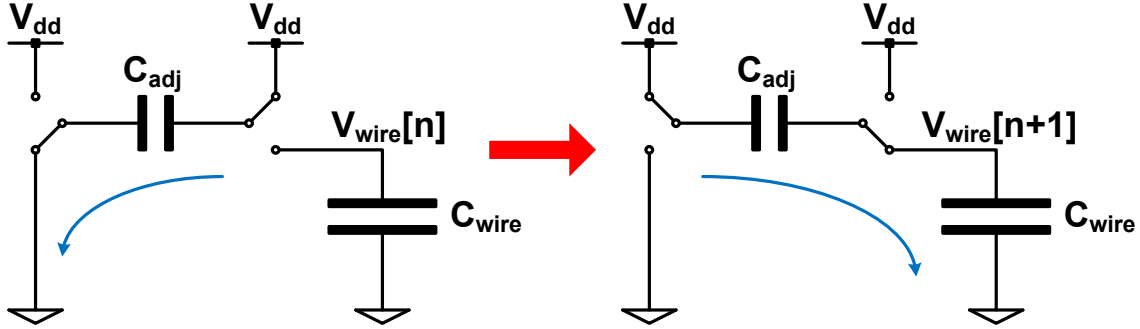


Figure 3.8: Bootstrapped charge pump for incrementing line voltage.

To find the step size we start with the change in charge stored in adjustment capacitor, Q_{adj} , which is given by

$$\Delta Q_{adj} = C_{adj} \Delta V_{adj} \quad (3.10)$$

$$= C_{adj} [(V_{wire}[n+1] - V_{dd}) - V_{dd}] \quad (3.11)$$

$$= C_{adj} (V_{wire}[n+1] - 2V_{dd}). \quad (3.12)$$

The step size can then be derived as

$$\Delta V_{wire} = \frac{\Delta Q_{wire}}{C_{wire}} = \frac{-\Delta Q_{adj}}{C_{wire}} \quad (3.13)$$

$$= \frac{C_{adj}}{C_{wire}} (2V_{dd} - V_{wire}[n+1]). \quad (3.14)$$

When we make the approximation that $V_{wire} \approx V_{dd}$, we get

$$\Delta V_{wire} \approx \frac{C_{adj}}{C_{wire}} V_{dd}, \quad (3.15)$$

where the step size is again proportional to the ratio of the two capacitances. For a given step size, bootstrapping allows the use of an adjustment capacitor that is comparable in size to that of the decremter. Figure 3.9 shows the circuit implementation of the bootstrapped incremter.

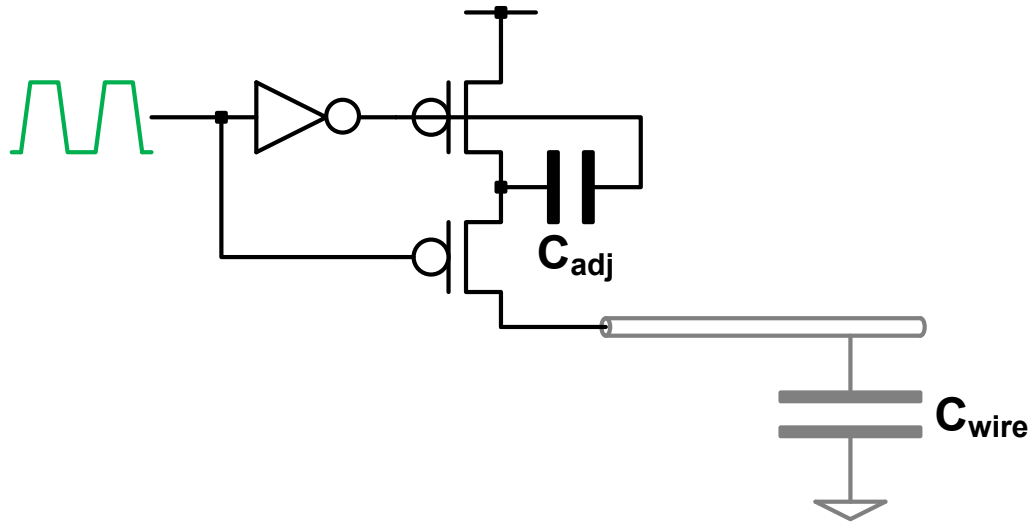


Figure 3.9: Bootstrapped incrementer schematics.

3.4 Leakage

The SCI relies on charge storage for proper operation. Over time, leakage cause the DC line voltages to drift from their ideal values, necessitating recalibration. As shown in figure 3.10, leakage on the lines can be either differential or common-mode. Differential leakage cause the voltage between the lines to diverge or converge and directly cuts into the noise margin. Common-mode leakage, on the other hand, does not change the voltage between the lines, so it does not deteriorate the noise margin. However, if left unchecked, it can cause the input devices at the receiver to fall out of their active region, creating timing errors.

The calibration mechanism of the SCI takes into account of both types of leakage. When the weaker line is precharged to the supply, common-mode drifts are essentially converted into differential drifts. The combined differential drift is then removed through the voltage adjustment process described in section 3.2. Since all voltage adjustments are performed on the stronger line, precharging the weaker line does not undo adjustments made in previous calibrations.

The magnitude of leakage determines the maximum calibration interval that can be used. As leakage increases, we need to calibrate more frequently to ensure that the

line voltages do not drift far enough to cause errors. As shown in section 3.2.1, a larger calibration interval improves availability. For optimal performance, it is important to account for and minimize all sources of leakage.

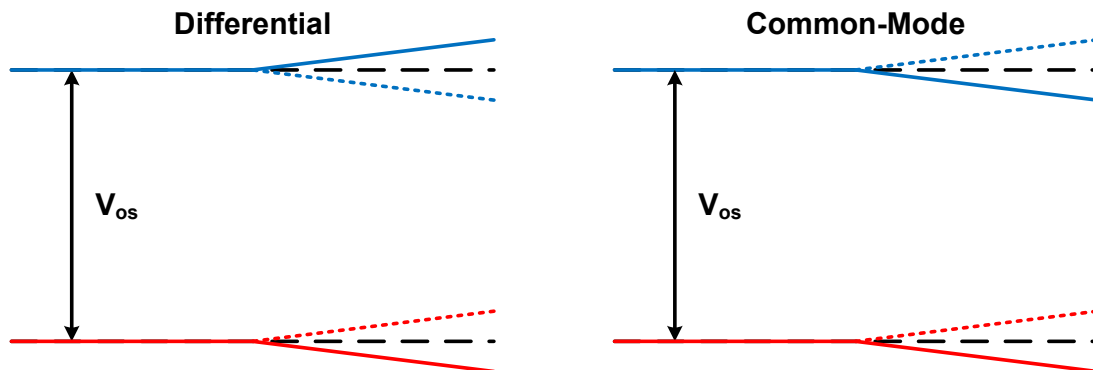


Figure 3.10: Effect of differential and common-mode leakage on line voltages.

3.4.1 Receiver Inputs

The gates of the input devices in the comparator is a source of leakage. Fortunately, these devices have very small dimensions compared to the wire, so their leakage currents are limited. Moreover, gate leakage has a small temperature coefficient, so they will also not be a problem at elevated temperatures.

3.4.2 Coupling Capacitors

The coupling capacitors at the transmitter is another source of leakage. Here gate leakage is again the dominant leakage mechanism. Since the coupling capacitors are sized in proportion to the wire capacitance, their dimensions, and hence leakage, can become significant. In our design the coupling capacitors are implemented out of thick-gate PFETs. Compared to the standard devices, the thick-gate variants require approximately twice the area per unit capacitance, but their gate leakage currents are orders of magnitude lower.

3.4.3 Direct Channel Connections

All devices with channels directly connecting to the wires are sources of leakage. This includes the precharge devices for bias detection and the access devices in the voltage adjustment circuits. For these devices, the drain-to-source leakage current is the dominant leakage mechanism. Since these devices are not on the critical path, we use the highest threshold voltage to minimize their leakage currents.

As discussed earlier, the lines in the SCI are held near the supply voltage during normal operations. This keeps a low drain-to-source voltage across the precharge device and the access device in the incrementer, further reducing their leakage. Unfortunately, for the same reason, the access device in the decrementer can have a large drain-to-source voltage if the implementation in figure 3.6 is used. Figure 3.11 shows the leakage-optimized decrementer used in the SCI. It is essentially the same decrementer from figure 3.6 augmented with three additional leakage control devices (M3-M5). When calibration is active (CAL high), M5 is OFF and M3 and M4 form a direct connection to ground, so this circuit behaves exactly as described before. However, when calibration is inactive (CAL low), M3 and M4 are OFF and M5 charges the node between M3 and M4 to the supply. This forces a negative gate bias across M3, throttling any leakage currents that flows from M1.

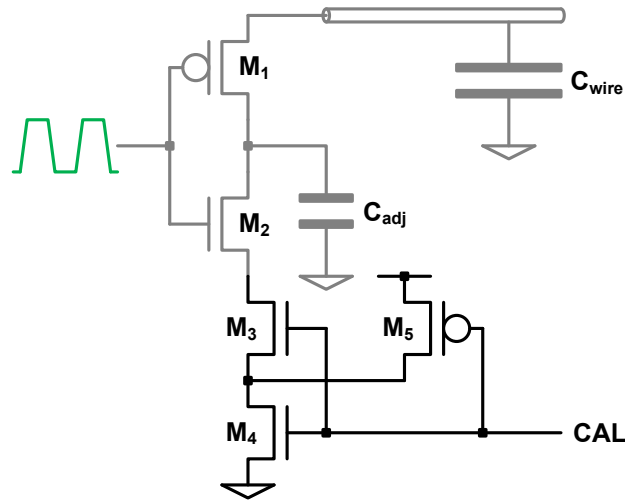


Figure 3.11: Leakage optimized voltage decrementer.

3.5 Complete Design

Putting everything together, the full SCI circuit is illustrated in figure 3.12. Apart from the bias detection circuits and the charge pumps for each line, four AND-gates are used to control the feedback paths from the comparator. This allows the majority of the calibration circuits to be switched off during data transmission, minimizing the load on the critical path. The pull-up PFETs are shared between the bias detection and the voltage adjustment circuits. During reset, both devices are on to keep the lines at the supply voltage. During calibrations, only the selected PFET is activated to restore the weaker line to the supply voltage.

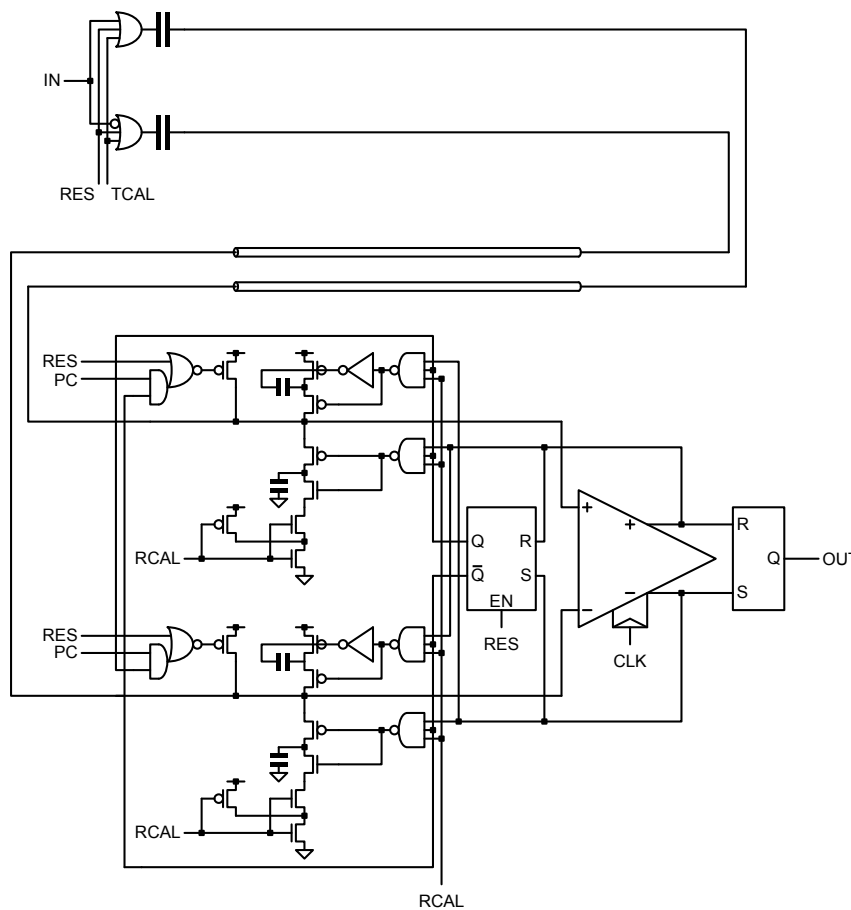


Figure 3.12: Full SCI Schematics.

3.6 Control Signals

As illustrated in figure 3.12, the calibration mechanism for the SCI is controlled by three closely-related signals: Transmitter calibration (TCAL), receiver calibration (RCAL) and precharge (PC).

3.6.1 Timing

Figure 3.13 shows the timing relationship between the different calibration control signals. At the beginning of a calibration, TCAL is asserted to force both transmitter outputs to a known state (logical-1). One cycle later, RCAL is asserted to enable voltage adjustments, and PC is pulsed high for one cycle to precharge the weaker line to the supply voltage. At the end of a calibration, RCAL is deasserted first, followed by TCAL a cycle later. Staggering TCAL and RCAL prevents adjustments from occurring while the line voltages are changing, leading to more accurate calibrations. Similarly, since supply noise can degrade the convergence of the voltage adjustments, pulsing PC is preferred to connecting the weaker line to the supply for the duration of the calibration.

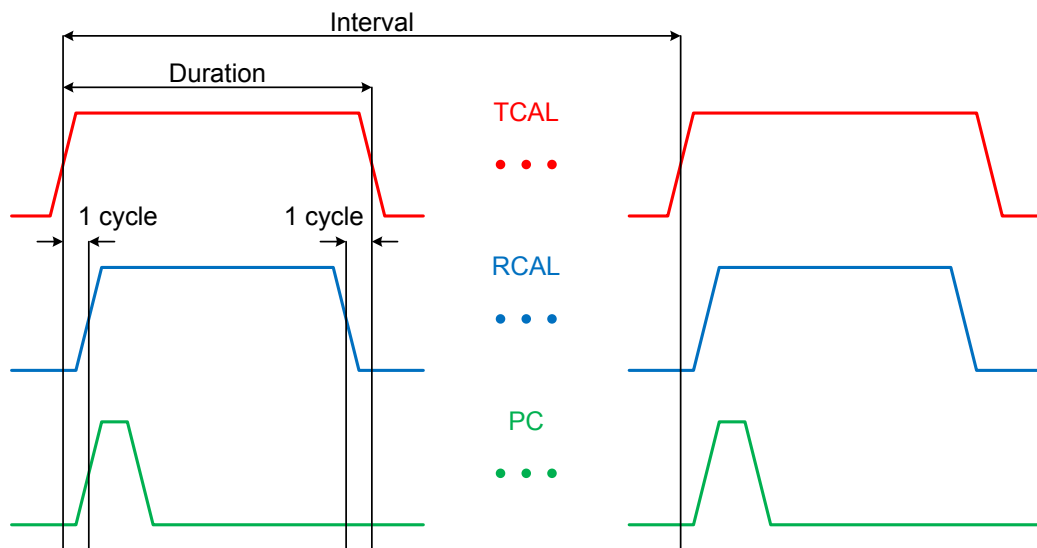


Figure 3.13: Timing relationship between the calibration control signals.

3.6.2 Generation

Designers can choose to generate the calibration control signals in a number different ways, provided that the timing relationship shown in figure 3.13 is preserved. One approach is to generate all three signals at a central counter and then distributing them globally on-chip. Unlike reset, the calibration control signals are easier to distribute globally because different SCIs are not required to enter calibration simultaneously. As long as all signals are still aligned to clock boundaries, different parts of the chip do not need to be synchronized with one another.

Alternatively, given the similarities between the signals, we can use only one global calibration signal and then locally generate the three control signals. Figure 3.14 illustrates this approach, where the global signal (CAL) is periodically asserted for one cycle less than the desired calibration duration. At the transmitters, CAL is logically ORed with a delayed version of itself, giving TCAL that is asserted for the duration of the calibration. At the receivers, CAL is logically ANDed with its delayed version, giving RCAL that is asserted one cycle after TCAL and deasserted one cycle before TCAL. An additional flip-flop with a complementary output is used to pulse PC high for one cycle. Generating control signals locally further simplifies global routing because control delays, measured as the number of cycles from the central counter, no longer need to be matched for multiple signals.

3.6.3 Energy Overhead

Apart from the interconnects themselves, the energy associated with generating and distributing the calibration control signals must also be accounted for. Care must be taken to ensure that control overheads are not excessive, as this will defeat the purpose of using SCIs in the first place. In practice, this means that SCIs should only be used in bulk, where the overhead of the central controller and any local generation logic, if applicable, can be amortized over a large number of interconnects. Most modern systems use either 32-bit or 64-bit words. The actual minimum transfer widths can be even greater in order to meet the bandwidth requirements. This makes the aforementioned constraint relatively easy to satisfy.

The activity factors of the control signals also have a tremendous influence on the energy overhead. Provided that leakage are kept under control, the calibration interval can range from a few thousands to hundreds of thousands of cycles. With such infrequent calibrations, the activity factors of the control signals are three to five orders of magnitude below that of the data signals, making their dynamic energy contributions completely inconsequential. This is yet another reason why leakage minimization techniques are so crucial for the SCI.

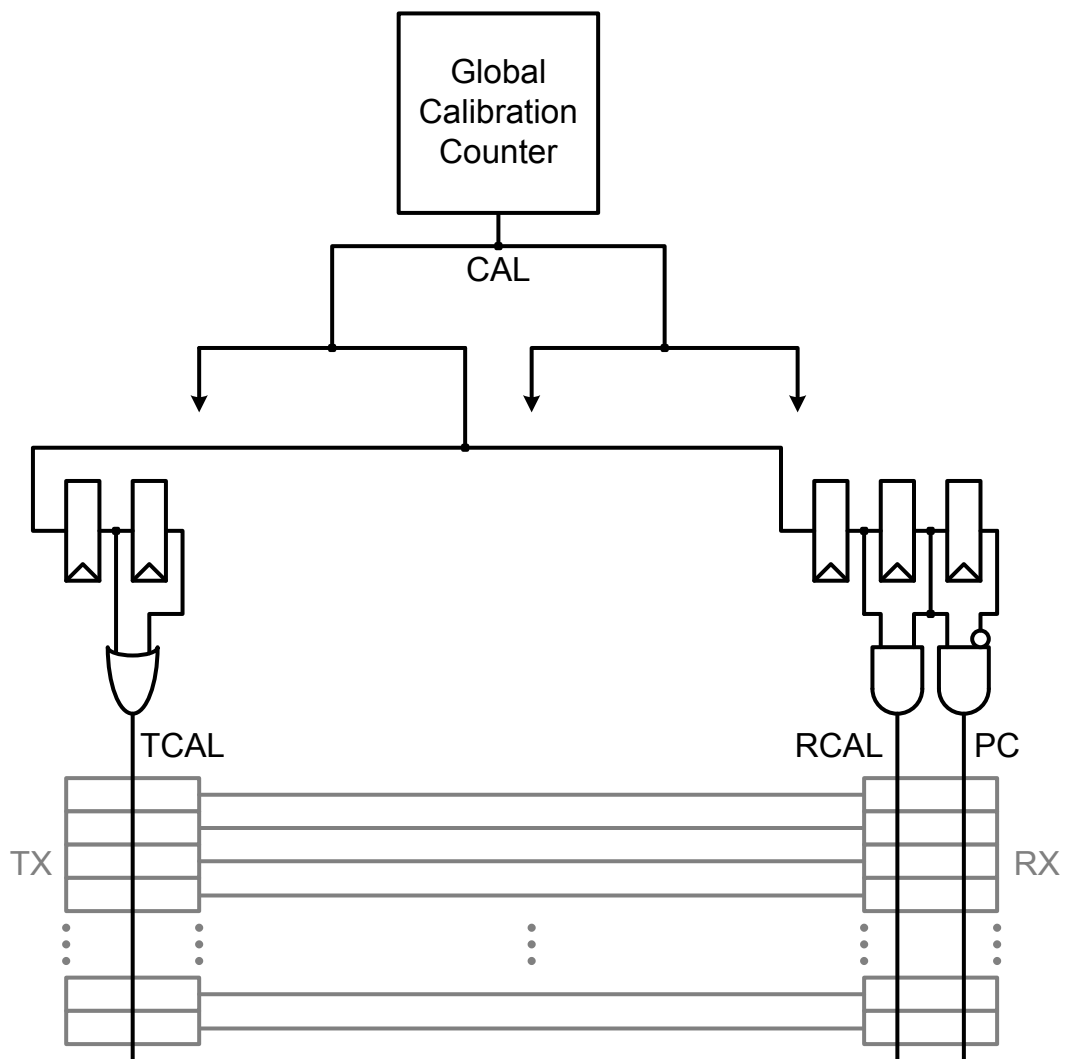


Figure 3.14: Generating control signals from one global calibration signal.

3.7 Operation

The complete operation of the SCI is shown in figure 3.15. In this example, the transmitter is sized to give a 20 mV signal swing, and device mismatch gives rise to approximately 70 mV of input offset at the receiver. A toggling data pattern that alternates between logical-0 and logical-1 every cycle is applied at the input (not shown). Data transmissions occur when TCAL is low, and calibrations are active when TCAL is high, during which the outputs are ignored. For clarity, the calibration interval is intentionally reduced from its typical value (thousands of cycles) to 16 so that the simulated waveforms can fit into the available space.

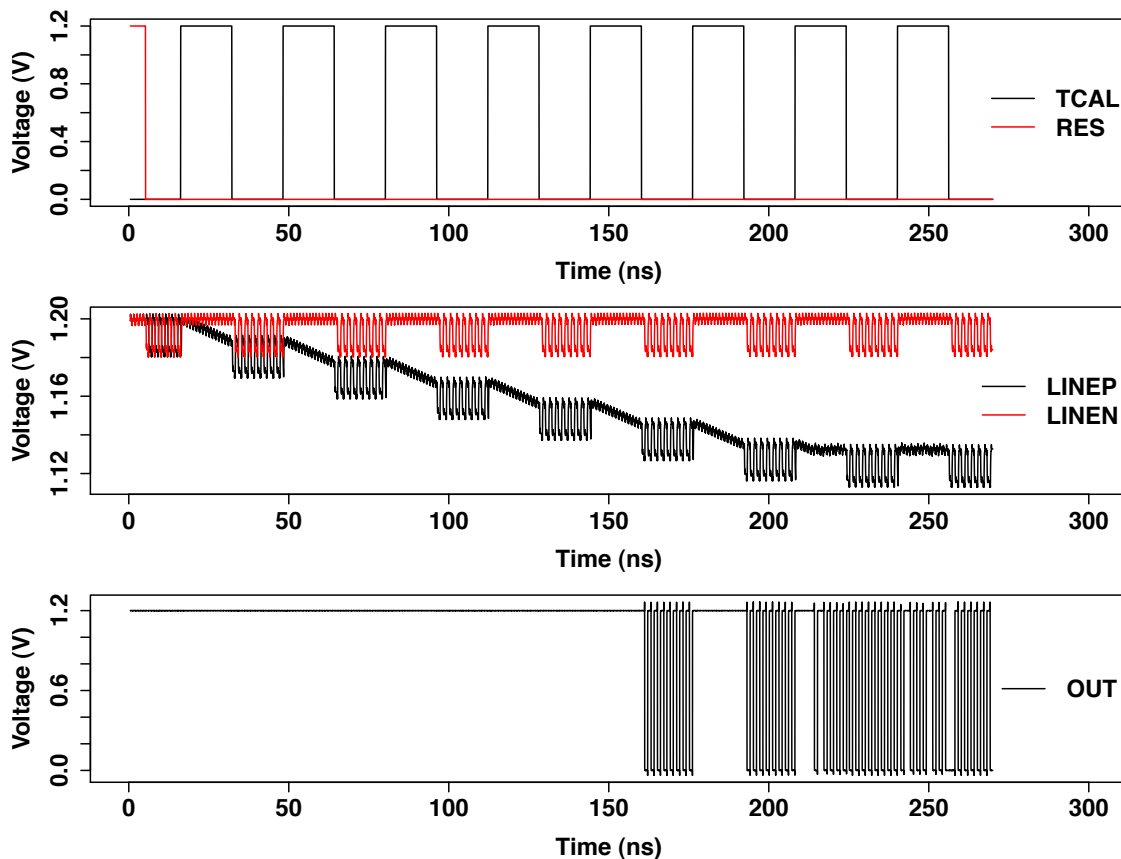


Figure 3.15: SCI operating under 70 mV of input offset with a 20 mV signal swing.

3.7.1 Reset Phase

In the beginning, the system is in reset (RES high) and the SCI undergoes bias detection as described in section 3.1. For our example, the receiver is one-biased, so LINEP is marked as the stronger line and LINEN as the weaker line.

3.7.2 Warm-Up Phase

Coming out of reset, the receiver offset initially overwhelms the signal. This causes the output to be pinned at logical-1 irrespective of the input. Over the next five calibration periods, the DC voltage on LINEP is gradually reduced. Each calibration starts at where its predecessor left off, in accordance with the concept of incremental calibration. By the end of the fifth calibration, the residual offset drops below the signal swing for the first time, and the output begins to track the input. Voltage adjustment continues through the sixth and seventh calibration, until finally during the seventh period the DC voltage on LINEP plateaus, indicating a full offset cancellation. When this happens the output also begins to toggle during the calibration period, reflecting oscillations in the line voltage around the true offset.

3.7.3 Maintenance Phase

After the seventh calibration, the SCI is fully warmed-up and ready for robust data transmission. From this point onwards, subsequent calibrations simply correct for any drifts caused by leakage as previously mentioned. In this example the drifts are not noticeable because the calibrations are so close together.

3.8 Performance

The ability to use a low signal swing in spite of large input offsets allows the SCI to realize significant energy savings. First, the reduced swing directly improves the wire energy. Moreover, with the use of calibration, the comparators no longer need to be aggressively up-sized to limit worst-case device mismatch. This is important because

comparators tend to have high activity factors and often dominate the overall energy in shorter interconnectsⁱⁱ. The SCI is also useful in timing-critical applications. For a given transmitter size, the SCI can trigger its receiver earlier than a design that does not have any offset cancellation, reducing communication latency.

Figure 3.16 compares the energy and frequency performance of the SCI to other interconnects discussed in chapter 2. All designs are in the IBM 90 nm LP CMOS process and 4 mm long. The SCI outperforms its predecessor, the CDI, in both energy and frequency. At moderate frequencies, it can achieve a 7-fold energy improvement over the FSI, which is comparable to what an ideal MSI can achieve at very low frequencies. At the other extreme, the SCI is the only design shown that can achieve a maximum frequency higher than the FSI.

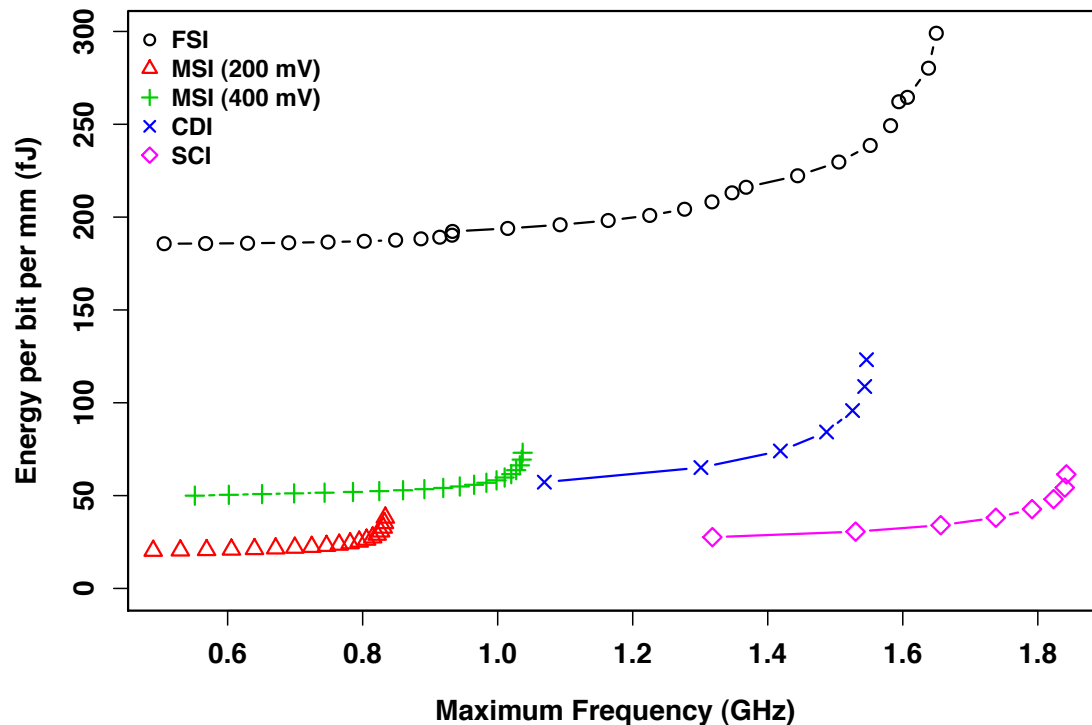


Figure 3.16: Comparison of energy and maximum operating frequency of the SCI against other interconnect architectures.

ⁱⁱThe activity factor of the comparator is the same as that of the clock net, which is 4-times the activity factor of the data lines under random traffic.

Chapter 4

Noise

The SCI is primarily designed to overcome the limitations imposed by device mismatch. In practice, noise also has a critical influence on the reliability of LSIs. Noise causes random fluctuations in device currents, creating decision errors when its magnitude exceeds that of the signal. However, unlike device mismatch where the value is deterministic, the magnitude and polarity of the noise currents vary randomly with time. This makes it virtually impossible to remove their effects through calibration. To improve reliability, we can either reduce the total amount of noise, increase the signal strength, or do both. Essentially, we want to maximize the signal-to-noise ratio (SNR), given by

$$SNR = \frac{P_{sig}}{P_n}, \quad (4.1)$$

where P_{sig} and P_n are the signal and noise powers respectively.

For an LSI, the most dominant source of noise is the thermal noise internal to the comparator. In this chapter we analyze comparator thermal noise in detail, and present techniques to minimize its impact. We focus on the StrongARM latch, the ubiquitous comparator used in LSIs (including the SCI) [24, 32], memory arrays [11], and analog-to-digital converters (ADCs) [46]. Nonetheless, much of our discussions will be equally applicable to other comparator architectures. For reference, the schematics of the StrongARM latch is shown again in figure 4.1.

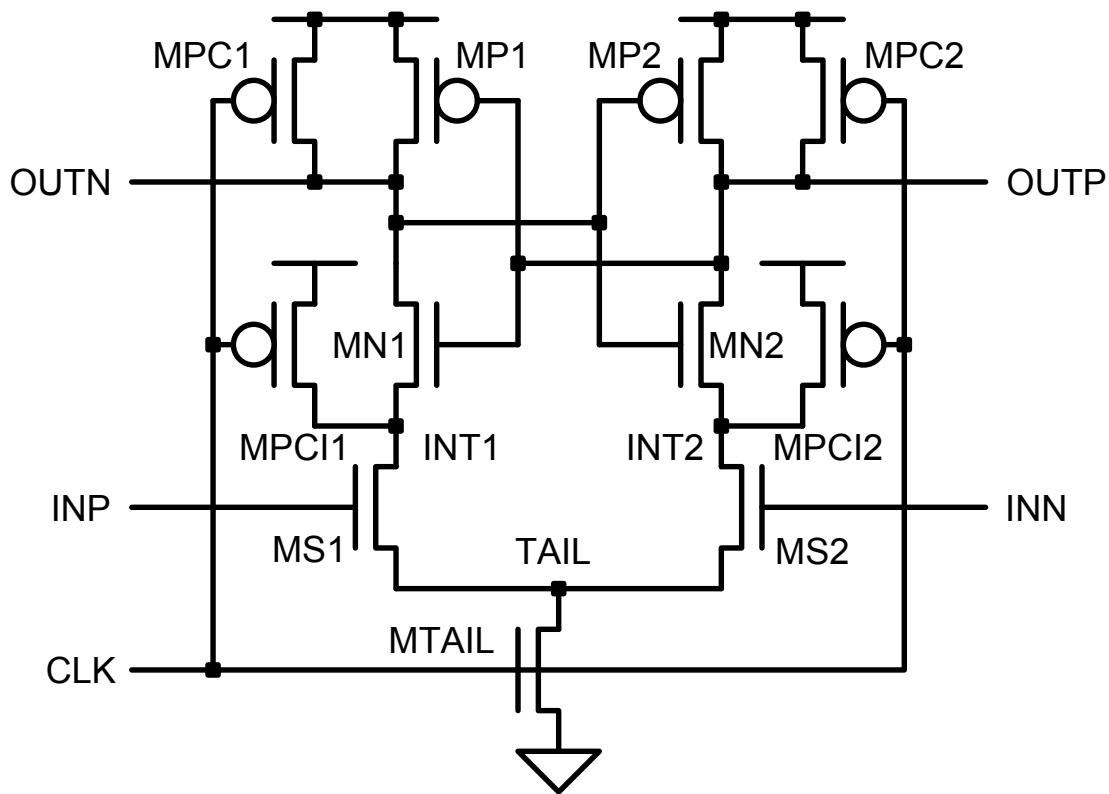


Figure 4.1: The StrongARM latch.

4.1 Theory

Figure 4.2 shows that the operation of the StrongARM latch can be conceptually divided into five distinct phases. During the reset phase, the outputs (OUTP, OUTN) and the internal nodes (INTP, INTN) are held at the supply by the precharge devices (MPC1-2, MPC11-2). At the rising clock edge, precharge turns off while the tail device (MTAIL) turns on, and the latch transitions into the sampling phase. Here the sensing devices (MS1-2) convert the differential input voltage into a differential current through the two branches, which in turn generates a differential voltage across the output nodes. Once the output voltages drop below a threshold from the supply, the latch enters the regeneration phase where the cross-coupled inverters (MN1-2, MP1-2) amplify the output voltage through positive feedback. Eventually, the latch

reaches the ready phase where the outputs stabilize at full-swing logical values. After the falling clock edge, the outputs and internal nodes are pulled up in the precharge phase and the cycle repeats.

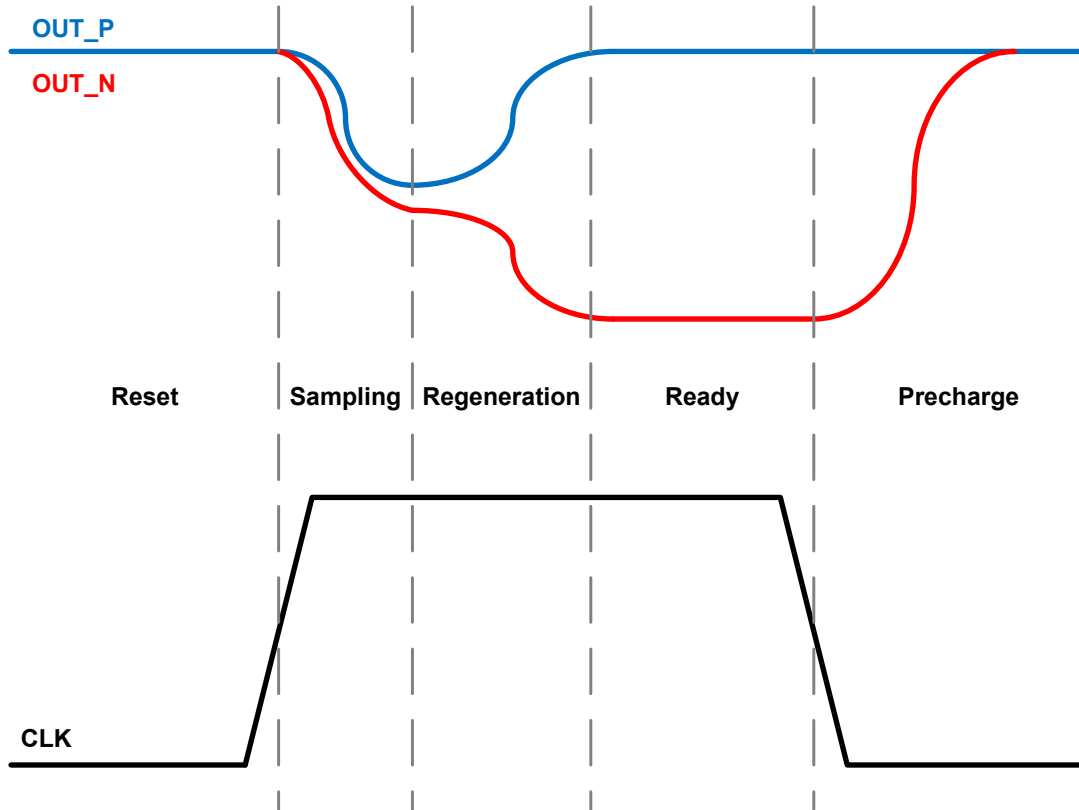


Figure 4.2: The five operating phases of the StrongARM latch.

From a noise perspective, the StrongARM latch is a time-varying system. Specifically, noise injections in the proximity of the sampling phase are more likely to affect the decision outcome than injections during other times. Prior to sampling, disturbances are largely restored by the precharge devices. Once significant regeneration has taken place, the latch is again immune to further noise injections because the differential output has already reached a critical value to sustain the positive feedback. Because of this time-varying characteristic, the well developed linear time-invariant (LTI) noise theory for analog circuits cannot be readily applied to analyze the noise behavior of the StrongARM latch.

Strictly speaking, the StrongARM latch is also nonlinear due to the large-signal compression that occurs during regeneration. Fortunately, studies have shown that the latch is relatively insensitive to noise during its non-linear phases, so we can model it as a linear-time-varying (LTV) system when we are only interested in its noise characteristics [28]. The use of an LTV model greatly simplifies the noise analysis of the StrongARM latch, which previously required solving complex stochastic differential equations (SDEs) [39]. We now derive the key equations that governs the thermal noise behavior in an LTV system. More background details on LTV systems can be found in [49].

4.1.1 Noise Analysis in LTV Systems

An LTV system is one where the principle of superposition holds but where time-invariance fails. Mathematically, this means that if the inputs $x_1(t)$ and $x_2(t)$ produce the outputs $y_1(t)$ and $y_2(t)$ independently, then the input $c_1x_1(t) + c_2x_2(t)$ always produces the output $c_1y_1(t) + c_2y_2(t)$, where c_1 and c_2 are arbitrary constants. However, for an arbitrary delay t_0 , the input $x_1(t - t_0)$ does not in general produce the output $y_1(t - t_0)$. An LTV system is fully characterized by a time-varying impulse response, $h(t, \tau)$, which is the response of the system across time t to a unit impulse arriving at time τ . For an arbitrary input $x(t)$ the output $y(t)$ is given by the superposition integral

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t, \tau) d\tau. \quad (4.2)$$

For an LTI system $h(t, \tau)$ reduces to $h(t - \tau)$, and equation 4.2 becomes the well-known convolution integral.

In the context of noise for comparators, we are usually not interested in the output response over all time. Instead, we want to know the total effective noise at some observation instant, t_{obs} , such as the rising edge of a clock. Under these circumstances,

we can work with a simpler version of the time-varying impulse response

$$\Gamma(\tau) \triangleq h(t_{obs}, \tau), \quad (4.3)$$

which is often referred to as the impulse sensitivity function (ISF) of the system at t_{obs} ⁱ. Figure 4.3 shows that the ISF can be interpreted as the dual of the impulse response for an LTI system. The impulse response (top) characterizes the response of a system over different times to an impulse arriving at a fixed time. In contrast, the ISF (bottom) characterizes the response of a system at a fixed time to impulses arriving at different times.

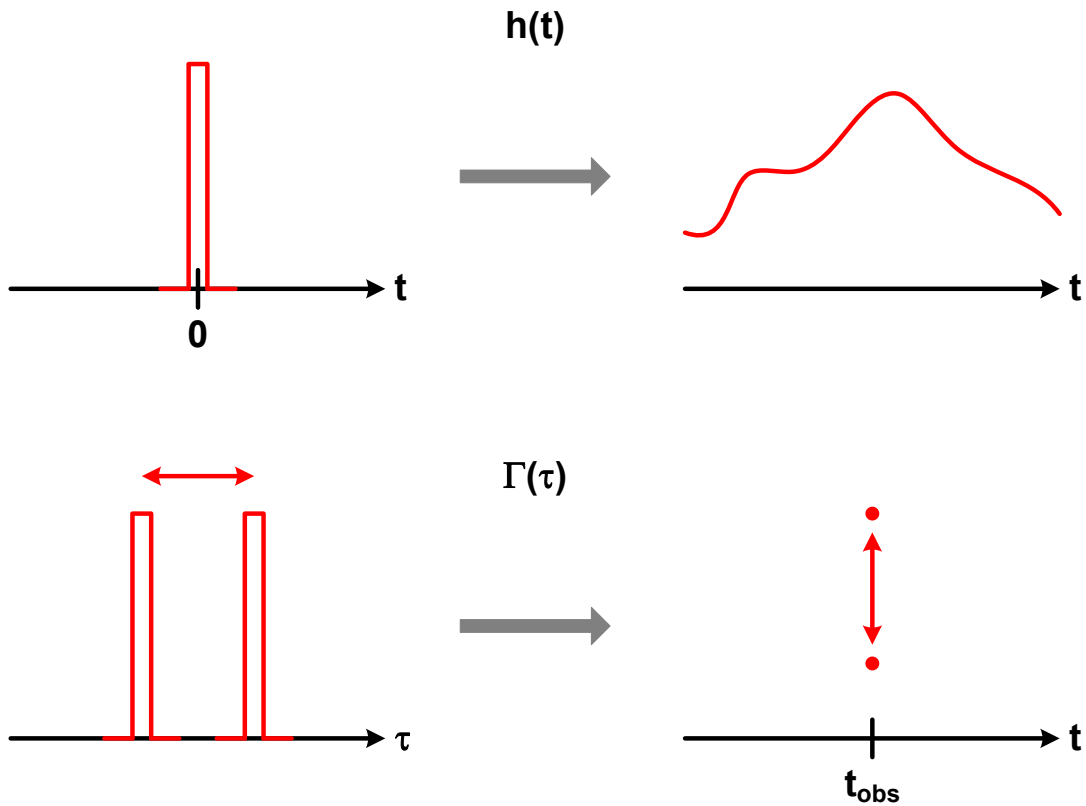


Figure 4.3: Duality between the impulse response and the ISF.

ⁱThe ISF was first used by Hajimiri et al. to characterize the response of oscillators to noise injections [20].

Using the ISF, the output noise variance at the observation instant, $\sigma_y^2(t_{obs})$, in response to an input noise process, $x(t)$ can be derived as

$$\sigma_y^2(t_{obs}) = E [y^2(t_{obs})] \quad (4.4)$$

$$= E \left[\int_{-\infty}^{\infty} x(\tau)\Gamma(\tau) d\tau \int_{-\infty}^{\infty} x(\lambda)\Gamma(\lambda) d\lambda \right] \quad (4.5)$$

$$= \int_{-\infty}^{\infty} \Gamma(\tau) \int_{-\infty}^{\infty} \Gamma(\lambda) E [x(\tau)x(\lambda)] d\lambda d\tau \quad (4.6)$$

$$= \int_{-\infty}^{\infty} \Gamma(\tau) \int_{-\infty}^{\infty} \Gamma(\lambda) R_{xx}(\tau, \lambda) d\lambda d\tau \quad (4.7)$$

where $R_{xx}(\tau, \lambda)$ is the autocorrelation of the input noise process. If $x(t)$ is a white noise process with variance σ_x^2 , as is the case for thermal noise, then

$$R_{xx}(\tau, \lambda) = R_{xx}(\lambda - \tau) = \sigma_x^2 \delta(\lambda - \tau), \quad (4.8)$$

and equation 4.7 simplifies to

$$\sigma_y^2(t_{obs}) = \int_{-\infty}^{\infty} \Gamma(\tau) \int_{-\infty}^{\infty} \Gamma(\lambda) \sigma_x^2 \delta(\lambda - \tau) d\lambda d\tau \quad (4.9)$$

$$= \sigma_x^2 \int_{-\infty}^{\infty} \Gamma^2(\tau) d\tau. \quad (4.10)$$

In other words, the output noise variance is fully determined by the variance of the input noise process and its associated ISF.

In practice, the input noise variance is not constant but changes as the node voltages in the comparator varies during its operation. To model this time dependence we update equation 4.10 to

$$\sigma_y^2(t_{obs}) = \int_{-\infty}^{\infty} \sigma_x^2(\tau) \Gamma^2(\tau) d\tau, \quad (4.11)$$

where $\sigma_x^2(\tau)$ is the time-varying input noise variance whose value at time τ is taken to be what the input noise variance of the same source would be if all the node voltages are biased at their respective transient values at the same instant.

Finally, when multiple noise sources are present, the total output noise variance can be found as

$$\sigma_y^2(t_{obs}) = \sum_{k=1}^N \int_{-\infty}^{\infty} \sigma_{xk}^2(\tau) \Gamma_k^2(\tau) d\tau, \quad (4.12)$$

where we simply sum the individual noise contributions from each source according to the principle of superposition.

4.1.2 Input-Referred Noise

Equation 4.12 is a general expression that can be applied to any output in an LTV system, provided that a valid ISF can be defined with respect to each noise source. In the design of a comparator, we want to find the minimum differential input voltage required to reliably signal both logical valuesⁱⁱ. It is therefore useful to look at the total input-referred noise voltage, V_n , as shown in figure 4.4. Here, the LTV system "output" is taken to be the signal input, and we use equation 4.12 to find the variance of V_n , σ_n^2 , that has the same effect as the all the individual noise sources within the comparator combinedⁱⁱⁱ.

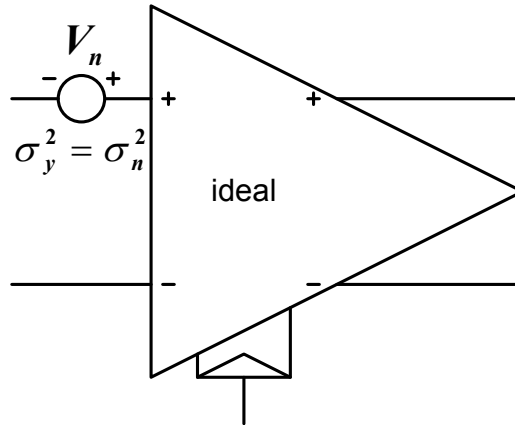


Figure 4.4: Modeling the effect of noise with input-referred noise voltage.

ⁱⁱIn this context, we are referring to the input to the comparator, not the noise sources.

ⁱⁱⁱNote the similarity to modeling device mismatch with the input offset voltage.

4.1.3 Bit Error Rate

Theoretically, the magnitude of thermal noise is unbounded. No matter what the signal swing is, there is always a finite probability that the noise will overpower the signal and lead to decision errors. Assuming the input-referred noise is normally distributed with zero mean with standard deviation σ_n , the bit error rate (BER) can be expressed as

$$\boxed{BER = P(V_n > V_{sig}) = Q(VSNR)}, \quad (4.13)$$

where $VSNR = \frac{V_{sig}}{\sigma_n} = \sqrt{SNR}$ is the voltage signal-to-noise ratio, and $Q(x)$ is the tail probability of the standard normal distribution given by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{x^2}{2}\right). \quad (4.14)$$

Figure 4.5 shows how the BER changes as VSNR is increased.

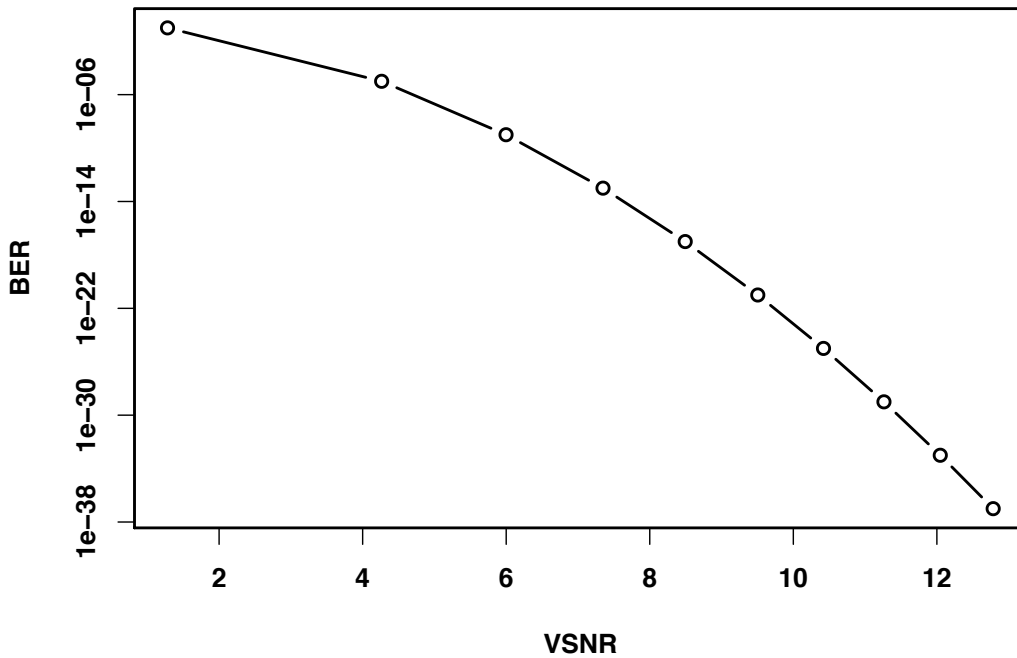


Figure 4.5: Effect of increasing VSNR on BER.

The maximum acceptable BER, or equivalently the minimum acceptable VSNR, varies with the application. To get a handle on the requirement for LSIs consider the 64-node network-on-chip shown in figure 4.6. The total number of interconnects on this chip, N , can be calculated as

$$N = [(6 \times 6 \times 4) + (6 \times 4 \times 3) + (4 \times 2)] \times 32 = 7,168. \quad (4.15)$$

If each interconnect has a BER of r , the probability that there are no on-chip errors in a given cycle is given by

$$P(\text{No errors}) = (1 - r)^N \quad (4.16)$$

$$\approx 1 - Nr, \quad (4.17)$$

where we assume that that $r \ll 1$, and that equation 4.16 can be well approximated by keeping only the linear term in its binomial expansion. Let T denote the lifetime of the chip, and f denote its operating frequency. The probability that there will be at least one error over the lifetime can be expressed as

$$P(\text{At least one error over lifetime}) = 1 - P(\text{No errors over lifetime}) \quad (4.18)$$

$$\approx 1 - (1 - Nr)^{fT} \quad (4.19)$$

$$\approx NfTr, \quad (4.20)$$

where we again use the linear approximation to the binomial series in the final step. To meet a given failure rate requirement, R , we need

$$NfTr < R, \quad (4.21)$$

or more conveniently,

$$\boxed{r < \frac{R}{NfT}}. \quad (4.22)$$

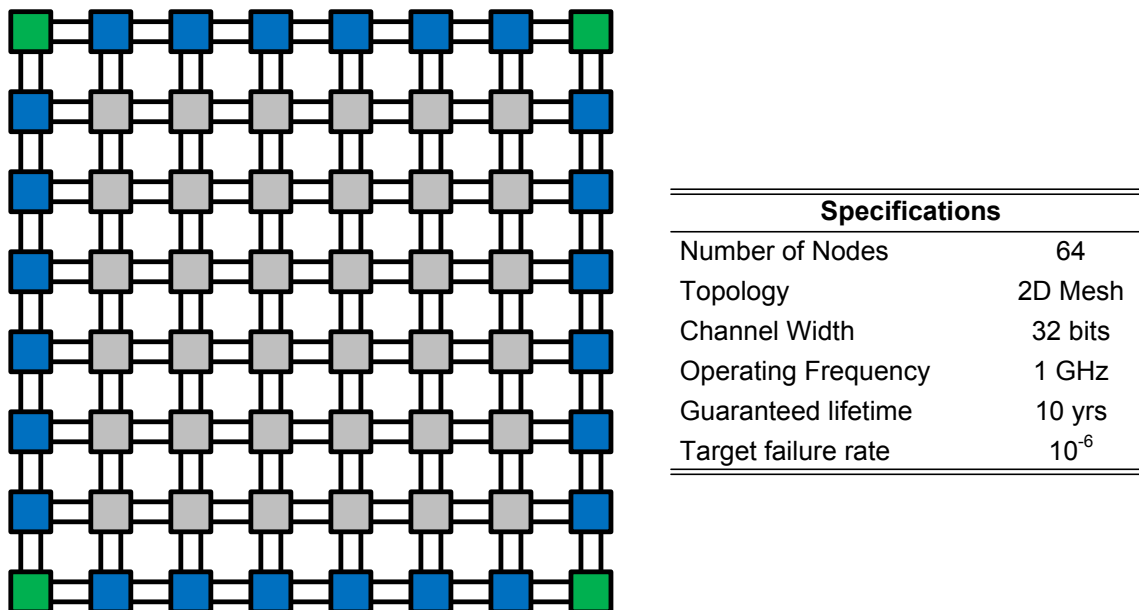


Figure 4.6: Representative 64-node network-on-chip.

Substituting the numerical values from our example into equation 4.22, the maximum acceptable BER comes to 4.4×10^{-28} . From figure 4.5, this requires a VSNR of just under 11. In other words, the signal swing needs to be about 11 times larger than the standard deviation of the total input-referred noise voltage^{iv}. In our SCI prototype, we design for a target BER of 1×10^{-30} .

4.2 Simulation

The SNR of a comparator can be analyzed with modern RF circuit simulators such as SpectreRF and HSPICE-RF [39, 28]. To do this, the operating point at an observation time is first determined through a periodic steady state (PSS) simulation. The output noise power density can then be obtained by running a periodic noise (PNoise) simulation. Finally, the SNR is calculated by integrating the noise spectral density across frequency and comparing it to the output signal power.

^{iv}If the system can tolerate additional latency, error-correction codes can be implemented to relax the BER requirement significantly.

Instead of relying on RF simulators, we find it more informative to simulate the input-referred thermal noise for the StrongARM latch through a direct implementation of equation 4.12. This approach is illustrated in figure 4.7.

For every device noise current source, i_{nk} , we define an associated ISF, $\Gamma_k(\tau)$, as the minimum DC differential input voltage that must be present to neutralize the effect of a unit charge injection from the source τ seconds from the observation time, t_{obs} . For convenience, we arbitrarily take t_{obs} to be the instant when the rising clock reaches 50% of its final value. Defining the ISF this way implicitly takes into account any signal gain achieved during sampling, making the calculation of input-referred noise more straightforward^v. To simulate $\Gamma_k(\tau)$, a narrow 1 ps wide current pulse^{vi} is injected at various delays while the differential input voltage is swept until metastability is observed at the output. For a given delay, $\Gamma_k(\tau)$ is taken to be the value of the input voltage at metastability normalized by the charge injected by the pulse.

In addition to finding the ISF, a transient simulation is performed to capture how voltages at the terminals of a device vary with time. For each delay, these transient node voltages are used as the bias point in an AC simulation to extract the noise current variance for that device. When running the transient simulation different differential inputs produce slightly different results. However, this has little effect on the final result because node voltages only diverge noticeably after significant regeneration has taken place, at which point the system is no longer sensitive to noise anyway ($\Gamma_k(\tau) \approx 0$). For symmetry, we arbitrarily set the differential input voltage to zero. Once the ISF and transient noise variance are determined for each device, the total input-referred noise is calculated according to equation 4.12.

Figure 4.8 shows a representative noise simulation of a StrongARM latch in a 90 nm process. For every device we plot its transient noise current variance, the associated ISF, and the total noise integrand against time. We also plot the input-referred

^vNote that the variance of the noise current source $\overline{i_{nk}^2}$ is being used as the input noise variance σ_{xk}^2 in equation 4.12.

^{vi}The exact width of the pulse is not important as long as it is at least 1-2 orders of magnitude smaller than the decision latency.

noise contributions from each device and the total input-referred noise. The tail device (MTAIL) is excluded because its noise injection is common to both branches and thus has negligible effect on the decision outcome.

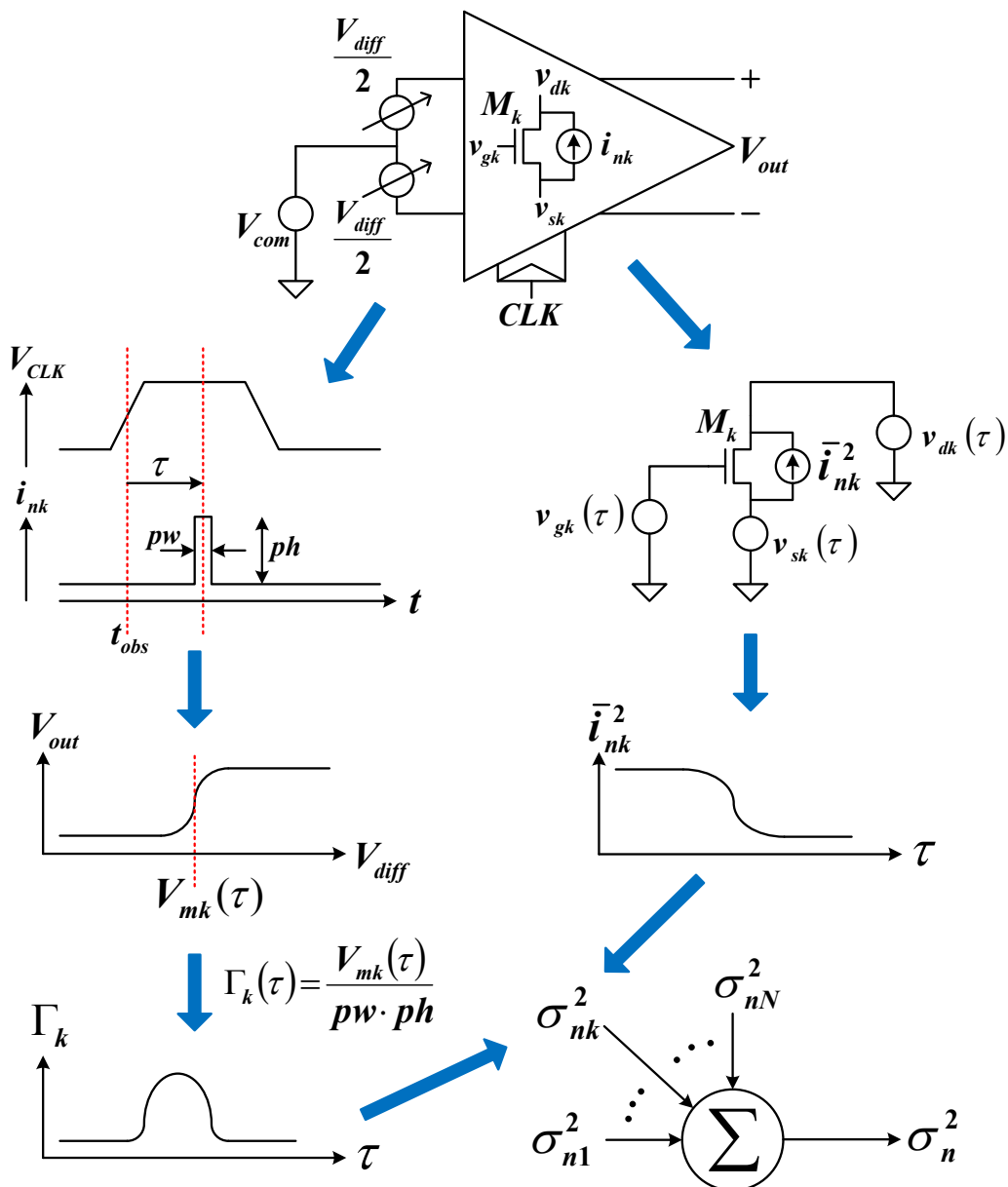


Figure 4.7: Noise simulation setup based on LTV circuits theory.

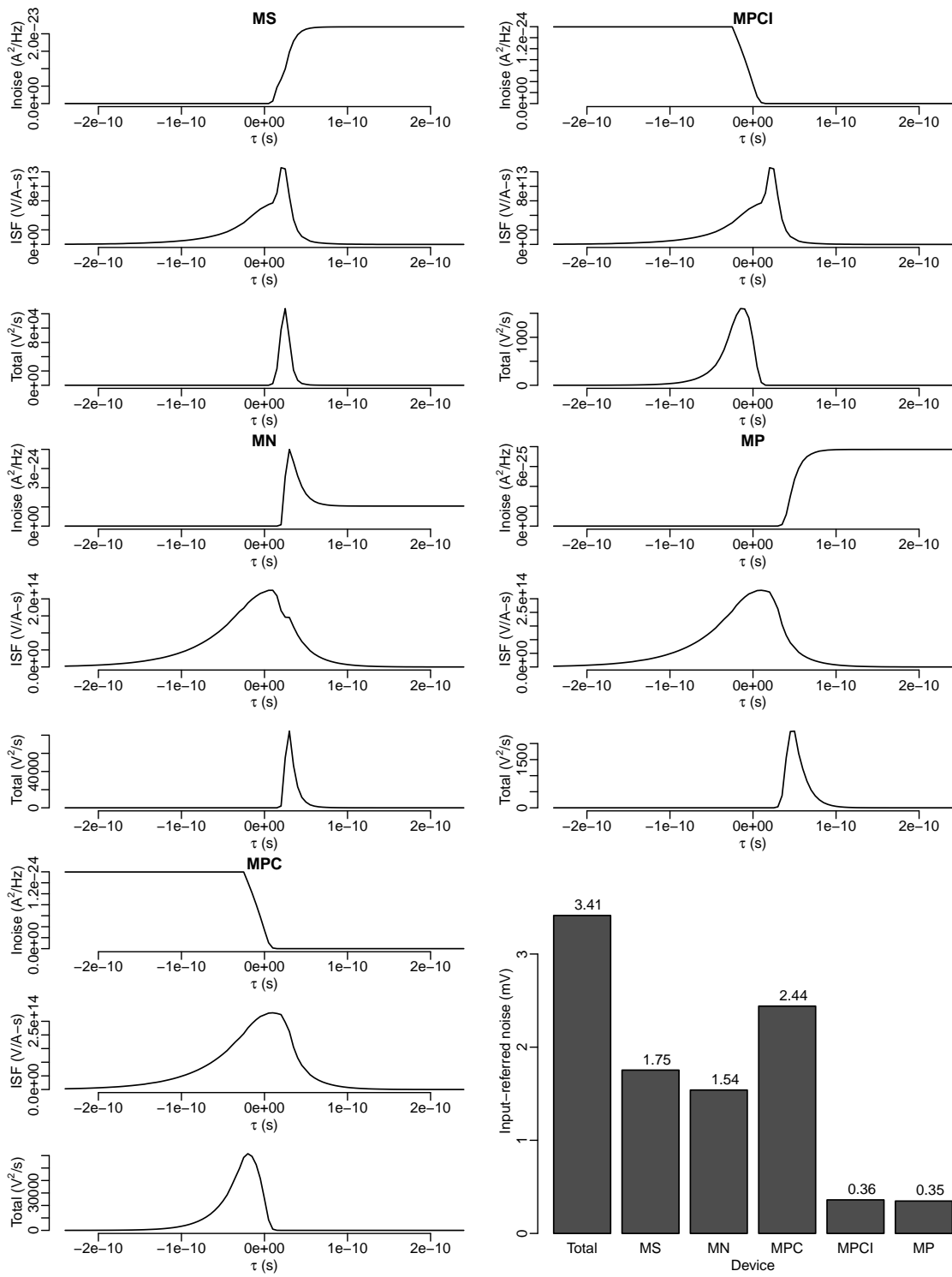


Figure 4.8: Representative outputs from a noise simulation run.

The simulations reveal insights into why certain devices contribute more noise than others. For example, since the output precharge devices (MPC) are normally much smaller than the sensing devices (MS), it can be tempting to conclude that their noise contribution will also be small. In reality, as shown in figure 4.8, the output precharge devices can actually surpass the sensing devices as the single largest contributor to total noise.

The apparent contradiction can be explained by comparing the ISFs. For the output precharge devices, noise is directly injected into the sensitive output nodes, where it gets amplified through the sampling and regeneration phases. In contrast, noise injected by the sensing devices is first attenuated by a current divider formed between the sensing devices themselves and the N-regeneration devices (MN), so only a fraction of the noise reach the output nodes. The result is that the ISF for the output precharge devices is significantly higher than that of the sensing devices, which often more than makes up for the lower noise current variance.

For the same reason, the internal precharge devices (MPCI) contribute significantly lower noise than the output precharge devices even though they are comparable in size. The P-regeneration devices (MP) contribute little noise even though their ISFs are high. One reason for this is that the drain-to-source voltage for these devices never gets high enough to generate large amount of noise current. Moreover, by the time they emerge from cutoff the system has already begun regeneration and thus no longer very sensitive to noise. This can be seen from the lack of overlap between the noise variance and the ISF for these devices.

4.3 Noise Reduction

In this section we summarize the common techniques used to improve the SNR of comparators.

4.3.1 Upsize

To first order, the average thermal noise power per unit bandwidth for a device, $\overline{i_n^2}$, can be modeled as

$$\overline{i_n^2} = 4kT\gamma g_m, \quad (4.23)$$

where $k = 1.38 \times 10^{-23} J/K$ is the Boltzmann constant, T is the absolute temperature in Kelvins, g_m is the small signal transconductance, and γ is an adjustment factor called the excess noise factor [21, 16]. The signal current, i_{sig} for a given gate-to-source bias, v_{gs} , is given by

$$i_{sig} = g_m v_{gs}. \quad (4.24)$$

At the device level, the SNR for the drain current can be derived as

$$SNR = \frac{P_{sig}}{P_n} = \frac{i_{sig}^2}{\overline{i_n^2}} = \frac{v_{gs}^2 g_m}{4kT\gamma} \quad (4.25)$$

which is proportional to g_m . Since g_m itself is directly proportional to the device width, making all devices wider will lead to a linear increase in SNR (a square-root increase in VSNR). As devices are made larger, the energy consumed per decision also increases, so there is a tradeoff between noise and energy.

4.3.2 Increase Aperture

Another way to improve SNR is to increase the aperture of the comparator, which can be interpreted as the period of time during which the comparator is sensitive to variations in its input voltage. More precisely, we can define the aperture, t_s , as the interval where the ISF of the comparator with respect to its input voltage is within 50% of its peak value. Figure 4.9 illustrates the ISF for two comparators, one with a higher sensitivity but smaller aperture, while the second has a larger aperture but lower sensitivity. There is an inverse relationship between sensitivity and aperture because more sensitive comparators make their decisions quicker, so

they do not linger in their sensitive phases as long. Increasing the aperture improves SNR because the high-frequency noise components have more time to average out, reducing their effective variances.

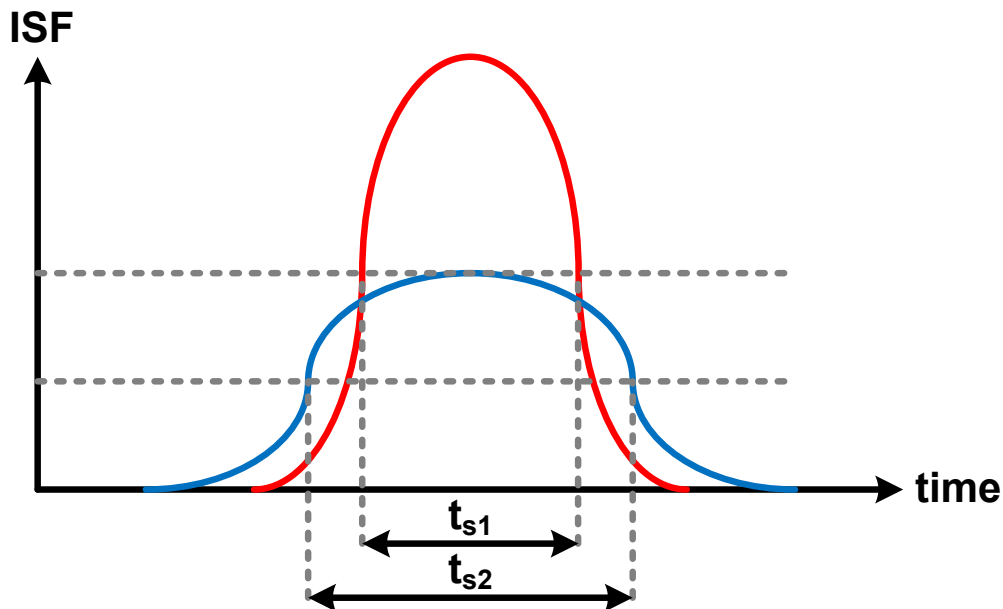


Figure 4.9: Apertures of comparators.

One way to increase the aperture of the StrongARM latch is to reduce the width of MTAIL, which reduces the currents in both branches and slows down the critical sampling phase of the operation. Unfortunately, reducing the width of MTAIL also slows down the regeneration phase, which is strictly not necessary because by then the latch is insensitive to noise anyway. This tends to create excessive delay penalty for a given improvement in SNR.

A slightly better approach increases the aperture by reducing the slew rate of the clock signal [43]. Initially, the lower gate voltage on MTAIL reduces the current during the sampling phase. However, once the clock reaches its maximum value the latch can use a larger current during its regeneration phase, reducing the penalty in decision latency.

4.3.3 Preamplifier

In some applications, SNR can be improved by using another amplifier stage, known as the preamplifier, before the inputs of the comparator. Figure 4.10 illustrates how the preamplifier improves SNR. The raw signals, v_{sig} , is first amplified by the preamplifier with a gain of A to generate the output, $v_o = Av_{sig}$, which is then fed into the comparator. In the process, the preamplifier adds its own sources of internal noise with variance σ_{np}^2 to the signal. The new SNR at the input of the comparator, including its input-referred noise, is given by

$$SNR_p = \frac{A^2 v_{sig}^2}{\sigma_n^2 + \sigma_{np}^2}, \quad (4.26)$$

assuming that the noise contributed by the preamplifier and the comparator itself are independent. Rearranging terms, this can be expressed as

$$SNR_p = \frac{A^2}{1 + \eta} SNR_0, \quad (4.27)$$

where $\eta = \frac{\sigma_{np}^2}{\sigma_n^2}$, and SNR_0 is the SNR without the preamplifier. Equation 4.27 suggests that as long as $A^2 > 1 + \eta$, the preamplifier improves the overall SNR. Intuitively, this just means that a good preamplifier has a high gain and low internal noise.

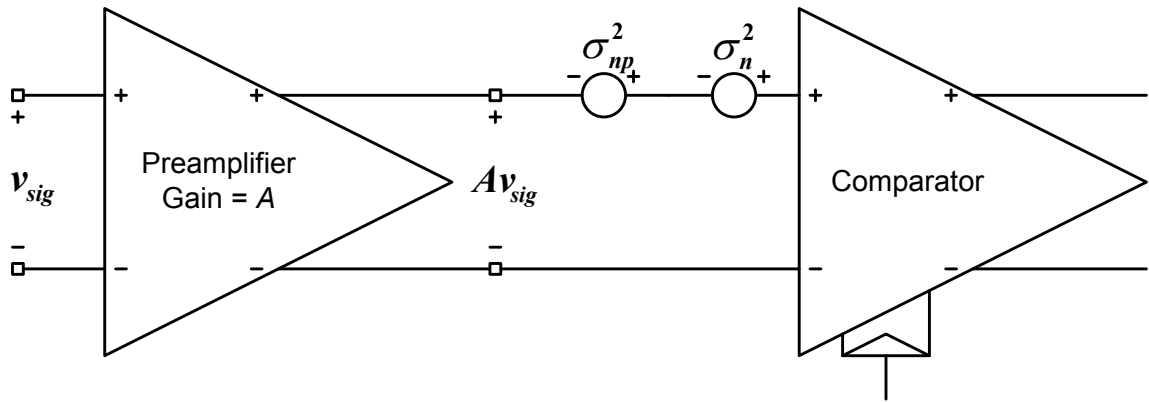


Figure 4.10: Improving comparator SNR with a preamplifier.

Preamplifiers increase the design complexity of LSI receivers. They also require DC bias currents which is undesirable for energy-efficiency. For these reasons, the use of preamplifiers only makes sense in very long interconnects where the energy savings from the swing reduction is large enough to justify the extra overhead.

4.3.4 Integrating Comparator

Figure 4.11 shows how an integration stage can be used to improve the SNR of comparators [15]. The differential input voltage is first converted into a differential current and then integrated on the input capacitance of the comparator. The integral, which in this case equals the differential input voltage to the comparator, grows with time. In contrast, since noise currents have zero mean, integration produce a random noise voltage whose variance decrease with time due to the effect of averaging. Provided that we integrate for a long enough period, the boost to the signal will more than compensate for any additional noise introduced, reducing the SNR at the input of the comparator.

Unlike preamplifiers the integrating stage does not dissipate static power. However, it does contribute additional dynamic energy and adds delay to the critical path of the interconnect.

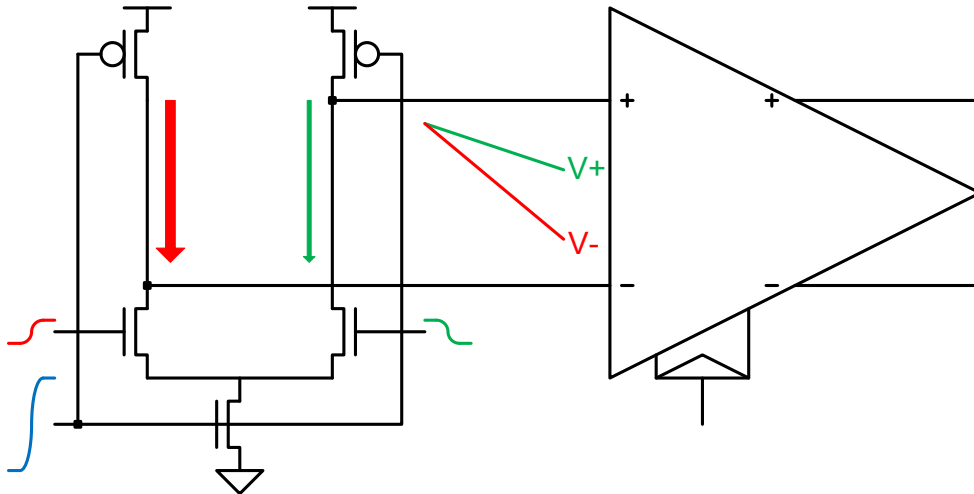


Figure 4.11: Improving comparator SNR with integrating stage.

4.4 Impact of Precharge Device Sizing

Up to now, the circuits community has paid little attention to the effect of the precharge devices on the noise characteristics of the StrongARM latch. Traditionally, the precharge devices were only used to restore the output and internal node voltages to supply during the precharge and reset phases. To simplify analyses, it is often assumed that these devices turn off instantaneously at the rising clock edge. Under such a paradigm, it is natural to assume that the precharge devices have little impact on the critical sampling and regeneration phases of the operation. This encourages the use of the smallest precharge devices that can still meet timing in order to reduce clock load and to minimize the noise current that they inject into the system.

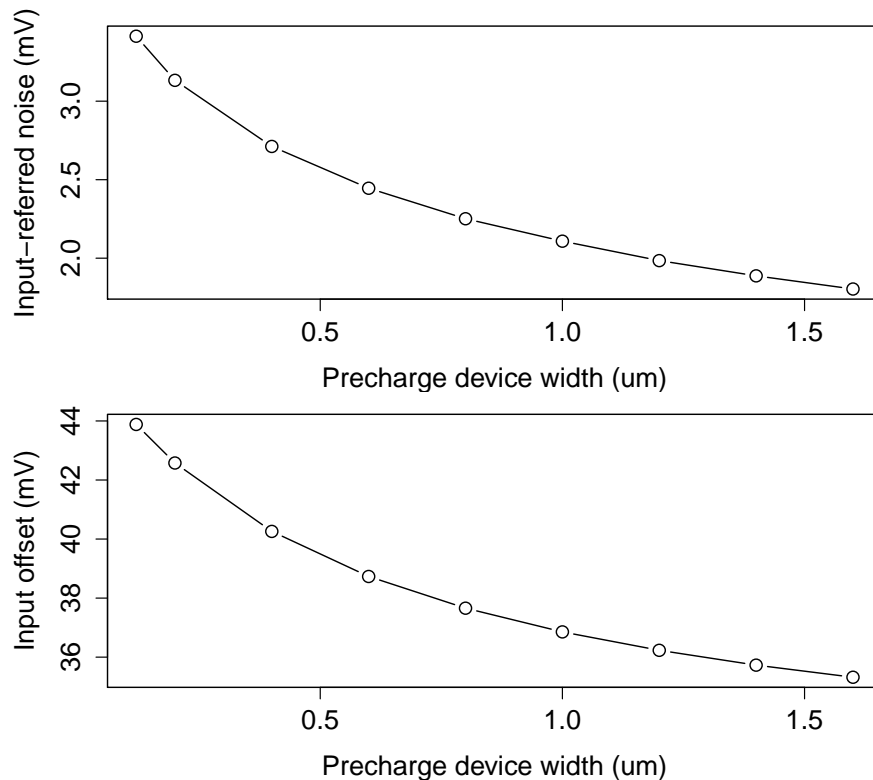


Figure 4.12: Variation of input-referred noise voltage and input offset with precharge device size.

In our studies, we found a strong correlation between the size of the precharge devices and the input-referred noise of the StrongARM latch. Figure 4.12 shows that, contrary to intuition, the input-referred noise actually decreases as the width of the precharge devices are increased. Not only so, the input offset voltage of the StrongARM latch also improves as the precharge devices are upsized.

To better understand this phenomenon figure 4.13 shows how the simulated noise contributions change for two different precharge device sizes. As expected, the noise current variances for the precharge devices increase significantly as their widths are increased from 0.12 μm to 1.00 μm ^{vii}. What is surprising, however, is that there is a system-wide reduction in both the width and height of the ISFs for all the devices. The reductions in the ISFs more than compensate for the increase in the noise current injections, lowering the overall input-referred noise. There is also a small shift in the peaks of the ISFs in the positive direction, which causes a proportionally larger decrease in the noise contribution of the output precharge devices compared to the sensing devices.

The ISF-shaping effect from upsizing the precharge devices can be traced to its effects on transient voltages and the small-signal parameters. Figure 4.14 shows how the drain-to-source voltage (V_{ds}), transconductance (g_m), and output conductance (g_{ds}) of the sensing devices vary with time for different sizing combinations. In general, a higher g_m is better since it generates more signal current for a given input voltage. Similarly, a higher output impedance (lower g_{ds}) is better because it allows more signal current to pass through to the output nodes. As the precharge devices get wider, they hold the node voltages longer during the clock transition. This allows the sampling phase to begin at a higher V_{ds} . Since both g_m and output impedance increase with V_{ds} , signal gain at the outputs improve and StrongARM latch becomes more tolerant to noise injections. The higher signal gain also explains the reduction in the input offset voltage.

^{vii}In all our experiments the size of the internal precharge devices (MPCI) are the same as the output precharge devices (MPC).

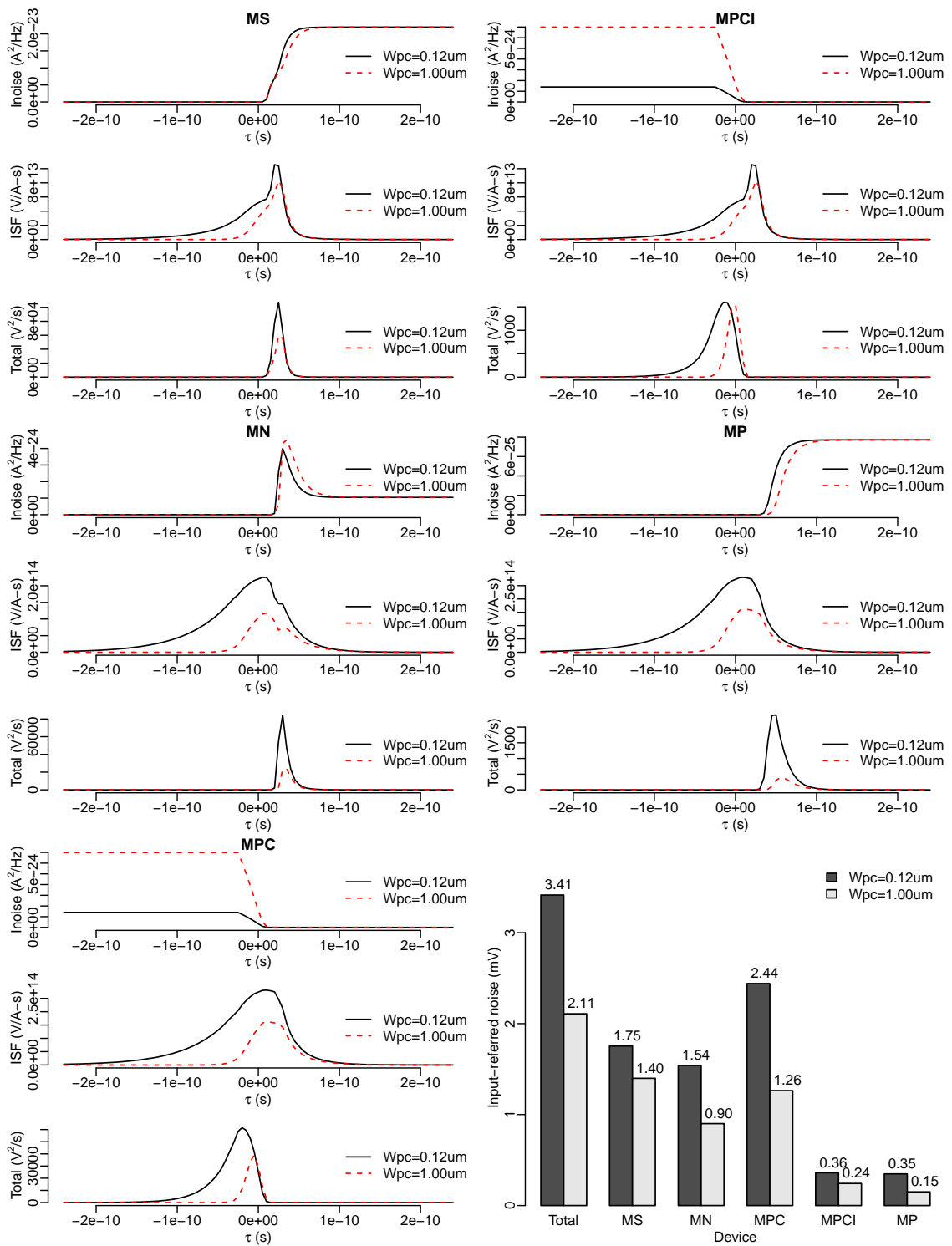


Figure 4.13: Impact of increasing precharge device width on noise contributions.

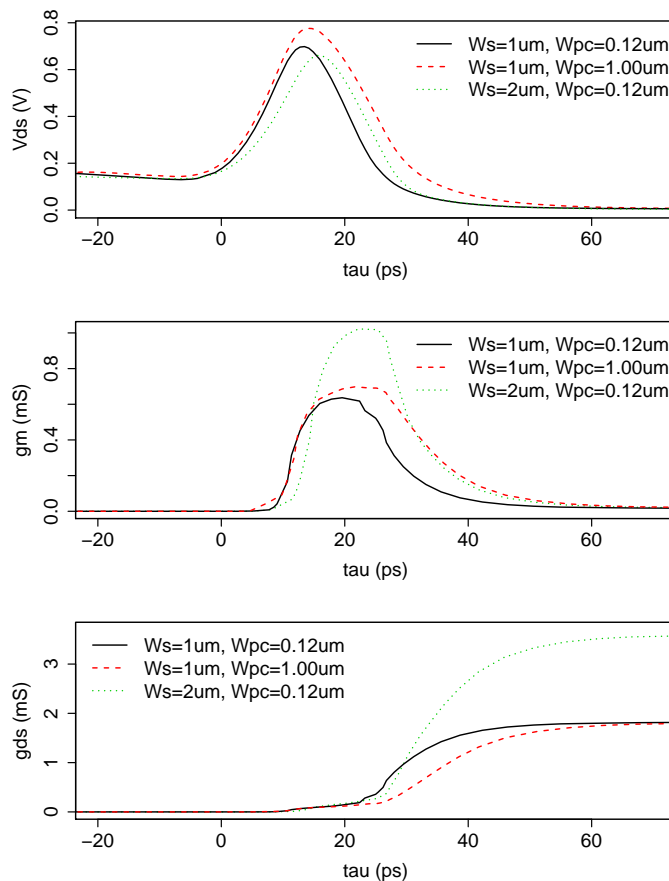


Figure 4.14: Effect of sizing on small signal parameters of the sensing device.

It should be emphasized that the benefit of using larger precharge stems primarily from their higher currents during clock transition, not the extra capacitance they add to the nodes. Adding output capacitance is less effective because disproportionate more capacitance is needed for the same noise reduction. This slows down all phases of operation, increasing decision latency. In contrast, the transient currents from the precharge devices disappear after the noise-critical sampling phase, allowing for faster regeneration.

Changes in small signal parameters also explains why upsizing some devices can produce less than expected noise improvements. Figure 4.14 shows that when only the width of the sensing devices are increased g_m improves but its transient V_{ds}

decreases to equalize the current through all the pull-down devices. This exacerbates the reduction in output impedance that already comes with the larger width, limiting the improvement in signal gain at the output nodes. Moving forward, as the voltage headroom continues to scale down and the pull-down devices are pushed further away from saturation, the dependence of g_m and g_{ds} on V_{ds} will become more pronounced. This makes the transient effects of the precharge devices even more critical.

The effect of precharge device sizing has important implications in the optimization of the StrongARM latch in energy-critical applications. Figure 4.15 plots the energy per decision against the achievable input-referred noise for the StrongARM latch under different upsizing strategies. The first scheme, also called the proportional scheme, increase all device widths by the same percentage at each step. The second scheme is a variation of the first where the precharge devices are fixed at their minimum sizes. In the third and fourth scheme, only the sensing devices and the precharge devices are upsized respectively. Comparing the two proportional schemes, we see that the scheme that allows precharge devices to be upsized performs significantly better. Moreover, for a given energy budget, upsizing the precharge devices alone can attain a noise level over 30% lower than the next best scheme.

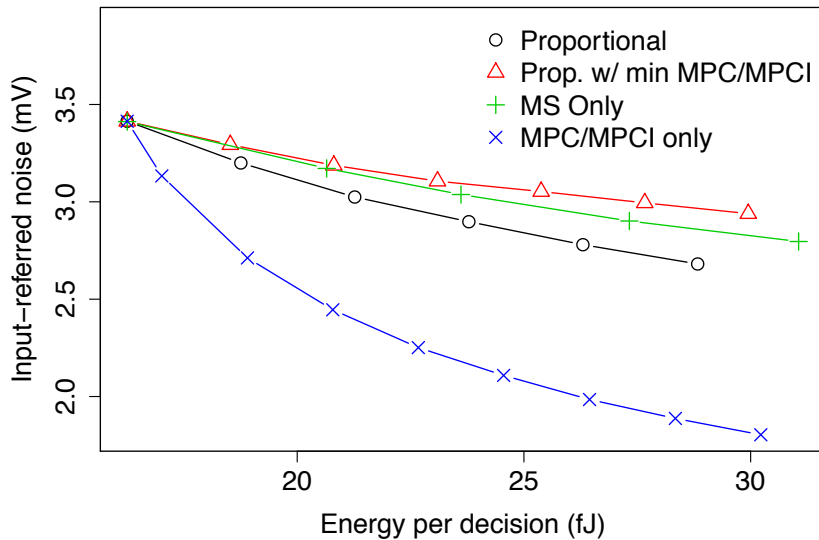


Figure 4.15: Energy and noise tradeoff of the StrongARM latch under different upsizing techniques.

For a noise reduction technique to be useful, it should not increase the decision latency of the comparator excessively. Figure 4.16 shows how the simulated decision latency for the StrongARM latch varies as the precharge devices are upsized. We define decision latency as the delay between the rising clock and the differential outputs reaching 50% and 90% of their final values respectively. A 1 fF load, which roughly corresponds to the input capacitance of a small inverter, is assumed for each output. While larger precharge devices have more parasitic capacitances, the improved signal gain at the outputs can cause comparisons to be resolved more quickly. For this reason, the decision latency actually improves initially until the precharge device width reaches around 0.5 μm . From there onwards, diminishing return sets in and the decision latency increases with further increases in the precharge device width. However, even at a width of 1.2 μm , which gives a 42% reduction in noise, the latency penalty is only at 2.7%.

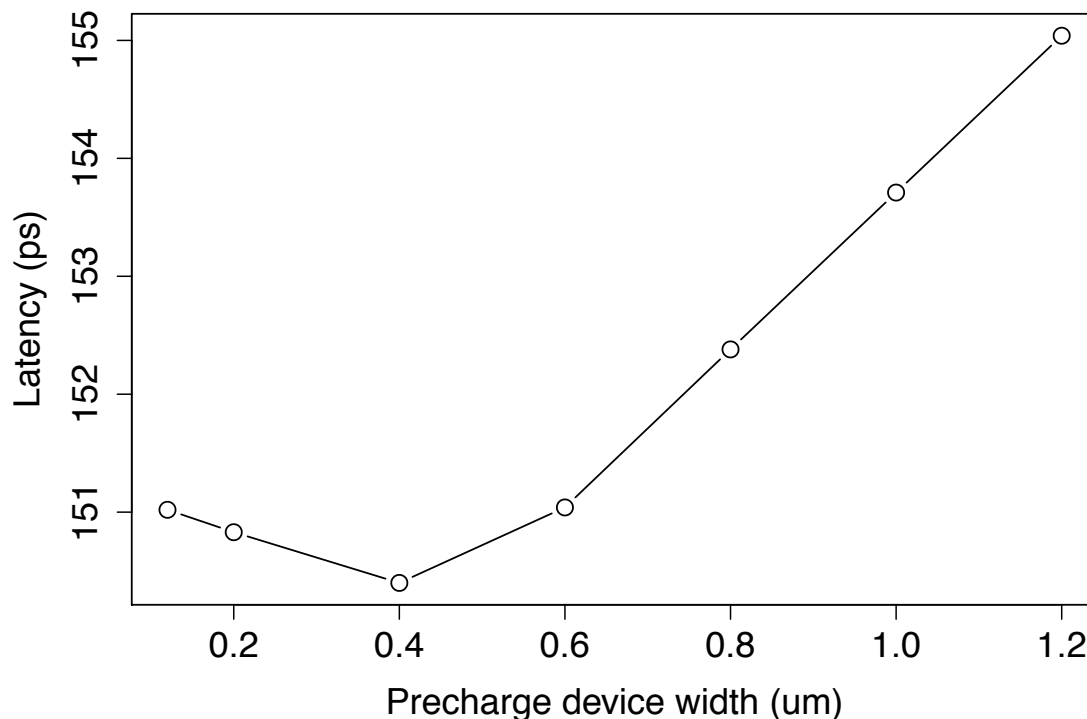


Figure 4.16: Variation of StrongARM latch decision latency with precharge device size.

Chapter 5

Prototype Evaluation

We fabricated a prototype chip in the IBM 90 nm low-power CMOS process. The chip, shown in figure 5.1, is 1 mm by 1 mm and wire-bonded in a 24-pin QFN package. It gives us the opportunity to benchmark the performance of the SCI under real process variations and operating conditions. In addition, it allows comparator noise to be directly measured and compared to simulations.

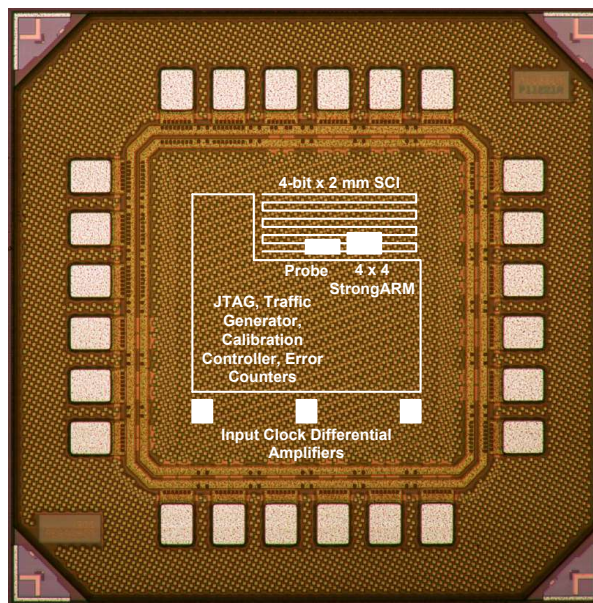


Figure 5.1: Die photograph of prototype chip in the IBM 90 nm low-power process.

5.1 Chip Design

The prototype chip consists of an implementation of the SCI and an array of StrongARM latches, complete with testing infrastructures. Figure 5.2 shows how the chip is organized at the top level. Three independent clocks are generated externally and brought on-chip through differential amplifiers. The transmitter clock (TCLK) launches data at the SCI transmitters, and the receiver clock (RCLK) captures the data at the receivers. The sampling clock (SCLK) is used to sample analog voltages and is also the main clock for all circuits used for noise measurements.

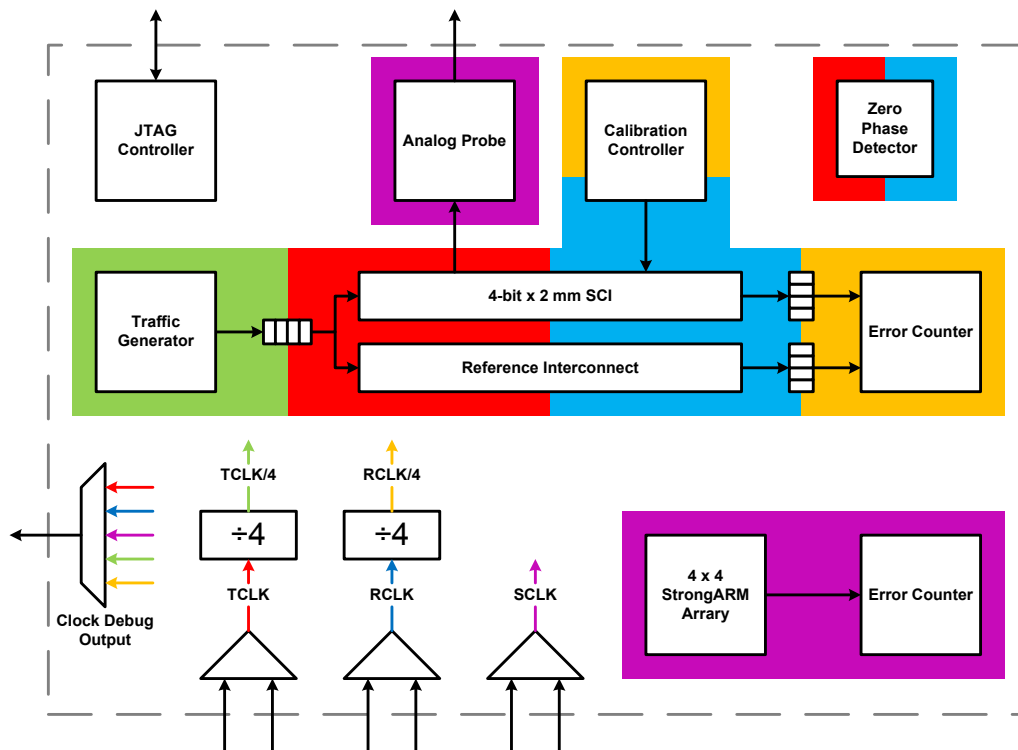


Figure 5.2: Prototype Chip Block Diagram.

In addition to the core clocks TCLK and RCLK are divided down by a factor of four to generate TCLK/4 and RCLK/4. Wherever possible, the testing infrastructures are run on these slower clocks and serializers or deserializers are used at the

boundaries to match data bandwidth. This minimizes the risk that the testing infrastructure will become the critical path at higher frequencies. Altogether this creates five different clock domains, which are color-coded on the diagram. Some modules operate purely in one domain while others span two different domains. Mesochronous synchronizers are used when crossing from one clock domain to another to preserve data integrity [?].

5.1.1 SCI

At the heart of the chip, we built a fully-custom, 4-bit wide, 2 mm long SCI in metal 7, which is a 2X wiring layer. In a production-level design die area is expensive and must be allocated efficiently. Using a 2X layer to build global interconnects approximately doubles the per-unit-length wire capacitance. However, this frees up the lower wiring layers for local routing and thus allows logic to be placed in the area that would otherwise be occupied solely by the interconnect. Since the SCI, like other LSIs, use two wires for every bit of communication, this organization leads to better overall area utilization. To simplify power measurements, the SCI is also placed on a custom supply (CVDD) that is separate from the rest of the chip.

The SCI presented in figure 3.12 is complete but does not lend itself easily to testing. Specifically, in order to find the lower bound on signal swing, it is desirable to have the swing itself to be adjustable. Unfortunately, with the capacitively-driven scheme, signal swing is fixed at design time by the capacitance ratios. Moreover, to quantify the effectiveness of calibration it is useful to identify the native offset for each receiver prior to calibration. As it stands, however, the SCI cannot maintain proper operation in the absence of periodic calibrations. This makes it difficult to separate improvements attained through calibration and those that are simply due to low underlying device mismatch.

We use a slightly modified SCI in our prototype as shown in figure 5.3. First, the final driver of the transmitter is placed on a separate signaling supply (SVDD) and an NFET is introduced that pulls-up in parallel with the PFET. The signaling supply allows the signal swing to be externally controlled while the NFET ensures

that a robust connection to SVDD exists even when it falls below the threshold of the PFET. The drivers are sized to give a signal swing of 120 mV at the maximum allowed value for SVDD. Next, a new control signal, DISABLE, is used to selectively turn off calibration in a non-destructive way. When asserted, DISABLE forces both outputs of the bias detection latch low, cutting off any feedback from the sense amplifier to the charge pumps. In addition, it allows the precharge PFET on both branches to be directly controlled by RCAL. Effectively, the SCI now operates as a pure CDI with dynamic refresh where both lines are periodically restored to the supply for the duration of the calibration.

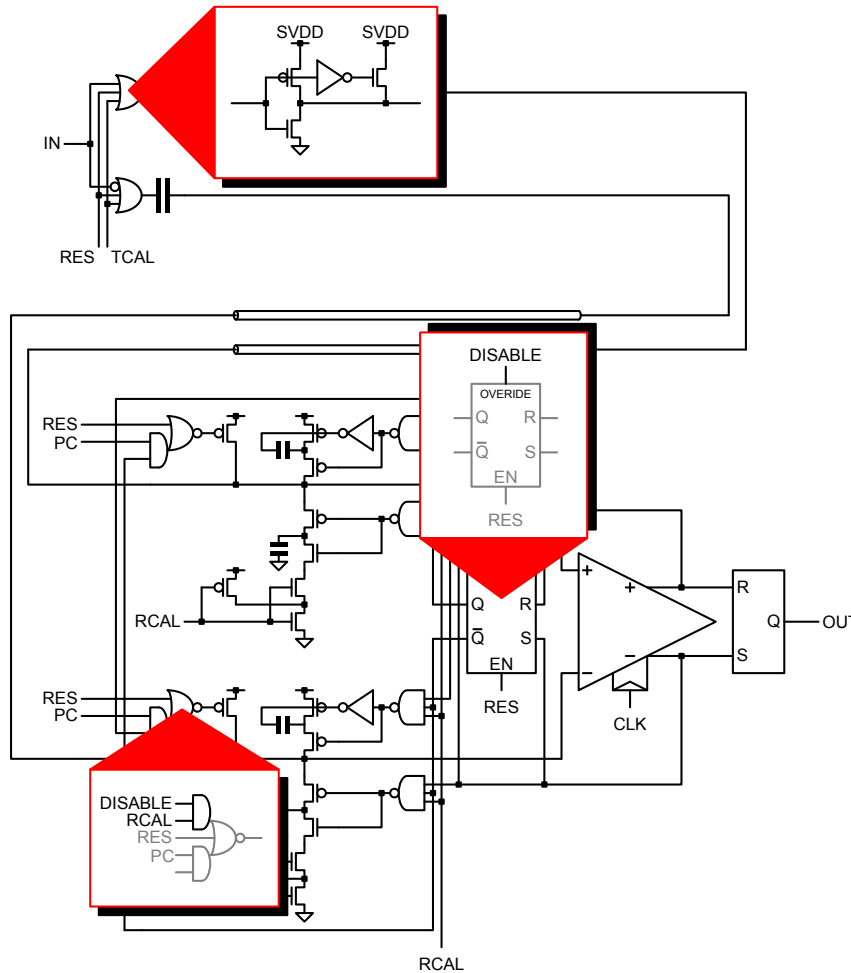


Figure 5.3: Modified SCI for testing.

5.1.2 Reference Interconnect

The reference interconnect (RI) is a simple, cycle-matched FSI that is implemented in parallel with the SCI. At each cycle the outputs of the SCI and the RI can be compared to determine if a transmission error has occurred over the SCI. In our design the SCI is serpentine to conserve chip area, so the transmitters and receivers are in close proximity in the final floorplan. This allows the RI to be trivially constructed with back-to-back flip-flops.

5.1.3 Zero Phase Detector

The advantage of using separate transmitter and receiver clocks is that the phase offset between the two clocks can be arbitrarily adjusted. The zero phase detector (ZPD) provides the capability to align the two clocks as in a perfectly synchronous system. As shown in figure 5.4, the ZPD makes use of two cross-coupled flip-flops, one clocked by TCLK with input tied to RCLK, and another clocked by RCLK with input tied to TCLK. If TCLK lags RCLK significantly, the first flip-flop will sample a one while the later will sample a zero. On the other hand, if TCLK leads RCLK significantly, the converse will be true. In-between the two extremes, metastability cause both flip-flops to sample a mixture of ones and zeros with some measurable probability. By interpolating the measured probabilities, the zero phase can be readily identified. Fully asynchronous synchronizers are used at the outputs to prevent the reading of a metastable state.

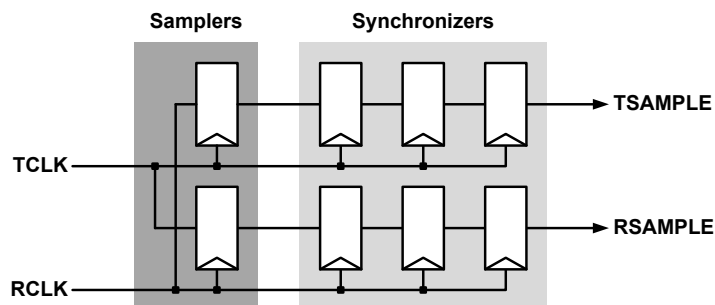


Figure 5.4: Zero phase detector built from cross-coupled flip-flops.

5.1.4 Traffic Generator

We implemented a programmable traffic generator (TG) to test the SCI under different data patterns. As shown in figure 5.5, the TG is based on a 16-bit LFSR, and can be operated under either the random or preset mode of operation. In the random mode, the LFSR is enabled and outputs a 16-bit PRBS at a quarter of the SCI frequency. The 16-bit pattern is multiplexed into the SCI in 4-bit segments and then updated at the end of every 4 TCLK cycles. In the preset mode, the LFSR is disabled and each bit of the SCI cycles through 4 of the 16 bit of the seed pattern. This allow any repeated 4-bit pattern to be transmitted.

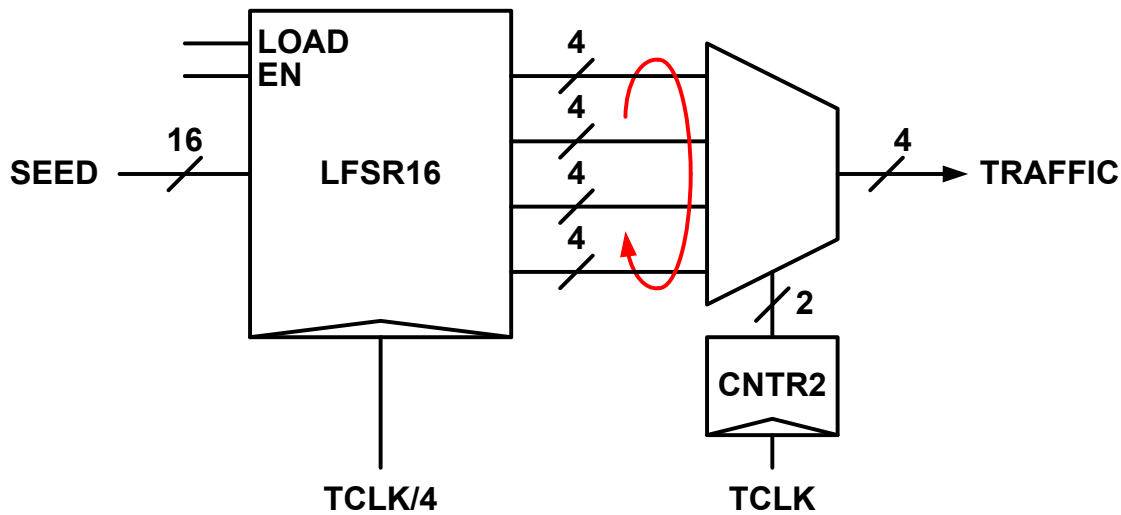


Figure 5.5: LFSR based traffic generator.

5.1.5 Calibration Controller

The calibration controller (CC) generates the calibration control signals according to the timing relationships outlined in section 3.6. To allow for more flexibility during testing, both the calibration interval and the calibration duration are designed to be programmable. Apart from a small frontend most of the CC runs on RCLK/4. This limits the minimum programmable resolution to 4 cycles.

5.1.6 Error Counter

The error counter (EC) compares the outputs of the SCI and RI over a programmable number of cycles and records the total number of transmission errors that are detected within that window. Since the EC runs at a quarter of the SCI frequency four bits are compared in parallel each cycle and the error count is incremented by a number between 0 and 4 each time. At the end of a run the BER can be determined by dividing the number of errors by the total number of cycles observed.

5.1.7 Analog Probe

For debugging purposes, an on-chip sampling scope [22] is used to monitor all power and ground rails as well as the line voltage on one bit of the SCI. The analog voltage is sampled with a switched capacitor and converted to an externally measurable current via a series of current mirrors.

5.1.8 StrongARM Array

An array of 16 StrongARM latches with increasing precharge device widths (4 different sizes, 4 samples per size) are fabricated alongside the SCI on our prototype chip. The supplies CVDD and SVDD are reused as input voltage sources to the comparators. By varying CVDD and SVDD independently, the common mode and differential input voltages can be set to arbitrary values.

The StrongARM array has its own error counter that compares the output of a selected comparator to a programmable reference bit. Unlike the one for the SCI, the counter runs in the same clock domain as the comparator so only one comparison needs to be made each cycle.

5.1.9 JTAG Controller

The JTAG controller implements the IEEE 1149.1 standard and provides a centralized testing and debugging interface for our prototype. Through the JTAG controller the user has full access to all the on-chip modules. This includes setting programmable

control registers, starting and stopping experiment runs, and collecting measurements from result registers. Appendix B summarizes all control and status registers that are accessible via the JTAG interface.

5.2 Test Board and Experimental Setup

To facilitate the testing of our prototype we built a custom circuit board as shown in figure 5.6. A zero-insertion-force (ZIF) socket is used so that multiple chips can be tested without the need for multiple boards.

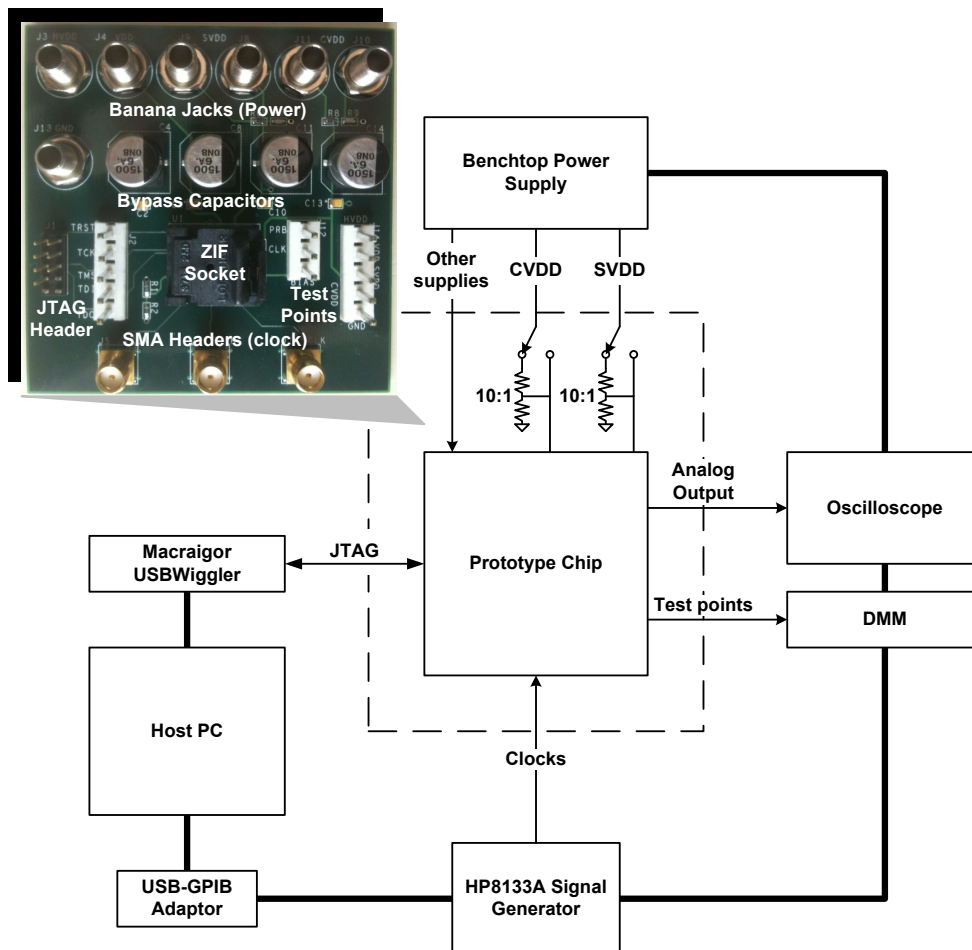


Figure 5.6: Board and lab setup for prototype testing.

The board is powered through programmable bench-top power supplies and clocks are generated from the HP8133A high frequency signal generator. For CVDD and SVDD, the board provides the additional option of delivering power to the chip through a 10:1 resistive voltage divider. This is designed to minimize supply ripples during comparator noise measurements, where sub-millivolt precision is required. By attenuating the supply output by a factor of 10 we also get a 10-fold reduction in the maximum output ripple. The extra supply resistance is not an issue here because the chip draws only a very small amount of leakage current from CVDD and SVDD when it is configured for noise measurements.

The analog probe output and other voltage test points are monitored with an oscilloscope and a digital multi-meter (DMM). All lab equipments are centrally controlled by a host PC through a USB-GPIB interface. A Macraigor USBWigglerTM is used to communicate with the JTAG interface in our prototype.

5.3 SCI Experiments

With the prototype and testing fixtures in place, we now turn our discussion to a series of experiments performed on the SCI.

5.3.1 Calibration Effectiveness

For our first experiment, we set out to evaluate the effectiveness of the SCI in countering input offset caused by device mismatch. To do this, we measure how the BER for the SCI changes with signal swing, both with and without calibration. The experiments are performed at 1.5 GHz, and the BER is taken to be the worst-case BER observed over 6 different traffic patterns: pseudo-random, toggle (1010...), constant-zero (0000....), constant-one (1111...), pulse-zero (1110...) and pulse-one (0001...). In order to minimize distortions due to statistical fluctuations, the average from 30 iterations are used for each data point.

Figure 5.7 shows the relationship between BER and signal swing for the best and worst samples in a group of 16 SCIs. Without calibration, there is a large spread in

the signal swing required to achieve a given BER. When device mismatch is small, a relatively low swing can be used. When device mismatch is large, a higher swing, up to 50 mV more, is required to maintain the same BER. This large variability poses a difficult challenge to an interconnect designer, who must determine the appropriate signal swing amidst the conflicting objectives of yield and energy-efficiency.

With calibration enabled, all samples achieve a BER below 10^{-9} with less than 20 mV of signal swing. Moreover, the performance gap between the best and worst sample is reduced significantly. For a given BER, the additional swing required by the worst sample over the best sample is no more than 3 mV. Overall, the swing reduction enabled by calibration reached 51 mV, with an average of 18 mV. As integration density rises and the worst case device mismatch continues to deteriorate, we expect the gains from calibration to become more pronounced.

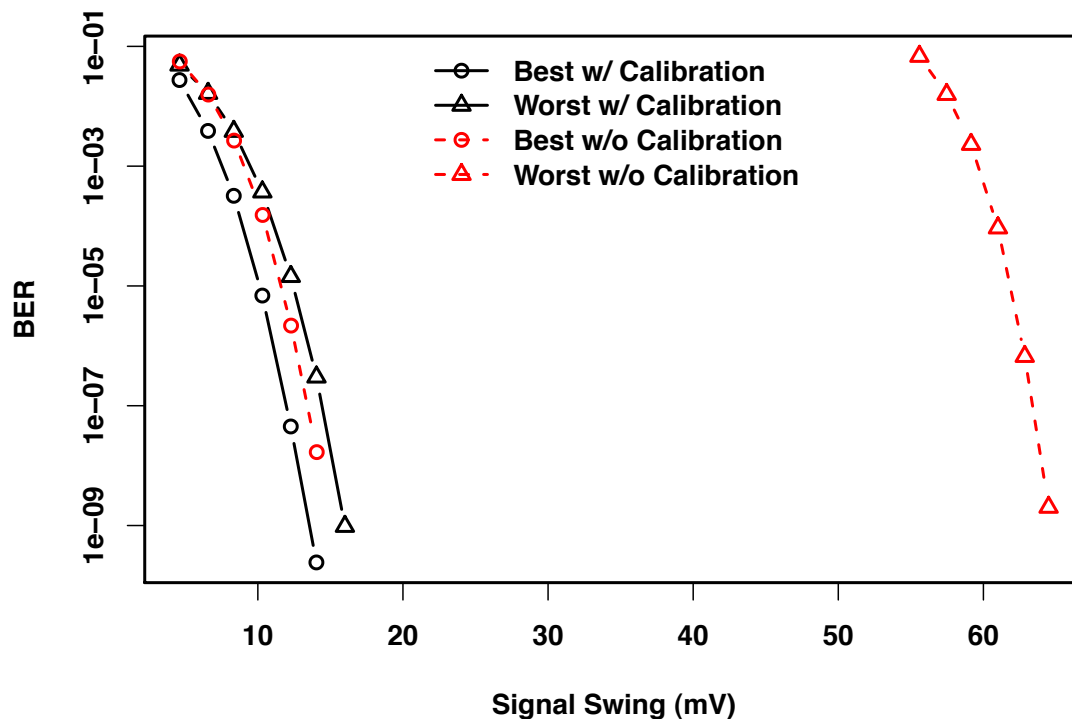


Figure 5.7: Dependence of BER on signal swing with and without calibration.

5.3.2 Eye Opening

Figure 5.8 shows how the BER of an SCI varies as the phase of RCLK is varied with respect to TCLK at a fixed signal swing of 32 mV. Given enough timing margin, all 16 samples have a projected BER below 10^{-30} at this swing. Without loss of generality, we define the zero phase as the point where the two extrapolated lines from the edges of the eye intersect. As before, the experiment is conducted at 1.5 GHz. However, constant-zero and constant-one traffic patterns are not used here since they do not produce any line voltage transitions that are sensitive to changes in clock timing.

Based on our measurements, the desired BER of 10^{-30} can be sustained over 75% of the width of the eye. This corresponds to a wire bandwidth of around 5.9 GHz. The excessive wire bandwidth suggests that we are not operating near the physical limit of the wire, and that significantly longer interconnects can be built at the same frequency if desired. Unfortunately, the maximum frequency for our 2 mm prototype could not be determined because it exceeds the limitations of our testing infrastructureⁱ.

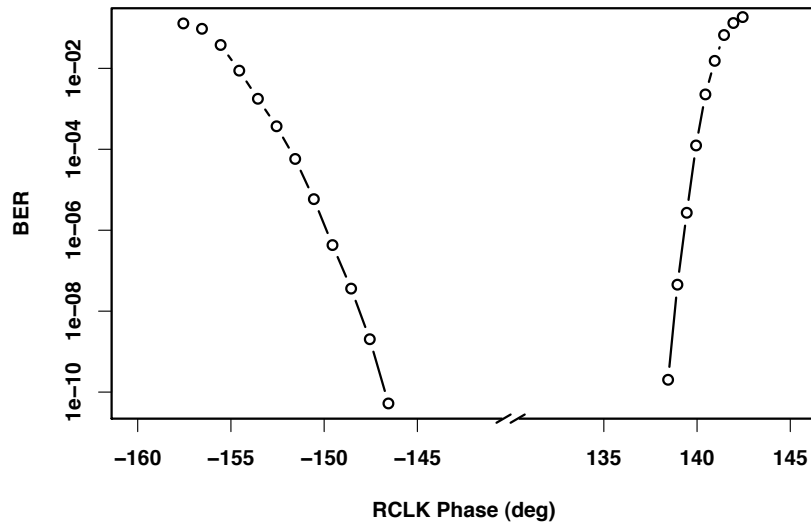


Figure 5.8: Variation of BER with the phase of the receiver clock.

ⁱIn our design the standard cell library and the pin bandwidth of the ZIF socket limited the maximum testable frequency to around 1.5 GHz.

5.3.3 Leakage

As described section 3.4, leakage on the lines degrades the performance of the SCI. To quantify the effects of leakage in our prototype, Figure 5.9 shows the signal swing an SCI requires to sustain a BER of $1\text{E-}9$ as the calibration interval is varied from 2,000 to 250,000 cycles. The experiment is conducted at 1.5 GHz at room temperature, and calibration duration is fixed at 16 cycles. As expected, a higher signal swing is required for longer intervals since more charge leaks away in-between calibrations. However, even at the maximum setting, where the interconnect is only in calibration 0.007% of the time, the additional swing required to maintain the same BER is less than 0.7 mV.

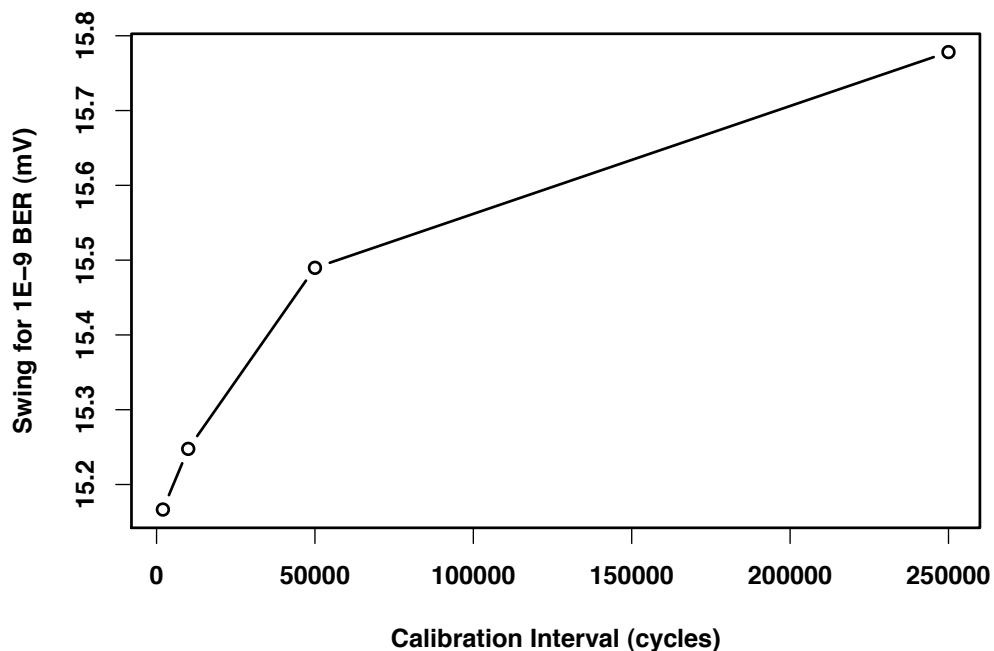


Figure 5.9: Signal swing required to sustain a BER of 10^{-9} for different calibration intervals.

The non-linear relationship between the required signal swing and the calibration interval is somewhat counter-intuitive and deserves further explanation. Given how little the line voltage changes between calibrations, we would normally expect

leakage to increase linearly with the length of the interval. Upon a closer examination of figure 3.11, we can trace the source of non-linearity to the transient leakage characteristics of the voltage decremter. Immediately upon the completion of a calibration, C_{adj} is fully discharged, so the drain-to-source voltage across M1, and hence its channel leakage, is high. As the voltage on C_{adj} rises, the throttle mechanism described earlier kicks in, reducing the leakage through M1. Since the other sources of leakage do not exhibit this degree of non-linearity, the combined average leakage current decreases as the length of the interval gets longer.

5.3.4 Energy and Area

To find the energy-efficiency of the SCI we set the traffic pattern to pseudo-random and directly measure the total current drawn from the power supplies dedicated to the SCI. The current is multiplied by the voltage then divided by the number of bits and the operating frequency to calculate the per bit energy of communication. This approach automatically includes the energy contribution of the calibration control signals. One slight complication is that at our target signal swing of 32 mV SVDD is at only 168 mV instead of its nominal 1.2 V. As a result, the wire energy observed benefits from a quadratic energy reduction that would typically not be present in the SCI. To correct for this, we scale our measured wire energy by the ratio of SVDD's nominal value to its final value.

Table 5.1 compares the energy and area of the SCI to other published designs in similar technologies. Mensink's design [36] is a straightforward extension of the CDI and serves as a good base for evaluating the incremental gains of using the SCI. Kim's design [27] use very aggressive feedforward and feedback equalization techniques to boost the operating frequency at the expense of area. The three designs are very different both in length and in their respective wiring structures. To make the comparisons more meaningful, we normalize measurements associated with the transmitter (which tend to scale with the wire) by the total wire capacitance of the interconnect, or where data is not available, by the wire length. The receivers, which do not depend on wire geometries, are compared separately.

With the very low signal swing enabled by calibration, the transmitter energy for the SCI is over 50% lower than its predecessors after normalization. Moreover, since the comparators no longer need to be aggressively upsized, the SCI receiver is also 37-58% more energy-efficient than the other designs. The SCI is able to realize these energy savings without significantly increasing area. After normalization, both the transmitter and receiver areas are quite comparable to Mensink’s design, which does not use any calibrationⁱⁱ. Adding the calibration circuits does not significantly increase receiver area because a smaller comparator can be used.

Table 5.1: Comparison of the SCI with other energy-efficient on-chip interconnects.

	SCI	Mensink et al. [36]	Kim et al. [27]
Technology	90 nm/1.2V	90 nm/1.2V	90 nm/1.2V
Wiring layer	M7 (2X)	M4 (1X)	M8 (2X)
Wire width/spacing	0.28 $\mu\text{m}/0.28 \mu\text{m}$	0.54 $\mu\text{m}/0.32 \mu\text{m}$	0.6 $\mu\text{m}/0.4 \mu\text{m}$
Length	2 mm	10 mm	10 mm
Signal swing	32 mV	100 mV	> 98 mV
BER	10^{-30}	-	-
<i>Energy</i>			
Total	77 fJ/b	280 fJ/b	356 fJ/b
TX (per mm)	13.7 fJ/b	16 fJ/b	27.7 fJ/b
TX (per pF)	13.4 fJ/b	28.6 fJ/b	-
RX	49.9 fJ/b	120 fJ/b	79 fJ/b
<i>Area</i>			
TX (total)	72 μm^2 (adj.)	226 μm^2	1120 μm^2
TX (per mm)	36 μm^2	22.6 μm^2	112 μm^2
TX (per pF)	70.3 μm^2	80.7 μm^2	-
RX	123 μm^2	117 μm^2	640 μm^2

ⁱⁱThe transmitter area for the SCI is adjusted to reflect the final 32 mV swing.

5.4 Comparator Noise Experiment

In this section, we provide experimental evidence of our prediction that increasing the width of the precharge devices improves the input-referred noise for the comparator. We also seek to verify the accuracy of the noise simulation technique that formed the basis of our analyses in chapter 4.

Figure 5.10 illustrates our approach to measuring the input-referred noise for each comparator. First, we sweep CVDD and SVDD and record how the BER changes as the differential input voltage is varied. This produces a series of data points of the form (V_i, ϵ_i) where ϵ_i is the log of the BER when the input voltage is at V_i (red points). Next, we interpolate the data points to find, V^* , the input voltage required to sustain a BER of 0.5. In a perfectly balanced comparator, the voltage required for a BER of 0.5 is zero, so here V^* is the extra input voltage needed due to device mismatch. Subtracting V^* from our voltage measurements gives a new set of data points (V'_i, ϵ_i) that reflects the relationship between BER and noise alone (blue points). Finally, we perform a least-squares curve fitting to an ideal log-BER-noise profile to find the standard deviation of the effective input referred noise voltage, σ_n , as

$$\sigma_n = \min_{\sigma} \sum_{i=0}^N [\epsilon_{ideal}(\sigma, V_i) - \epsilon_i]^2, \quad (5.1)$$

where

$$\epsilon_{ideal}(\sigma, V_i) = \log \left[Q \left(\frac{V_i}{\sigma} \right) \right] \quad (5.2)$$

is the log of the ideal BER at V_i for a normally-distributed input noise voltage with a standard deviation of σ .

Figure 5.11 shows how the measured and simulated input-referred noise changes with increasing precharge device widths. Each measured point is the average from 64 StrongARM samples and the error-bars represent the one standard deviation confidence interval. The measurements confirm the hypothesis that upsizing the precharge devices leads to lower input-referred noise. Moreover, while individual measurements

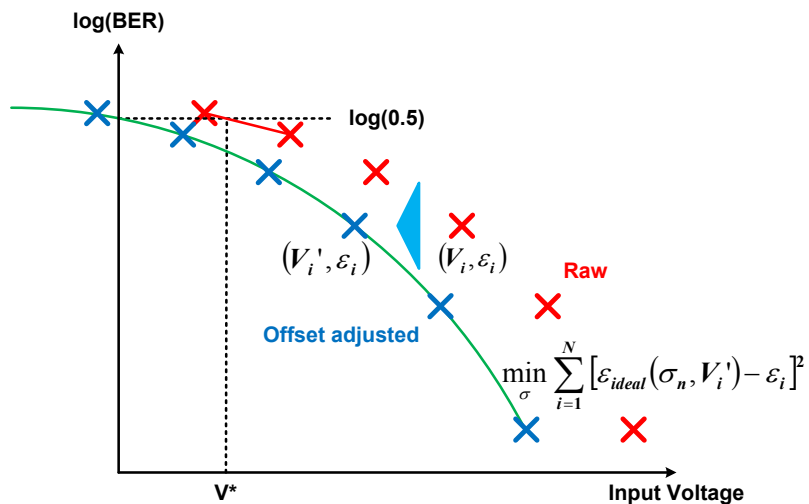


Figure 5.10: Measuring input-referred noise through curve-fitting.

can vary, the sample means match the simulated values well both in trend and absolute value. Across all sizes, the simulated noise is 10 to 15% lower than the measurements. This is expected since our simulations only consider thermal noise while the measurements include all sources of noise.

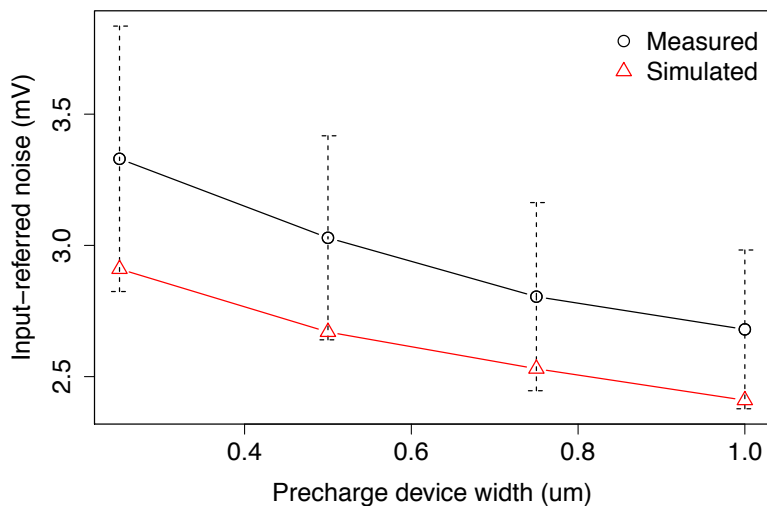


Figure 5.11: Comparison of measured and simulated input-referred noise for StrongARM latches with different precharge device sizes.

In our analyses, we predicted that the improved g_m and g_{ds} for the sensing devices as precharge devices are upsized will also bring about an improvement in the input-offset voltage. Figure 5.12 plots how the measured and simulated input-offset voltage changes with precharge device width. As expected, the input-offset voltage decreased as precharge device width increased. However, this time, the measured offset is consistently about 20-30% lower than what was predicted by the simulations. This can be attributed to a limitation in the Monte Carlo simulation parameters which were only characterized down to an area of 100 um^2 for our design kit. In practice, the devices in the StrongARM latch are only a few microns apart so they exhibit better matching.

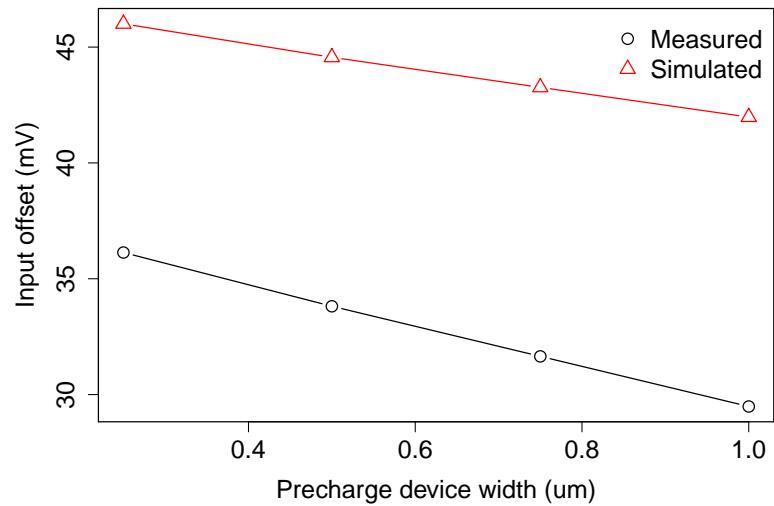


Figure 5.12: Comparison of measured and simulated input offset for StrongARM latches with different precharge device sizes.

Chapter 6

Conclusion

In the coming years, in order for us to realize our computing aspirations, it is critical to have circuits that provide robust, energy-efficient on-chip communications. In this dissertation, we critically examined the state-of-art low-swing interconnects (LSIs) and discovered that they are inadequate to meet our growing needs. Specifically, the deteriorating device mismatch, which limits signal swing, coupled with the decreasing supply voltage, prevents the LSIs from reaching their full energy-saving potentials. The capacitively-driven interconnect (CDI), while attractive in its simplicity and equalization benefits, is negatively impacted the most due to the linear relationship between wire energy and swing. It also requires dedicated biasing circuits to maintain the DC line voltages.

To improve the longevity of LSIs, we conceived and developed the self-calibrating interconnect (SCI). Built on top of a CDI, the SCI use feedback and charge-pumps to neutralize receiver offset and establish DC bias in one unified mechanism. This allows the SCI to use extremely low signaling voltages, which maximizes energy-efficiency, without sacrificing robustness and yield. In our prototype, we demonstrated reliable communication at BER below 10^{-30} with only a 32 mV signal swing, and we achieved over 50% energy improvement over a previously published CDI in a comparable process.

As signal swing declines, the reliability of an LSI becomes increasingly influenced by the random noise in the system, most of which is in the thermal noise internal to

the comparators. To that end, we have developed an intuitive thermal noise simulation technique for comparators that exposes the contribution of individual devices. Using this tool, we uncovered the previously undetected relationship between the precharge device sizes and the noise performance of the popular StrongARM latch. More precisely, we were able to show that upsizing the precharge devices, contrary to popular beliefs, improves the input-referred noise of the comparator. In fact, for a given energy budget, using larger precharge device sizes can lower noise by over 30% from what was previously thought possible.

The invention of the SCI is a step in the right direction, but there is more work ahead to fully close the energy-gap we seek. While we have demonstrated that the SCI performs well in a low-power process, it remains to be seen if the same holds true in high-performance processes. On the one hand, the relatively smaller and faster devices will allow the calibration circuits to be implemented with a smaller overhead. On the other hand, the increase in leakage will outpace any gains in speed, requiring more frequent calibrations. The recent emergence of fin-FETs [25, 5], which leak significantly less than traditional planar devices, could make the use of SCIs more attractive moving forward.

Like all LSIs, the SCI trades off a lower wire energy with higher transceiver energy. An SCI is most effective when the distance between its transceivers is large. Unfortunately, with the per unit length wire resistance on the rise and the push for higher bandwidth, there is pressure to decrease the transceiver spacing in order to meet the frequency requirements. As devices become more plentiful in the newer processes, another interesting study will be looking at how various feed-forward equalizers (FFE) and decision-feedback equalizers (DFEs) can be incorporated into the SCI design. This will allow for great transceiver spacing for a given frequency and thus maintain the amortization benefits of longer interconnects.

The need for periodic calibration necessitates some architectural support before the SCI can be productively deployed. The SCI is initially developed for the channel circuits in a network-on-chip (NOC). Here the system already have well-established flow control mechanisms to hold back data transmission if the downstream router does not have enough buffer spaces for the flits at the source router. Integrating

the SCI into a NOC is straightforward because we can leverage existing hardware and throttle data transmission when there are no available credits or if the SCI is in calibration. To use the SCI in another context, such as the data bus between the cache and the processor, more investigation is needed to ensure that the calibration does not disrupt critical operations.

Finally, as with many custom circuits, the SCI is a significant departure from the standard cell design that is prevalent in most commercial ICs. Compared to the FSI, which just involves inverters, there is a very steep increase in design complexity. As it stands, designing, implementing and testing SCI in a large scale chip is a very formidable task, especially if there is time pressure. More research into how the SCI design process can be automated within the existing CAD tool flow will prove to be extremely valuable.

Appendix A

Transfer Function Derivation for Capacitor-Coupled Wire Segment

To find the transfer function of the capacitor coupled wire segment shown in figure 2.4b, we start by applying KCL at the intermediate and output nodes, which gives

$$sC_c (V_s - V_i) + \frac{1}{2}sC_w V_s + \frac{1}{R_w} (V_s - V_o) = 0, \quad (\text{A.1})$$

$$\frac{1}{R_w} (V_o - V_s) + \frac{1}{2}sC_w V_o = 0. \quad (\text{A.2})$$

Solving (A.2) for V_s , we get

$$V_s = \left(1 + \frac{1}{2}R_w C_w\right) V_o. \quad (\text{A.3})$$

Substituting this into (A.1) and collecting like terms, the relationship between V_o and V_i is given by

$$\left(sC_c + sC_w + \frac{1}{2}s^2 R_w C_w C_c + \frac{1}{4}s^2 R_w C_w^2\right) V_o = sC_c V_i. \quad (\text{A.4})$$

Dividing both side by sC_c , and letting $\rho \triangleq \frac{C_w}{C_c}$ and $\tau \triangleq \frac{1}{2}R_wC_w$, this becomes

$$[(1 + \rho) + (1 + 0.5\rho)\tau s] V_o = V_i. \quad (\text{A.5})$$

It follows that

$$\frac{V_o}{V_i} = \frac{1}{a_1 s + a_0}, \quad (\text{A.6})$$

where $a_0 = 1 + \rho$ and $a_1 = (1 + 0.5\rho)\tau$.

Appendix B

Registers Accessible via JTAG

Table B.1: Control and status registers accessible via the JTAG interface on the prototype chip.

Addr	Name	N	R/W	Description
0	ZPD_OUT	2	R	Output of the zero-phase detector: 10 if TCLK leads RCLK by more than t_{setup} , 01 if TCLK lags RCLK by more than t_{setup} , 00 or 11 in-between.
1	TG_PATTERN	16	W	Seed for pseudo-random traffic or bit-pattern for preset traffic.
2	TG_LOAD	1	W	Loads the pattern specified by TG_PATTERN into the traffic generator.
3	TG_EN	1	W	Selects the pseudo-random traffic by enabling the internal LFSR.
4	SER_PHASE	1	W	Selects the sampling clock phase of the serializer at the boundary of TCLK/4 and TCLK: 0 for positive TCLK edge, 1 for negative TCLK edge.

continued

Addr	Name	N	R/W	Description
5	CC_CAL_INTERVAL	16	W	Length of the calibration interval in RCLK/4 cycles.
6	CC_CAL_DURATION	8	W	Length of each calibration in RCLK/4 cycles.
7	CC_DISABLE_CAL	1	W	Disables calibration.
8	CC_DOUBLE_PC	1	W	Increase the width of the PC pulse to 2 RCLK cycles.
9	CC_PHASE	1	W	Selects the sampling clock phase of the calibration controller at the boundary of RCLK/4 and RCLK: 0 for positive RCLK edge, 1 for negative RCLK edge.
10	CC_EN	1	W	Starts the calibration controller (performs dynamic refresh when calibration is disabled).
11	AP_SOURCE	5	W	One-hot selection of the analog voltage being probed: SVDD (bit 0), VSS (bit 1), CVDD (bit 3), Negative SCI Line (bit 3), Positive SCI Line (bit 4).
12	RI_PHASE	1	W	Selects the sampling clock phase of the reference interconnect at the boundary of TCLK and RCLK: 0 for positive RCLK edge, 1 for negative RCLK edge.
13	DES_BIT_SELECT	2	W	Selects which bit of SCI to measure the BER for.
14	DES_PHASE	1	W	Selects the launching clock phase of the deserializer at the boundary of RCLK and RCLK/4: 0 for positive RCLK edge, 1 for negative RCLK edge..

continued

Addr	Name	N	R/W	Description
15	DES_PATH_DELAY	2	W	Selectively delay the SCI output or the RI output by one additional cycle: no delay (00), delay SCI (01), delay RI (10), delay both (11)
16	EC_RUN_DURATION	48	W	The total number of RCLK/4 cycles to observe for a given SCI error-count measurement.
17	EC_EN	1	W	Starts the error counter to record the number of SCI transmission errors.
18	EC_TIME_IS_UP	1	R	When asserted it signifies that the maximum number of observation cycles have elapsed and that the SCI error count is no longer updated.
19	EC_ERROR_COUNT	48	R	Total number of SCI transmission errors recorded in the observation time window.
20	SA_REF_VAL	1	W	The reference value for the comparators.
21	SA_BIT_SELECT	4	W	Selects one of the 16 StrongARM latches for measurement: smallest precharge devices (0-4) to largest precharge devices (12-15).
22	SA_RUN_DURATION	48	W	The total number of SCLK cycles to observe for a given comparator error-count measurement.
23	SA_EN	1	W	Starts the error counter to record the number of the comparator decision errors.

continued

Addr	Name	N	R/W	Description
24	SA_TIME_IS_UP	1	R	When asserted it signifies that the maximum number of observation cycles have elapsed and that the comparator error count is no longer updated.
25	SA_ERROR_COUNT	48	R	Total number of comparator decision errors recorded in the observation time window.
26	CM_SOURCE	3	W	Select the clock source to monitor: TCLK (001), RCLK (010), SCLK (011), TCLK/4 (100), RCLK/4 (101), NONE (all other patterns).

Bibliography

- [1] International Technology Roadmap for Semiconductors 2010 Tables: Interconnect.
- [2] International Technology Roadmap for Semiconductors 2010 Tables: Process Integration, Devices and Structures.
- [3] <http://ark.intel.com/ProductCollection.aspx?familyId=59142&MarketSegment=DT>.
- [4] http://en.wikipedia.org/wiki/List_of_CPU_power_dissipation.
- [5] <http://spectrum.ieee.org/tech-talk/semiconductors/design/intels-new-transistors-enter-the-third-dimension>.
- [6] <http://www.geforce.com/#/Hardware/GPUs/geforce-gtx-590/specifications>.
- [7] http://www.mosis.com/ibm/ibm_processes.html.
- [8] E. Alon. Measurement and Regulation of On-Chip Power Supply Noise. *Ph.D. Dissertation, Stanford University*, 2006.
- [9] A. Asenov. Simulation of statistical variability in nano mosfets. In *VLSI Technology, 2007 IEEE Symposium on*, pages 86–87. IEEE, 2007.
- [10] A. Asenov, S. Kaya, and J.H. Davies. Intrinsic threshold voltage fluctuations in decanano mosfets due to local oxide thickness variations. *Electron Devices, IEEE Transactions on*, 49(1):112–119, 2002.

- [11] M. Bhargava, M.P. McCartney, A. Hoeffler, and K. Mai. Low-overhead, digital offset compensated, sram sense amplifiers. In *Custom Integrated Circuits Conference, 2009. CICC'09. IEEE*, pages 705–708. IEEE.
- [12] T. Burd. Energy efficient processor system design. *Ph.D. Dissertation, UC Berkeley*, 1998.
- [13] L. Capodieci. From optical proximity correction to lithography-driven physical design (1996-2006): 10 years of resolution enhancement technology and the roadmap enablers for the next decade. In *Proceedings of SPIE*, volume 6154, page 615401, 2006.
- [14] W.J. Dally and J. Poulton. Transmitter equalization for 4-gbps signaling. *Micro, IEEE*, 17(1):48–56, 1997.
- [15] W.J. Dally and J.W. Poulton. *Digital systems engineering*. Cambridge Univ Pr, 1998.
- [16] M.J. Deen, Chih-Hung Chen, and Yuhua Cheng. Mosfet modeling for low noise, rf circuit design. In *Custom Integrated Circuits Conference, 2002. Proceedings of the IEEE 2002*, pages 201 –208, may 2002.
- [17] W. Ellersick, C.K.K. Yang, M. Horowitz, and W. Dally. Gad: A 12-gs/s cmos 4-bit a/d converter for an equalized multi-level link. In *VLSI Circuits, 1999. Digest of Technical Papers. 1999 Symposium on*, pages 49–52. IEEE, 1999.
- [18] H. Fukutome, Y. Momiyama, T. Kubo, Y. Tagawa, T. Aoyama, and H. Arimoto. Direct evaluation of gate line edge roughness impact on extension profiles in sub-50-nm n-mosfets. *Electron Devices, IEEE Transactions on*, 53(11):2755–2763, 2006.
- [19] T. Furuyama, S. Saito, and S. Fujii. A new sense amplifier technique for vlsi dynamic ram's. In *Electron Devices Meeting, 1981 International*, volume 27, pages 44 – 47, 1981.

- [20] A. Hajimiri and T.H. Lee. A general theory of phase noise in electrical oscillators. *Solid-State Circuits, IEEE Journal of*, 33(2):179–194, 1998.
- [21] Kwangseok Han, Hyungcheol Shin, and Kwyro Lee. Analytical drain thermal noise current model valid for deep submicron mosfets. *Electron Devices, IEEE Transactions on*, 51(2):261 – 269, feb. 2004.
- [22] R. Ho, B. Amrutur, K. Mai, B. Wilburn, T. Mori, and M. Horowitz. Applications of on-chip samplers for test and measurement of integrated circuits. In *VLSI Circuits, 1998. Digest of Technical Papers. 1998 Symposium on*, pages 138–139. IEEE, 1998.
- [23] R. Ho, K. Mai, and M. Horowitz. Efficient on-chip global interconnects. In *VLSI Circuits, 2003. Digest of Technical Papers. 2003 Symposium on*, pages 271–274, 2003.
- [24] R. Ho, I. Ono, F. Liu, R. Hopkins, A. Chow, J. Schauer, and R. Drost. High-speed and low-energy capacitively-driven on-chip wires. In *IEEE International Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers*, pages 412–612, 2007.
- [25] Xuejue Huang, Wen-Chin Lee, Charles Kuo, D. Hisamoto, Leland Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Yang-Kyu Choi, K. Asano, V. Subramanian, Tsu-Jae King, J. Bokor, and Chenming Hu. Sub 50-nm finfet: Pmos. In *Electron Devices Meeting, 1999. IEDM Technical Digest. International*, pages 67 –70, 1999.
- [26] S. Keckler. GPU Computing and the Road to Extreme-Scale Parallel Systems, NVIDIA Presentation, 2011.
- [27] B. Kim and V. Stojanovic. A 4gb/s/ch 356fj/b 10mm equalized on-chip interconnect with nonlinear charge-injecting transmit filter and transimpedance receiver in 90nm cmos. In *Solid-State Circuits Conference-Digest of Technical Papers, 2009. ISSCC 2009. IEEE International*, pages 66–67. IEEE, 2009.

- [28] J. Kim, B.S. Leibowitz, J. Ren, and C.J. Madden. Simulation and analysis of random decision errors in clocked comparators. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 56(8):1844–1857, Aug. 2009.
- [29] Peter Kogge, Keren Bergman, Shekhar Borkar, Dan Campbell, William Carlson, William Dally, Monty Denneau, Paul Franzon, William Harrod, Kerry Hill, Jon Hiller, Sherman Karp, Stephen Keckler, Dean Klein, Robert Lucas, Mark Richards, Al Scarpelli, Steven Scott, Allan Snavely, Thomas Sterling, R. Stanley Williams, and Katherine Yelick. ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems. 2008.
- [30] T. Krishna, J. Postman, C. Edmonds, Li-Shiuan Peh, and P. Chiang. Swift: A swing-reduced interconnect for a token-based network-on-chip in 90nm cmos. In *Computer Design (ICCD), 2010 IEEE International Conference on*, pages 439–446, oct. 2010.
- [31] H. Lan, R. Schmitt, and C. Yuan. Simulation and measurement of on-chip supply noise in multi-gigabit i/o interfaces. In *Quality Electronic Design, 2008. ISQED 2008. 9th International Symposium on*, pages 670–675. IEEE, 2008.
- [32] M.J.E. Lee, W.J. Dally, and P. Chiang. Low-power area-efficient high-speed i/o circuit techniques. *Solid-State Circuits, IEEE Journal of*, 35(11):1591–1599, 2000.
- [33] K. Lim, P. Ranganathan, J. Chang, C. Patel, T. Mudge, and S. Reinhardt. Understanding and designing new server architectures for emerging warehouse-computing environments. In *International Symposium on Computer Architecture*, pages 315–326. IEEE, 2008.
- [34] CT Liu, FH Baumann, A. Ghetti, HH Vuong, CP Chang, KP Cheung, JI Colonell, WYC Lai, EJ Lloyd, JF Miner, et al. Severe thickness variation of sub-3 nm gate oxide due to si surface faceting, poly-si intrusion, and corner stress. In *VLSI Technology, 1999. Digest of Technical Papers. 1999 Symposium on*, pages 75–76. IEEE, 1999.

- [35] H. Mahmoodi, S. Mukhopadhyay, and K. Roy. Estimation of delay variations due to random-dopant fluctuations in nanoscale cmos circuits. *Solid-State Circuits, IEEE Journal of*, 40(9):1787–1796, 2005.
- [36] E. Mensink, D. Schinkel, E. Klumperink, E. van Tuijl, and B. Nauta. A 0.28 pj/b 2gb/s/ch transceiver in 90nm cmos for 10mm on-chip interconnects. In *Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers. IEEE International*, pages 414–612. IEEE, 2007.
- [37] T. Mizuno, J. Okumtura, and A. Toriumi. Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in mosfet's. *Electron Devices, IEEE Transactions on*, 41(11):2216–2221, 1994.
- [38] J. Montanaro, R.T. Witek, K. Anne, A.J. Black, E.M. Cooper, D.W. Dobberpuhl, P.M. Donahue, J. Eno, W. Hoepfner, D. Kruckemyer, et al. A 160-mhz, 32-b, 0.5-w cmos risc microprocessor. *Solid-State Circuits, IEEE Journal of*, 31(11):1703–1714, 1996.
- [39] P. Nuzzo, F. De Bernardinis, P. Terreni, and G. Van der Plas. Noise analysis of regenerative comparators for reconfigurable adc architectures. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 55(6):1441–1454, Jul. 2008.
- [40] M.J.M. Pelgrom, A.C.J. Duinmaijer, and A.P.G. Welbers. Matching properties of mos transistors. *Solid-State Circuits, IEEE Journal of*, 24(5):1433–1439, 1989.
- [41] D. Schinkel, E. Mensink, E. Klumperink, E. van Tuijl, and B. Nauta. Low-power, high-speed transceivers for network-on-chip communication. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 17(1):12–21, 2009.
- [42] K. Seno, K. Knorpp, L.L. Shu, N. Teshima, H. Kihara, H. Sato, F. Miyaji, M. Takeda, M. Sasaki, Y. Tomo, et al. A 9-ns 16-mb cmos sram with offset-compensated current sense amplifier. *Solid-State Circuits, IEEE Journal of*, 28(11):1119–1124, 1993.

- [43] R. Singh and N. Bhat. An offset compensation technique for latch type sense amplifiers in high-speed low-power srams. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 12(6):652–657, 2004.
- [44] P.A. Stolk, F.P. Widdershoven, and DBM Klaassen. Modeling statistical dopant fluctuations in mos transistors. *Electron Devices, IEEE Transactions on*, 45(9):1960–1971, 1998.
- [45] S. Suzuki and M. Hirata. Threshold difference compensated sense amplifier. *Solid-State Circuits, IEEE Journal of*, 14(6):1066–1070, 1979.
- [46] G. Van der Plas, S. Decoutere, and S. Donnay. A 0.16pJ/conversion-step 2.5mW 1.25GS/s 4b ADC in a 90nm digital CMOS process. *ISSCC Dig. Tech. Papers*, page 2310, Feb. 2006.
- [47] Y. Watanabe, N. Nakamura, and S. Watanabe. Offset compensating bit-line sensing scheme for high density dram’s. *Solid-State Circuits, IEEE Journal of*, 29(1):9–13, 1994.
- [48] K.L.J. Wong and C.K.K. Yang. Offset compensation in comparators with minimum input-referred supply noise. *Solid-State Circuits, IEEE Journal of*, 39(5):837–840, 2004.
- [49] L.A. Zadeh. Frequency analysis of variable networks. *Proceedings of the IRE*, 38(3):291–299, Mar. 1950.
- [50] H. Zhang, V. George, and JM Rabaey. Low-swing on-chip signaling techniques: effectiveness and robustness. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 8(3):264–272, 2000.